

УЗБЕКСКОЕ АГЕНТСТВО СВЯЗИ И ИНФОРМАТИЗАЦИИ
ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
Кафедра «Программное обеспечение информационных технологий»

Камилов М.М., Эргашев А.К.

КОНСПЕКТ ЛЕКЦИЙ

**По дисциплине
«МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ»**

**для студентов направления 5521900.
«Информатика и информационные технологии»**



ТАШКЕНТ-2008

АННОТАЦИЯ

В данном сборнике лекций представлены современные принципы, подходы и методы моделирования сложно формализуемых объектов. Описаны задачи структурной и параметрической идентификации. Рассмотрены практические задачи управления сложными объектами. Для усвоения материалов лекций достаточно знания основ высшей математики в объеме обычного курса ВТУЗа. Курс рассчитан для подготовки бакалавров по направлению образования В-522600 "Информатика и информационные технологии".

КУРС РАСЧИТАН НА 100 ЧАСОВ ЗАНЯЙ, ИЗ НИХ:

- **50 ЧАСОВ — ЛЕКЦИОННЫХ ЗАНЯТИЙ;**
- **25 ЧАСОВ — ПРАКТИЧЕСКИХ ЗАНЯТИЙ;**
- **25 ЧАСОВ — ЛАБОРАТОРНЫХ ЗАНЯТИЙ.**

СОДЕРЖАНИЕ

Лекция 1. ОБЩИЕ ВОПРОСЫ ТЕОРИИ МОДЕЛИРОВАНИЯ (2 часа).....	6
1. Предмет теории моделирования	6
2. Роль и место моделирования в исследованиях систем	9
3. Классификация моделей	11
4. Моделирование в процессах познания и управления.....	13
5. Классификация объектов моделирования	17
6. Основные этапы моделирования	18
7. Этапы моделирования объектов (процессов, явлений)	18
Лекция 2. ТЕХНОЛОГИЯ МОДЕЛИРОВАНИЯ (2 часа).....	22
1. Создание концептуальной модели	22
2. Подготовка исходных данных	27
3. Разработка математической модели.....	29
4. Выбор метода моделирования	33
Лекция 3. ТЕХНОЛОГИЯ МОДЕЛИРОВАНИЯ (2 часа).....	37
1. Выбор средств моделирования	37
2. Проверка адекватности и корректировка модели	40
3. Планирование экспериментов с моделью.....	42
4. Анализ результатов моделирования.....	43
Лекция 4. ПРОБЛЕМА МОДЕЛИРОВАНИЯ (2 часа).....	48
1. Объект моделирования	48
2. Сведения об объекте	49
3. Априорная информация.....	49
4. Апостериорная информация	50
Лекция 5. ЗАДАЧА ИДЕНТИФИКАЦИИ (2 часа).....	52
1. Постановка задачи идентификации.....	52
2. Трудности идентификации.....	55
Лекция 6. ЗАДАЧИ ИДЕНТИФИКАЦИИ И КЛАССИФИКАЦИЯ МЕТОДОВ ИДЕНТИФИКАЦИИ (2 часа)	58
1. Идентификация структуры и параметров объекта	58
2. Классификация методов идентификации	59
Лекция 7. СТРУКТУРНАЯ ИДЕНТИФИКАЦИЯ (2 часа)	62
1. Выделение объекта из среды.....	62
2. Ранжирование входов и выходов объекта (Метод экспертных оценок)	64
3. Метод непосредственного ранжирования	66
Лекция 8. ИССЛЕДОВАНИЕ ЗАДАЧ СТРУКТУРНОЙ ИДЕНТИФИКАЦИИ (2 часа).....	69
1. Метод парных сравнений	69
2. Определение рационального числа входов и выходов объекта, учитываемых в модели.....	72
3. Определение характера связи между входом и выходом модели объекта	72
Лекция 9. АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 часа)	74

1. Потоки заявок	74
2. Марковские модели.....	77
Лекция 10. ХАРАКТЕРИСТИКИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 часа)	82
1. Характеристики вычислительных систем как систем массового обслуживания	82
2. Характеристики вычислительных систем как сложных систем массового обслуживания.....	86
3. Методы приближённой оценки характеристик вычислительных систем	91
Лекция 11. НЕСТАЦИОНАРНЫЕ РЕЖИМЫ ФУНКЦИОНИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 часа)	96
1. Нестационарные режимы функционирования вычислительных систем	96
2. Характеристики вычислительных систем как стохастических сетей.....	98
Лекция 12. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 часа)	102
1. Процедура имитационного моделирования	102
2. Обобщенные алгоритмы имитационного моделирования	107
Лекция 13. МЕТОДЫ ИССЛЕДОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 часа).....	111
1. Методы определения характеристик вычислительных систем.....	111
2. Метод повторных экспериментов.....	114
3. Методы генерации случайных величин и последовательностей	120
Лекция 14. МОДЕЛИ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 часа)	125
1. Модели линейной дискретной оптимизации с булевыми переменными	125
2. Преобразование задачи с дискретными переменными к задаче с булевыми переменными	126
3. Преобразование задачи линейного булева программирования к задаче нелинейного булева программирования	127
Лекция 15. НОВЫЕ МОДЕЛИ ЗАДАЧ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 часа)	130
1. Задача и модель оптимизации работы насосной станции.....	130
2. Модель задачи автоматической классификации.....	134
3. Задача об оптимизации размещения букв алфавита на клавиатуре ЭВМ	135
Лекция 16. РЕШЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 часа)	138
1. Моделирование и оптимизация работы насосной станции	138
2. Проверка адекватности математической модели.....	140
3. Алгоритм оптимального управления работы насосной станции	144
Лекция 17. ИССЛЕДОВАНИЕ МЕТОДОВ ПОСТРОЕНИЯ ИНФОРМАТИВНЫХ НАБОРОВ ПРИЗНАКОВ (2 часа)	147
1. Сравнительный анализ методов построения информативных наборов признаков	147

2. Аналитический подход к формированию информативной подсистемы признаков в задаче распознавания	153
3. Упрощенный метод классификации с использованием аналитического подхода формирования информативной подсистемы признаков при наличии обучающей выборки	159

ЛЕКЦИЯ 1. ОБЩИЕ ВОПРОСЫ ТЕОРИИ МОДЕЛИРОВАНИЯ (2 ЧАСА)

План

1. Предмет теории моделирования
2. Роль и место моделирования в исследованиях систем
3. Классификация моделей
4. Моделирование в процессах познания и управления
5. Классификация объектов моделирования
6. Основные этапы моделирования
7. Этапы моделирования объектов (процессов, явлений)

1. ПРЕДМЕТ ТЕОРИИ МОДЕЛИРОВАНИЯ

Мысленные модели как форма теоретического осмыслиения и отражения действительности играют большую роль в физическом познании. В этой связи важное теоретико-познавательное и методологическое значения приобретает вопрос о формировании моделей, использовании их в познании, возможности их включения в более общие представления и их связи с другими формами познавательной деятельности, мысленным и реальным экспериментами, гипотезой, теорией.

Модели, как научные гипотезы, мы можем рассматривать как форму развития науки, неокончательно разработанные теории, согласно мнениям некоторых представителей конкретных наук, можно рассматривать в качестве моделей будущих совершенных теорий.

Представляется, что можно дать следующее рабочее определение мысленным и материальным моделям.

Моделирование — это замещение одного объекта (оригинала) другим (моделью) и фиксация или изучение свойств оригинала путем исследования свойств модели. Замещение производится с целью упрощения, удешевления, ускорения фиксации или изучения свойств оригинала.

В общем случае объектом-оригиналом может быть любая естественная или искусственная, реальная или воображаемая система.

Метод моделирования применяется все большим числом ученых.

Примеры из механики, физики (твердого тела), химии, биологии, медицины, экономии и др.

Концепция моделирования прежде всего преследует цель включения моделей в процесс создания теорий, поскольку идеальные модели могут быть предварительной ступенью в построении или моделью интерпретации теории.

Гипотезы отличаются от идеальных моделей, являющихся идеализированными объектами теории, и могут быть представлены как предварительная ступень или модель интерпретации теории, прежде всего благодаря своей функции в процессе познания в качестве научно обоснованного предложения о до сих пор неизвестных и недоступных явлениях. Гипотезы могут рассматриваться как предварительные ступени формирующихся моделей.

При разработке исходной модели интуиция исследователя играет большую роль. В начале может выдвигаться большое число моделей, однако в ходе исследования их число сокращается.

Формой работы смоделью является мысленный эксперимент. Иногда он называется идеализированным, что вскрывает тесную его связь с реальными экспериментами и основное различие между ними. В известной мере мысленный эксперимент представляет образное мысленное реконструирование определенных сторон реального эксперимента «мышление есть не более чем продукт опыта в уме», -«мысленный опыт, безусловно удобнее, чем действительный опыт: мысли у нас всегда имеются, и легче накопить опыт в уме, чем в действительности» [по Энгельмайеру].

Эксперимент остается критерием адекватности отражения в модели определенных сторон объекта оригинала. Эксперимент играет роль судьи, который выносит решение, жить или не жить представлению, полученному с помощью модели.

Представление изучаемого явления, процесса или объекта с помощью математических соотношений и формул называется математической моделью. При моделировании объекта исследования дело начинается с формализации объекта, т.е с построения соответствующей математической модели. Для этого выделяются его наиболее существенные черты и свойства и описываются с помощью математических соотношений.

После того, как построена математическая модель, т.е задача придана математическая форма, мы можем воспользоваться для ее изучения математическими методами.

Примеры математических моделей.

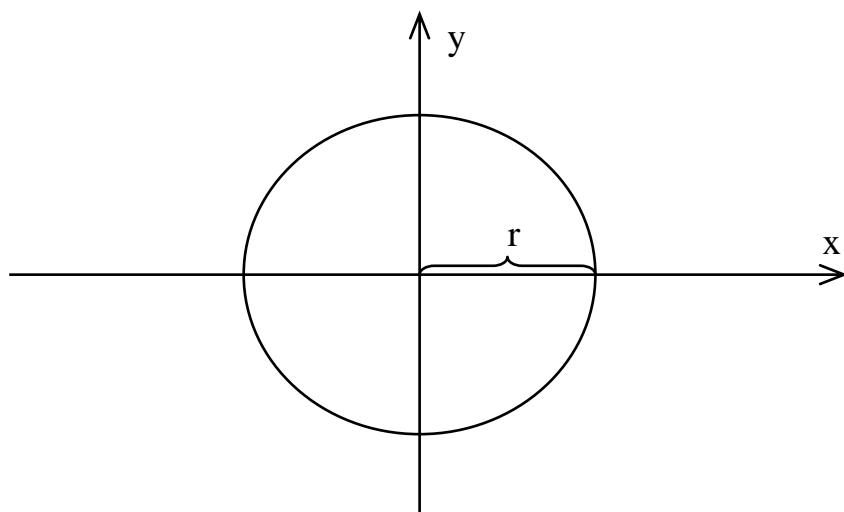


Рис.1. $x^2+y^2=r^2$ -уравнение (модель) окружности.

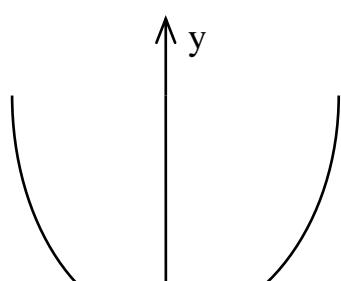
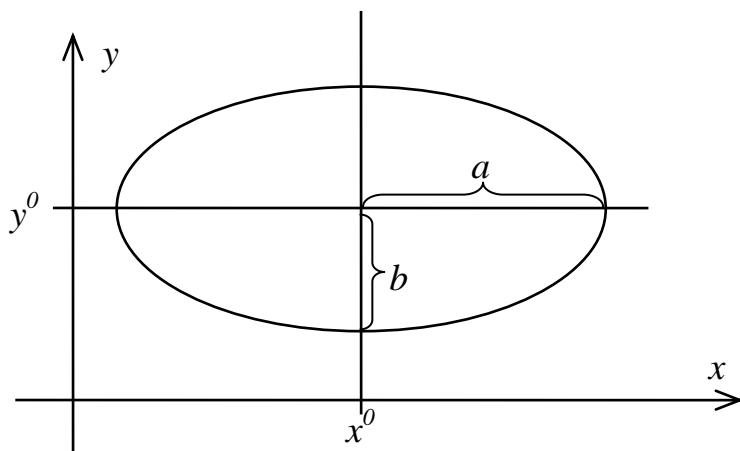


Рис.2 $y=ax^2$ -уравнение (модель) параболы.Рис.3. $\frac{(x - x^0)^2}{a^2} + \frac{(y - y^0)^2}{b^2} = 1$ - уравнение (модель) эллипса.

Задача 1. Необходимо определить площадь поверхности письменного стола.

$$S = a \cdot b$$

Это означает, реальный объект (письменный стол) заменяется абстрактной математической моделью прямоугольника. Прямоугольнику присваиваются размеры, полученные в результате измерения, и площадь такого прямоугольника приближенно принимается за искомую площадь.

Выбор модели прямоугольника для поверхности стола мы обычно делаем, полагаясь на свое зрительное восприятие. Однако, человеческий глаз как измерительный инструмент не отличается высокой точностью. Поэтому при более серьезном подходе к задаче, прежде чем воспользоваться моделью прямоугольника для определения площади, эту модель, т.е. объект исследования, нужно проверить на предмет описания его моделью прямоугольника. Для этого можно измерить противоположные стороны и обе диагонали прямоугольника. Если они попарно равны, то поверхность стола действительно можно рассматривать как прямоугольник. В

противном случае от модели прямоугольника надо отказаться, и следует перейти к модели четырехугольника общего вида.

Моделирование целесообразно, когда у модели отсутствуют те признаки оригинала, которые препятствуют его исследованию, или имеются отличные от оригинала параметры, способствующие фиксации или изучению свойств модели.

Теория моделирования представляет собой взаимосвязанную совокупность положений, определений, методов и средств создания и изучения моделей. Эти положения, определения, методы и средства, как и сами модели, являются предметом теории моделирования.

Основная задача теории моделирования заключается в том, чтобы вооружить исследователей технологией создания таких моделей, которые достаточно точно и полно фиксируют интересующие свойства оригиналов, проще или быстрее поддаются исследованию и допускают перенесение его результатов на оригиналы.

Теория моделирования является основной составляющей общей теории систем — системологии, где в качестве главного принципа постулируются осуществимые модели: система представима конечным множеством моделей, каждая из которых отражает определенную грань ее сущности.

В данной книге в качестве объектов-оригиналов рассматриваются ВС, т. е. ВС составляют предметную область моделирования. Понятие ВС здесь трактуется в широком смысле—от однопроцессорных систем обработки данных до распределенных сетей ЭВМ с различным программным обеспечением и функциональным назначением.

Поскольку ВС — это искусственные, инженерные системы, все их параметры известны, по крайней мере, они известны создателям ВС, а значит могут быть изучены, познаны их исследователями. Это предопределяет принципиальную возможность моделирования ВС.

2. РОЛЬ И МЕСТО МОДЕЛИРОВАНИЯ В ИССЛЕДОВАНИЯХ СИСТЕМ

Трудно переоценить роль моделирования в научных изысканиях, инженерном творчестве и, вообще, в жизни человека. Познание любой системы сводится, по существу, к созданию ее модели. Перед изготовлением каждого устройства или сооружения разрабатывается его модель-проект. Любое произведение искусства является моделью, фиксирующей действительность. Человек, прежде чем совершить что-либо, обдумывает возможную последовательность действий или интуитивно руководствуется определенными установившимися апробированными моделями поведения.

Особую ценность имеют конструктивные модели, т. е. такие, которые допускают не только фиксацию свойств, но и исследование зависимостей характеристик от параметров системы. Такие модели позволяют оптимизировать функционирование систем. **Оптимизационные модели** — основа теории сложных систем.

Роль моделирования как метода научного познания и метода решения технических задач всегда оценивалась достаточно высоко. С развитием техники нашло

широкое применение **физическое моделирование** сооружений, машин и механизмов.

Достижения математики привели к распространению **математического моделирования** различных объектов и процессов. Подмечено, что динамика функционирования разных по физической природе систем описывается однотипными зависимостями, т.е. может быть представлена одинаковыми моделями. Расчетные формулы, которые используют инженеры для анализа и синтеза всевозможных систем, зачастую выведены из математических моделей этих систем.

На качественно новую ступень поднялось моделирование в результате разработки методологии **имитационного моделирования**. Это обусловлено тем, что существенно расширился класс систем, которые могут быть исследованы с помощью моделирования.

Появление вычислительных машин способствовало углублению и расширению сфер применения моделирования. Особое значение приобретает моделирование в современных условиях всемерного ускорения научно-технического прогресса, при требованиях достижения больших эффектов с ограниченными материальными, трудовыми, энергетическими и временными ресурсами.

Сейчас трудно указать область человеческой- деятельности, где не применялось бы моделирование. Разработаны, например, модели производства автомобилей, выращивания пшеницы, функционирования отдельных органов человека, жизнедеятельности Азовского моря, последствий атомной войны. В перспективе для каждой системы могут быть созданы свои модели, перед реализацией каждого технического или организационного проекта. должно проводиться моделирование.

Специалисты считают, что моделирование становится основной функцией вычислительных систем. Действительно, в настоящее время основные усилия направлены на внедрение вычислительной техники в автоматизированные системы управления технологическими процессами, организационно-экономическими комплексами и процессами проектирования, для создания банков данных и знаний. Но любая система управления нуждается в информации об управляемом объекте или процессе. Поэтому справочные системы, банки данных и знаний следует рассматривать как подспорье для управляющих систем. В свою очередь, любая система управления нуждается в модели управляемого объекта или процесса, в моделировании последствий тех или иных управляющих решений. В связи с этим применение вычислительных систем для моделирования приобретает первостепенное значение.

Сами ВС как сложные и дорогостоящие технические системы могут и должны стать объектами моделирования. Моделирование целесообразно использовать на этапе проектирования ВС, и для анализа функционирования действующих систем в экстремальных условиях или при изменении их состава, структуры, способов управления или рабочей нагрузки. Применение моделирования на этапе проектирования позволяет анализировать варианты проектных решений, определять работоспособность и производительность, выявлять дефицитные и малозагруженные ресурсы, вычислять ожидаемые времена реакции и принимать решения по рацио-

нальному изменению состава и структуры ВС или по способу организации вычислительного процесса.

В том случае, когда не удается создать модель, допускающую использование аналитических методов оптимизации, процесс разработки ВС принимает итерационный характер. Предварительно выбранный вариант проектного решения анализируют путем моделирования. Это дает возможность определить ожидаемые характеристики будущей системы, выявить ее сильные и слабые стороны. Если характеристики не удовлетворяют предъявляемым требованиям, по результатам анализа выполняют корректировку проекта, затем снова проводят моделирование. Этот процесс повторяется до достижения требуемого качества функционирования разрабатываемой ВС.

При анализе действующих систем с помощью моделирования определяют границы работоспособности системы, выполняют имитацию экстремальных условий, которые могут возникнуть в процессе функционирования системы. Искусственное создание таких условий на действующей системе затруднено и может привести к катастрофическим последствиям, если система не справится со своими функциональными обязанностями. Применение моделирования может быть полезным при разработке стратегии развития ВС, ее усовершенствовании и образовании связей с другими ВС при создании многомашинных комплексов и сетей ЭВМ, при изменении количества, номенклатуры или частоты решаемых задач.

Целесообразнее использовать моделирование для действующих ВС, поскольку можно опытным путем проверить адекватность модели и оригинала и более точно определить те параметры системы и внешних воздействий на нее, которые служат исходными данными для моделирования. Моделирование реальной ВС позволяет выявить ее резервы и прогнозировать качество функционирования при любых изменениях, поэтому полезно иметь модели всех развивающихся систем.

3. КЛАССИФИКАЦИЯ МОДЕЛЕЙ

Физические модели. В основу классификации положена степень абстрагирования модели от оригинала. Предварительно все модели можно подразделить на две группы: **материальные (физические) и абстрактные (математические)**.

Физической моделью обычно называют систему, которая эквивалентна или подобна оригиналу, либо у которой процесс функционирования такой же, как у оригинала, и имеет ту же или другую физическую природу. Можно выделить следующие виды физических моделей: **натурные, квазинатурные, масштабные и аналоговые.**

Натурные модели — это реальные исследуемые системы. Их называют макетами и опытными образцами. Натурные модели имеют полную адекватность с системой-оригиналом, что обеспечивает высокую точность и достоверность результатов моделирования. Процесс проектирования ВС завершается зачастую испытанием опытных образцов.

Квазинатурные модели представляют собой совокупность натурных и математических моделей. Этот вид моделей используется в случаях, когда мате-

матическая модель части системы не является удовлетворительной (например, модель человека-оператора) или когда часть системы должна быть исследована во взаимодействии с остальными частями, но их еще не существует, либо их включение в модель затруднено или дорого. Примерами квазинатурных моделей могут служить вычислительные полигоны, на которых отрабатывается программное обеспечение различных систем, или реальные АСУ, исследуемые совместно с математическими моделями соответствующих производств.

Масштабная модель — это система той же физической природы, что и оригинал, но отличающаяся от него масштабами. Методологической основой масштабного моделирования является теория подобия, которая предусматривает соблюдение геометрического подобия оригинала и модели и соответствующих масштабов для их параметров. При проектировании ВС масштабные модели могут использоваться для анализа вариантов компоновочных решений по конструкции системы и ее элементов.

Аналоговыми моделями называются системы, имеющие физическую природу, отличающуюся от оригинала, но сходные с оригиналом процессы функционирования. Обязательным условием при этом является однозначное соответствие между параметрами изучаемого объекта и его модели, а также тождественность безразмерных математических описаний процессов, протекающих в них. Для создания аналоговой модели требуется наличие математического описания изучаемой системы.

В качестве аналоговых моделей' используются механические, гидравлические, пневматические системы, но наиболее широкое применение получили электрические и электронные аналоговые модели, в которых сила тока или напряжение являются аналогами физических величин другой природы. Особенностью аналоговых моделей является их гибкость и простота адаптации к изменению и измерению количественных значений параметров и характеристик моделируемой системы.

Аналоговые модели используют при исследовании средств вычислительной техники на уровне логических элементов и электрических цепей, а также на системном уровне, когда функционирование системы описывается, например, дифференциальными или алгебраическими уравнениями.

Математические модели. Математическая модель представляет собой формализованное описание системы с помощью абстрактного языка, в частности с помощью математических соотношений, отражающих процесс функционирования системы. Для составления модели можно использовать любые математические средства — алгебраическое, дифференциальное и интегральное исчисление, теорию множеств, теорию алгоритмов и т. д. По существу вся математика создана для составления и исследования моделей объектов или процессов.

К средствам абстрактного описания систем относятся также языки химических формул, схем, чертежей, карт, диаграмм и т. п. Выбор вида модели определяется особенностями изучаемой системы и целями моделирования, так как исследование модели позволяет получить ответы на определенную группу вопросов. Для получения другой информации может потребоваться модель другого вида.

Цели моделирования и характерные черты оригинала определяют в конечном счете ряд других особенностей моделей и методы их исследования. Например, математические модели можно классифицировать на детерминированные и вероятностные (стохастические). Первые устанавливают однозначное соответствие между параметрами и характеристиками модели, а вторые — между статистическими значениями этих величин. Выбор того или иного вида модели обусловлен степенью необходимости учета случайных факторов. Среди математических моделей можно выделить по методу их исследования аналитические, численные и имитационные модели.

Аналитической моделью называется такое формализованное описание системы, которое позволяет получить решение уравнения в явном виде, используя известный математический аппарат.

Численная модель характеризуется зависимостью такого вида, который допускает только частные численные решения для конкретных начальных условий и количественных параметров модели.

Имитационная модель — это совокупность описания системы и внешних воздействий, алгоритмов функционирования системы или правил изменения состояния системы под влиянием внешних и внутренних возмущений. Эти алгоритмы и правила не дают возможности использования имеющихся математических методов аналитического и численного решения, но позволяют имитировать процесс функционирования системы и производить измерения интересующих характеристик.

Имитационные модели могут быть созданы для гораздо более широкого класса объектов и процессов, чем аналитические и численные модели. Поскольку для реализации имитационных моделей используются, как правило, ВС, средствами формализованного описания имитационных моделей служат зачастую универсальные или специальные алгоритмические языки. Имитационные модели в наибольшей степени подходят для исследования ВС на системном уровне.

4. МОДЕЛИРОВАНИЕ В ПРОЦЕССАХ ПОЗНАНИЯ И УПРАВЛЕНИЯ

С проблемой моделирования мы сталкиваемся в двух случаях: во-первых, в процессах познания, когда стараются познавательные модели объектов и явлений, с которыми приходится сталкиваться человеку, и во-вторых, в процессах управления, связанных с целенаправленным изменением объекта, т.е. с достижением целей, поставленных человеком.

Рассмотрим оба типа моделирования подробнее.

В процессах познания создается **познавательная модель** объекта, отображающая в необходимой мере механизм его функционирования. Примером такого рода моделирования является изучение окружающей нас природы. Объяснение феноменов природы, их взаимная увязка и обусловленность, анализ механизмов и т.д. - вот основные задачи моделирования этого рода. Такое моделирование по сути дела мало чем отличается от познания вообще, целью которого, как известно, является синтез моделей; отражающих важную для человека специфику объек-

тов окружающего его мира. Эта специфика отражается в своеобразии причинно-следственных связей каждого объекта или явления, которые удобно представить в виде некоторого «преобразователя» причины в следствие (рис 4.)

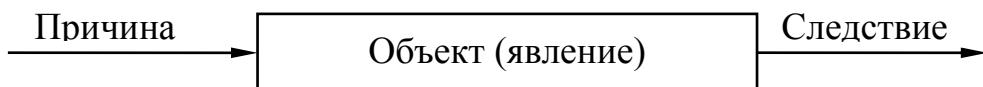


Рис 1. Представление объекта познания.

Описание «работы» такого преобразователя на каком-либо языке мы будем подразумевать моделью.

Таким образом под моделью мы будем подразумевать рассуждения (на любом языке - математическом, графическом, алгоритмическом, разговорном и т.д.), позволяющие имитировать наблюдаемое явления. Очевидно, что конкретные цели конкретизируют и язык на котором описывается модель. Так, языком большого числа физических и технических моделей является математика.

Формализуем сказанное. Будем обозначать причину буквой X, а следствие-Y. Связь между ними запишем условно в виде

$$Y=F(X),$$

где F-правило преобразования причины X в следствие Y. Это и есть модель. Назовем F оператором модели.

На рис.5 показано взаимодействие моделируемого объекта со средой.

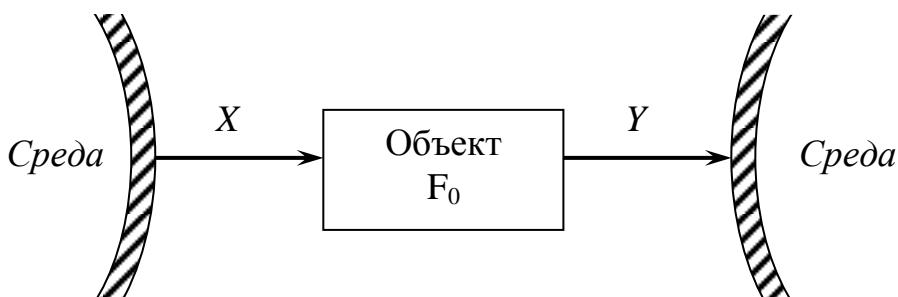


Рис.5. Взаимодействие объекта со средой.

Это взаимодействие происходит по каналам X и Y. По каналу X среда воздействует на объект, а по каналу Y объект воздействует на среду.

Задача моделирования сводится к определению оператора F, связывающего вход и выход объекта.

Пусть x₁,x₂, ... ,x_N - наблюдения входа объекта, y₁,y₂, ... ,y_N - соответствующие им наблюдения его выхода в дискретные моменты времени 1,2, ... , N эти наблюдения связаны неизвестным оператором объекта F₀, т.е.

$$Y_i=F_0(x_i) \quad (i=1,2,\dots,N)$$

задача моделирования заключается в построении (синтезе) модельного оператора F , т.е. в получении оценки F_0 по наблюдениям x_i и y_i . Естественно потребовать, чтобы F был близок к F_0 в смысле некоторого критерия, т.е. $F \sim F_0$.

Существенной особенностью познавательных моделей является отражения механизма объекта или явления в структуре оператора F , т.е. всех причинно - следственных связей, имеющихся у объекта и выявленных в процессе моделирования. При не учете этих связей познавательная сторона модели существенно пострадала бы, так как для познания необходимо представлять не только как, но и почему это происходит. Мы эти вопросы не будем затрагивать.

Другой тип моделирования, который мы будем изучать **связан непосредственно с потребностями управления объектом** и по отношению к управлению имеет вспомогательный характер. Действительно, чтобы управлять, нужно прежде всего знать, чем управляешь, т.е. иметь модель объекта, на которой можно «разигрывать» последствия предполагаемого управления и выбрать наилучшее. Поэтому в процессе моделирования такого рода должна быть создана модель, которая прежде всего обязана удовлетворять потребностям управления.

Следует отметить, что такая модель, синтезированная специально для потребностей управления, может и не отражать внутренних механизмов явления, что совершенно необходимо для познавательной модели. Ей достаточно лишь констатировать наличие определенной формальной связи между входом и выходом объекта.

В связи с этим целесообразно выяснить, что следует подразумевать под понятием «управление» и какие требования оно накладывает на модель управляемого объекта, получаемую в процессе моделирования.

Под управлением будем понимать процесс такого целенаправленного воздействия на объект, в результате которого объект оказывается в определенном смысле «ближе» к выполнению поставленных целей, чем до управления. На рис. 6 показана общая схема управления объектом.

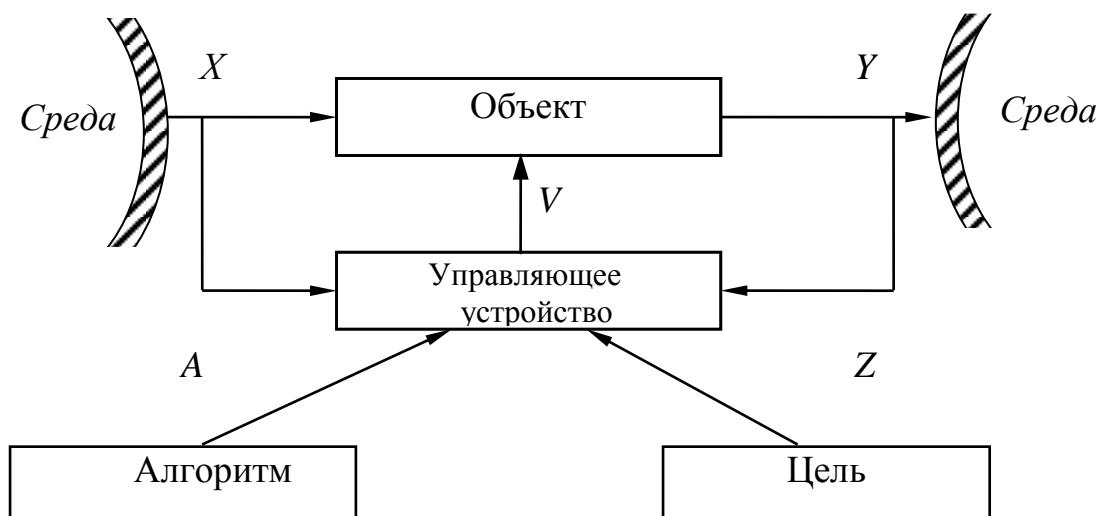


Рис. 6. Общая схема управления объектом.

Здесь X -неуправляемая, но контролируемая составляющая;

U -управляемая составляющая; Y -информация о состоянии объекта, доступная управляющему устройству.

Для синтеза управления необходимо прежде всего определить цель Z , т.е. то, к чему должно «стремиться» управляющее устройство при воздействии на объект, каким должен быть объект с точки зрения управления. Однако, этого мало, необходимого еще иметь алгоритм управления A , который указывает, как достигнуть этой цели.

Таким образом управление реализуется четверкой

$$\langle U, I = \langle X, Y \rangle, A, Z \rangle,$$

где U -управляющее воздействие; $I = \langle X, Y \rangle$ - информация о состоянии среды и объекта; A - алгоритм; Z -цель управления.

Цель Z определяет требования, выполнение которых обеспечивается и организацией управляющего воздействия U с помощью алгоритма A и сбором информации по каналу Y . Не зная, как X и U влияют на состояние Y , т.е. не имея модели $Y = F(X, U)$, нельзя определить управление U , достаточно лишь констатировать наличие определенной формальной связи.

5. КЛАССИФИКАЦИЯ ОБЪЕКТОВ МОДЕЛИРОВАНИЯ

Задача моделирования, как задача построения оператора модели, отражающего качественные и количественные стороны объекта, может быть сформулирована и соответственно решаться по схеме, приведенной на рис 7. Исторически эти подходы возникли независимо друг от друга и в связи с решением различных задач. Классификационных задач, породивших эти методы, удобно произвести на основе понятий динамического и статического объектов.

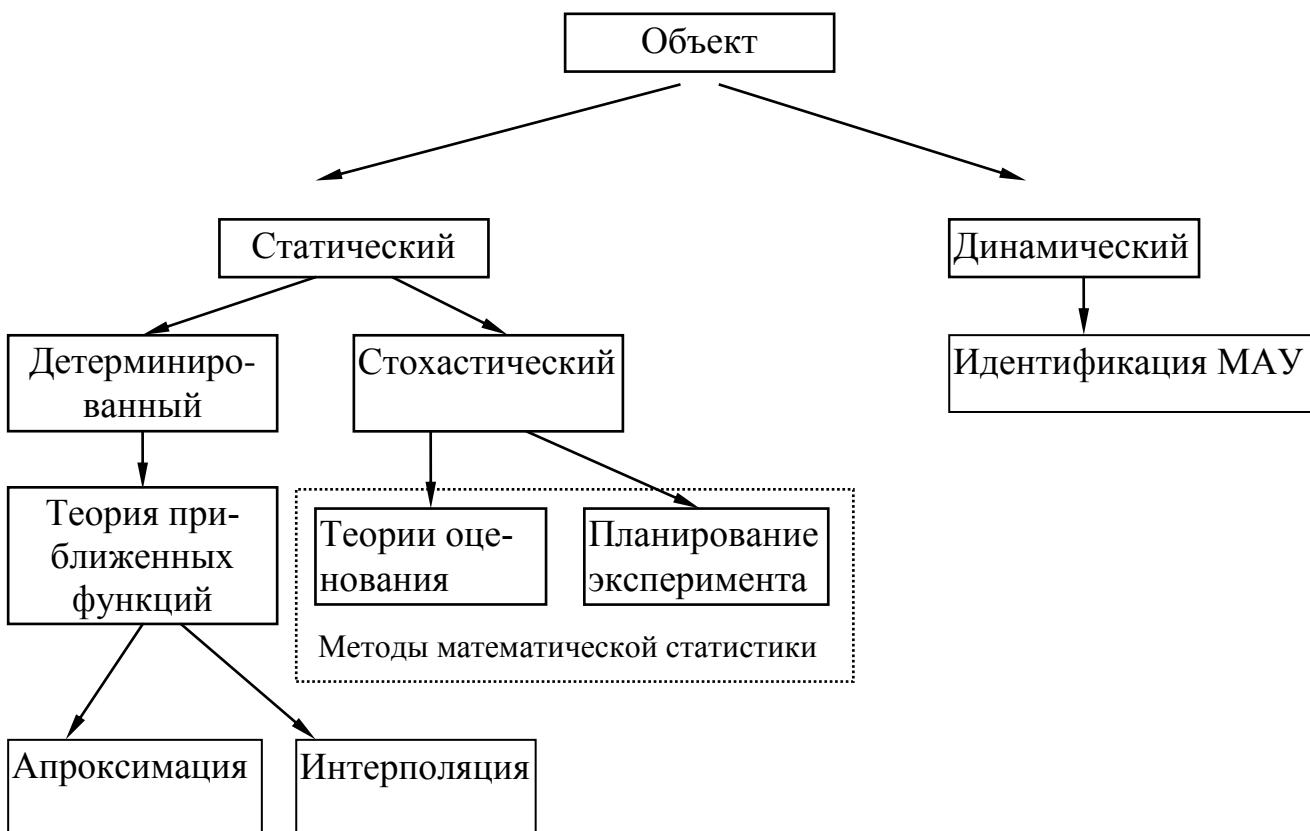


Рис. 7. Классификация объектов моделирования.

Первыми и простейшими объектами, которые были подвергнуты процедуре модулирования, оказались статические детерминированные (не стохастические) объекты, т.е. регулярные функции, связывающие вход и выход объекта. Это обстоятельство и породило первый подход в теории моделирования, который известен в математическом анализе в виде теории приближения функций многочленами и ведет свое начало от работ П.Л. Чебышева. Это направление связано с представлением функции в виде разложения по некоторой системе функций (чаще всего по системе полиномов). В этой теории известны два направления - теория аппроксимации и теория интерполяции. Для идентификации стохастических объектов используются методы математической статистики. В этом направлении известны две теории - теория оценивания и планирования эксперимента. Основной задачей теории оценивания является оценка неизвестных параметров статического объекта по

наблюдениям в обстановке помех и случайных возмущений. В теории планирования экспериментов рассматриваются активные эксперименты в целях определения неизвестных параметров стохастического объекта.

Третьим подходом моделирования динамических объектов являются методы теории систем автоматического управления. В этой теории разработаны специальные методы моделирования динамических объектов управления в режиме нормальной эксплуатации, т.е. в обстановке помех и случайных возмущений.

6. ОСНОВНЫЕ ЭТАПЫ МОДЕЛИРОВАНИЯ

Для моделирования необходимо создать модель и провести ее исследование. Перед созданием модели требуется конкретизировать цели моделирования. После исследования надо выполнить анализ результатов моделирования. Процесс создания модели проходит несколько стадий. Он начинается с изучения объекта и внешних воздействий и завершается разработкой или выбором математической модели или программы для ВС, если моделирование будет проводиться с ее помощью. Некоторые математические модели могут быть исследованы без применения средств вычислительной техники, но в дальнейшем, как правило, рассматриваются такие модели, исследование которых надо проводить с помощью ВС. Таким образом, моделирование на ВС предполагает наличие следующих укрупненных этапов: формулировка целей, изучение объекта, описательное моделирование, математическое моделирование, выбор (или создание) метода решения задачи, выбор или написание программы для решения задачи на ЭВМ, решение задачи на ЭВМ, анализ получаемого решения.

7. ЭТАПЫ МОДЕЛИРОВАНИЯ ОБЪЕКТОВ (ПРОЦЕССОВ, ЯВЛЕНИЙ)

1. Формулировка целей. В основе всякой задачи, проблемы моделирования лежит информация о том, чего собственно добывается, что хочет субъект от объекта, т.е. каковы его цели $\{Z\}$. Именно эта информация определяет объект. Существует своеобразный парадокс: цель определяется объектом, а объект целью. Этот парадокс разрешается довольно просто. Субъект, формулируя цель, всегда имеет какие-то представления об объекте. Эти представления могут быть очень приближенными, но всегда отражают некоторые его свойства, достаточные для эффективной формулировки целей моделирования. Обычно в задачах моделирования цель достигается путем максимизации или минимизации некоторого критерия, задаваемого в виде целевой функции.

2. Изучение объекта. При этом требуется понять происходящий процесс, определить границы объекта с окружающей его средой, если таковые имеются. Кроме того, на данном этапе определяются перечень всех входных и выходных параметров объекта исследования и их влияние на достижение целей моделирования.

3. Описательное моделирование - установление и словесная фиксация основных связей входных и выходных параметров объекта.

4. Математическое моделирование - перевод описательной модели на формальный математический язык. Цель записывается в виде функции, которую обычно называют целевой. Поведение объекта описывается с помощью соотношений, входные и выходные параметры объекта на данном этапе в зависимости от сложности исследуемой проблемы могут возникать ряд задач чисто математического характера. Такими задачами являются задачи математического программирования, линейной алгебры, задачи дифференциального и интегрального исчисления и многие другие.

5. Выбор (или создание) метода решения задачи. На данном этапе для возникшей математической задачи подберется подходящий метод. При выборе такого метода необходимо будет обратить внимание на сложность метода и потребляемые вычислительные ресурсы. Если подходящего метода по предъявленным критериям не окажется, то требуется разработать новый метод решения задачи. Мы делаем упор на разработку новых эффективных методов, не уступающих известным методам по основным вычислительным характеристикам.

6. Выбор или написание программы для решения задачи на ЭВМ. На данном этапе выбирается подходящая программа, реализующая выбранный метод решения. Если такая программа отсутствует, то необходимо написать такую программу.

7. Решение задачи на ЭВМ. Вся необходимая информация для решения задачи вводится в память ЭВМ вместе с программой. С использованием подходящей программы производится обработка целевой информации и получение результатов решения задач в удобной форме.

8. Анализ получаемого решения. Анализ решения бывают двух видов: формальный (математический), когда проверяется соответствие полученного решения построенной математической модели (в случае несоответствия проверяется программа, исходные данные, работа ЭВМ и др.) и содержательный (экономический, технологический и т.п.), когда проверяется соответствие полученного решения тому объекту, который моделировался. В результате такого анализа в модель могут быть внесены изменения или уточнения, после чего весь рассмотренный процесс повторяется. Модель считается построенной и завершенной, если она с достаточной точностью характеризует деятельность объекта по выбранному критерию. Только после этого модель можно использовать при расчетах.

Перечисленные этапы вытекают из общей методологии моделирования систем. При моделировании различных систем трудоемкости одних и тех же этапов могут быть разными. В процессе моделирования конкретной системы могут иметь место некоторые изменения технологии. В частности, может быть заранее предопределен метод моделирования или выбрано конкретное средство моделирования. Математическая модель окажется настолько простой, что не потребуется проведения машинных экспериментов, разработка программной модели исключит необходимость создания математической модели.

Следует обратить внимание на первоочередную необходимость постановки, формулирования цели моделирования. В этом вопросе должно быть достигнуто взаимопонимание между заказчиком, ответственным за создание или модернизацию

цию системы, и разработчиком модели. Важность корректного выполнения этого этапа определяется тем, что все последующие этапы проводятся с ориентацией на определенную цель моделирования.

На этом же этапе конкретизируются, в каких единицах измерения (относительных или абсолютных) должны быть представлены результаты моделирования. Под относительными единицами здесь понимаются качественные градации, сравнительные оценки разных вариантов системы (типа «лучше—хуже», «больше—меньше»). При необходимости представления результатов в абсолютных единицах должен быть решен вопрос о точности измерения. Этот вопрос зачастую не имеет однозначного ответа, но крайне важен для выполнения всех этапов моделирования.

Проверка адекватности указана выше в виде одного из этапов моделирования. Не надо это понимать буквально, так как на адекватность модели оказывает влияние качество выполнения практически всех этапов. Поэтому проверка адекватности должна проводиться в том или ином виде, начиная от разработки концептуальной модели и кончая анализом результатов моделирования.

Под разработкой математической модели подразумевается создание полностью формализованного описания динамики функционирования системы. Однако не для всех систем, внешних условий и целей моделирования может быть подобран известный метод формализации или конструктивный математический аппарат. Тем не менее и для таких систем следует разработать однозначные зависимости выходных характеристик от параметров и воздействий для каждой составляющей системы, алгоритмы взаимодействия между составляющими, логические условия изменения состояний.

Результаты машинного моделирования должны быть проанализированы с целью проверки их достоверности и выработки рекомендаций о способах повышения качества исследуемой системы. На всех этапах моделирования следует обратить особое внимание на документирование принимаемых решений, допущений, ограничений и выводов.

Из организационных аспектов моделирования следует выделить необходимость непосредственного участия в работе квалифицированных представителей заказчика на первых этапах (вплоть до разработки математической модели) и на этапе анализа результатов моделирования. Ответственный за систему заказчик должен четко понимать цели моделирования, разработанную концептуальную модель, программу исследований и полученные результаты. Это будет способствовать внедрению выработанных рекомендаций.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. В чем заключается сущность моделирования?
2. Роль и место моделирования в процессах познания.
3. Какие разновидности моделей используются в исследованиях систем?
4. Классификация моделей.
5. Определение модели.
6. Определение математической модели.
7. Роль моделирования в задачах управления.

-
- 8. Классификация объектов моделирования.
 - 9. Основные этапы моделирования объектов (процессов, явлений).

ЛИТЕРАТУРА

- 1.
- 2.
- 3.
- 4.
- 5.

ЛЕКЦИЯ 2. ТЕХНОЛОГИЯ МОДЕЛИРОВАНИЯ (2 ЧАСА)

ПЛАН

1. Создание концептуальной модели
2. Подготовка исходных данных
3. Разработка математической модели
4. Выбор метода моделирования

1. СОЗДАНИЕ КОНЦЕПТУАЛЬНОЙ МОДЕЛИ

Определение и ориентация. В процессе разработки модели можно условно выделить такие этапы описания, как концептуальный, математический и программный. На этих этапах создается соответствующая модель.

Концептуальная (содержательная) модель — это абстрактная модель, определяющая состав и структуру системы S_o , свойства элементов и причинно-следственные связи, присущие исследуемой системе и существенные для достижения цели моделирования. В концептуальной модели обычно в словесной форме приводятся сведения о природе и параметрах элементарных явлений исследуемой системы, о виде и степени взаимодействия между ними, о месте и значении каждого элементарного явления в общем процессе функционирования системы.

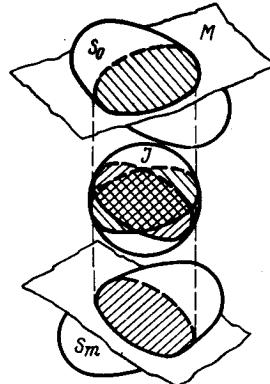


Рис. 1. Отображение оригинала S_o и модели S_m в сознании исследователя

Первоначально концептуальная модель системы S_o возникает в сознании исследователя j (рис. 1). Модель ориентируется на выявление определенных свойств системы в соответствии с целями моделирования. Для этого исследователь делает как бы мысленный срез системы в «плоскости» той метасистемы M , в качестве элемента которой представляет интерес система S_o , т. е. выполняет M -ориентацию. Затем исследователь выявляет основные признаки ориентированной модели и может добавить некоторые признаки и условия, которые облегчат исследование модели или позволяют представить ее в виде некоторого среза моделирующей системы S_m . Концептуальная модель — это субстрат системы с позиций достижения целей моделирования.

Разработка концептуальной модели требует достаточно глубоких знаний системы S_o , так как надо обосновать не только то, что должно войти в модель, но и

то, что может быть отброшено без существенных искажений результатов моделирования. Последнее является наиболее проблематичным, поскольку возникает замкнутый круг: для точного определения влияния исключения какого-либо элемента или явления из модели на степень искажения результатов необходимо создать и исследовать две модели — с учетом и без учета этого элемента или явления. Выполнить это по каждому сомнительному элементу и явлению не представляется возможным в связи со значительным увеличением объема работ.

Основная проблема при создании модели заключается в нахождении компромисса между простотой модели и ее адекватностью с исследуемой системой. Имеются теоретические проработки решения данной проблемы, но практически их трудно реализовать. Поэтому разработчик модели, руководствуясь своими знаниями системы, оценочными расчетами, опытом, должен принять решение об исключении какого-то элемента или явления из модели без достаточно полной уверенности в том, что это не внесет существенных погрешностей в результаты моделирования. Процесс создания концептуальной модели, очевидно, никогда не может быть полностью формализован. Именно в связи с этим иногда говорят, что моделирование является не только наукой, но и искусством. При создании ориентированной модели уточняются множества полезных и возмущающих внешних воздействий.

Стратификация. Следующим шагом на пути создания концептуальной модели служит выбор уровня детализации модели.

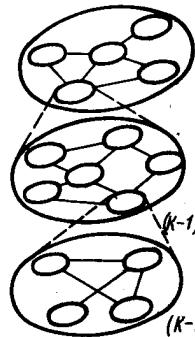


Рис. 2. Уровни модели

Известно, что любая система, в том числе и вычислительная, — это прежде всего целостная совокупность элементов. Непременным свойством каждой системы является ее членимость. Модель системы представляется в виде совокупности частей (подсистем, элементов). В эту совокупность включаются все части, которые обеспечивают сохранение целостности системы. Исключение каких-либо элементов из модели не должно приводить к потере основных свойств системы при выполнении функций по отношению к метасистеме.

С другой стороны, каждая часть системы тоже состоит из совокупности элементов, которые, в свою очередь, могут быть расчленены на элементы. С учетом этого проблема выбора уровня детализации может быть разрешена путем построения иерархической последовательности моделей. Система представляется семейством моделей, каждая из которых отображает ее поведение на различных уровнях детализации (рис. 2). На каждом уровне существуют характерные особенности си-

стемы, переменные, принципы и зависимости, с помощью которых описывается поведение системы.

Уровни детализации иногда называются стратами, а процесс выделения уровней—стратификацией. Выбор страт зависит от целей моделирования и степени предварительного знания свойств элементов. Для одной и той же системы могут быть использованы различные страты. Обычно в модель включаются элементы одного уровня детализации — K -страта. Однако может представлять интерес построение модели из элементов разных страт. В том случае, когда общесистемные (функциональные) свойства отдельных элементов мало известны или вызывает затруднение их описание, можно для каждого такого элемента включить в модель его детализированное описание из нижестоящего ($K - 1$)-страта. Некоторые элементы и этого уровня можно расчленить, т. е. использовать их описание из следующего уровня — ($K - 2$)-страта.

При построении ориентированной и стратифицированной концептуальной модели необходимо руководствоваться следующим. В модель должны войти все те параметры системы S_{ok} и, в первую очередь, параметры $\{s_{oj}\}$, допускающие варьирование в процессе моделирования, которые обеспечивают определение интересующих исследователя характеристик Y_{ok} при конкретных внешних воздействиях $\{x_{on}\}$ на заданном временном интервале T функционирования системы. Остальные параметры должны быть, по возможности, исключены из модели.

Детализация. При расчленении системы на элементы можно поступать следующим образом. Функционирование любой системы представляет собой выполнение одного или нескольких технологических процессов преобразования вещества, энергии или информации. Каждый процесс складывается из последовательности элементарных операций. Выполнение каждой элементарной операции обеспечивается определенным ресурсом — элементом. Поэтому в модели должны присутствовать все элементы, которые реализуют выполнение всех технологических процессов. Кроме них в модель могут быть включены элементы, которые служат для управления ресурсами и процессами и для хранения объектов преобразования в промежутках времени между выполнением элементарных операций, а также для хранения информации, необходимой для управления. Применение этого правила требует предварительного определения понятия элементарной операции.

Детализация системы должна производиться до такого уровня, чтобы для каждого элемента были известны или могли бы быть получены зависимости параметров выходных воздействий элемента, существенных для функционирования системы и определения ее выходных характеристик, от параметров воздействий, которые являются входными для этого элемента.

Если по результатам ориентации, стратификации и расчленения получается модель большой размерности, т. е. с большим числом параметров, в частности, с большим числом элементов (несколько сотен или даже тысяч), то ее следует упростить, поскольку с громоздкой моделью работать неудобно. Это можно сделать разными способами изоморфных преобразований модели без снижения степени адекватности, в том числе путем декомпозиции системы на подсистемы, интеграции элементарных операций и соответствующей интеграции элементов, исключе-

ния или усечения второстепенных технологических процессов с исключением обеспечивающих эти процессы элементов.

Локализация. Последующий шаг создания концептуальной модели — ее локализация, которая осуществляется путем представления внешней среды в виде генераторов внешних воздействий, включаемых в состав модели в качестве элементов. При необходимости они дифференцируются на генераторы рабочей нагрузки, поставляющие на вход системы основные исходные объекты — вещество (сырье, полуфабрикаты, комплектующие), энергию для энергетических систем или данные для информационных систем, в том числе для ВС; генераторы дополнительных обеспечивающих объектов и энергии; генераторы управляющих и возмущающих воздействий. Генераторы возмущающих воздействий нарушают процесс функционирования системы (рис. 3).

Приемники выходных воздействий системы обычно не включают в модель. Считается, что результаты функционирования системы,

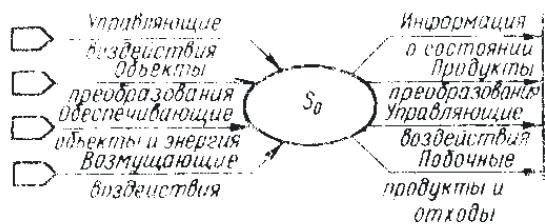


Рис. 3. Локализованная модель

включая основные продукты преобразования, побочные продукты и отходы, информацию о состоянии системы и управляющие воздействия на другие системы, внешняя среда потребляет (принимает) полностью и без задержек.

Структуризация.

Управление.

Завершается построение структуры модели указанием связей между элементами. Связи могут быть подразделены на вещественные и информационные. Вещественные связи отражают возможные пути перемещения продукта преобразования от одного элемента к другому. Информационные связи обеспечивают передачу между элементами управляющих воздействий и информации о состоянии. Отметим, что как информационные, так и вещественные связи не обязательно должны быть представлены в системе некоторым материальным каналом связи. В простых системах, составленных из однофункциональных элементов, имеющих не более чем по одной выходной вещественной связи, информационные связи могут вообще отсутствовать. Управление процессом функционирования в таких системах определяется самой структурой, т. е. в них реализован принцип структурного управления. Примерами таких систем могут служить логические элементы и аналоговые вычислительные машины.

В более сложных системах, включающих многофункциональные элементы или элементы, которые имеют больше чем по одной выходной вещественной связи, имеются управляющие средства (решающие элементы) и соответствующие информационные связи. Управление требуется для указания, какому элементу какой ис-

ходный объект когда и откуда взять, какую операцию по преобразованию выполнить и куда передать. О таких системах можно говорить, что они функционируют в соответствии с программным или алгоритмическим принципом управления. В концептуальной модели должны быть конкретизированы все решающие правила или алгоритмы управления рабочей нагрузкой, элементами и процессами.

Выделение процессов. Рассмотренные выше действия направлены на создание модели, отражающей статику системы — состав и структуру. Поскольку основной интерес представляют динамические системы, следует дополнить эту модель описанием работы системы.

Функционирование системы заключается в выполнении технологического процесса преобразования вещества, энергии или информации. В сложных системах зачастую одновременно протекает несколько технологических процессов. В частности, все современные универсальные ВС рассчитаны на мультипрограммный режим работы. Технологический процесс представляет собой определенную последовательность отдельных элементарных операций. Часть операций может выполняться параллельно разными элементами (ресурсами) системы. Задается технологический процесс маршрутной картой, путевым листом, программой и т. п., другими словами — одним из видов изображения алгоритма.

Алгоритм однозначно определяет, какие ресурсы системы, в какой последовательности и какие операции должны выполнить для достижения некоторого целевого назначения системы. В системах с программным принципом управления, обеспечивающих параллельное выполнение нескольких технологических процессов, имеются алгоритмы управления совокупностью процессов. Их основное назначение заключается в разрешении конфликтных ситуаций, возникающих, когда два или более процесса претендуют на один и тот же ресурс. Совокупность алгоритмов управления A_0 совместно с параметрами входных воздействий X_0 и элементов S_0 отражают динамику функционирования системы.

Обычно алгоритмы преобразовываются к виду A_m , удобному для моделирования. Основные приемы преобразования алгоритмов изложены в гл. 4. Данный подход к описанию динамики работы системы особенно удобен для имитационного моделирования и является естественным способом определения множества характеристик системы:

$$Y = \Phi(X, S, A, T),$$

где Φ — множество операторов вычисления выходных характеристик (здесь и в дальнейшем индексы o , m, k множеств, указывающие на интересующие (k) элементы оригинала (o) и модели (m), опущены в целях упрощения записи).

Отражение состояний. В ряде случаев, в частности для систем со структурным принципом управления, получил распространение другой подход. Для каждого элемента выбирается определенный параметр s (иногда несколько параметров), значение которого изменяется в ходе функционирования элемента и отражает его состояние в текущий момент времени $z(t)$. Множество таких параметров по всем $n = \overline{1N}$ элементам системы $\{z_n\}$ отражает состояние системы $Z(t)$. Функционирование системы представляется в виде последовательной смены состояний: $Z(t_0), Z(t_1), \dots, Z(T)$. Множество $\{Z\}$ возможных состояний системы называют про-

странством состояний. Текущее состояние системы в момент времени $t(t_0 < t \leq T)$ отражается в виде координаты точки в га-мерном пространстве состояний, а вся реализация процесса функционирования системы за время T — в виде некоторой траектории.

Если известно начальное состояние системы $Z^0 = Z(t_0)$, то можно определить ее состояние в любой момент t , принадлежащий интервалу T , когда известна зависимость

$$Z(t) = H(X, Z^0, Z, t).$$

Тогда выходные характеристики определяются по формуле

$$V = G(Z, T). \quad (3)$$

Созданная концептуальная модель должна быть проверена на адекватность исследуемому объекту. Поскольку на данном этапе возможен только умозрительный анализ и эксперимент, желательно, чтобы такую проверку выполняли эксперты, а не разработчик модели.

2. Подготовка исходных данных

Сбор фактических данных. При создании концептуальной модели выявляются качественные (функциональные) и количественные параметры объекта и внешних воздействий X . Для количественных параметров необходимо определить их конкретные значения, которые будут использованы в виде исходных данных при моделировании. Это трудоемкий и ответственный этап работы. Он существенно влияет на успех моделирования. Очевидно, что достоверность результатов моделирования однозначно зависит от точности и полноты исходных данных.

На ранней стадии создания концептуальной модели зачастую выявляется часть параметров, которые определено войдут в модель. По этим параметрам сбор исходных данных можно вести параллельно с разработкой концептуальной модели. По мере уточнения концептуальной модели определяются остальные параметры. Сбор исходных данных осложняется по следующим причинам. Во-первых, значения параметров могут быть не только детерминированными, но и стохастическими. Во-вторых, не все параметры оказываются стационарными. Особенно это относится к параметрам внешних воздействий. В-третьих, всегда идет речь о моделировании несуществующей (проектируемой, модернизируемой) системы или системы, которая должна функционировать в новых условиях.

Большая часть параметров — это случайные величины по своей природе. Однако в целях упрощения модели часто многие из них представляются детерминированными средними значениями. Это можно делать, если случайная величина имеет небольшой разброс, или в случае, когда для достижения цели моделирования достаточно вести расчет по средним значениям. Например, производительность процессора может быть задана определенным количеством операций, выполняемых в единицу времени. Но это количество детерминировано только для определенной смеси операций, которые может выполнять процессор. Подмена в расчетах случайных значений параметров детерминированными величинами должна производиться обдуманно, так как она может привести к погрешностям моделирования.

Под воздействием случайных факторов результаты функционирования системы не только подвергаются рассеиванию, но могут также получить смещение своих средних значений.

При создании модели может иметь место и обратное явление — детерминированные параметры представляются случайной величиной. Делается это при интеграции элементов системы или внешних воздействий с целью сокращения размерности модели. Например, при выполнении программы ВС обрабатывается вполне определенное количество данных. При следующем выполнении этой программы может обрабатываться другое, но тоже определенное количество данных. Для моделирования многократного выполнения программы можно задать всю совокупность количеств данных или подменить это множество значений случайной величиной с определенным законом распределения.

Подбор закона распределения. Для случайных параметров организуется сбор статистики и последующая ее обработка. В процессе обработки выявляется возможность представления параметра некоторым теоретическим законом распределения. Это необходимо в связи с тем, что при определенных законах распределения основных параметров системы и нагрузки появляется возможность создания аналитической модели, а при имитационном моделировании может оказаться проще задать вид закона распределения и основные статистические характеристики, чем представлять случайную величину, например, в виде таблицы.

Процедура подбора вида закона распределения заключается в следующем. По совокупности численных значений параметра строится гистограмма относительных частот — эмпирическая плотность распределения. Гистограмма аппроксимируется плавной кривой. Полученная кривая последовательно сравнивается с кривыми плотности распределения различных теоретических законов распределения. Выбирается один из законов по наилучшему совпадению вида сравниваемых кривых. По эмпирическим значениям вычисляют параметры этого распределения. Затем выполняют количественную оценку степени совпадения эмпирического и теоретического распределения по тому или другому критерию согласия, например, Пирсона (хи-квадрат), Колмогорова, Смирнова, Фишера или Стьюдента. Вопросы подбора вида закона распределения детально разработаны в математической статистике.

Особую сложность представляет сбор данных по случайным параметрам, которые являются функциями времени. В первую очередь такие параметры характерны для внешних воздействий. Пренебрежение фактами нестационарности параметров, которое зачастую имеет место в практике моделирования, приводит к существенным нарушениям адекватности модели.

Аппроксимация функций. Для каждого элемента системы существует функциональная связь между параметрами входных воздействий на этот элемент и его выходными характеристиками. Вид функциональной зависимости для одних элементов бывает очевиден, для других может быть легко выявлен исходя из природы функционирования. Однако для некоторых элементов может быть получена только совокупность экспериментальных данных о количественных значениях выходных характеристик при различных значениях параметров. В этом случае возникает

необходимость ввести некоторую гипотезу о характере функциональной зависимости, т. е. аппроксимировать ее определенным математическим уравнением. Поиск математических зависимостей между двумя или более переменными по собранным опытным данным- может выполняться с помощью методов регрессионного, корреляционного или дисперсионного анализа.

Предварительно для описания определенного элемента вид уравнения задает исследователь. При двух переменных это делается достаточно просто по результатам сравнения графика, на который нанесены экспериментальные точки, с графиками наиболее распространенных аппроксимирующих функций, таких как прямая, парабола, гипербола, экспонента и т. д. Затем методами регрессионного анализа вычисляются константы выбранного уравнения таким образом, чтобы обеспечить наилучшее приближение кривой к экспериментальным данным независимо от того, насколько хорошо выбран вид кривой. Зачастую приближение оценивается по критерию наименьших квадратов.

Для выяснения того, насколько точно выбранная зависимость согласуется с опытными данными, используется корреляционный анализ. Коэффициент корреляции лежит в пределах от 0 до ± 1 , что соответствует изменению степени согласования от полного отсутствия корреляции до случая, когда все экспериментальные точки лежат точно на кривой.

Выдвижение гипотез. По части параметров, которые отражают новые элементы будущей системы или новые условия функционирования, отсутствует возможность сбора фактических данных. Для таких параметров выдвигаются гипотезы об их возможных значениях. Важно, чтобы гипотезы выдвигали эксперты-специалисты, которые достаточно хорошо представляют создаваемую систему или новые внешние воздействия на систему. Большой успех может быть достигнут, если представляется возможность получить сведения от группы специалистов. В этом случае можно уменьшить степень субъективности и воспользоваться хорошо отработанными методиками экспертных оценок. При проведении данной работы определенные сведения можно получить в результате анализа функционирования аналогичных, систем или прототипов будущей системы.

Заканчивается этап сбора и обработки исходных данных классификацией на внешние и внутренние, постоянные и переменные, непрерывные и дискретные, линейные и нелинейные, стационарные и нестационарные, детерминированные и стохастические. Для переменных количественных параметров, которыми может варьировать исследователь в ходе моделирования, определяются границы их изменений, а для дискретных — возможные значения.

3. РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Обобщенные модели. Концептуальная модель и количественные исходные данные служат основой для разработки математической модели. Создание математической модели преследует две основные цели: 1) дать формализованное описание структуры и процесса функционирования системы для однозначности их по-

нимания; 2) попытаться представить процесс функционирования в виде, допускающем аналитическое исследование системы.

Разработка единой методики создания математических моделей, очевидно, не представляется возможной. Это обусловлено большим разнообразием классов систем. Системы могут быть статические и динамические, со структурным или программным управлением, с постоянной или переменной структурой, с постоянным (жестким) или сменным (гибким) программным управлением. По характеру входных воздействий и внутренних состояний системы подразделяются на непрерывные и дискретные, линейные и нелинейные, стационарные и нестационарные, детерминированные и стохастические. При исследовании ВС может быть получено такое же разнообразие моделей в зависимости от ориентации, а также от степени стратификации и детализации.

Для определенных классов систем разработаны формализованные схемы и математические методы, которые позволяют описать функционирование системы, а в некоторых случаях—выполнять аналитические исследования.

Средствами формализованного описания процессов функционирования систем с программным принципом управления служат определенные языки и системы имитационного моделирования. Некоторые из них описаны в книге.

Агрегативные системы. Одной из наиболее общих формализованных схем является описание в виде агрегативных систем. Этот метод позволяет представить функционирование непрерывных и дискретных, детерминированных и стохастических систем. Он в наибольшей мере приспособлен для описания систем, у которых характерно представление входных и выходных воздействий в виде «сообщений», составленных из совокупностей «сигналов».

В основе метода лежит понятие агрегата как элемента системы. Математическая модель агрегата выражается в виде зависимостей с конкретизацией входных воздействий, состояний и операторов переходов и выходов. В частности, выделяют особые состояния агрегата, к которым относятся состояния в моменты получения входного или управляющего сигнала либо выдачи выходного сигнала. Из особого состояния агрегат скачкообразно может переходить в новое состояние. Агрегативная система образуется при расчленении системы на элементы, каждый из которых представляет собой агрегат.

Единообразное математическое описание исследуемых объектов в виде агрегативных систем позволяет использовать универсальные средства имитационного моделирования.

Кусочно-линейные агрегаты. Дальнейшая конкретизация структуры пространств состояний, входных и выходных воздействий, а также операторов переходов и выходов приводит к понятию кусочно-линейных агрегатов, удобных для формализации широкой совокупности разнообразных процессов и явлений материального мира. В основе подхода лежит кусочно-линейный закон изменения состояния системы, что обеспечивает простоту вычисления опорных моментов времени и, как следствие, простоту реализации модели кусочно-линейного агрегата и системы, составленной из таких агрегатов. В частных случаях для кусочно-линейных агрегативных систем результаты могут быть получены аналитическим методом.

Совместно с формализованным описанием системы в виде совокупности кусочно-линейных агрегатов может применяться метод управляющих последовательностей. Суть метода заключается в том, что функционирование системы определяется управляющими последовательностями, которые имеют определенный физический смысл, а также алгоритмами, описывающими управление системой с помощью введенных последовательностей. Управляющие последовательности и алгоритмы позволяют составлять рекуррентные соотношения для описания функционирования кусочно-линейного агрегата.

Стохастические сети. Для описания стохастических систем с дискретными множествами состояний, входных и выходных воздействий, функционирующих в непрерывном времени, широко используются стохастические сети. Стохастическая сеть представляет собой совокупность систем массового обслуживания, в которой циркулируют заявки, переходящие из одной системы в другую.

Большая группа языков имитационного моделирования основана на формализованном представлении систем в виде стохастических сетей. При определенных условиях стохастическая сеть может рассматриваться как совокупность независимых систем массового обслуживания. Это открывает возможность применения достижений теории массового обслуживания для проведения аналитического моделирования.

Системы массового обслуживания. В основе системы массового обслуживания лежит понятие прибора, который может выполнять конечное множество операций. Прибор выполняет операцию, когда возникает заявка — требование на выполнение операции. Если прибор выполняет любую операцию, то считается, что он занят (работает), в противном случае прибор свободен. Ограничение числа состояний прибора приводит к большей степени абстрактности, чем понятие агрегата.

Временная последовательность заявок называется потоком заявок. Общий поток заявок может состоять из нескольких потоков. В случаях независимости потоков, случайных моментов поступления или завершения обслуживания заявок в системе могут возникать очереди. Очередь — это заявки, ожидающие обслуживания, когда прибор занят. Прибор может состоять из нескольких элементов (каналов), каждый из которых способен обработать любую заявку. Совокупность прибора, потоков заявок и очередей к нему называют системой массового обслуживания (СМО).

Теория массового обслуживания хорошо разработана. Поэтому она нашла широкое применение для создания математических моделей, в частности, при моделировании ВС. Применение теории марковских процессов и теории диффузионных процессов для исследования СМО при определенных ограничениях и допущениях позволило получить ряд важных аналитических зависимостей.

Непрерывные детерминированные системы. Если в модели системы не учитывается воздействие случайных факторов, а операторы переходов и выходов непрерывны (это означает, что малые изменения входных воздействий приводят к такого же порядка малым изменениям выходного воздействия и состояния системы), то состояния системы и выхода соответственно могут быть представлены в виде дифференциальных уравнений

$$\frac{dz}{dt} = h(z(t), (t), t); \\ y = g(z(t), t),$$

где h, g — вектор функции состояний и выходов соответственно;

x, z, y — векторы входных воздействий, состояний и выходных воздействий соответственно.

В случае линейности таких систем, когда операторы переходов и выходов обладают свойствами однородности и аддитивности, вид уравнений (4) и (5) упрощается, что дает возможность аналитического решения или исследования известными методами с помощью вычислительных машин.

Построение математических моделей непрерывных линейных детерминированных систем в виде дифференциальных уравнений используется при анализе функционирования элементов и электрических цепей ВС.

Автоматы. Рассмотренные выше формализованные математические схемы применимы для систем, функционирующих в непрерывном времени. Системы, состояния которых определены в дискретные моменты времени t_0, t_1, t_2, \dots , получили название автоматов. Если за единицу времени выбран тakt $\Delta t = t_i - t_{i-1}$, то просто пишут: 0, 1, 2, В каждый дискретный момент времени, за исключением $\langle e \rangle$ в автомат поступает входной сигнал $x(t)$, под действием которого автомат переходит в новое состояние в соответствии с функцией переходов

$$z(t) = \varphi_a(z(t-1), x(t))$$

и выдает выходной сигнал, определяемый функцией выходов

$$y(t) = \psi_a(z(t-1), x(t)).$$

Если автомат характеризуется конечными множествами состояний z , входных сигналов x и выходных сигналов y , он называется конечным автоматом. Функции переходов и выходов конечного автомата задаются таблицами, матрицами или графиками.

Стochasticеские системы, функционирующие в дискретном времени, можно представлять вероятностными автоматами. Функция переходов вероятностного автомата определяет не одно конкретное состояние, а распределение вероятностей на множестве состояний, а функция выходов — распределение вероятностей на множестве выходных сигналов. Функционирование вероятностных автоматов изучается при помощи аппарата цепей Маркова. Для оценки характеристик систем, представляемых в виде автоматов, могут использоваться аналитические или имитационные методы.

Кроме приведенных математических схем для формализованного описания функционирования систем используются исчисление высказываний, тензорная алгебра, сети Петри, Е-сети и др..

Таким образом, построение математической модели предусматривает анализ концептуальной модели и исходных данных в целях выбора одной из подходящих формализованных схем, подбора необходимых множеств и конкретизации операторов. Если это не удается сделать для всей системы, то формализованные схемы

могут быть применены для описания отдельных элементов, а вся система описывается с использованием программного или структурного подхода.

4. ВЫБОР МЕТОДА МОДЕЛИРОВАНИЯ

Аналитические методы. Разработанная математическая модель функционирования системы может быть исследована различными методами — аналитическими или имитационными. С помощью аналитических методов анализа можно провести наиболее полное исследование модели. В некоторых случаях наличие аналитической модели делает возможным применение математических методов оптимизации. Для использования аналитических методов необходимо математическую модель преобразовать к виду явных аналитических зависимостей между характеристиками и параметрами системы и внешних воздействий. Однако это удается лишь для сравнительно простых систем. Применение аналитических методов для более сложных систем связано с большей по сравнению с другими методами степенью упрощения реальности и абстрагирования. Поэтому аналитические методы исследования используются обычно для первоначальной грубой оценки характеристик всей системы или отдельных ее подсистем, а также на ранних стадиях проектирования систем, когда недостаточно информации для построения более точной модели. Они могут использоваться для анализа параллельных процессов в сложных системах.

Ряд аналитических моделей не поддается аналитическим решениям известными математическими методами. Для их исследования могут быть использованы численные методы. Они применимы к более широкому классу систем, для которых математическая модель представляется в виде системы уравнений, допускающей решение численными методами. Использование численных методов особенно эффективно с помощью быстродействующих ВС. Для исследования ВС, функционирование которых описывается марковскими процессами, разработано, например, программное средство для автоматизированного составления уравнений и их решения на ВС. Результатом исследования систем численными методами являются таблицы значений искомых величин для конечного набора значений параметров системы и нагрузки.

Если полученные уравнения не удается решить аналитическими или численными методами, то прибегают к **качественным методам**. Качественные методы позволяют в ряде случаев оценить асимптотические значения искомых величин, устойчивость, а также судить о поведении траектории системы в целом. Перечисленные свойства относятся к поведению отдельных траекторий. Рассматриваются и такие качественные свойства, которые характеризуют поведение совокупностей траекторий. Примером такого свойства является непрерывность, наличие которой говорит о том, что при малых изменениях параметров характеристики системы также мало изменяются. Следует отметить, что для сложных систем важность качественных методов возрастает.

Имитационные методы. Имитационное моделирование является наиболее универсальным методом исследования систем и количественной оценки характе-

ристик их функционирования. При имитационном моделировании динамические процессы системы-оригинала подменяются процессами, имитируемыми в абстрактной модели, но с соблюдением таких же соотношений длительностей и временных последовательностей отдельных операций. Поэтому метод имитационного моделирования мог бы называться алгоритмическим или операционным. В процессе имитации, как при эксперименте с оригиналом, фиксируют определенные события и состояния или измеряют выходные воздействия, по которым вычисляют характеристики качества функционирования системы.

Имитационное моделирование позволяет рассматривать процессы, происходящие в системе, практически на любом уровне детализации. Используя алгоритмические возможности ВС, в имитационной модели можно реализовать любой алгоритм управления или функционирования системы. Модели, которые допускают исследование аналитическими методами, также могут анализироваться имитационными методами. Все это является причиной того, что имитационные методы моделирования становятся основными методами исследования сложных систем.

Методы имитационного моделирования различаются в зависимости от класса исследуемых систем, способа продвижения модельного времени и вида количественных переменных параметров системы и внешних воздействий.

В первую очередь можно разделить методы имитационного моделирования дискретных и непрерывных систем. Если все элементы системы имеют конечное множество состояний и переход из одного состояния в другое осуществляется мгновенно, то такая система относится к системам с дискретным изменением состояний, или дискретным системам. Если переменные всех элементов системы изменяются постепенно и могут принимать бесконечное множество значений, то такая система называется системой с непрерывным изменением состояний, или непрерывной системой. Системы, у которых имеются переменные того и другого типа, считаются дискретно-непрерывными. У непрерывных систем могут быть искусственно выделены определенные состояния элементов. Например, некоторые характерные значения переменных фиксируются как достижение определенных состояний. При моделировании ВС на системном уровне их зачастую удобно рассматривать как системы с дискретным изменением состояний.

Одним из основных параметров при имитационном моделировании является модельное время, которое отображает время функционирования реальной системы. В зависимости от способа продвижения модельного времени методы моделирования подразделяются на методы с приращением временного интервала и методы с продвижением времени до особых состояний. В первом случае модельное время продвигается на некоторую величину Δt . Определяются изменения состояний элементов и выходных воздействий системы, которые произошли за это время. После этого модельное время снова продвигается на величину Δt , и процедура повторяется. Так продолжается до конца периода моделирования T_m . Шаг приращения времени Δt зачастую выбирается постоянным, но в общем случае он может быть и переменным. Этот метод называют "принципом Δt ".

Во втором случае в текущий момент модельного времени t сначала анализируются те будущие особые состояния — поступление дискретного входного воз-

действия (заявки), завершение обслуживания и т. п., для которых определены моменты их наступления $t_i > t$. Выбирается наиболее раннее особое состояние, и модельное время продвигается до момента наступления этого состояния. Считается, что состояние системы не изменяется между двумя соседними особыми состояниями. Затем анализируется реакция системы на выбранное особое состояние. В частности, в ходе анализа определяется момент наступления нового особого состояния. Затем анализируются будущие особые состояния, и модельное время продвигается до ближайшего. Процедура повторяется до завершения периода моделирования T_m . Данный метод называют «принципом особых состояний», или «принципом δz ». Благодаря его применению экономится машинное время моделирования. Однако он используется только тогда, когда имеется возможность определения моментов наступления будущих очередных особых состояний.

Количественные параметры системы и внешних воздействий могут быть детерминированными или случайными. По этому признаку различают детерминированное и статистическое моделирование. При статистическом моделировании для получения достоверных вероятностных характеристик процессов функционирования системы требуется их многократное воспроизведение с различными конкретными значениями случайных факторов и статистической обработкой результатов измерений. В основу статистического моделирования положен метод статистических испытаний, или метод Монте-Карло.

Особое значение имеет стационарность или нестационарность случайных независимых переменных системы и внешних воздействий. При нестационарном характере переменных, в первую очередь — внешних воздействий, что часто наблюдается на практике, должны быть использованы специальные методы моделирования, в частности метод повторных экспериментов.

Еще одним классификационным параметром следует считать схему формализации, принятую при создании математической модели. Здесь прежде всего необходимо разделить методы, ориентированные на алгоритмический (программный) или структурный (агрегатный) подход. В первом случае процессы управляют элементами (ресурсами) системы, а во втором — элементы управляют процессами, определяют порядок функционирования системы.

Из вышеизложенного следует, что выбор того или иного метода моделирования полностью определяется математической моделью и исходными данными.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под сбором фактических данных для построения модели?
2. Как решается подбор вида закона распределения?
3. Что понимается под аппроксимацией функций?
4. Какие виды средств используются для формализования описания функционирования систем?
5. Что вы понимаете под системой массового обслуживания?

ЛИТЕРАТУРА

- 6.
- 7.
- 8.
- 9.
- 10.

ЛЕКЦИЯ 3. ТЕХНОЛОГИЯ МОДЕЛИРОВАНИЯ (2 ЧАСА)

План

1. Выбор средств моделирования
2. Проверка адекватности и корректировка модели
3. Планирование экспериментов с моделью
4. Анализ результатов моделирования

1. ВЫБОР СРЕДСТВ МОДЕЛИРОВАНИЯ

Технические средства моделирования. После выбора метода моделирования необходимо выбрать технические и программные средства для проведения исследования модели с помощью ВС. В качестве программных средств могут быть использованы процедурно-ориентированные алгоритмические языки, проблемно-ориентированные языки или автоматизированные системы моделирования.

Для исследования моделей применяются универсальные или специализированные ВС. Для проведения аналитического моделирования с помощью универсальных ВС зачастую не предъявляется каких-либо особых требований к техническим средствам. Основным требованием к универсальным ВС, которые используются для имитационного моделирования, является наличие оперативной памяти достаточно большой емкости. Это объясняется тем, что в процессе модельного эксперимента постоянно производятся чередующиеся обращения к параметрам элементов и действий (к атрибутам статических и динамических составляющих), поэтому все они должны находиться в оперативной памяти.

Каждый модельный эксперимент при статистическом моделировании требует существенных затрат машинного времени, поэтому желательно использовать для моделирования высокопроизводительные ВС. Остальные требования к составу и техническим характеристикам универсальных ВС не являются существенными.

К специальным техническим средствам аналитического моделирования относятся аналоговые вычислительные машины, используемые для исследования непрерывных детерминированных систем.

В связи с широким применением имитационного моделирования в различных областях все более актуальными становятся разработка и выпуск специализированных ВС. К таким средствам относятся стохастические машины, машины имитационного моделирования и гибридные моделирующие комплексы. Наиболее мощными специализированными техническими средствами моделирования призваны стать распределенные системы моделирования.

Алгоритмические языки. Для создания программных моделей могут использоваться универсальные процедурно-ориентированные алгоритмические языки высокого уровня такие, как Pascal, Delphi, C++, Java и др. Известны примеры применения алгоритмических языков для составления программ имитационного моделирования ВС. При создании имитационных моделей на языках общего назначения возникает ряд трудностей, не типичных для практики программирования традиционных языков.

онных задач обработки данных. Эти трудности связаны с двумя основными особенностями алгоритмов имитационного моделирования.

Первая особенность заключается в том, что алгоритмы поведения сложных систем относятся к параллельным алгоритмам, т. е. предполагающим выполнение более чем одного преобразования в каждый момент времени. Трудности программирования параллельных алгоритмов состоят в том, что алгоритмические языки ориентированы на описание чисто последовательных процессов. Программная имитация параллельных процессов при использовании языков высокого уровня сводится к Организации псевдопараллельного развития параллельных процессов, что достаточно сложно для программирования.

Вторая особенность состоит в том, что в процессе моделирования необходима обработка данных, объем которых весьма трудно оценить априорно. Это обусловлено динамическим характером имитационных моделей и их направленностью на изучение массовых процессов в системах. При программировании таких алгоритмов первостепенное внимание уделяется динамическому распределению оперативной памяти.

Достоинства применения процедурно-ориентированных языков для составления программ имитационного моделирования состоят в возможности использования стандартного программного обеспечения ВС, написания экономичных по затратам памяти и быстродействующих программ, учета детальных особенностей функционирования моделируемых систем.

Языки моделирования. При создании программ имитационного моделирования возникают задачи, общие для широкого класса моделей. Это — организация псевдопараллельного выполнения алгоритмов; динамическое распределение памяти; операции с модельным временем, отображающим астрономическое время функционирования оригинала; имитация случайных процессов; ведение массива событий; сбор и обработка результатов моделирования. Для облегчения решения этих и некоторых других задач созданы специальные проблемно-ориентированные средства (программные системы), которые называют языками моделирования. Решение перечисленных выше задач осуществляется полностью или частично внутренними средствами языка.

Описательные средства языков моделирования позволяют идентифицировать и задавать параметры моделируемой системы и внешних воздействий, алгоритмы функционирования и управления, режимы и требуемые результаты моделирования. По структуре и правилам программирования языки моделирования подобны процедурно-ориентированным алгоритмическим языкам высокого уровня. Они имеют тот или иной набор операторов, сопровождаемых соответствующими • операндами. Но операторы языков моделирования предопределяют выполнение более сложных процедур, поэтому языки моделирования имеют более высокий уровень по сравнению с уровнем алгоритмических языков, что упрощает составление программ. Языки моделирования следует рассматривать как формализованный базис создания математических моделей.

В настоящее время известно более 500 языков моделирования. Такое множество языков частично обусловлено разнообразием классов моделируемых систем,

методов их формализованного математического описания, целей и методов моделирования. По классу систем языки подразделяются на семейства, ориентированные на моделирование дискретных, непрерывных и комбинированных систем. В отдельное семейство выделяются языки, предназначенные для автоматизированного составления схем соединения блоков аналоговых ЭВМ. Другим классификационным признаком может служить алгоритмический или структурный подход к описанию процессов функционирования систем. Можно подразделить языки и по другим признакам.

Автоматизированные системы моделирования. Желание дальнейшего упрощения и ускорения процесса создания машинных моделей привело к реализации идей по автоматизации программирования имитационных моделей. Создан ряд систем, которые избавляют исследователя от программирования. Программа создается автоматически по одной из формализованных схем на основании задаваемых исследователем параметров системы, внешних воздействий и особенностей функционирования. Исходные данные представляются в той или иной канонической форме или в ходе диалога с ВС. По результатам машинного эксперимента основные выходные данные вычисляются и выводятся автоматически, дополнительные — по указанию исследователя. Такие системы называют еще универсальными автоматизированными имитационными моделями, или генераторами имитационных программ.

Перед исследователями систем, использующими имитационное моделирование, неизбежно возникает задача выбора соответствующих программных средств моделирования. Обилие этих средств, в большинстве своем реализованных на разнотипных ВС, отсутствие исчерпывающей документации, единой методики сравнения существующих систем значительно усложняет решение этой задачи. Усилиями рабочей группы Международной ассоциации по применению математических методов и вычислительных машин в имитационном моделировании разработаны единые классификационные таблицы для представления средств программного обеспечения машинного моделирования, которые позволяют в компактной форме описать различные системы моделирования, особенности их реализации и применения.

Программные и технические средства моделирования выбираются с учетом ряда критериев. Непременное условие при этом — достаточность и полнота средств для реализации концептуальной и математической модели. Среди других критериев можно назвать доступность средств, наличие у исследователя информации о тех или других средствах. Немаловажное значение имеет простота и легкость освоения программных средств моделирования, скорость и корректность создания программной модели, существование методики использования средств для моделирования систем определенного класса.

После выбора языка разрабатывают программную модель. Этот процесс включает разработку алгоритма, конкретизацию форм представления входных данных и результатов, написание и отладку программы. Это важный и трудоемкий этап, но по технологии он практически не отличается от всякого другого программирования и поэтому здесь детально не рассматривается.

2. ПРОВЕРКА АДЕКВАТНОСТИ И КОРРЕКТИРОВКА МОДЕЛИ

Проверка адекватности. Проверка адекватности модели системе заключается в анализе ее соразмерности с исследуемой системой, а также равнозначности системе. Однако модель не должна быть полным отображением системы, иначе теряется смысл ее создания. Адекватность нарушается из-за идеализации внешних условий и режимов функционирования, исключения тех или других параметров, пренебрежения некоторыми случайными факторами. Отсутствие точных сведений о внешних воздействиях, определенных нюансах структуры системы, принятые аппроксимации, интерполяции, предположения и гипотезы тоже ведут к уменьшению соответствия между моделью и системой. Перечисленные и другие факторы могут стать причиной того, что результаты моделирования будут существенно отличаться от реальных.

Естественной простейшей мерой адекватности может служить отклонение некоторой характеристики y_0 оригинала и y_m модели:

$$\Delta y = |y_0 - y_m|$$

или, что лучше, отношение отклонения к характеристике оригинала

$$\frac{\Delta y}{y_0} = \left| \frac{y_0 - y_m}{y_0} \right|$$

Тогда можно считать, что модель адекватна с системой, если вероятность того, что отклонение Δy не превышает предельной величины Δ , больше допустимой вероятности P_Δ :

$$P(\Delta y < \Delta) \geq P_\Delta$$

Однако практическое использование данного критерия адекватности зачастую невозможно по следующим причинам.

Во-первых, для проектируемых или модернизируемых систем отсутствует информация о значении характеристики Y_0 , а моделируются, как правило, именно такие системы. Можно сравнивать характеристики модели и некоторой системы-аналога, но тогда будет одинаковая степень недоверия к этим характеристикам. Во-вторых, система оценивается не по одной, а по множеству характеристик, у которых может быть разная величина отклонения. В-третьих, характеристики могут быть случайными величинами и функциями, а часто и нестационарными функциями. Для стохастических систем может оказаться, что статистические характеристики, полученные на модели с высокой степенью адекватности, более точны, чем соответствующие характеристики, вычисленные по результатам измерений на реальной системе. Это объясняется тем, что результаты моделирования определяются по большому числу реализаций, в то время как количества измерений на реальной системе всегда ограничены. В-четвертых, отсутствует возможность априорного точного задания предельных отклонений Δ и допустимых вероятностей P_Δ .

Тем не менее проверять адекватность необходимо, так как по неверным результатам моделирования могут быть приняты неправильные решения. На практике оценка адекватности обычно проводится путем экспертного анализа разумности результатов моделирования. Можно выделить следующие виды проверок:

- проверка моделей элементов (в сомнительных случаях следует детализировать элемент или провести дополнительный анализ);
- проверка модели внешних воздействий (принятые предположения, аппроксимации и гипотезы необходимо оценить математическими методами);
- проверка концептуальной модели функционирования системы (выявляются ошибки постановки задачи);
- проверка формализованной и математической модели;
- проверка способов измерения и вычисления выходных характеристик; выявляются ошибки решения;
- проверка программной модели (анализируется соответствие операций и алгоритмов функционирования программной и математической модели, проводятся контрольные расчеты при типовых и предельных значениях переменных, выявляются инструментальные ошибки программирования).

Некоторые приемы проверок адекватности конкретных моделей приведены в работах.

Корректировка модели. Если по результатам проверки адекватности выявляется недопустимое рассогласование модели и системы, возникает необходимость в корректировке или калибровке модели. При этом могут быть выделены следующие типы изменений:

глобальные, локальные и параметрические.

Необходимость в глобальных изменениях возникает в случае обнаружения методических ошибок в концептуальной или математической модели. Устранение таких ошибок приводит к разработке новой модели. Локальные изменения связаны с уточнением некоторых параметров или алгоритмов. Они выполняются, например, путем замены моделей компонентов системы и внешних воздействий на эквивалентные, но более точные модели. Локальные изменения требуют частичного изменения математической модели, но могут привести к необходимости разработки новой программной модели. Для уменьшения вероятности таких изменений рекомендуется сразу разрабатывать модель с большей степенью детализации, чем это необходимо для достижения цели моделирования.

К параметрическим относятся изменения некоторых специальных параметров, называемых калибровочными. Для обеспечения возможности повышения адекватности модели путем параметрических изменений следует заранее выявить калибровочные параметры и предусмотреть простые способы варьирования ими.

Стратегия корректировки модели должна быть направлена на первоочередное введение глобальных изменений, затем — локальных и, наконец, параметрических изменений. Общая методика корректировки приведена в работе. Для выработки тактики параметрических изменений большое значение имеет анализ чувствительности модели к вариациям ее параметров.

Завершается этап проверки адекватности и корректировки модели определением и фиксацией области пригодности модели. Под областью пригодности понимается множество условий, при соблюдении которых точность результатов моделирования находится в допустимых пределах.

3. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ С МОДЕЛЬЮ

Стратегическое планирование. Цели моделирования достигаются путем исследования разработанной модели. Исследования заключаются в проведении экспериментов, в результате которых определяются выходные характеристики системы при разных значениях управляемых переменных параметров модели. Эксперименты следует проводить по определенному плану. Особую важность приобретает планирование экспериментов при численном и статистическом имитационном моделировании на универсальных ВС. Это обосновывается большим числом возможных сочетаний значений управляемых параметров, а каждый машинный эксперимент проводится при определенном сочетании значений параметров. Например, при пяти управляемых параметрах, каждый из которых может иметь три значения, количество сочетаний параметров равно 243, при десяти параметрах (по пять значений каждого) число сочетаний приближается к 10 млн. При ограниченных вычислительных и временных ресурсах обычно не представляется возможным провести все эксперименты. Возникает необходимость в выборе определенных сочетаний параметров и последовательности проведения экспериментов. Это называется стратегическим планированием.

Разработка плана начинается на ранних этапах создания модели, когда выявляются характеристики качества и параметры, с помощью которых предполагается управлять качеством функционирования системы. Эти параметры называют в теории планирования экспериментов факторами. Затем выделяются возможные значения количественных параметров и варианты качественных (функциональных) параметров. Их называют уровнями q . При этом число сочетаний

$$N_s = q_1 \times q_2 \times \dots \times q_k,$$

где k — число факторов.

Если число факторов велико, то для проведения исследований системы используется один из методов составления плана по неполному факторному анализу. Эти методы хорошо разработаны в теории планирования экспериментов и частично рассмотрены в книге. Особую важность приобретает тщательное планирование экспериментов при исследовании нестационарных систем в связи с необходимостью существенного увеличения общего количества экспериментов.

Тактическое планирование. Совокупность методов уменьшения длительности машинного эксперимента при обеспечении статистической достоверности результатов имитационного моделирования получила название тактического планирования. На длительность одного эксперимента (периода моделирования T_m) влияет степень стационарности системы, взаимозависимости характеристик и значения начальных условий моделирования.

Данные, собранные в эксперименте, можно рассматривать как временные ряды, состоящие из замеров определенных характеристик. Ряд замеров характеристики u может рассматриваться как выборка из стохастической последовательности. Если эта последовательность стационарна, ее среднее \bar{u} не зависит от времени. Оценкой \bar{u}_N является среднее по временному ряду u_1, \dots, u_N . Для эргодической последовательности точность этой оценки возрастает с ростом N .

Если заданы максимальная допустимая ошибка оценки (доверительный интервал) и минимальная вероятность того, что истинное среднее у лежит внутри этого интервала, то существует минимальный размер исследуемой выборки. Этот размер соответствует минимальной длительности эксперимента. Для оценки нескольких характеристик период моделирования определяется по максимальному значению.

Требуемый размер выборки существенно зависит от дисперсии оцениваемой характеристики. Чем больше дисперсия, тем больше должен быть размер выборки. Для коррелированных случайных характеристик следует оценивать дисперсии. Имеются специальные приемы обработки результатов моделирования, которые получили название методов уменьшения дисперсии. Они используют априорную информацию о системе и позволяют уменьшить размер выборки при сохранении заданной точности оценок. К ним относятся методы коррелированных, стратифицированных выборок и др.

Большинство имитационных моделей используются для изучения установившихся равновесных режимов функционирования. Но в начальный период работы системы или модели существует переходный режим даже при неизменных значениях параметров входных воздействий. Как показали исследования, длительность переходного режима может быть весьма большой. Значения выходных характеристик, измеренные в переходный период, смещают их общие оценки.

Существует три основных метода уменьшения ошибки, обусловленной начальными условиями. Первый состоит в достаточном увеличении периода моделирования. С увеличением числа замеров влияние начального смещения на статистическую оценку стремится к нулю. Второй метод состоит в том, чтобы начинать сбор статистики не с начального момента, а по истечении некоторого времени. Третий метод заключается в инициализации модели не с «нулевого», а специально заданного состояния, близкого к установившемуся.

Первые два метода приводят к увеличению длительности эксперимента и не дают гарантии уменьшения ошибки, так как априорно неизвестна длительность переходного режима. Третий метод можно применять при наличии информации о подходящем начальном состоянии. В последующих экспериментах для задания начальных состояний могут использоваться уточненные сведения из предшествующих экспериментов.

При моделировании нестационарных систем установившийся режим может полностью отсутствовать. Естественным методом определения характеристик имитационного моделирования нестационарных систем является метод повторных экспериментов. В этом случае число экспериментов существенно увеличивается, что приводит к особым требованиям по их планированию.

4. АНАЛИЗ РЕЗУЛЬТАТОВ МОДЕЛИРОВАНИЯ

Обработка измерений имитационного эксперимента. При статистическом моделировании в ходе имитационного эксперимента измеряются множества значений по каждой выходной характеристике. Эти выборки необходимо отрабатывать для удобства последующего анализа и использования. Поскольку выходные ха-

теристики зачастую являются случайными величинами или функциями, обработка заключается в вычислении оценок математических ожиданий, дисперсий и корреляционных моментов. Оценки, полученные в результате статистической обработки измерений, должны быть состоятельными, несмещенными и эффективными.

Для того чтобы исключить необходимость хранения в машине всех измерений, обработку проводят по рекуррентным формулам, когда оценки вычисляют в процессе эксперимента методом нарастающего итога по мере появления новых измерений.

Для стохастических характеристик можно построить гистограмму относительных частот — эмпирическую плотность распределения. С этой целью область предполагаемых значений характеристики Y разбивается на интервалы. В ходе эксперимента по мере измерений определяют число попаданий характеристики в каждый интервал и подсчитывают общее число измерений. После завершения эксперимента для каждого интервала вычисляют отношение числа попаданий характеристики к общему числу измерений и длине интервала. Для построенной гистограммы можно попытаться подобрать теоретический закон распределения. Делается это так же, как и при подготовке исходных данных моделирования.

Если искомая характеристика является стационарной случайной функцией времени $y(t)$ и обладает свойством эргодичности, то для ее оценки вычисление среднего по времени заменяется вычислением среднего по множеству измерений при одном достаточно продолжительном эксперименте.

Для случайных нестационарных характеристик период моделирования T разбивается на отрезки с постоянным шагом T_m (прогоны или сечения), и запоминаются значения характеристики в конце каждого прогона. Проводится серия экспериментов с разными последовательностями случайных параметров модели. Затем измерения каждого сечения обрабатываются как при оценке случайных величин. В книге рассмотрена методика этих вычислений.

Процессы обработки измерений имитационного эксперимента направлены на получение интегральных характеристик, т. е. на сжатие данных.

Определение зависимостей характеристик от параметров системы. По результатам статистического моделирования может быть проведен анализ зависимостей характеристик от параметров системы и внешних воздействий. Для этого можно воспользоваться корреляционным, дисперсионным или регрессионным методами.

С помощью корреляционного анализа можно установить наличие связи между двумя или более случайными величинами. Оценкой связи служит коэффициент корреляции при наличии линейной связи между величинами и нормальном законе их совместного распределения. Коэффициент корреляции, равный единице по абсолютной величине, свидетельствует о наличии функциональной нестохастической линейной связи между анализируемыми величинами. При равенстве нулю коэффициента корреляции связь отсутствует. Промежуточные значения коэффициента корреляции соответствуют наличию линейной связи с рассеянием или нелинейной корреляции.

Дисперсионный анализ можно использовать для установления относительного влияния различных факторов на значения выходных характеристик. При этом общая дисперсия характеристики разлагается на компоненты, соответствующие рассматриваемым факторам. По значениям отдельных компонентов делают вывод о степени влияния того или другого фактора на анализируемую характеристику.

Когда все факторы в эксперименте являются количественными, можно найти аналитическую зависимость между характеристиками и факторами. Для этого используются методы регрессионного анализа. Найденная зависимость называется эмпирической моделью. Регрессионный анализ заключается в том, что выбирается вид соотношения между зависимыми и независимыми переменными, по экспериментальным данным вычисляются параметры выбранной зависимости и оценивается качество аппроксимации экспериментальных данных моделью. Если качество неудовлетворительное, берется зависимость другого вида, и процедура повторяется.

К анализу результатов моделирования можно отнести задачу анализа чувствительности модели к вариациям ее параметров. Под анализом чувствительности понимают проверку устойчивости характеристик процесса функционирования системы к возможным отклонениям значений параметров.

Анализ результатов моделирования позволяет уточнить множество информативных параметров модели, что может привести к существенному изменению первоначального вида концептуальной модели; найти функциональные зависимости характеристик и параметров, что иногда дает возможность создать аналитические модели системы, или определить весовые коэффициенты критерия эффективности.

Использование результатов моделирования. В конечном счете результаты моделирования используются для принятия решения о работоспособности системы, для выбора лучшего проектного варианта или для оптимизации системы. Решение о работоспособности принимается по тому, выходят или не выходят характеристики системы за установленные границы при любых допустимых изменениях параметров. При выборе лучшего варианта из всех работоспособных вариантов выбирается тот, у которого максимальное значение критерия эффективности. Наиболее общей и сложной является оптимизация системы: требуется найти такое сочетание значений переменных параметров системы или рабочей нагрузки из множества допустимых, которое максимизирует значение критерия эффективности:

$$E_{opt} = \max(\min) E(y)$$

при соблюдении ограничений на все n характеристик

$$y_{i_{\min}} \leq y_i \leq y_{i_{\max}}, i = 1, \dots, n$$

Если выходная характеристика y_i , является случайной величиной с некоторой плотностью распределения $f(y_i)$, целесообразно ввести в задачу оптимизации стochастические ограничения следующего вида:

$$P(y_{i_{\min}} \leq y_i \leq y_{i_{\max}}) = \int_{y_{i_{\min}}}^{y_{i_{\max}}} f(y_i) dy_i \geq \beta_i,$$

где β_i — минимально допустимая вероятность того, что конкретные значения y , не выйдут за ограничивающие пределы.

Для нестационарных систем выходная характеристика y_i зачастую является случайной функцией с плотностью распределения, изменяющейся во времени $f(y_i, t)$. В этом случае можно задать ограничения следующим образом:

$$P(y_{i_{\min}} \leq y_i \leq y_{i_{\max}}) = \frac{1}{T} \int_0^T \int_{y_{i_{\min}}}^{y_{i_{\max}}} f(y_i) dy_i dt \geq \beta_i,$$

где T — длительность анализируемого периода функционирования системы.

Приведенная постановка задачи оптимизации стохастических систем допускает выход за установленные границы не только отдельных значений характеристики y_i , но и ее математического ожидания $M[y_i, t]$ в пиковые моменты даже при величинах β_i , близких к единице. Это показано на рис. 1. для случая ограничения y_i только сверху.

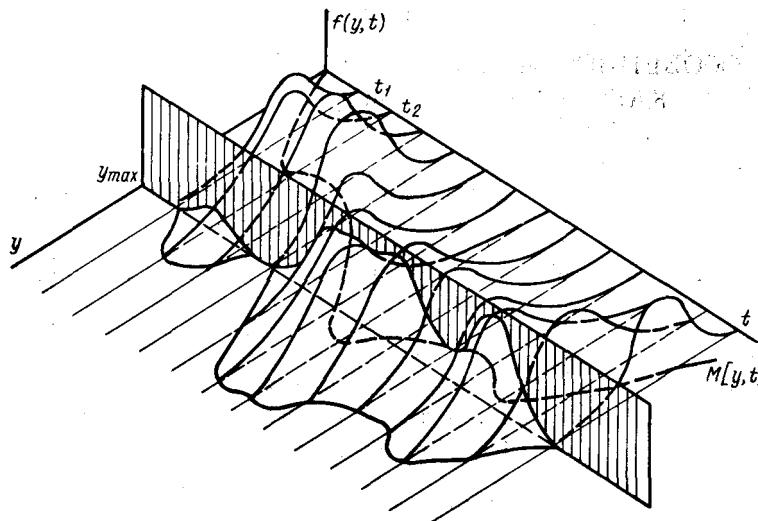


Рис. 1. Ограничение на выходную характеристику, представленную нестационарным случайным процессом

При статистическом моделировании вычисление приведенного двойного интеграла, которое часто оказывается невыполнимой, задачей в случае использования других методов исследования, сводится к вероятностной оценке, которая выполняется путем подсчета числа реализаций y_i , не выходящих за допустимые значения, и его деления на величину выборки за время эксперимента T_m

В случае большого количества сочетаний независимых переменных поиск оптимального варианта требует использования специальных процедур, например, методов математического программирования, в частности, метода наискорейшего спуска и др. Значительно сократить затраты машинного времени на моделирование позволяет проведение зависимых, экспериментов. Применение этих методов следует учитывать на этапе планирования экспериментов.

После создания системы целесообразна апостериорная проверка результатов моделирования и измерения характеристик функционирования. Такая проверка помогает уточнить модель и повысить эффективность системы. Наличие, модели действующей системы дает возможность прогнозирования качества функционирования при развитии системы или изменении внешних воздействий.

Контрольные вопросы

1. Какие виды методов используются для исследования сложных систем?
2. Предмет имитационного моделирования.
3. Что вы понимаете под средствами моделирования и какие виды знаете?
4. Что вы понимаете под проверкой адекватности модели?
5. Что означает планирование экспериментов с моделью?

ЛИТЕРАТУРА

- 11.
- 12.
- 13.
- 14.
- 15.

ЛЕКЦИЯ 4. ПРОБЛЕМА МОДЕЛИРОВАНИЯ (2 ЧАСА)

План

1. Объект моделирования
2. Сведения об объекте
3. Априорная информация
4. Апостериорная информация

При постановке и решении проблемы моделирования исследователь сталкивается с различными вопросами, одним из основных является вопрос – что называть объектом моделирования и какие общие свойства, черты модели объекта? Поэтому проблему моделирования начнем с изучения объекта моделирования.

1. ОБЪЕКТ МОДЕЛИРОВАНИЯ

Объект моделирования удобно представлять в виде многополюсника, изображенного на рис.1.а, где x_1, \dots, x_n —наблюдаемые входы объекта, e_1, \dots, e_k — его ненаблюдаемые входы; y_1, \dots, y_m —наблюдаемые выходы объекта.

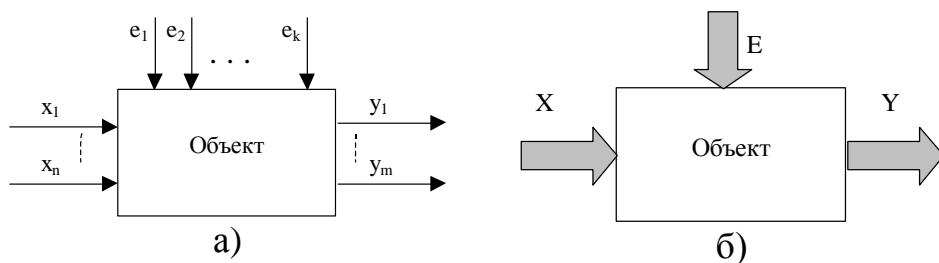


Рис. 1. Изображение объекта моделирования.

Многомерный объект удобно описывать в векторной форме (рис. 1.б), где

$$X = (x_1, \dots, x_n);$$

$$Y = (y_1, \dots, y_m);$$

$$E = (e_1, \dots, e_k);$$

Все входы объекта представляют собой воздействия внешней среды на объект и являются какими-то определенными функциями состояния среды и времени. Так как отсутствует модель среды, воздействующей на объект моделирования, то входы объекта естественно рассматривать как случайные функции времени, т. е.

$$X = X(t), E = E(t),$$

статистические свойства которых в общем случае неизвестны. Однако известны наблюдения входа и выхода объекта, т. е. реализации функций $X(t)$ и $Y(t)$ в непрерывной или дискретной форме. Относительно ненаблюдаемого входа $E(t)$ предполагается известной его структурой, т. е. характер этой случайной функции. В данной книге мы ограничимся случаем, когда $E(t)$ является нормальным случайным процессом, непосредственное наблюдение которого невозможно.

Объект связывает входы X и E с выходом Y некоторым априори неизвестным оператором F_0

$$Y=F_0(X, E).$$

Однако идентифицируется не он, а оператор модели F , связывающий наблюдаемые входы и выходы:

$$Y=F(X).$$

Ненаблюдаемый фактор $E(t)$ рассматривается как случайная помеха, затрудняющая определение оператора F .

Резюмируя, можно сказать, что объект идентификации в общем случае представляется в виде многополюсника, часть входов которого ненаблюдана (это и есть E)

2. СВЕДЕНИЯ ОБ ОБЪЕКТЕ

Все сведения об объекте, которые необходимо иметь для того, чтобы начать процедуру идентификации, как сказано выше, подразделяются на два вида: априорные A и апостериорные B . Так, что двойка

$$\langle A, B \rangle$$

характеризует всю информацию об объекте. Рассмотрим оба вида сведений в отдельности.

3. АПРИОРНАЯ ИНФОРМАЦИЯ

Априорная информация, которой необходимо располагать еще до наблюдения входов и выходов объекта, должна ответить на вопрос, что представляет собой структура идентифицируемого объекта. Структуру прежде всего мы будем характеризовать значениями четырех признаков:

$$A = \langle \alpha, \beta, \gamma, \delta \rangle,$$

которые кодируют объект по четырем признакам. Следует сразу отметить, что структура объекта полностью далёко не исчерпывается этими четырьмя признаками.

Рассмотрим указанные признаки вида объекта подробнее и уточним их смысл.

1. **Признак, динамичности α .** Будем объект называть динамическим ($\alpha=1$), если поведение его выхода зависит не только от значений входа в текущий момент времени, но и от предыдущих значений входа. Это означает, что объект обладает памятью (или инерционностью), которая и определяет зависимость выхода от истории входа.

В противном случае объект будем называть статическим ($\alpha=0$).

2. Признак стохастичности β . Будем объект называть стохастическим ($\beta=1$), если поведение его выхода зависит от неконтролируемых входов объекта или (что то же) сам объект содержит неконтролируемый источник случайных факторов возмущений. В противном случае будем объект называть детерминированным ($\beta=0$).

Заметим, что, строго говоря, нестохастических объектов не существует в природе, так как всякое измерение неизбежно вносит свою погрешность в результат наблюдения. Поэтому правильнее говорить о “малой” и “большой” стохастичности объекта, подразумевая, что с малой стохастичностью можно не считаться и называть такой объект детерминированным.

3. Признак нелинейности γ . Объект будем называть нелинейным ($\gamma=1$), если его реакция на два различных возмущения входа не эквивалентна сумме реакций на каждое из этих возмущений в отдельности. Для случая без помех нелинейность определяется условием

$$F_0(X_1 + X_2) \neq F_0(X_1) + F_0(X_2).$$

При невыполнении этого условия, т. е. при равенстве в этом выражении, объект будем называть линейным ($\gamma=0$).

4. Признак дискретности δ . Будем объект называть дискретным ($\delta=1$), если состояние его входов и выходов изменяется или измеряется лишь в дискретные моменты времени $t=1, 2, \dots, n$. Если же вход и выход изменяются или измеряются непрерывно, то объект назовем непрерывным ($\delta=0$). Таким образом, способ измерения может изменить этот признак объекта.

Как видно, A в значительной степени проясняет вид модели, а для ее полной определенности следует сказать о характере динамики (при $\alpha=1$), вероятностных свойствах стохастичности (при, $\gamma=1$) и виде нелинейности (при $\beta=1$).

Естественно, что представления о виде модели, определяемые A , могут измениться после анализа апостериорной информации, т. е. после наблюдения за поведением входа и выхода объекта.

4. АПОСТЕРИОРНАЯ ИНФОРМАЦИЯ

Если априорная информация A имеет качественный характер, то апостериорная—количественный, т. е. результат (протокол) наблюдений входа и выхода объекта. Этот протокол имеет вид:

$$B = \langle X, Y \rangle,$$

где X — результаты всех измерений 'входов объекта; Y — результаты этих измерений его выходов за тот же период наблюдений.

Для непрерывных объектов ($A=\alpha\beta\gamma 0$) имеем записи непрерывных данных $X=X(t)$, $Y=Y(t)$ в интервале $0 \leq t \leq T$. Таким образом, получаем:

$$B_0 = (\langle X(t), Y(t) \rangle \mid 0 \leq t \leq T).$$

Это означает, что поведение объекта зарегистрировано в виде $n+m$ различных кривых: $x_1(t), \dots, x_n(t)$, $y_1(t), \dots, y_m(t)$ в этом интервале.

Заметим, что X и $X(t)$ в данном случае не тождественны, так как X представляет собой всю зависимость $X(t)$ в заданном интервале, а $X(t)$ может выражать только конкретное значение этой зависимости в момент t . Аналогична не тождественность Y и $Y(t)$.

В дискретном случае ($A=\alpha\beta\gamma l$) имеем $X=(X_1, \dots, X_N)$, $Y=(Y_1, \dots, Y_N)$ и протокол записывается в виде

$$B_1 = (\langle X_i, Y_i \rangle \quad (i=1, \dots, N)),$$

который представляет собой таблицу чисел из $n+m$ столбцов и N строк:

$$B_1 = \begin{vmatrix} x_{11} & x_{21} & \dots & x_{n1} & y_{11} & y_{21} & \dots & y_{m1} \\ x_{12} & x_{22} & \dots & x_{n2} & y_{12} & y_{22} & \dots & y_{m2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \dots & x_{nN} & y_{1N} & y_{2N} & \dots & y_{mN} \end{vmatrix},$$

Очевидна преемственность этих двух форм записи. Так, протокол B_1 может быть получен из B_0 путем фиксации дискретных моментов времени $t=0, \delta, 2\delta, \dots, (N-1)\delta$, где δ - интервал дискретности ($\delta=T/N$).

(Заметим, что обратный переход возможен далеко не всегда.)

Таким образом двойка (1) достаточно полно характеризует объект для целей его идентификации. Она и будет использоваться при изложении соответствующих идей, методов и подходов идентификации.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Общее представление объекта моделирования в виде многополюсника.
2. Априорная и апостериорная информация об объекте моделирования.
3. По каким признакам классифицируются объекты моделирования?

ЛИТЕРАТУРА

- 16.
- 17.
- 18.
- 19.

ЛЕКЦИЯ 5. ЗАДАЧА ИДЕНТИФИКАЦИИ (2 ЧАСА)

ПЛАН

1. Постановка задачи идентификации.
2. Трудности идентификации

1. ПОСТАНОВКА ЗАДАЧИ ИДЕНТИФИКАЦИИ.

Задачей идентификации является определение оператора F_0 объекта, т.е. построения такого оператора модели F , которой был определенном смысле близок к оператору объекта F_0 т.е.

$$F \approx F_0$$

(Заметим, что указанная «близость» весьма относительно, так как оператора F_0 и F могут иметь разные структуру, могут быть сформулированы на разных языках и иметь разные число входов. Именно поэтому близость операторов непосредственно оценить трудно или просто невозможно, тем более что часто об операторе объекта F_0 мало что известно.) В связи с этим естественно оценивать близость операторов по их реакциям на одно и тоже входное воздействие X , т.е. по выходом объекта $Y(t)=F_0[X, E(t)]$ и модели $Y^M=F(X)$. Степень близости этих реакций в каждый момент времени можно оценить, например, значением квадрата модуля разности векторов выхода:

$$q(t) = |Y(t) - Y^M(t)|^2 = \sum_{i=1}^m [y_i(t) - y_i^M(t)]^2, \quad (2)$$

где $Y^M = (y_1^M, \dots, y_m^M)$ векторов выхода модели.

В общем случае близость объекта и модели оценивается так называемой функцией невязки ρ . Это скалярная функция двух векторных аргументов – выходов объекта и модели:

$$\rho(Y(t), Y^M(t)),$$

которая обладает следующими свойствами:

- 1) не отрицательна для любых $Y(t)$ и $Y^M(t)$, т.е.

$$\rho(Y(t), Y^M(t)) \geq 0$$

- 2) равно нулю при $Y(t) \equiv Y^M(t)$, т.е.

$$\rho(Y(t), Y^M(t)) = 0;$$

- 3) непрерывна и выпукла вниз по обоим аргументам, т.е.

$$\begin{aligned} \rho((1-\lambda)Y_1 + \lambda Y_2, Y^M) &\leq (1-\lambda)\rho(Y_1, Y^M) + \lambda\rho(Y_2, Y^M) \\ \rho(Y(1-\lambda)Y_1^M + \lambda Y_2^M, Y^M) &\leq (1-\lambda)\rho(Y, Y_1^M) + \lambda\rho(Y, Y_2^M) \end{aligned} \quad (4)$$

где $0 \leq \lambda \leq 1$.

Говоря проще, эта функция всегда лежит ниже отрезка прямой, соединяющей две любые точки (Y_1, Y^M_1) и (Y_2, Y^M_2) , где Y_i, Y^M_i – произвольные векторы. Удовлетворить этим требованиям не сложно. Так, соотношение (2') соответствует им. Именно оно и будет чаще всего применяться в дальнейшем.

Теперь сформулируем задачу идентификации. Она заключается в том, чтобы построить такой оператор модели F , который бы реагировал на возмущение X аналогично реакции объекта Y . Реакция оператора модели на вход X имеет вид:

$$Y^M = F(X)$$

Следовательно модельный оператор F должен быть таким, чтобы:

$$Y^M \sim Y$$

где \sim знак эквивалентности, т.е. выходы модели и объекта при одинаковых входных воздействиях X должен быть эквивалентны. Этого можно добиться, если ввести единую меру близости на всем интервале наблюдения, а не только в каждой точке, как (3).

Такой мерой в непрерывном случае (объект $A=\alpha\beta\gamma 0$) может быть интеграл

$$Q = \int_0^T \rho(Y(t), Y^M(t)) dt.$$

Действительно, в соответствии с определением функции $\rho(\cdot, \cdot)$ величина Q выражает степень близости функций $Y(t)$ и $Y^M(t)$ в интервале $0 \leq t \leq T$. Значение явно зависит от F :

$$Q(F) = \int_0^T \rho\{Y(t), F(X)\} dt$$

и задача идентификации заключается в ее минимизации путем соответствующего выбора оператора модели F . Если по физическому смыслу задачи важность информации B в различные моменты времени не одинакова, то целесообразно введение переменного веса $h(t) > 0$:

$$Q(F) = \int_0^T \rho\{Y(t), F(x)\} h(t) dt$$

с естественным нормированием

$$\int_0^T h(t) dt = const.$$

Выбор функции $h(t)$ определяется ценностью информации. Например, для стохастического непрерывного объекта ($A=\alpha\beta\gamma 0$) при неравноточных наблюдениях, т.е. когда дисперсия ошибки наблюдения выхода зависит определенным образом от времени

$$\sigma^2 = f(t) \neq 0,$$

где $f(t)$ – заданная функция, вес $h(t)$ должен изменяться следующим образом:

$$h(t) = k / f(t),$$

где, k – нормирующий член, обеспечивающий выполнение условия (6). Это означает, что ценность информации обратно пропорционально уровню дисперсии случайных помех.

Величину $Q(F)$ часто называют невязкой выходов объекта и модели. Эта невязка является функционалом, зависящим от оператора модели F . По своей конструкции эта невязка неотрицательно и равно нулю при $Y(t) \equiv F(X)(0 \leq t \leq T)$, т. е. при совпадение выходов объекта и модели на исследуемом интервале.

Если объект является статическим и непрерывным ($A=0\beta\lambda 0$ т. е., $F(\cdot)$ есть функция то невязка (5) принимает вид:

$$Q(F) = \int_0^T \rho \{Y_i, F(X)\} h(t) dt$$

Для дискретного объекта ($A=\alpha\beta\lambda 1$) функционал невязки записывается в очевидной форме:

$$Q(F) = \sum_{i=1}^N \rho \{Y_i, F(X)\} h_i,$$

а статическим дискретный объект ($A=0\beta\lambda 1$) имеет функционал невязки в виде:

$$Q(F) = \sum_{i=1}^N \rho \{Y_i, F(X_i)\} h_i,$$

где, $h_i > 0 (i = 1, \dots, N; \sum_{i=1}^N h_i = N)$ - вес информация в i -й момент времени. Если объект стохастический и дискретный ($A=\alpha\lambda\lambda 1$) и измерения, например, зашумлены случайной помехой с изменяющейся дисперсией $\sigma_i^2 (i=1, \dots, N)$, то вес следует определять как

$$h_i = k / \sigma_i^2,$$

где k – нормирующий член.

Таким образом степень несоответствия (степень невязки) операторов модели и объекта можно выразить в виде функционалов типов (5) и (7), зависящих явно от оператора модели F .

Естественно, процесс идентификации, т. е. процесс определения оператора модели, строит так, чтобы минимизировать указанную невязку, т. е. решать задачи минимизации функционала $Q(F)$ по оператору F :

$$Q(F) \rightarrow \min_{F \in \Omega} \Rightarrow F^*$$

Эта символическая запись выражает следующую простую мысль: нужно минимизировать функционал $Q(F)$, варьируя оператором (или в простейшим случае

функцией) F не произвольно , а в некотором определенном классе операторов (или функцией) Ω . Это обозначается с отношение $F \in \Omega$, т. е. F принадлежит классу Ω , где Ω – заданный класс операторов или функций. Результатом процедуры минимизации является некоторый оператор (или функция) F^* (не обязательно единственно), обладающий свойством:

$$Q^* = Q(F^*) = \min_{F \in \Omega} Q(F),$$

т. е. невязка Q^* на этом операторе минимально (точнее, не превышает всех возможных невязок, которые можно получить в классе Ω).

Говоря еще проще, для идентификации в заданном классе надо найти оператор F , минимизирующий функционал невязки $Q(F)$ на этом классе.

Утверждения, что идентификация всегда сводиться к операции отыскания минимума, естественно, преувеличено. Действительно легко представить себе статический объект , который идентифицируется путем решения системы линейных или в общем случае нелинейных уравнений. Однако утверждения о сведении задания идентификации к задаче минимизации имеют общий характер для всех случаев идентификации с любыми классами допустимых операторов и функций . Таким образом, использование процедуры минимизация для решение задачи идентификации объектов является принципиальным и важным обстоятельством, свойственным обычно решению сложных задач идентификации.

2. ТРУДНОСТИ ИДЕНТИФИКАЦИИ

Отметим две трудности постановки и решения задачи идентификации.

Первая трудность заключается в определении класса оператора Ω , в котором ищется это решение. Преодоление этой трудности едва ли в настоящая время возможно формальным образом.

Действительно, на стадии определение класса Ω должна быть использована априорная информации об объекте как предмета идентификации для целей управления. Этот этап крайне трудно формализуем и нуждается в эвретических решениях. Пока такие решения может принимать только человек.

Для принятия решения о классе Ω необходимо учесть следующее:

- 1) структур объекта управления ;
- 2) механизм работы объекта ;
- 3) цель управления ;
- 4) алгоритм управления.

Последние два пункта связывают класс Ω с будущим управлением, для которого и идентифицируется объект.

Вторая трудность, которую нужно преодолевать при идентификации, заключается в решении поставленной задачи минимизации (8) с наименьшим ущербом для потребителя . дело в том, что процесс решения всякий задачи связан с

определенными потерями (времени, средств, оборудование, энергии и т. д.). Это обстоятельство накладывает определенные ограничение на выбор алгоритма идентификации.

Действительно, это алгоритм должен решать поставленную задачу в определенном смысле наилучшим образом. Например, задачи идентификации должна быть решена за минимальное время или затраты средств на ее решение должны быть минимальные и т. д.

Как видно, всегда должен быть определен критерий эффективности процесса решения задачи идентификации. Чаще всего это будет ущерб, наносимый в процессе решение задачи идентификации, т. е. потери на идентификацию. Очевидно, что эти потери зависят от сложности задачи (8), необходимого объема экспериментальных данных и способа решения задачи, т. е. от алгоритма минимизации функционала $Q(F)$.

Обозначим через A – алгоритм решения задачи идентификации (9), а I – потери на идентификацию, которые представляют собой функционал, зависящий от алгоритма A . Очевидно, что алгоритм следует выбирать таким, чтобы потери на идентификацию I были минимальны. Отсюда очевидным образом следует задача определения алгоритма как задача минимизации:

$$I\left\{ \left[Q(F) \rightarrow \min_{F \in \Omega} \right], A \right\} \rightarrow \min_{A \in \Xi} \Rightarrow A^* \quad (10)$$

где $I\{B, A\}$ – потери на идентификацию, т. е. на решение задачи $B = [Q((F) \rightarrow \min_{F \in \Omega})]$ с помощью алгоритма A . Этот функционал должен быть задан. К примеру это может быть временно или стоимость решения задачи идентификации, сложность ее программирования, сложность применяемой при этой операторе, ее амортизация в процессе работе и т.д.

Задача (10) формулируется следующим образом: нужно в классе Ξ найти алгоритме идентификации A^* , которой минимизирует потери на идентификацию I заданного объекта. Такой алгоритм A^* естественно называть оптимальном в указанном смысле. Класс Ξ алгоритмов идентификации при этом должен быть задан. Если множество Ξ состоит из конечного и не слишком большого числа алгоритмов т. е.

$$\Xi = \{A_1, \dots, A_k\},$$

$$\min_{i=1, \dots, k} I[B, A_i] = I[B, A^*] \Rightarrow A^*$$

Заметим сразу, что, как правило, для идентификации выбирается не оптимальный алгоритм A^* , а некоторый рациональный алгоритм \tilde{A}^* , который обеспечивает решения задачи при допустимых потерях на идентификацию. Пусть I – допустимые потери. Тогда задача отыскание рационального алгоритма сводится к следующей:

$$I\left[\left(Q(F) \rightarrow \min_{F \in \Omega} \right), A \in \Xi \right] \leq \tilde{I} \Rightarrow \tilde{A}^* \quad (11)$$

Любой алгоритм из множества Ξ , удовлетворяющий этому неравенству, следует считать рациональным.

Таким образом, вторая трудность решение задачи идентификации сводится к отысканию алгоритма решения этой задачи. При этом искомый алгоритм не может быть любым и должен удовлетворять определенным требованиям – оптимальности (10) или рациональности (11). Не последнюю роль здесь играет задание класса алгоритмов идентификации Ξ . процесс определение класса Ξ является эвристическим и пока доступным лишь человеку.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Определение задачи идентификации.
2. Как ставиться задача идентификация?
3. Что понимается под функцией невязки выходов объекта и модели?
4. Сведение задачи идентификация к задаче оптимизации?
5. Какие трудности возникают в задачах идентификации?

ЛИТЕРАТУРА

- 20.
- 21.
- 22.
- 23.
- 24.

ЛЕКЦИЯ 6. ЗАДАЧИ ИДЕНТИФИКАЦИИ И КЛАССИФИКАЦИЯ МЕТОДОВ ИДЕНТИФИКАЦИИ (2 ЧАСА)

План

1. Идентификация структуры и параметров объекта
2. Классификация методов идентификации

1. ИДЕНТИФИКАЦИЯ СТРУКТУРЫ И ПАРАМЕТРОВ ОБЪЕКТА

Будем называть структурной идентификацией процесс определения структуры оператора модели F. Если же структура этого оператора F определена или априори известно, то процесс идентификации сводится к определению параметров этой структуры, т. е. задаче более простой чем предыдущая. Назовем ее параметрической идентификацией (иногда первый процесс называет идентификацией широком смысле, а второй – в узком).

Таким образом , идентификация структуры связана прежде всего с предварительном выбором структуры модели, а идентификация параметров – лишь с определением параметров этой модели при заданной структуре. Как видно, первый этап структурной идентификации предшествует второму и часто включает в себя второй как составную часть.

К сожалению, понятия «структур» не имеет четкого определения, хотя по видимому, интуитивно понимается всеми примерно одинаково. Будем под структурой модели понимать вид оператора с точностью до его коэффициентов. Заметим, что структура объекта , кодируемая A, вообще говоря, может не совпадать со структурой модели . Так, стохастический свойства объекта обычно не отражаются модели, а лишь определяют выбор метода идентификации ее параметров. Кроме того, модель может заведомо иметь меньше входов и выходов, чем их имеет объект. Это часто делает при малом объеме наблюдений (иначе не определить параметры модели).

Теперь уточним задачу идентификации. В (8) проблема сформулирована в самом общем виде, когда идентифицируется и структура и параметры модели. Пусть структуры и модели известна , т. е. задача структурной идентификации решена. Тогда оператор F(x) может быть представлен в виде

$$F(x)=f(x, c),$$

где $f(., .)$ – заданный оператор, а $C=(c_1, \dots, c_k)$ –вектор неизвестных параметров модели. В этом случае задача идентификации параметров модели может быть записана, вообще говоря в виде задачи минимизации функции (а не функционала) невязки:

$$Q(C) \rightarrow \min_{C \in R^r} \Rightarrow C^*$$

решением которой является вектор $C^*=(c^*_1, \dots, c^*_k)$. Здесь

$$Q(C) = \int_0^T \rho[Y(t), f[X(t), C]]h(t)dt$$

функция невязки выходов объекта и модели ; R^k – k- мерное евклидово пространства векторов С . Здесь трудности решения задачи заключается в организации эффективного процесса минимизации заданных функций многих (к) переменных. Заметим, что так, как структура модели известна, то число переменных k определено заранее.

Очень часто структуру можно закодировать, введя структурный параметры. Такими структурными параметрами является числа k и l в примере 2. В общем случае обозначим эти параметры вектором.

$$D=(d_1, \dots, d_q),$$

Это означает, что структура кодируется q величинами d_1, \dots, d_q . Оператор модели теперь представляется в виде

$$F(X) = f(X, C, D),$$

Где f—заданный оператор. Здесь оператор модели определяется двумя типами параметров структурными D и параметрами объекта С. Функция невязки выходов объекта и модели (5) здесь принимает вид:

$$Q(C, D) = \int_0^T \rho\{Y(t), f[X, C, D]\}h(t)dt$$

Тогда задачи идентификации в широком смысле сводится к решению следующий задачи минимизации функции k+q переменных:

$$Q(C, D) \rightarrow \min_{\substack{C \in R_k \\ D \in S}} \Rightarrow C^*, D^*. \quad (13)$$

Здесь S – область определения структурных параметров.

В заключение отметим, что сведение общей задачи идентификации (8) к параметрической идентификации (12) и (13) естественно имеет условный характер. Целью такого представления является упрощения задачи и сведение ее к известно ранее с хорошо разработанными методами решения. Такой задачей является задача математической программирования: минимизация функции многих переменных, принадлежащих заданному множеству. Именно так мы сформулируем задачи параметрических идентификации.

Однако не следует думать, что такое сведение задачи идентификации к задаче математического программирования решает все проблемы идентификации. Здесь возникает ряд новых проблем, например как это сведение сделать в конкретном случае как решить полученную задачу минимизации? Эти проблемы порождают другие и т. д. Но связь идентификации с математическим программированием, отмеченную выше, следует всегда иметь виду.

2. КЛАССИФИКАЦИЯ МЕТОДОВ ИДЕНТИФИКАЦИИ

Будем различать методы идентификации по трем классификационным признакам и характеризовать метод значениями этих признаков:

$$C = \langle \xi, \eta, \zeta \rangle,$$

которые кодируют метод. Здесь ξ , η , ς – структурные признаки, которые могут принимать два значения. Естественно, что структура метода никак не исчерпывается этими тремя признаками. Тройка (14) служит, скорее, для обозначения метода, чем для его описания. Рассмотрим и охарактеризуем эти признаки.

1. Признак активности. ξ . Будем метод идентификации называть активным ($\xi=1$), если при его реализации возможно задавать и изменять определенным образом состояния входов объекта и т. е. как бы изменять состояние среды. Это типичное управление объектом, но для достижения целей идентификации. **Если объект не позволяет управлять состоянием его входа, то метод его идентификации мы будем называть пассивным ($\xi=0$)**, т. е. опирающийся на данные β , полученные в режиме нормальной эксплуатации объекта.

2. Признак адаптивности η . **Если информация β о поведении объекта используется в процессе идентификация не сразу, а по мере ее поступления или циклически и при этом значения идентифицируемых параметров корректируется на каждом шаге или непрерывно, то такой метод будем называть адаптивным.** В противном случае метод будем называть неадаптивным.

Если адаптивный метод параметрической идентификации применяет в реальном масштабе времени использую непосредственно измерения входа и выхода объекта то в этом случае его называют методов самонастраивающейся модели. Суть этого метода состоит в следующем.

В каждый момент времени сопоставляются выходы объекта и модели, при этом квадрат разности выходов минимизируется путем соответствующего выбора параметров с оператора модели. Для повышения эффективности процесса минимизация используется информация о состоянии среды X . Как видно модель таким образом все время подстраивается к объекту, чтобы их реакции на один и тот же вход в каждый момент времени различались минимально.

Адаптивный метод для дискретных объектов всегда описывается рекуррентной формулой вида

$$C_i = I(C_{i-1}, X_i, Y_i),$$

Где C_i – вектор идентифицируемых параметров на i -м шаге адаптации; I – алгоритм адаптации. Выражение (15) удобно записать в виде

$$C_i = C_{i-1} + \Delta C_i$$

Где ΔC_i – приращение, реализуемое алгоритмом адаптации

$$\Delta C_i = \phi(C_{i-1}, X_i, Y_i).$$

В k -мерном пространстве идентифицируемых параметров $C = (c_1, \dots, c_k)$ процесс адаптации иллюстрируется ломаной $C_0 \dots C_{i-1} C_i C_{i+1} \dots$, которая стремится к $C^* = (c_1^*, \dots, c_k^*)$ – точному значению параметров.

Для прерывного объекта ($A = \alpha\beta\gamma 0$) процесс адаптивной идентификации реализуется дифференциальным уравнением

$$\frac{dC}{dt} = \varphi(C, X(t), Y(t)).$$

Однако режим адаптивной идентификации может реализоваться не только по схеме самонастраивающейся модели, т. е. в режиме реального масштаба времени. Если объем наблюдений мал, т. е. малы N (в дискретном случае) и T (в непрерыв-

ном), то однократное использование информации β может не решить задачи идентификации. В этом случае целесообразно образовать цикл
 $B = BBB \dots,$

которые решает поставленную задачу .

Следует отметить такую особенность адаптивного метода. Он почти никогда не решает задачу идентификации абсолютно точно, во всяком случае в пассивном варианте. Но зато он позволяет все время улучшать значения идентифицируемых параметров. По этому его целесообразно применять для идентификации “дрейфующих” объектов параметры которых медленно изменяются. В этом случае адаптивной метод позволяет отслеживать медленные изменения.

3. Признак шаговости ξ . Если идентифицируемые параметры в процессе адаптивной идентификации изменяются дискретно, то такой метод будем называть шаговым ($\xi=1$). В противном случае метод непрерывный ($\xi=0$). Так, адаптивный метод (15) имеет шаговый характер, а (17) – непрерывный.

Как видно, хотя этими тремя признаками метод идентификации описать просто невозможно, они характеризирует структурные особенности метода которые определяются спецификой объекта.

Контрольные вопросы

1. Что понимается под структурной идентификацией?
2. Как решается задача параметрической идентификации?
3. По каким признакам можно классифицировать методы идентификации?
4. Адаптивный метод идентификации.

ЛИТЕРАТУРА

25.
26.
27.
28.
29.

ЛЕКЦИЯ 7. СТРУКТУРНАЯ ИДЕНТИФИКАЦИЯ (2 ЧАСА)

План

1. Выделение объекта из среды
2. Ранжирование входов и выходов объекта (Метод экспертных оценок)
3. Метод непосредственного ранжирования

К задачам идентификации структуры объекта будем относить следующие:

1) выделенные объекта из среды.;

2) ранжирование входов и выходов объекта, по степени их влияния на выполнение целей управления в объекте;

3) определение рационального числа входов и выходов объекта, учитываемых в модели;

4) определения характера связи между входом и выходом модели объекта, т. е. определение вида оператора F (это есть собственно структура объекта.)

Рассмотрим каждую из этих задач в отдельности и укажем возможные пути их решения.

1. ВЫДЕЛЕНИЕ ОБЪЕКТА ИЗ СРЕДЫ

Как указано выше, процесс выделения объекта из среды полностью определяется целями и алгоритмом управления. Это утверждение, однако, нуждается в пояснении. Действительно, если процесс определения объекта зависит от целей, которые будут реализоваться в нем, то естественно спросить, а от чего зависят эти цели? Цель по отношению к управлению имеет внешний характер. Она формулируется на более высоком иерархическом уровне и выражает требования этого уровня, предъявляемые к объекту управления. Однако формулировка и определения множества целей связано с представлениями об объекте, в котором должны реализовать эту цель. Действительно, нельзя эффективно сформулировать цель, если не иметь какой-то модели объекта управления. Поэтому еще до формулировки цели некоторая модель объекта должна быть. Естественно, что эта первая модель очень приближена, но именно она должна лежать в основу определения объекта управления, т. е. процесса выделения объекта из среды.

Пусть цель T в виде множества состояний $T = \{Y^*\}$, которые должны быть достигнуты в процессе управления. Какую именно цель $Y^* \in \{Y^*\}$ необходимо будет реализовать в определенный момент времени определит верхний уровень управления. Это означает, что управляющее воздействие U в этот момент должно быть таким, чтобы состояние объекта Y совпало с заданным целевым, т. е.

$$Y^U \rightarrow Y^*.$$

Однако всякое управляющее воздействие U ограничено определенным ресурсом R . Пусть $\{U\}_R$ – ресурсное множество управлений, которым располагает управлений т. е.

$$U \in \{U_R\}$$

Таким образом, для определения объекта необходимо знать множества целей $\{Y^*\}$ и располагаемых управляющих воздействий $\{U\}_R$.

Рассмотрим теперь процесс выделение объекта из среды как последовательный переход от простейших форм объекта к более сложным.

Простейшей формой объекта назовем такую минимальную часть среды, которая несет информацию, необходимую для проверки выполнимости поставленной цели, причем

$$Y=F(X, U).$$

Если этот объект и среда X таковы, что для любой цели $Y^* \in \{Y^*\}$ и любого состояния среды $X \in \{X\}$ всегда найдется управление $U^* \in \{U\}_R$ такое, что

$$Y^*=F(X, U^*),$$

то этот объект удовлетворяет целям управления и процесс его определения можно считать законченным.

В противном случае, когда ресурса R не хватает для обслуживания всех целей ввиду того, что среда X при этом слишком сильно изменяется, приходится расширять рамки объекта, включая в него ту часть среды, которой можно управлять. Так, как действие X в значительной степени образуется каким-то определенным элементом среды, то естественно попытаться этот элемент присоединить к объекту и управлять им с тем, чтобы изменить в нужную сторону X .

Если сделанное расширение объекта не удовлетворят целям управления, то следует воздействовать на X или X' , т. е. далее расширять объект.

Легко себе представить, что для любого множества заданных целей $\{Y^*\}$ описанный процесс расширения объекта может либо закончиться на каком – то шаге, либо никогда не закончится . В первом случае мы получаем объект управления, во втором – констатируем, что поставленная цель $T=\{Y^*\}$ не достижима вообще, т. е. сформулирована ошибочна , без учета возможностей объекта и среды. Такую цель надо менять или увеличить ресурс R .

Таким образом, процесс расширения объекта при его выделении должен подчиняться следующему принципу: присоединять к объекту лишь те элементы среды, которые действует на реализацию цели и которыми можно эффективно управлять для достижения этой цели.

Заметим, что формальные реализации описанной процедуры выделения объекта возможно лишь в том случае, если для каждого варианта выделения объекта идентификация проводиться до конца т. е. до составления полной модели объекта. Но для этого необходимо располагать измерениями B , которые производятся при выбранной структуре (иначе не известно что именно измерять .) Столь громоздкую процедуру с выходом на эксперимент едва ли в ближайшее время удастся формализовать. По этому на стадии структурной идентификации широко пользуются методом экспертного опроса, по результатам которого и принимаются решения о структуре объекта, его взаимоотношения со средой, возможным состоянием среды и т.д.

2. РАНЖИРОВАНИЕ ВХОДОВ И ВЫХОДОВ ОБЪЕКТА (МЕТОД ЭКСПЕРТНЫХ ОЦЕНОК)

Для определения структуры модели объекта, представленного в виде многополюсника, прежде всего необходимо выяснить, какие именно входы и выходы объекта будут включены в его модель. Для этого прежде всего выявляют всех возможных реальных «претендентов» на роль входов и выходов и из них выделяют наиболее существенные, которые и образуют многополюсник модели $n \times m$. Но здесь сразу возникает вопрос о том, какие входы называть существенными. Ответить здесь однозначный. Так как модель создается для целей управления, то существенным является тот фактор, который наибольшим образом оказывает влияние на осуществление цели управление в объекте, а значит, и в его модели. Это связано с тем, что при формировании структуры системы управления прежде всего необходимо знать, какие воздействия может и будет испытывать объект управления и как результаты этого воздействия связаны с достижением цели управление в объекте. Поэтому при выборе структуры приходится варьировать числом учитываемых входов и оценивать эффективность выбранной комбинации. Для этого следует проранжировать выбираемые факторы. В этом случае выбор наиболее существенных факторов очевиден – они должны быть расположены в начале ранжированного ряда.

Поэтому первым этапом в задаче идентификации объекта следует считать задачу ранжирования его входов. Так как модели объекта еще нет, то ранжирование его можно сделать, например, методом экспертных оценок.

Для этого прежде всего определяются все входы и выходы, состояния которых в какой-то степени можно влиять на выполнение цели в объекте

Как и выше будем различать два вида входных переменных – воздействие среды X и управления U и выход Y . Рассмотрим их подробнее.

Входы $X=(x_1, \dots, x_n)$ подразумеваются прежде всего контролируемыми. В противном случае их не имеет смысла вводить в модель. Отбор «претендентов» на входы X должен подчиняться следующим требованиям:

- 1) отбираются те и только те входы x_1, \dots, x_n , состояние которых влияет на реализацию цели в объекте;
- 2) состояния каждого входов x_i должно эффективно, т. е. легко и надежно, контролироваться (измеряться);

Исходя из этих соображений, составляется ряд «претендентов». Однако это совсем не означает, что все они войдут в модель. Действительно некоторые из них почти не влияют на цель и ими можно пренебречь. А другие хотя и влияют на цель, но трудно измерямы и поэтому также могут быть отброшены.

Однако прежде, чем производить селекцию входов необходимо их проранжировать по степени их влияние на реализацию цели управления в объекте. Это означает что **каждому входу $x_i (i=1, \dots, n)$ следует постановить в соответствие некоторое целое число k_i – его ранг**:

$$x_i \rightarrow k_i (i = 1, \dots, n),$$

где единичный ранг ($k=1$) имеет вход, влияющие наибольшим образом на реализацию цели в объекте. Второй ранг ($k_1=2$) и т. д. имеют входы, влияющие не столь существенно, как единичный. Здесь как видно, индексы при рангах определяют номер ранжированного входа от первого до n -го.

Расположим теперь входы в порядке возрастания их рангов

$$x_{i_1}, x_{i_2}, \dots, x_{i_n},$$

где индекс i_j равен номеру фактора с рангом j . Этот ряд будем называть ранжированным рядом. Здесь на первом месте стоит самый существенный вход, а далее следуют остальные в порядке уменьшения их влияния на цели управления. Теперь, если в модели следует по каким-то соображениям оставить лишь q входов, ими будет факторы с номерами от i_1 до i_q т. е. имеющие первые q рангов.

Составить ранжированную последовательность при отсутствии модели объекта можно например, с помощью специалистов – экспертов, хорошо осведомленных об особенностях среды объекта F_0 а также цели и способах ее достижения, т. е. имеющих представление о будущем алгоритме управления этим объектом.

Таким образом с помощью экспертов составляется последовательность:

$$k_1, k_2, \dots, k_n,$$

где k_i — ранг i -го входа x_i . Построить из нее ранжированный ряд (2) не представляет труда. Например, при $n=5$ последовательность рангов (3) может иметь вид:

$$3, 1, 5, 4, 2.$$

Это означает, что для данной задачи наибольшее влияния на цели управления с учетом возможности измерения имеет второй вход x_2 . Ему приписывается единичный ранг ($k_2=1$). Второй ранг имеет пятый вход ($k_5=2$) и т. д., т. е. $k_1=3$, $k_4=4$, $k_3=5$.

Теперь рассмотрим входы управления $U=(u_1, \dots, u_q)$. Эти входы также нужно проранжировать, учитывая степени их влияния на достижения целей управления и простоту организации изменения этих входов, т. е. простоту реализации управления. Этот последний фактор управляемости очень важен, так как далеко не всегда желание управлять каким-то определенным входом согласуется с возможностями. Поэтому ранжирования входов управления должно производиться экспертами, не только осведомленными об особенностях объекта управления, но знакомыми со способами организации управляющих воздействий.

Аналогично каждому входу управления $u_j (j=1, \dots, q)$ эксперто ставиться в соответствие ранг k_j целое число в интервале $[1, q]$):

$$u_j \rightarrow k_j (j=1, \dots, q),$$

причем единичный ранг соответствует самому существенному и легко изменяемому входу управления, а самому не существенному и самому трудноизменяемо приписываться ранг q .

Выходы объекта также должны быть проранжированы. Здесь критерием может служить количество информации который несет данный выход о близости к реализации целей управления в объекте. Не имя модели объекта, это может сделать эксперто, получить соотношения

$$y_z \rightarrow k_z (z=1, \dots, m),$$

где k_z – ранг выхода y_z .

Как видно, все три случае, по сути дела одинаковы. Нужно путем опроса экспертов-специалистов присвоить определенным параметрам (входа и выхода будущего объекта управления) различные ранги по степени их влияния на один или несколько различных критериев. Это процедура получила названия метода экспертных оценок. Рассмотрим только две модификации этого метода:

- 1) метод непосредственного ранжирования;
- 2) метод парных сравнений.

В первом случае эксперта сразу присваивают ранги фактором, которые представлены для ранжирования. Второй метод использует парное ранжирование факторов, что упрощает задачу эксперта, но требует дальнейшей обработки для получения ранжированного ряда.

3. МЕТОД НЕПОСРЕДСТВЕННОГО РАНЖИРОВАНИЯ

Пусть N экспертов ранжируют n факторов. Каждому фактору каждый эксперт присваивает ранг – целое число от 1 до n . Так, i -му фактору j -й эксперт присваивает ранг k_{ij} . В результате получается матрица $N \times n$ мнений экспертов где номера строк соответствуют номерам экспертом, а номера столбец – номером

$$K = \begin{vmatrix} k_{11} & k_{21} & \dots & k_{n1} \\ \dots & \dots & \dots & \dots \\ k_{1N} & k_{2N} & \dots & k_{nN} \end{vmatrix} \quad (4)$$

ранжируемых факторов. Это означает что j -я строка представляет собой имени j -го эксперта, а i -й столбец – мнений всех экспертов по поводу i -го фактора.

При назначении рангов экспертами нужно соблюдать следующие условия:

1. Сумма рангов, назначенных всем факторам, должна быть одинакова для каждого эксперта и равна:

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}.$$

Это означает, что сумма элементов любой строки матрицы (4)

$$\sum_{i=1}^n k_{ij} = \frac{n(n+1)}{2} \quad j = (1, \dots, N).$$

2. Если эксперт какие-то q факторов считает одинаковыми, то он присваивает им один ранг. Этот ранг равен среднему из q целых рангов, которые получены при условии, что эксперту удалось их проранжировать. Например, равноценность четырех факторов ($q=4$):

x_1, x_2, x_3, x_4 , стоящих на пятом месте в ранжированном ряду, приводит к тому, что их ранги равны:

$$k_1 = k_2 = k_5 = k_6 = \frac{5+6+7+8}{4} = 6,5$$

Как видно, в этом случае ранги могут быть дробными. Как легко убедиться, дробные ранги кратны $1/2$.

Для определения результирующих рангов следует вычислить средние ранги каждого фактора

$$\bar{k}_i = \frac{1}{N} \sum_{j=1}^N k_{ij} (i = 1, \dots, n).$$

Эти ранги и дают возможность проранжировать факторы. На первом месте ставится фактор, имеющий минимальный средний ранг

$$\bar{k}_l = \min_{i=1, \dots, n} \{\bar{k}_i\},$$

т. е. фактор x_l , на втором — фактор, имеющий следующий по малости средний ранг, и т. д. Полученные ранги позволяют построить ранжированный ряд факторов, который и будет осредненным мнением коллектива из N экспертов.

Очевидно, что далеко не всякий результат экспертного опроса следует считать удовлетворительным. Действительно, если эксперты сильно противоречат друг другу (например, половина экспертов фактору x_i присвоила первый ранг, а другая половина — последний), то такое ранжирование не может быть положено в основу решающих процедур. Поэтому для оценки всякого экспертного опроса引进ется критерий, характеризующий согласованность экспертов. Чем выше эта согласованность, тем более можно “верить” результатам экспертного опроса, и наоборот.

Согласованность экспертов удобно определять как степень рассеяния средних рангов $\bar{k}_i (i = 1, \dots, n)$. Действительно, если эксперты полностью согласованы, то средние ранги представляют собой целые, не равные друг другу числа (случай одинаковых рангов пока не рассматриваем).

Если же эксперты полностью не согласованы, то средние ранги примерно равны $(n+1)/2$. В промежуточном случае (при частично согласованных экспертах) ранги сгруппируются вокруг среднего значения $(n+1)/2$.

Вычислим дисперсию средних рангов. Она, по определению, равна:

$$D(\bar{k}) = \frac{1}{n} \sum_{i=1}^n (\bar{k}_i - \bar{\bar{k}})^2,$$

где

$$\bar{\bar{k}} = \frac{1}{n} \sum_{i=1}^n \bar{k}_i = \frac{n+1}{2}$$

— математическое ожидание среднего ранга. Определим максимальную дисперсию (она бывает при полностью согласованных экспертах)

$$D_{\max}(\bar{k}) = \frac{1}{n} \sum_{i=1}^n \left(i - \frac{n+1}{2} \right)^2 = \frac{n^2 - 1}{12},$$

Критерий согласованности экспертов удобно представить в виде отношения

$$W = \frac{D(\bar{k})}{D_{\max}(\bar{k})} = \frac{12}{n(n^2 - 1)} \sum_{i=1}^n \left(\bar{k}_i - \frac{n+1}{2} \right)^2.$$

Легко заметить, $0 \leq W \leq 1$. При $W=0$ эксперты полностью несогласны, а при $W=1$ они высказываются как один, т. е. единогласно. Таким образом, значение W характеризует степень согласованности экспертов.

Чем ближе W к единице, тем более единодушны эксперты и тем более достоверен результат ранжирования. (Следует отметить, что эксперты должны высказывать свое мнение независимо друг от друга, т. е. до ранжирования они не должны знать мнение других экспертов. В противном случае возможно появление корреляции мнений, что повышает критерий согласованности W , хотя в действительности эксперты не столь единодушны).

Для того чтобы знать, велико или мало конкретное значение критерия согласованности, который никогда не бывает равным ни нулю, ни единице, можно предложить следующий подход. Предположим, что m из N экспертов абсолютно компетентны, а остальные $N-m$ совершенно некомпетентны, т. е. принимают свое решение чисто случайно (хотя такого, как правило, не бывает). Тогда дисперсия средних рангов

$$D(\bar{k}) = \frac{mD_{\max} + (N-m) \cdot 0}{N} = \frac{m}{N} D_{\max}.$$

Разделив все на D_{\max} , получим:

$$W = m/N.$$

Это значит, что W выражает долю абсолютно компетентных экспертов. Так, при $W=0,3$ можно считать, что 30% экспертов были вполне компетентны, а остальные 70% принимали свое решение случайно, что, естественно, могло оказаться роковое влияние на окончательную ранжировку (а могло и не оказаться!).

Отсутствие согласованности мнений экспертов может иметь двоякое объяснение. Во-первых, это возможно из-за некомпетентности экспертов, связанной с новизной или слабой изученностью объекта идентификации. Во-вторых сложность объекта затрудняют эксперта в ответах о рангах факторов. Эксперту в этом случае проще сопоставить важность некоторых факторов попарно, т. е. указать, чей ранг одного из двух факторов будет больше. Именно в таких ситуациях обращаются к методу парных сравнений, который мы и рассмотрим ниже.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие задачи структурной идентификации вы знаете?
2. Что понимается под ранжированием входов и выходов объекта?
3. В чем заключается сущность метода непосредственного ранжирования?

ЛИТЕРАТУРА

- 30.
- 31.
- 32.
- 33.
- 34.

ЛЕКЦИЯ 8. ИССЛЕДОВАНИЕ ЗАДАЧ СТРУКТУРНОЙ ИДЕНТИФИКАЦИИ (2 ЧАСА)

План

1. Метод парных сравнений
2. Определение рационального числа входов и выходов объекта, учитываемых в модели
3. Определение характера связи между входом и выходом модели объекта

1. МЕТОД ПАРНЫХ СРАВНЕНИЙ

Эксперту предлагается проранжировать факторы попарно, т. е. каждой паре факторов x_i и x_l поставить в соответствие число q_{il} :

$$q_{il} = \begin{cases} 1, & \text{если } x_i \succ x_l; \\ 0, & \text{если } x_i \sim x_l; \\ -1, & \text{если } x_i \prec x_l; \end{cases}$$

где знак “ \succ ” обозначает предпочтительность. Так, выражение $x_i \succ x_l$ следует читать: “ i -й фактор более предпочтителен при ранжировании, чем l -й”. Знак \sim является знаком эквивалентности факторов с точки зрения ранжирования. Числа q_{il} обладают очевидным свойством

$$q_{il} + 0 = q_{il}.$$

Таким образом, каждый (j -й) эксперт свое мнение представляет в виде таблицы $n \times n$ (табл. 1)

или

$$Q^j = \|q_{il}^j\|,$$

где верхний индекс определяет номер эксперта.

Таблица 2-1

	x_1	X_2	...	x_n
x_1	0	q_{12}^j	...	q_{1n}^j
x_2	q_{21}^j	0	...	q_{2n}^j
⋮	⋮	⋮	⋮⋮⋮	⋮
x_n	q_{n1}^j	q_{n2}^j	...	0

Осредним мнение экспертов. Для этого достаточно построить осредненную таблицу ($n \times n$)

$$\bar{Q} = \|\bar{q}_{il}\|,$$

где

$$\bar{q}_{il} = \frac{1}{N} \sum_{j=1}^N q_{il}^j$$

— среднее предпочтение i -го фактора перед l -м. Это и есть мнение экспертов. Определим согласованность экспертов. В качестве меры согласованности аналогично предыдущему естественно выбрать дисперсию величин \bar{q}_{il} . В силу того, что среднее значение \bar{q}_{il} равно нулю, для дисперсии получаем:

$$D(\bar{q}_{il}) = \frac{1}{n(n-1)} \sum_{i \neq l}^n (\bar{q}_{il})^2,$$

где суммирование производится по всей таблице \bar{Q} . Максимальное значение дисперсии будет иметь место при полной согласованности экспертов и равно:

$$D_{\max}(\bar{q}) = 1.$$

Тогда, вводя критерий согласованности как отношение дисперсии средних предпочтений к максимальной дисперсии, получаем:

$$W = \frac{1}{n(n-1)} \sum_{i \neq l}^n (\bar{q}_{il})^2.$$

Однако эксперты могут противоречить. Примером такой таблицы \bar{Q} может служить табл. 2, где противоречие имеет вид $x_1 \succ x_2 \succ x_3 \succ x_1$, т.е. оказалось что $x_1 \succ x_3$ и $x_3 \succ x_1$ одновременно.

Таблица 2

	x_1	x_2	x_3
x_1	0	1	-1
x_2	-1	0	1
x_3	1	-1	0

Выявление подобных противоречий совершенно необходимо не только в осредненной таблице, но и в мнении каждого эксперта. Это можно сделать на основе следующего, довольно очевидного правила, которое называется правилом транзитивности.

Для предпочтений оно имеет вид:

$$\text{если } x_1 \succ x_2 \text{ и } x_2 \succ x_3, \text{ то } x_1 \succ x_3. \quad (5)$$

Для эквивалентности:

$$\text{если } x_1 \sim x_2 \text{ и } x_2 \sim x_3, \text{ то } x_1 \sim x_3$$

Таблицы, представляющие собой мнение каждого эксперта, должны удовлетворять указанной транзитивности и при обнаружении противоречий возвращаются соответствующему эксперту для разрешения отмеченных противоречий.

Для определения рангов ранжируемых факторов следует определить правило назначения рангов по таблице Q. Таких правил может быть много. Рассмотрим два из них,

Правило 1. Определим суммарные предпочтения каждого фактора $q_i = \sum_{l=1}^n (\bar{q}_{il})$.

Естественно считать, что первый ранг имеет фактор, суммарное предпочтение которого максимально, т. е. при

$$q_z = \max_{i=1,\dots,n} \{q_i\}$$

первый ранг имеет фактор x_z , т. е. $k_z=1$. Аналогично образуются ранги остальных факторов.

Рассмотренное правило, однако, излишне осредняет предпочтения. Так, фактор, имеющий ряд явных предпочтений, которые легко обнаруживаются эксперты, получит первый ранг только потому, что его второстепенность по отношению к другим факторам будет не столь ярко выражена. Именно в этом случае часто приходится обращаться к другому правилу.

Правило 2. Основная мысль этого правила опирается на идею усиления контраста. Для этого вводится порог δ . Если предпочтение выше этого порога, то оно имеет явный характер, а если ниже, то оно сомнительно, т. е. больше соответствует эквивалентности. Получаем следующее преобразование матрицы средних предпочтений \bar{Q} в контрастную матрицу U , которую легко преобразовать в ранжированный ряд:

$$u_{il} = \varphi(q_{il}) \quad (i \neq l = 1, \dots, N),$$

где

$$\varphi(q) = \begin{cases} -1, & \text{если } q \leq -\delta \\ 0, & \text{если } |q| < \delta \\ 1, & \text{если } q \geq \delta \end{cases}$$

Как видно, это преобразование целиком и полностью определяется порогом δ ($0 < \delta < 1$). При $\delta=1$ контрастная матрица U становится нулевой и все факторы эквивалентны. При $\delta=0$ она полностью заполняется единицами, но при этом почти неизбежно появление противоречий в виде нарушений транзитивности предпочтений (5).

Поэтому при выборе порога δ следует помнить, что его увеличение приводит к отказу от ранжирования, а уменьшение — к увеличению числа явных предпочтений и к опасности появления противоречий.

Одной из возможных рекомендаций по определению оптимального порога является выбор порога δ на “пороге противоречий”, т. е. такого значения δ , незначительное уменьшение которого приводит к противоречиям.

2. ОПРЕДЕЛЕНИЕ РАЦИОНАЛЬНОГО ЧИСЛА ВХОДОВ И ВЫХОДОВ ОБЪЕКТА, УЧИТЫВАЕМЫХ В МОДЕЛИ

Описанным выше образом получаются ранжированные ряды всех претендентов на входы и выходы модели:

$$\left. \begin{array}{c} x_1 \succ x_2 \succ \dots \succ x_n \\ u_1 \succ u_2 \succ \dots \succ u_q \\ y_1 \succ y_2 \succ \dots \succ y_m \end{array} \right\}$$

(здесь для удобства произведена перенумерация ранжированных факторов таким образом, чтобы их номер стал равен рангу, т. е. $k_i=i$).

Выбор рациональных чисел n^* , q^* и m^* , характеризующих модель, т. е. размерность ее входов и выходов, следует также производить экспортно. Для этого нужно начать с минимального числа входов и выходов (n_1, q_1, m_1 , т. е. с простейшей модели, например, с $n_1=0; q_1=m_1=1$). Назовем эту модель F_1 [характер связи $Y=F_1(X, U)$, где $Y=(y_1, \dots, y_{m1})$; $X=(x_1, \dots, x_{n1})$; $U=(u_1, \dots, u_{q1})$: при этом выяснить не нужно, модель рассматривается как “черный ящик”]. Таким образом, первая, простейшая модель характеризуется тройкой $F_1=\langle n_1, q_1, m_1 \rangle$. Вторая модель $F_2=\langle n_2, q_2, m_2 \rangle$ должна быть выбрана экспертами из трех:

$$\left. \begin{array}{c} F'_2 = \langle n_1 + 1, q_1, m_1 \rangle \\ F''_2 = \langle n_1, q_1 + 1, m_1 \rangle \\ F'''_2 = \langle n_1, q_1, m_1 + 1 \rangle \end{array} \right\}$$

Здесь мы воспользовались ранжированными рядами (6), что позволяет усложнять объект введением фактора, имеющего очередной ранг.

Эксперты ранжируют тройки (7) по критериям, заранее оговоренным. Тройка, получившая первый ранг, образует F_2 . Аналогично $(i+1)$ -я модель F_{i+1} определяется ранжированием трех полученных из F_i троек:

$$\begin{aligned} F'_{i+1} &= \langle n_i + 1, q_i, m_i \rangle \\ F''_{i+1} &= \langle n_i, q_i + 1, m_i \rangle \\ F'''_{i+1} &= \langle n_i, q_i, m_i + 1 \rangle \end{aligned}$$

Таким образом, получим последовательность моделей объекта F_1, F_2, \dots, F_l , расположенных в порядке их уточнения и усложнения. Теперь остается в этом ряду установить предпочтение, т. с. выбрать ту модель, которая и будет идентифицироваться. Это можно также сделать с помощью экспертов. Пусть в результате получен ряд предпочтений:

$$F_1 \prec F_2 \prec \dots \prec F_z \succ F_{z+1} \succ \dots \succ F_l$$

Это означает, что следует остановиться на модели

$$F_z = (n_z, q_z, m_z)$$

и таким образом $n^*=n_z, q^*=q_z, m^*=m_z$.

3. ОПРЕДЕЛЕНИЕ ХАРАКТЕРА СВЯЗИ МЕЖДУ ВХОДОМ

И ВЫХОДОМ МОДЕЛИ ОБЪЕКТА

Мы уже говорили, что структура модели определяется видом оператора модели F . Этот оператор, прежде всего, характеризуется кодом A . С него и следует начинать.

Определение кода A требует четырех двоичных выборов:

$$\alpha = ?; \beta = ?; \gamma = ?; \delta = ?,$$

где каждый из признаков может принимать одно из двух значений: 0 или 1. Анализ следует начинать с простейшего (нулевого) случая. Действительно, код A построен так, что наличие трудностей в процессе идентификации отражается единицами кода. Так, динамический объект ($\alpha=1$) труднее идентифицировать, чем статический ($\alpha=0$); стохастический ($\beta=1$) труднее детерминированного ($\beta=0$), нелинейный ($\gamma=1$) сложнее линейного ($\gamma=0$) и т. д.

В процессе выбора кода A модели следует иметь в виду, что, учитывая эти трудности, вполне можно намеренно “снизить” модель, т.е. сделать ее значительно проще объекта. Так, поведение заведомо динамического объекта можно описывать статической моделью, если динамика объекта не слишком ярко выражена; нелинейный объект можно аппроксимировать линейным и т. д. Разумеется, что при этом эффективность управления, построенного на основе такой модели, снижается. Но если это снижение невелико, а выигрыш в идентификации значителен, то такой выбор следует считать оптимальным.

После выбора кода A модели следует определить конкретную форму ее оператора F .

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Чем отличается два правила для определения рангов ранжируемых факторов в методе парных сравнений?
2. Как определяется рациональное число входов в выходов объекта?
3. Как определяется характер связи между входом и выходом модели объекта?

ЛИТЕРАТУРА

- 35.
- 36.
- 37.
- 38.
- 39.

ЛЕКЦИЯ 9. АНАЛИТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 ЧАСА)

План

1. Потоки заявок
2. Марковские модели

1. ПОТОКИ ЗАЯВОК

Простейший поток. При аналитическом моделировании характеристики системы вычисляются наиболее просто для потока заявок, называемого простейшим. Простейший поток — это поток заявок, который обладает следующими свойствами: 1) стационарность; 2) отсутствие последействия; 3) ординарность.

Стационарность означает постоянство вероятности того, что в течение определенного временного интервала поступит одинаковое количество заявок вне зависимости от расположения интервала на оси времени.

Отсутствие последействия заключается в том, что поступившие заявки не оказывают влияния на будущий поток заявок, т. е. заявки поступают в систему независимо друг от друга.

Ординарность — это значит, что в каждый момент времени в систему поступает не более одной заявки.

Любой поток, обладающий этими свойствами, является простейшим.

У простейшего потока интервалы времени τ между двумя последовательными заявками — независимые случайные величины с функцией распределения:

$$F(\tau) = 1 - e^{-\lambda\tau}.$$

Такое распределение называется экспоненциальным (показательным) и имеет плотность

$$f(\tau) = \lambda e^{-\lambda\tau},$$

математическое ожидание длины интервала

$$M[\tau] = \int_0^{\infty} \tau f(\tau) d\tau = 1/\lambda,$$

дисперсию

$$D[\tau] = \int_0^{\infty} (\tau - M[\lambda])^2 f(\tau) d\tau = 1/\lambda^2 \quad (4)$$

и среднеквадратическое отклонение, равное математическому ожиданию. **Экспоненциальное распределение характеризуется одним количественным параметром — интенсивностью.**

Простейшие потоки заявок обладают следующими особенностями:

1. Сумма M независимых, ординарных, стационарных потоков с интенсивностями $\lambda_i (i=1, \dots, M)$ сходятся к простейшему потоку с интенсивностью

$$\lambda = \sum_{i=1}^M \lambda_i$$

при условии, что складываемые потоки оказывают примерно одинаковое малое влияние на суммарный поток.

2. Поток заявок, полученный в результате случайного разрежения исходного стационарного ординарного потока, имеющего интенсивность λ , когда каждая заявка исключается из потока с определенной вероятностью p независимо от того, исключены другие заявки или нет, образует простейший поток с интенсивностью $p\lambda$.

3. Интервал времени между произвольным моментом времени и моментом поступления очередной заявки имеет экспоненциальное распределение с таким же математическим ожиданием $1/\lambda$, что и интервал времени между двумя последовательными заявками.

4. Простейший поток создает тяжелый режим функционирования системы, поскольку, во-первых, большее число (63 %) промежутков времени между заявками имеет длину меньшую, чем ее математическое ожидание $1/\lambda$, и, во-вторых, коэффициент вариации,

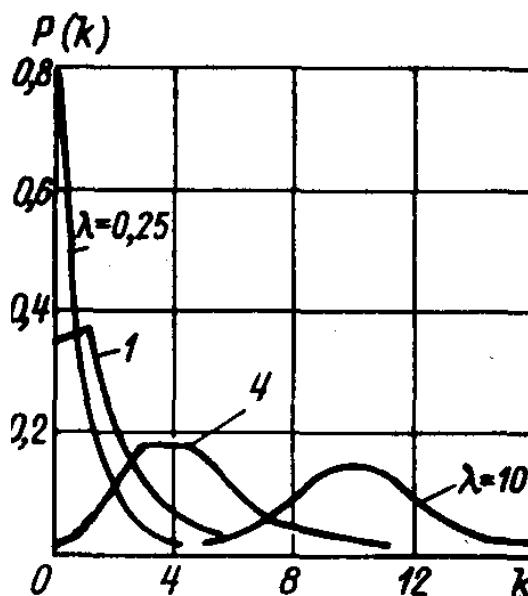


Рис. 1. Распределение Пуассона

равный отношению среднеквадратического отклонения к математическому ожиданию:

$$\vartheta = \frac{\sigma[\tau]}{M(\tau)}$$

и характеризующий степень нерегулярности потока, равен единице, в то время как у детерминированного потока коэффициент вариации $\vartheta = 0$, а для большинства законов распределения $0 < \vartheta < 1$.

Простейший поток имеет широкое распространение не только из-за аналитической простоты связанной с ним теории, но и потому, что большое количество реально наблюдаемых потоков статистически не отличимы от простейшего. Этот эмпири-

ческий факт подтвержден рядом математических моделей, в которых при довольно общих условиях доказывается, что поток близок к простейшему.

Пуассоновский поток. Пуассоновским потоком называется ординарный поток заявок с отсутствием последействия, у которого' число заявок, поступивших в систему за промежуток времени t , распределено по закону Пуассона:

$$P(k, \lambda) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \lambda > 0,$$

где $P(k, \tau)$ — вероятность того, что за время τ в систему поступит точно k заявок; λ — интенсивность потока заявок.

Математическое ожидание и дисперсия распределения Пуассона равны $\lambda \tau$. Вид зависимостей этого распределения при $\tau=1$ для разных λ , показан на рис. 1.

Следует подчеркнуть, что распределение Пуассона дискретно. Стационарный пуассоновский поток является простейшим.

Если нестационарный поток, интенсивность которого представляет собой функцию времени $\lambda=\lambda(t)$, описывается законом распределения Пуассона, то такой поток называется пуассоновским, но не простейшим. В распределении Пуассона длительности интервалов между двумя последовательными заявками — это случайные величины с экспоненциальным распределением.

Эрланговский поток. В общем случае интервалы времени между поступлением заявок могут иметь функцию распределения общего вида $G(\tau)$. Если интервалы независимы, то говорят, что заявки образуют рекуррентный поток или поток с ограниченным последействием.

Поток называется рекуррентным (потоком Пальма), если он стационарен, ординарен, а интервалы времени между заявками представляют собой независимые случайные величины с одинаковым произвольным распределением. Тогда простейший поток рассматривается как частный случай рекуррентного потока. Примером рекуррентного потока может служить поток Эрланга.

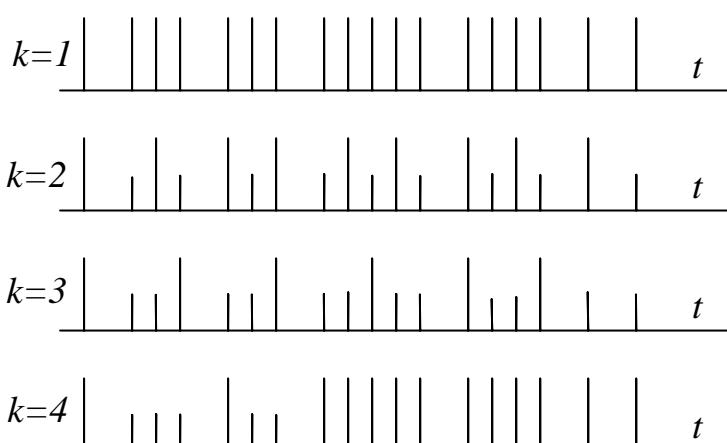


Рис. 2. Потоки Эрланга

Потоком Эрланга k -го порядка называется поток, у которого интервалы времени между моментами поступления двух последовательных заявок представляют собой сумму k независимых случайных величин, распределенных по экспоненциальному закону с параметром λ . Поток Эрланга получается из простейшего путем исключения ($k - 1$) заявок с сохранением каждой k -й заявки (рис. 2). Плотность распределения интервала времени между двумя соседними заявками в потоке Эрланга k -го порядка

$$f(\tau) = \frac{\lambda(\lambda\tau)^{k-1}}{(k-1)!} e^{-\lambda\tau}, k=1,2,\dots$$

Поток Эрланга превращается в простейший при $k = 1$. Приведенные потоки наиболее широко используются в теории массового обслуживания, в том числе при аналитическом моделировании ВС.

2. МАРКОВСКИЕ МОДЕЛИ

Общие сведения. В теории массового обслуживания к наиболее изученным и исследованным относятся модели, у которых случайный процесс функционирования относится к классу марковских процессов, т. е. марковские модели.

Случайный процесс, протекающий в системе, называется марковским, если для любого момента времени вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент и не зависят от того, когда и как система пришла в это состояние.

При исследовании ВС аналитическим моделированием наибольшее значение имеют марковские случайные процессы с дискретными состояниями и непрерывным временем.

Процесс называется процессом с дискретными состояниями, если его возможные состояния z_1, z_2, \dots можно заранее перечислить, т. е. состояния системы принадлежат конечному множеству, и переход системы из одного состояния в другое происходит мгновенно.

Процесс называется процессом с непрерывным временем, если смена состояний может произойти в любой случайный момент.

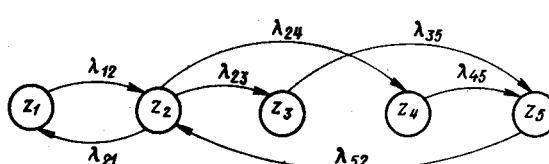


Рис. 3. Пример графа состояний

Описание марковской модели. Для описания поведения системы в виде марковской модели следует определить понятие состояния системы; выявить все состояния, в которых может находиться система; указать, в каком состоянии находится система в начальный момент; построить граф состояний, т. е. изобразить все состояния, например, кружками и возможные переходы из состояния в состояние — стрелками, соединяющими состояния (на рис. 3 выделено пять состояний); разметить граф, т. е. для каждого перехода указать интенсивность $\lambda_{ij}(t)$ потока событий, переводящих систему из состояния z_i в состояние z_j

$$\lambda_{ij}(t) = \lim_{\Delta t \rightarrow 0}$$

где $p_{ij}(t, t + \Delta t)$ — вероятность перехода из состояния z_i в состояние z_j за время от t до $t + \Delta t$.

Для стационарных марковских процессов интенсивности переходов не зависят от времени: $\lambda_{ij}(t) = \lambda_{ij}$, тогда $p_{ij}(t, t + \Delta t) = p_{ij}(\Delta t)$.

Понятие состояния зависит от целей моделирования. В одном случае, например, оно может быть определено по состояниям элементов, каждый из которых может быть "свободен" или "занят" в другом случае состояние системы определяется числом заявок, находящихся на обслуживании и в очередях.

Уравнения Колмогорова для вероятностей состояний. Исчерпывающей количественной характеристикой марковского процесса является совокупность вероятностей состояний, т. е. вероятностей $p_i(t)$ того, что в момент t процесс будет находиться в состоянии z_i ($i = 1, \dots, n$).

Рассмотрим, как определяются вероятности состояний по приведенному на рис. 3 графу состояний, считая все потоки простейшими. В случайный момент времени t система может находиться в одном из состояний z_i с вероятностью $p_i(t)$. Придадим t малое приращение Δt и найдем, например, $p_2(t + \Delta t)$ — вероятность того, что в момент $t + \Delta t$ система будет в состоянии Z_2 . Это может произойти, во-первых, если система в момент t была в состоянии Z_2 и за время Δt не вышла из него; во-вторых, если в момент t система была в состоянии z_1 или z_3 и за время Δt перешла в состояние Z_2 .

В первом случае надо вероятность $p_2(t)$ умножить на вероятность того, что за время Δt система не перейдет в состояния z_1 , z_2 или z_4 . Суммарный поток событий, выводящий систему из состояния z_2 , имеет интенсивность $\lambda_{21} + \lambda_{23} + \lambda_{24}$. Значит, вероятность того, что за время Δt система выйдет из состояния z_2 , равна $(\lambda_{21} + \lambda_{23} + \lambda_{24})\Delta t$. Отсюда вероятность первого варианта

$$p'_2(t + \Delta t) = p_2(t)[1 - (\lambda_{21} + \lambda_{23} + \lambda_{24})\Delta t]$$

Найдем вероятность перехода в состояние z_2 . Если в момент t система находилась в состоянии z_1 с вероятностью $p_1(t)$, то вероятность перехода в состояние z_1 за время Δt равна

$$p''_2(t + \Delta t) = p_1(t)\lambda_{12}\Delta t$$

Аналогично для состояния Z_5

$$p'''_2(t + \Delta t) = p_5(t)\lambda_{52}\Delta t$$

Складывая вероятности $p'_2(t + \Delta t)$, $p''_2(t + \Delta t)$ и $p'''_2(t + \Delta t)$, получим

$$p_2(t + \Delta t) = p_2(t)[1 - (\lambda_{21} + \lambda_{23} + \lambda_{24})\Delta t] + p_1(t)\lambda_{12}\Delta t + p_5(t)\lambda_{52}\Delta t,$$

Раскроем квадратные скобки, перенесем $p_2(t)$ в левую часть и разделим обе части на Δt :

$$\frac{p_2(t + \Delta t) - p_2(t)}{\Delta t} = p_1(t)\lambda_{12} + p_5(t)\lambda_{52} - p_2(t)(\lambda_{21} + \lambda_{23} + \lambda_{24}).$$

Если устремить Δt к нулю, то слева получим производную функции $p_2(t)$:

$$\frac{dp_2(t)}{dt} = p_1(t)\lambda_{12} + p_5(t)\lambda_{52} - p_2(t)(\lambda_{21} + \lambda_{23} + \lambda_{24}).$$

Аналогичные уравнения можно вывести для всех остальных состояний. Получается система дифференциальных уравнений:

$$\left. \begin{aligned} \frac{dp_1(t)}{dt} &= p_2(t)\lambda_{21} - p_1(t)\lambda_{12}; \\ \frac{dp_2(t)}{dt} &= p_1(t)\lambda_{12} + p_5(t)\lambda_{52} - p_2(t)(\lambda_{21} + \lambda_{23} + \lambda_{24}); \\ \frac{dp_3(t)}{dt} &= p_2(t)\lambda_{23} - p_3(t)\lambda_{35}; \\ \frac{dp_4(t)}{dt} &= p_2(t)\lambda_{24} - p_4(t)\lambda_{45}; \\ \frac{dp_5(t)}{dt} &= p_3(t)\lambda_{35} + p_4(t)\lambda_{45} - p_5(t)\lambda_{52}. \end{aligned} \right\} \quad (8)$$

Эта система линейных дифференциальных уравнений дает возможность найти вероятности состояний, если задать начальные условия. В левой части каждого уравнения стоит производная вероятности i -го состояния, а в правой — сумма произведений вероятностей всех состояний, из которых ведут стрелки в данное состояние, на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность i -го состояния.

Представим уравнения Колмогорова в общем виде:

$$\frac{dp_i(t)}{dt} = \sum_{\substack{j=1 \\ j \neq i}}^n p_j(t)\lambda_{ji} - p_i(t)\sum_{\substack{j=1 \\ j \neq i}}^n \lambda_{ij}, \quad i = 1, \dots, n. \quad (9)$$

Здесь учтено, что для состояний, **не имеющих** непосредственных переходов, можно считать $\lambda_{ij} = \lambda_{ji} = 0$

Предельные вероятности состояний. В теории случайных процессов доказывается, что если число n состояний системы конечно и из каждого состояния можно перейти в любое другое за конечное число шагов, то существуют предельные (финальные) вероятности состояний:

$$\lim_{t \rightarrow \infty} p_i(t) = p_i, i = 1, \dots, n. \quad (1)$$

Сумма вероятностей всех возможных состояний равна единице. При $t \rightarrow \infty$ в системе S устанавливается стационарный режим, в ходе которого система случайным образом меняет свои состояния, но их вероятностные характеристики уже не зависят от времени. Предельную вероятность состояния z_i можно трактовать как среднее относительное время пребывания системы в этом состоянии.

Для вычисления предельных вероятностей нужно все левые части в уравнениях Колмогорова (9) положить равными нулю и решить полученную систему линейных алгебраических уравнений:

$$p_i \sum_{\substack{j=1 \\ j \neq i}}^n \lambda_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^n p_j \lambda_{ji}, i = 1, \dots, n. \quad (1)$$

В связи с тем, что эти уравнения однородные, т. е. не имеют свободного члена и, значит, позволяют определить неизвестные только с точностью до произвольного множителя, следует воспользоваться нормировочным условием

$$\sum_{i=1}^n p_i = 1$$

и с его помощью решить систему уравнений.

Модель размножения и гибели. Разновидностью марковской модели с дискретным числом состояний и непрерывным временем является модель размножения и гибели. Она характерна тем, что ее граф состояний имеет вид цепи (рис. 4). Особенность этого графа состоит в том, что каждое из средних состояний z_1, \dots, z_{n-1} связано прямой и обратной стрелками с каждым из соседних состояний — п. z_j правым и левым, а крайние состояния z_0 и z_n — только с одним соседним состоянием.

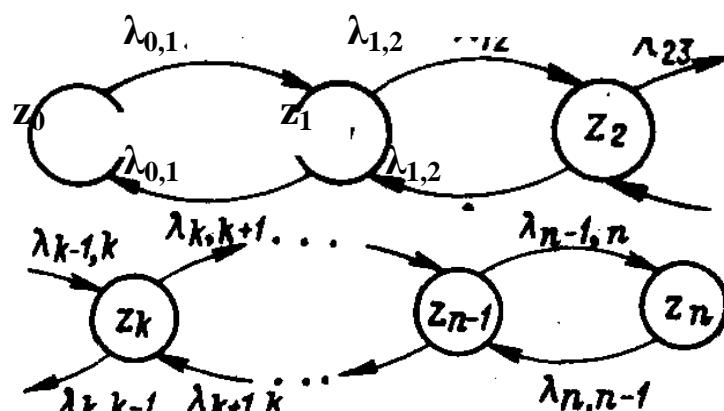


Рис. 4. Граф состояний модели размножения и гибели

В этой модели формулы для определения вероятностей состояний, полученные в результате решения уравнений Колмогорова, имеют вид:

$$p_k = \frac{\lambda_{k-1,k} \dots \lambda_{12} \lambda_{01}}{\lambda_{k,k-1} \dots \lambda_{21} \lambda_{10}} p_0; \quad (14)$$

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12} \lambda_{01}}{\lambda_{21} \lambda_{10}} + \dots + \frac{\lambda_{n-1,n} \dots \lambda_{12} \lambda_{01}}{\lambda_{n,n-1} \dots \lambda_{21} \lambda_{10}} \right)^{-1}$$

Эти формулы часто, используют при решении задач теории массового обслуживания.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Простейший поток и его свойства.
2. Основная характеристика экспоненциального распределения.
3. Пуассоновский поток.
4. Поток Эрланга и его основные свойства.
5. Какой процесс называется Марковским описание Марковской модели?
6. Для чего используется уравнение Колмогорова?
7. Граф состояний модели размножения и гибели и основные формулы.

ЛИТЕРАТУРА

- 40.
- 41.
- 42.
- 43.
- 44.

ЛЕКЦИЯ 10. ХАРАКТЕРИСТИКИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 ЧАСА)

План

1. Характеристики вычислительных систем как систем массового обслуживания
2. Характеристики вычислительных систем как сложных систем массового обслуживания
3. Методы приближённой оценки характеристик вычислительных систем

1. ХАРАКТЕРИСТИКИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ КАК СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

Описание системы. Предположим, что моделью ВС является одноканальная СМО с однородным бесконечным простейшим потоком заявок и неограниченной очередью. Интенсивность потока заявок равна λ . Длительность обслуживания заявки — это случайная величина с математическим ожиданием μ . Наряду с понятием средней длительности обслуживания используется понятие интенсивности обслуживания μ — величины, обратной $\mu: \mu = 1/\mu$, и характеризующей число заявок, которое может обработать прибор в единицу времени.

Поток обслуживания тоже будем считать простейшим с интенсивностью μ . В соответствии с символикой, принятой в теории массового обслуживания, такая система обозначается $M/M/1$.

Выделим состояния СМО по числу заявок, находящихся в системе:

- z_0 — прибор свободен, очереди нет;
- z_1 — прибор занят (обслуживает заявку), очереди нет;
- z_2 — прибор занят, одна заявка в очереди;
- z_k — прибор занят, $(k - 1)$ заявок стоит в очереди.

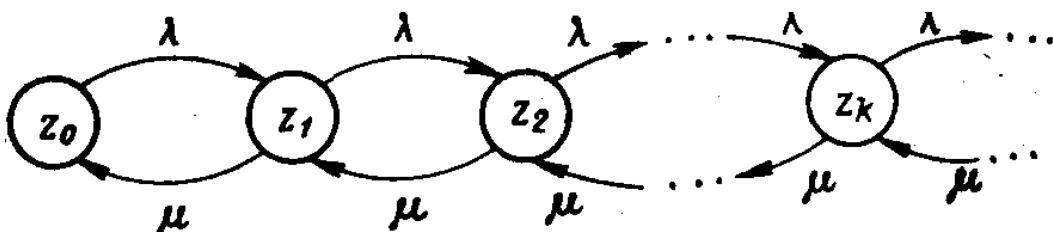


Рис. 5. Граф состояний СМО

Граф состояний такой системы изображен на рис. 5. Это модель размножения и гибели, но с бесконечным количеством состояний, поскольку, очередь неограничена.

Коэффициент загрузки. Предельная вероятность состояния

$$p_0 = [1 + \lambda / \mu + (\lambda / \mu)^2 + \dots + (\lambda / \mu)^k + \dots]^{-1}. \quad (1)$$

Обозначая $\lambda / \mu = \rho$, получаем

$$p_0 = (1 + \rho + \rho^2 + \dots + \rho^k + \dots)^{-1}. \quad (2)$$

Ряд в этой формуле представляет собой геометрическую прогрессию. Известно, что при $\rho < 1$ ряд сходится. Сумма членов прогрессии при этом равна $1/(1 - \rho)$, откуда $p_0 = 1 - \rho$.

Это вероятность того, что прибор свободен и очередь отсутствует. Значит, вероятность того, что прибор занят обслуживанием заявки,

$$1 - p_0 = 1 - (1 - \rho) = \rho.$$

Это означает, что **отношение**

$$\rho = \lambda / \mu$$

служит мерой загрузки СМО и является коэффициентом загрузки. Тогда $\eta = p_0 = 1 - \rho$ – коэффициент простоя.

Число заявок в СМО. Вероятности состояний z_1, \dots, z_k, \dots определяются из общей формулы размножения и гибели:

$$p_1 = \rho p_0, p_2 = \rho^2 p_0, \dots, p_k = \rho^k p_0, \dots$$

Определим среднее число заявок в системе n . В текущий момент времени в системе может быть $0, 1, 2, \dots, k, \dots$ заявок с вероятностями $p_0, p_1, p_2, \dots, p_k, \dots$ Математическое ожидание количества заявок равно

$$n = 0p_0 + 1p_1 + 2p_2 + \dots + kp_k + \dots = \sum_{k=0}^{\infty} kp_k.$$

Подставим значение p_k и p_0 , исключив первое слагаемое, равное нулю:

$$n = \sum_{k=1}^{\infty} k\rho^k p_0 = \sum_{k=1}^{\infty} k\rho^k (1 - \rho).$$

Вынесем за знак суммы $\rho(1 - \rho)$:

$$n = \rho(1 - \rho) \sum_{k=1}^{\infty} k\rho^{k-1}.$$

Но $k\rho^{k-1}$ – это производная по ρ от ρ^k :

$$n = \rho(1 - \rho) \sum_{k=1}^{\infty} \frac{d}{d\rho} \rho^k.$$

Меняя местами операции дифференцирования и суммирования, получим

$$n = \rho(1 - \rho) \frac{d}{d\rho} \sum_{k=1}^{\infty} \rho^k.$$

Сумма в этой формуле — это сумма бесконечно убывающей прогрессии при $\rho < 1$; она равна $\rho/(1-\rho)$, а ее производная $1/(1-\rho)^2$. Следовательно, число заявок в системе в установившемся стационарном режиме

$$n = \rho/(1 - \rho).$$

Длина очереди. Найдем среднее число заявок в очереди к обслуживающему прибору — среднюю длину очереди l . Она равна среднему числу заявок в системе за вычетом среднего числа заявок, находящихся под обслуживанием. Число заявок под обслуживанием может быть равно нулю, если прибор свободен, или единице, если прибор занят. В установившемся режиме математическое ожидание такой случайной величины равно вероятности того, что прибор занят. А эта вероятность определена ранее — ρ . Откуда получается средняя длина очереди в СМО:

$$l = n - \rho = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

Зависимость средней длины очереди от коэффициента загрузки изображена на рис. 2. При $\rho > 0,6—0,7$ очередь стремительно увеличивается и при $\rho \geq 1$ уходит в бесконечность. У детерминированной системы коэффициенты вариации интенсивностей потоков заявок и обслуживания равны нулю, при $\rho < 1$ очередь отсутствует, а при $\rho \geq 1$ — уходит в бесконечность. Для систем, которые имеют промежуточные коэффициенты вариации $0 < v < 1$, зависимости средней длины очереди от коэффициента загрузки лежат в области, заштрихованной на рис. 2.

Время реакции. Для определения среднего времени реакции рассмотрим поток заявок, прибывающих в СМО, и поток заявок, покидающих систему. Если в системе устанавливается предельный стационарный режим при $\rho < 1$, то среднее число заявок, прибывающих в единицу времени, равно среднему числу заявок, покидающих ее: оба потока имеют интенсивность λ .

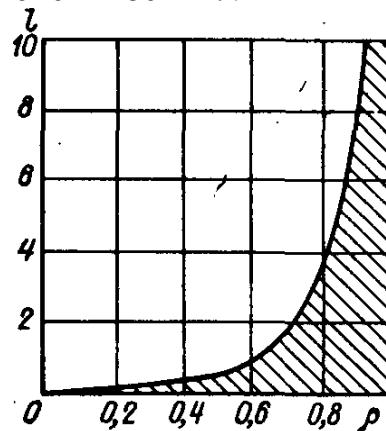


Рис. 2. Зависимость средней длины очереди от коэффициента загрузки для простейшей СМО

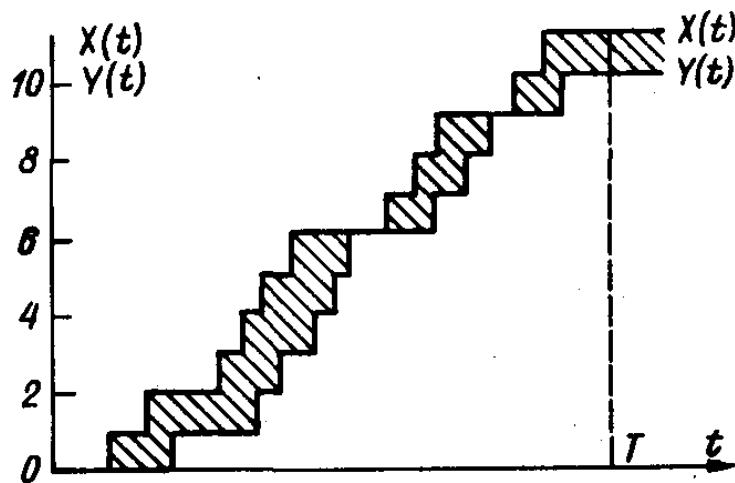


Рис. 3. Временная диаграмма процессов поступления и ухода заявок

Обозначим через $X(t)$ число заявок, поступивших в СМО до момента времени t , а через $Y(t)$ — число заявок, покинувших СМО до момента t . Та и. другая функции являются случайными и меняются скачком — увеличиваются на единицу в моменты прихода или ухода заявок (рис. 3).

Очевидно, что для любого момента времени t разность функций $n(t) = X(t) - Y(t)$ есть число заявок, находящихся в СМО. Рассмотрим большой промежуток времени T и вычислим среднее число заявок, находящихся в системе:

$$n = \frac{1}{T} \int_0^T n(t) dt.$$

Интеграл изображен в виде заштрихованной фигуры на рис. 3. Она состоит из прямоугольников, каждый из которых имеет высоту, равную единице, и основание, равное времени пребывания в системе i -й заявки t_i :

$$\int_0^t n(t) dt = \sum_i t_i,$$

где сумма распространяется на все заявки, поступившие в систему за время T . Разделим правую и левую части на T .

$$n = \frac{1}{T} \sum_i t_i.$$

Разделим и умножим правую часть на λ :

$$n = \frac{1}{T\lambda} \sum_i t_i \lambda = \frac{\sum_i t_i}{T\lambda} \lambda.$$

Произведение $T\lambda$ — это среднее количество заявок, пришедших за время T . Если разделить сумму всех времен t_i на среднее число заявок, то получится среднее время пребывания заявки в системе, т. е. среднее время реакции u :

$$n = u\lambda, u = n / \lambda.$$

Это формула Литтла: для каждой СМО при любом характере потока заявок и при любом распределении времени обслуживания среднее время реакции равняется среднему числу заявок в системе, деленному на интенсивность потока заявок. Отсюда получается:

$$u = \rho / [\lambda(1 - \rho)].$$

Вторая формула Литтла связывает среднее время пребывания заявки в очереди и среднее число заявок в очереди подобным соотношением:

$$\omega = l / \lambda = \rho^2 / \lambda(1 - \rho).$$

Среднее время реакции равняется сумме среднего времени пребывания заявки в очереди и средней длительности обслуживания заявки:

$$u = \omega + \vartheta.$$

Важно отметить, что в системе $M/M/1$ времена ожидания и реакции, а также периоды между моментами ухода следующих друг за другом заявок распределены по экспоненциальному закону. Для других систем при аналитическом моделировании не всегда представляется возможным определить законы распределения выходных характеристик.

При $\rho > 1$ в системе не устанавливается стационарный режим. В пределе длина очереди, а значит, и времена ожидания и реакции стремятся к бесконечности.

2. ХАРАКТЕРИСТИКИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ КАК СЛОЖНЫХ СИСТЕМ МАССОВОГО ОБСЛУЖИВАНИЯ

Многомерный поток. На вход обслуживающего прибора может поступать многомерный поток заявок, состоящий из заявок типов $1, \dots, M$, у которых интенсивности равны $\lambda_1, \dots, \lambda_M$. Предположим, что каждый из потоков заявок одного типа является простейшим. Загрузка прибора потоком заявок типа i будет составлять

$$\rho_i = \lambda_i \vartheta_i,$$

где ϑ_i — средняя длительность обслуживания заявок типа i . Суммарная загрузка прибора со стороны всех потоков

$$P = \sum_{i=1}^M \rho_i$$

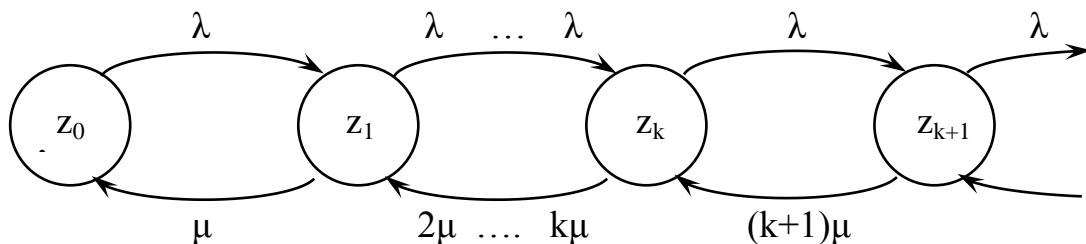


Рис. 4. Граф состояний многоканальной СМО

Условие существования стационарного режима: $P < 1$. Остальные характеристики обслуживания n_i, l_i, u_i, ω_i определяются для каждого i -го потока в отдельности по формулам (5) — (9).

Средние времена ожидания ω_{cp} и реакции u_{cp} по одной заявке из суммарного потока в системе связаны со средними количествами заявок в очереди и в системе следующими соотношениями:

$$\omega_{cp} = \sum_{i=1}^M \frac{\lambda_i}{\Lambda} \omega_i = \frac{1}{\Lambda} \sum_{i=1}^M l_i = \frac{l_{cp}}{\Lambda}; \quad (12)$$

$$u_{cp} = \sum_{i=1}^M \frac{\lambda_i}{\Lambda} u_i = \frac{1}{\Lambda} \sum_{i=1}^M n_i = \frac{n_{cp}}{\Lambda}, \quad (13)$$

где $\Lambda = \sum_{i=1}^M \lambda_i$ — суммарная интенсивность потоков; λ_i / Λ — вероятность того, что поступившая заявка является заявкой i -го типа;
 l_{cp} — средняя длина очереди заявок всех типов; n_{cp} — среднее число заявок всех типов в системе.

Многоканальная СМО. Предположим, что система имеет m обслуживающих каналов с одинаковой интенсивностью обслуживания μ , при общем простейшем потоке заявок с интенсивностью λ . Такая система условно обозначается $M/M/m$. Граф состояний этой системы (рис. 4) подобен графу одноканальной СМО. Интенсивности перехода в соседнее правое состояние определяются, как и у одноканальной СМО, интенсивностью входного потока: с приходом очередной заявки система переходит в следующее правое состояние. Иначе обстоит дело с интенсивностями у нижних стрелок. Пусть система находится в состоянии z_1 — работает один канал. Он производит μ обслуживания в единицу времени. Тогда $\lambda_{10} = \mu$. Представим, что система находится в состоянии z_2 . Для перехода в состояние z_1 надо, чтобы закончил обслуживание первый или второй канал. Значит, суммарная интенсивность их обслуживания $\lambda_{21} = 2\mu$. Суммарный поток обслуживания k каналами имеет интенсивность $k\mu$. При $k \geq m$ интенсивность обслуживания сохраняется $m\mu$. Получается модель размножения и гибели. Делая выкладки, как для одноканальной СМО, получим

$$p_0 = (1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^m}{m!} + \frac{\rho^{m+1}}{m!(m-\rho)})^{-1};$$

$$p_1 = \frac{\rho}{1!} p_0, \dots, p_{m+r} = \frac{\rho^{m+r}}{m^r \cdot m!} p_0, \dots$$

Средняя длина очереди

$$l = \sum_{r=1}^{\infty} r p_{n+r} = \frac{\rho^{m+1} p_0}{m m! (1 - \rho/m)^2}. \quad (14)$$

Прибавляя к ней среднее число заявок, находящихся под обслуживанием, равное среднему числу занятых каналов

$$k = \lambda / (\mu m) = \rho / m,$$

получим среднее число заявок в системе:

$$n = l + \rho / m$$

По формулам Литтла определяется среднее время пребывания заявки в очереди:

$$\omega = l / \lambda$$

и в системе — время реакции:

$$u = n / \lambda$$

В теории массового обслуживания имеются аналитические формулы для простейших СМО при одномерном и многомерном потоке заявок для одноканальных и многоканальных систем без ограничений и с ограничениями длин очередей.

Потоки обслуживания. При моделировании конкретных ВС потоки заявок и обслуживания могут отличаться от простейших. Потоки заявок могут быть, например, пуассоновскими или эрланговскими. Длительности обслуживания можно представить в общем случае гаммапределением. Это распределение с плотностью вероятности

$$f(\tau) = \frac{(k/\vartheta)^k}{\Gamma(k)} e^{-k\tau/\vartheta} \tau^{k-1}, \tau > 0, \quad (18)$$

где ϑ — математическое ожидание длительности обслуживания $M[\tau]$; k — параметр распределения ($k \geq 1$); $\Gamma(k)$ — гамма-функция.

Дисперсия гамма-распределения

$$D[\tau] = \vartheta^2 / k$$

При $k = 1$ гамма-распределение вырождается в экспоненциальное. В пределе при $k \rightarrow \infty$ это распределение становится детерминированным с постоянной длительностью обслуживания ϑ . Параметр распределения k может быть определен по математическому ожиданию и дисперсии:

$$k = \vartheta^2 / D[\tau].$$

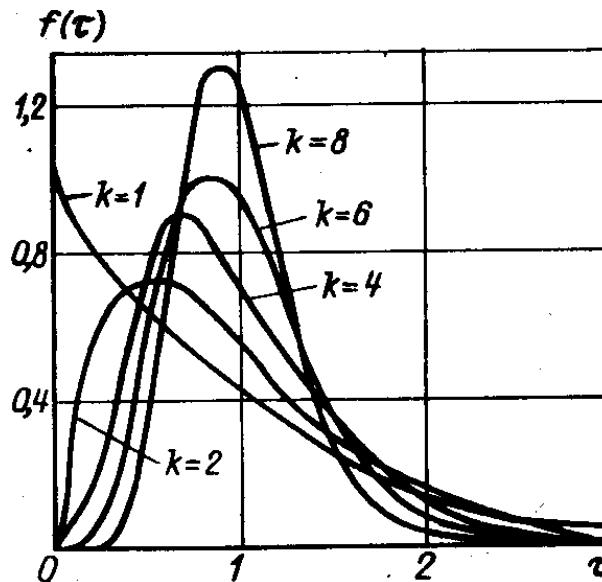


Рис. 5. Нормированное распределение Эрланга

При целочисленном k $\Gamma(k) = (k-1)!$. Тогда из уравнения (18) имеем

$$f(\tau) = \frac{(k/\vartheta)^k}{(k-1)!} e^{-k\tau/\vartheta} \tau^{k-1}, \tau > 0 \quad (2)$$

Это плотность нормированного распределения Эрланга k -го порядка. Вид распределения изображен на рис. 5. В данном распределении в отличие от потока Эрланга математическое ожидание не зависит от k и при $k \rightarrow \infty$ это распределение стремится к детерминированному, а не к нормальному.

В частных случаях длительности обслуживания могут быть распределены по экспоненциальному, равномерному, нормальному и другим законам. Для некоторых сочетаний законов распределений потоков заявок и обслуживания получены аналитические зависимости характеристик от параметров системы.

Системы с произвольным распределением длительности обслуживания. Представим, что моделью ВС является одноканальная СМО с неограниченной очередью. В эту систему поступает простейший поток заявок с интенсивностью λ . Заявки обслуживаются в порядке поступления. Длительность обслуживания имеет произвольное распределение с математическим ожиданием $\vartheta = 1/\mu$ и коэффициентом вариации v . Такая система обозначается $M/G/1$. В этой системе в установившемся режиме среднее число заявок в очереди

$$l = \rho^2 (1 + v)^2 / [2(1 - \rho)], \quad (3)$$

среднее число заявок в системе

$$n = \rho^2 (1 + v)^2 / [2(1 - \rho)] + \rho. \quad (2)$$

Последние два выражения называются формулами Поллачека—Хинчина. Средние времена пребывания заявок в очереди и в системе определяются по формулам Литтла.

Для системы $M/G/1$ могут быть аналитически определены дисперсии выходных характеристик. Подобные формулы известны также для случая многомерного простейшего потока заявок.

Системы с отказами. Предположим, что на ВС, представленную как m -канальная СМО, поступает простейший поток заявок с интенсивностью K . Поток обслуживания имеет произвольный закон распределения с интенсивностью ρ . Это система $M/G/m$. При этом очередная заявка, поступившая в систему, когда все каналы заняты, покидает ее без обслуживания. Это означает, что очереди в системе отсутствуют. Характеристиками такой системы могут служить пропускная способность, вероятность обслуживания и среднее число занятых каналов. Данная система соответствует модели размножения и гибели. На основании формул (13) и (14) (лекция 9) можно вывести вероятность того, что в СМО находится m заявок, т. е. все каналы заняты:

$$p_m = \frac{\rho^m}{m!} p_0; \quad p_0 = (1 + \rho + \frac{\rho^2}{2!} + \dots + \frac{\rho^m}{m!})^{-1}. \quad (28)$$

Вероятность того, что очередная заявка будет обслужена,

$$P_z = 1 - p_m = 1 - \frac{\rho^m}{m!} p_0. \quad (29)$$

Пропускная способность системы определяется как среднее число заявок, обслуживаемых в единицу времени;

$$S_z = \lambda P_z = \lambda \left(1 - \frac{\rho^m}{m!} p_0\right) \quad (30)$$

а среднее число занятых каналов определяется по формуле

$$k = S_z / \mu = \rho \left(1 - \frac{\rho^m}{m!} p_0\right) \quad (31)$$

Системы с приоритетными дисциплинами диспетчеризации.

В теории вычислительных систем детально изложены и исследованы аналитические зависимости характеристик от параметров ВС, представленных моделями СМО с ординарными и неординарными, одномерными и многомерными потоками заявок, обслуживаемых одноканальными и многоканальными приборами с произвольными законами распределения длительности обслуживания и различными дисциплинами диспетчеризации, включая системы с относительным, абсолютным, смешанным и динамическим приоритетами.

Например, допустим, что в СМО поступает M типов простейших потоков с интенсивностями $\lambda_1, \dots, \lambda_M$ и длительности обслуживания заявок каждого потока имеют математические ожидания $\vartheta_1, \dots, \vartheta_M$ и дисперсии $\sigma^2_1, \dots, \sigma^2_M$. В системе используется смешанная дисциплина диспетчеризации с тремя классами: 1) заявкам типов $1, \dots, M_1$ присвоены абсолютные приоритеты по отношению к заявкам второго и третьего классов; 2) заявкам типов M_1+1, \dots, M_1+M_2 — относительные приоритеты по отношению к заявкам третьего класса; 3) заявки типов M_1+M_2+1, \dots, M обслу-

живаются в порядке поступления. Среднее время ожидания заявок различных типов определяется из выражения:

$$\omega_k = \begin{cases} \frac{P_{k-1}g_k}{1-P_{k-1}} + \frac{\sum_{i=1}^k \lambda_i \sigma^2_i}{2(1-P_{k-1})(1-P_k)}, & k = 1, \dots, M_1; \\ \frac{P_{M_1}g_k}{1-P_{M_1}} + \frac{\sum_{i=1}^M \lambda_i \sigma^2_i}{2(1-P_{k-1})(1-P_k)}, & k = M_1 + 1, \dots, M_1 + M_2; \\ \frac{P_{M_1}g_k}{1-P_{M_1}} + \frac{\sum_{i=1}^M \lambda_i \sigma^2_i}{2(1-P_{M_1+M_2})(1-P_k)}, & k = M_1 + M_2 + 1, \dots, M, \end{cases} \quad (32)$$

где $P_J = \sum_{j=1}^J \rho_j$, $J = k-1, k, M_1, M_1 + M_2, M$,

Из формулы (32) могут быть получены как частные случаи характеристики систем с абсолютными ($M_2 = M_3 = 0$), относительными ($M_1 = M_3 = 0$) и смешанными приоритетами с двумя классами заявок: с абсолютными и относительными приоритетами ($M_3 = 0$), с абсолютными приоритетами и без приоритетов ($M_2 = 0$), с относительными приоритетами и без приоритетов ($M_1 = 0$).

3. МЕТОДЫ ПРИБЛИЖЁННОЙ ОЦЕНКИ ХАРАКТЕРИСТИК ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Оценка при большой нагрузке. Аналитические зависимости, позволяющие определить параметры распределений выходных характеристик, имеются только для ограниченного круга систем. У более широкого круга систем могут быть вычислены средние значения в стационарном установившемся режиме. Однако остается ряд систем и режимов, для которых отсутствуют точные формулы даже по определению средних значений. К таким системам относятся, в первую очередь, системы с произвольными распределениями периодов поступления и длительностей обслуживания заявок. Это системы $G/G/1$ и $G/G/m$. При отсутствии точных зависимостей приходится довольствоваться приближенными оценками.

Одним из методов приближений является оценка характеристик при близких к единице значениях коэффициента загрузки, как наиболее важных с практических позиций. В частности, для системы $G/G/1$ время ожидания заявки в очереди определено по экспоненциальному закону и среднее время ожидания можно определить по следующей формуле:

$$\omega \approx \frac{\lambda(\sigma^2_\lambda + \sigma^2_\mu)}{2(1-\rho)}, \quad \rho \rightarrow 1, \quad (33)$$

где $\sigma^2_\lambda, \sigma^2_\mu$ — дисперсия периодов поступления и длительностей обслуживания заявок соответственно.

Используя зависимость (10) и формулы Литтла, можно вычислить средние значения других характеристик.

Определение границ. При значениях $0 \leq \rho < 1$ для оценки характеристик используется несколько различных подходов определения верхней и нижней границ, в пределах которых находится истинное значение той или иной характеристики. Например, для системы G/G/1 приводятся следующие формулы расчета границ среднего времени ожидания заявки в очереди:

$$\frac{\rho^2(\nu^2_\lambda + \nu^2_\mu)}{2(1-\rho)\lambda} - \frac{\rho(1-\nu^2_\lambda)}{2\lambda} \leq \omega \leq \frac{\rho^2(\nu^2_\lambda + \nu^2_\mu)}{2(1-\rho)\lambda} + \frac{(1-\rho)(1-\nu^2_\lambda)}{2\lambda} \quad (24)$$

где ν_λ, ν_μ — коэффициенты вариации периодов поступления и длительностей обслуживания заявок соответственно.

Средняя длина очереди $l = \omega\lambda$, и среднее число заявок в системе $n = l + \rho$.

Дискретное и непрерывное приближения. Другие методы оценки ориентированы не на поиск приближенного решения исходной задачи, а на точное решение упрощенно сформулированной задачи. Уравнения, описывающие работу системы G/G/1, преднамеренно преобразуются к такому виду, при котором полученная система уравнений может быть решена.

В теории массового обслуживания показано, что если исходные величины являются дискретными, можно определить точное распределение времени ожидания. На этом основывается метод дискретного приближения, при котором промежутки времени между моментами поступления заявок и длительности обслуживания аппроксимируются дискретными распределениями.

Для исследования нестационарных систем и режимов перегрузки оказывается полезным метод непрерывного приближения. Процессы поступления и ухода заявок — это ступенчатые вероятностные процессы (рис. 7). Но когда длины очередей значительно больше единицы, а времена ожидания существенно больше среднего времени обслуживания, становится разумной замена этих ступенчатых процессов сглаженными непрерывными функциями времени, поскольку величины отдельных ступенек малы по сравнению со средними значениями (рис. 6).

Когда $X(f)$ становится значительно больше единицы, на основании закона больших чисел можно ожидать лишь небольшого относительного отклонения этой величины от ее среднего значения

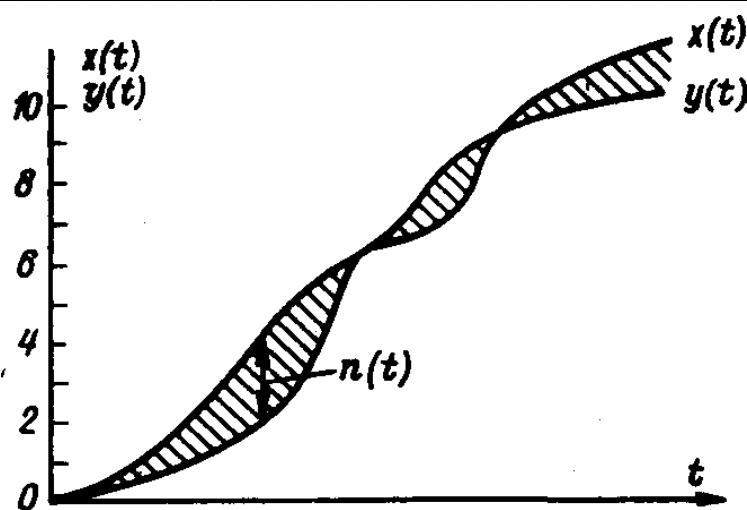


Рис. 6. Временная диаграмма процессов поступления и ухода заявок с непрерывной аппроксимацией зависимостей количества заявок от времени $M[x(t)] = \overline{x(t)}$. На этом основывается приближение первого порядка, которое заключается в замене вероятностного процесса его средним значением, зависящим от времени, т. е. детерминированным процессом $\overline{x(t)}$. Это относится и к процессу ухода заявок $\overline{y(t)}$. Тогда число заявок в системе тоже представляет собой детерминированную непрерывную функцию времени:

$$n(t) = \overline{x(t)} - \overline{y(t)}.$$

Функции $x(t)$ и $\overline{y(t)}$ определяются из зависимостей:

$$\overline{x(t)} = x(0) + \int_0^t \lambda(t) dt;$$

$$\overline{y(t)} = y(0) + \int_0^t \mu(t) dt,$$

—число поступивших и покинувших систему заявок к нулевому моменту времени.

Диффузионная аппроксимация. Непрерывное приближение является довольно грубым, поскольку не учитывает случайный характер процессов поступления и ухода заявок. По методу диффузионной аппроксимации непрерывное приближение усовершенствуется путем учета флюктуаций относительно среднего значения. С этой целью случайный процесс (42) заменяется марковским процессом диффузионного типа x (/) с непрерывным временем и непрерывным множеством состояний. Диффузионный процесс определяется коэффициентом сноса

$$\beta(x) = \lim_{\Delta t \rightarrow 0} \frac{M[x(t + \Delta t) - x(t)][x(t) = x]}{\Delta t}$$

и коэффициентом диффузии

$$\alpha^2(x) = \lim_{\Delta t \rightarrow 0} \left(\frac{M[(x(t + \Delta t) - x(t))^2 | x(t) = x]}{\Delta t} - \right. \\ \left. - \frac{(M[(x(t + \Delta t) - x(t)) = x])^2}{\Delta t} \right).$$

Эти коэффициенты выражаются через параметры исходной модели. Для системы G/G/1 считается, что при больших t распределение $X(t)$ является приближенно нормальным с математическим ожиданием λt и дисперсией $\lambda^3 \sigma^2 \lambda t$ и распределение $Y(f)$ тоже приближено нормальное с математическим ожиданием μt и дисперсией $\mu^3 \sigma^2 \mu t$. Тогда коэффициент сноса

$$\beta(x) = \lambda - \mu.$$

и коэффициент диффузии

$$\alpha^2(x) = \lambda^3 \sigma^2 \lambda + \mu^3 \sigma^2 \mu$$

В связи с тем, что процесс $n(t)$ не может принимать отрицательных значений, для аппроксимирующего процесса $x(t)$ задается граничное условие, удерживающее его траекторию на неотрицательной полуоси. Например, при достижении $x(t) = 0$ совершается скачок в целочисленные точки положительной полуоси, выполняемый с заданным распределением вероятностей после экспоненциальной задержки в нуле.

Применение диффузионной аппроксимации дает возможность получения оценок различных характеристик СМО. В частности, для системы G/G/1 средняя длина очереди в стационарном режиме определяется по формуле

$$l = \rho \left(\frac{1}{2} + \frac{\rho \lambda^2 \sigma^2 \lambda + \mu^2 \sigma^2 \mu}{2(1 - \rho)} \right) \quad (4)$$

при постоянных коэффициентах сноса и диффузии, соответствующих случаю независимых от длины очереди вероятностных характеристик поступления и обслуживания заявок.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Одноканальная СМО для описания ВС.
2. Как определяется коэффициент загрузки ВС?
3. Как определяется число заявок в СМО?
4. Как определяется длина очереди в СМО?
5. Как определяется время реакции в СМО?
6. Формулы Литтла.
7. Многоканальная СМО для описания ВС.
8. Основные характеристики многоканальной СМО.
9. Для каких систем используются методы приближенной оценки характеристик?

ЛИТЕРАТУРА

- 45.
- 46.
- 47.
- 48.
- 49.

ЛЕКЦИЯ 11. НЕСТАЦИОНАРНЫЕ РЕЖИМЫ ФУНКЦИОНИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 ЧАСА)

План

1. Нестационарные режимы функционирования вычислительных систем
2. Характеристики вычислительных систем как стохастических сетей

1. НЕСТАЦИОНАРНЫЕ РЕЖИМЫ ФУНКЦИОНИРОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Переходные процессы. До сих пор рассматривались характеристики ВС в стационарном установившемся режиме. Однако на практике не менее важным является анализ нестационарных режимов. Значения выходных характеристик в нестационарных режимах функционирования ВС можно определить для одних систем путем численного решения уравнений Колмогорова (9) задавая в них интенсивности как функции времени $\lambda_{ij}(t)$ для других систем — в результате непрерывной или диффузионной аппроксимации процессов.

Частным случаем нестационарных режимов является переходный процесс, когда, например, в начальный момент времени в системе отсутствуют очереди и начинают поступать заявки с постоянной интенсивностью X . Важно уметь определять, когда установится стационарный режим.

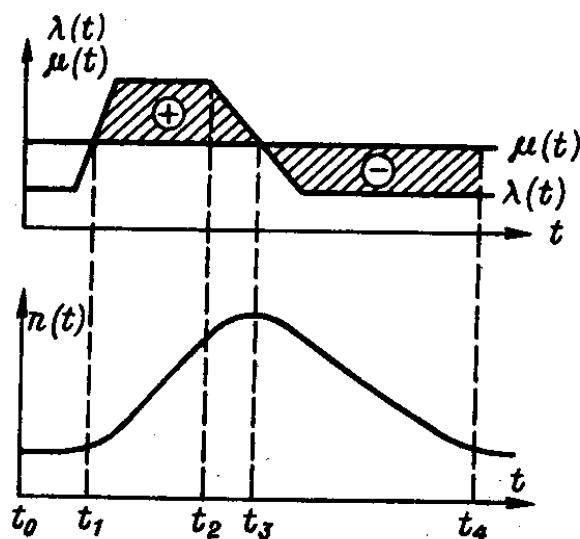


Рис. 12. Временная диаграмма поведения системы в режиме перегрузки

Рис. 11. График зависимости среднего времени ожидания от текущего времени в переходном режиме

В работе приводятся результаты анализа переходных процессов для системы $M/G/1$. Интенсивность входящего пуассо-новского потока λ принималась равной

0,95 заявок в минуту, а среднее время обслуживания $v = 1$ мин. Коэффициент загрузки $p = 0,95$. Рассматривались следующие распределения длительности обслуживания:

1) экспоненциальное распределение

$$f_1(\tau) = e^{-\tau};$$

2) нормированное распределение Эрланга 8-го порядка со средним, равным восьми,

$$f_2(\tau) = \frac{8(8\tau)^7}{7!} e^{-8\tau};$$

3) комбинация экспоненциального распределения со средним значением 4 и распределения Эрланга 4-го порядка со средним значением 2/3

$$f_3(\tau) = 0,1(0,25e^{-\tau/4}) + 0,9((6\tau)3e^{-6\tau}).$$

Зависимости среднего времени ожидания заявок в очереди в течение переходного процесса для этих случаев показаны на рис. 11. Установившиеся значения соответственно равны 19; 10,69 и 35,2 мин. Если принять длительность переходного процесса равной бремени, в течение которого среднее время ожидания достигает 0,8 от установленногося значения, то для этих случаев она соответственно составит 15,2; 8,55 и 28,2 ч. За эти времена система успевает обслужить 867, 487 или 1605 заявок. Можно утверждать, что ВС с суточным циклом никогда не работают практически в установленном режиме при большой загрузке. Этот вывод можно распространить на ВС с большей длительностью цикла, если они имеют соответственно меньшие интенсивности поступления и обслуживания заявок.

Режимы перегрузок. Методы непрерывной и диффузионной аппроксимации дают возможность проанализировать поведение системы при изменяющихся во времени интенсивностях прихода и обслуживания заявок. С практических позиций наибольшую важность представляет анализ режима перегрузок, когда в течение некоторого интервала времени коэффициент загрузки $p > 1$.

Рассмотрим этот режим на упрощенном примере (рис. 12). Предположим, что в одноканальную систему поступает одномерный поток заявок с интенсивностью $\lambda(t)$. Заявки обслуживаются в порядке поступления с постоянной интенсивностью $\mu(t) = \mu$.

В начальный момент t_0 коэффициент загрузки $p < 1$. В системе имеются заявки, накопление которых обусловлено случайным характером их поступления и обслуживания.

Затем интенсивность поступления заявок начинает расти, достигая максимального значения. С момента t_1 становится $p > 1$ и увеличивается число заявок в системе. При максимальном $\lambda(t)$ число заявок $n(t)$ растет линейно, стремясь к бесконечности. Но в связи с тем, что в момент t_2 интенсивность поступления начинает уменьшаться, рост $n(t)$ замедляется и достигает максимума в момент t_3 при $p = 1$.

В режиме перегрузки накопление заявок в системе определяется в основном не случайными факторами, а превышением средней интенсивности поступления над интенсивностью обслуживания. С момента t_3 число заявок уменьшается до мо-

мента t_4 принимая установившееся значение. Длительность рассасывания числа заявок может быть приближено оценена по равенству заштрихованных на рисунке площадей, обозначенных плюсом и минусом.

Важно подчеркнуть, что такая система может быть вполне работоспособна, если максимальное число заявок в системе, а соответственно и время реакции не превышают допустимых значений. При правильном задании стохастических ограничений, как показано в п. 2.9, систему можно считать работоспособной даже при кратковременном превышении допустимых значений математическим ожиданием времени реакции. Этот подход дает возможность выбрать производительность ВС не по максимальной интенсивности поступления заявок, обеспечивая $p < 1$, а на существенно более низком уровне. Отдельные вопросы анализа нестационарных ВС рассматриваются в работах.

2. ХАРАКТЕРИСТИКИ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ КАК СТОХАСТИЧЕСКИХ СЕТЕЙ

Описание стохастической сети. Обычно ВС представляется не отдельной СМО, а стохастической сетью. Для описания ВС в виде стохастической сети определяются следующие параметры:

- 1) число СМО, образующих сеть (S_1, S_2, \dots, S_n);
- 2) число каналов каждой СМО (m_1, \dots, m_n);
- 3) матрица вероятностей передач $P = [p_{ij}]$, где p_{ij} — вероятность того, что заявка, покидающая систему S_i , поступает в систему S_j ($i, j=0, 1, \dots, n$);
- 4) интенсивность λ_0 источника заявок S_0 в разомкнутой сети или число M заявок в замкнутой сети;
- 5) средние длительности обслуживания заявок $\vartheta_1, \dots, \vartheta_n$ в системах S_1, \dots, S_n .

Рассмотрим характеристики экспоненциальных сетей, как это сделано в работе. Экспоненциальная стохастическая сеть имеет простейшие входные потоки и распределенные по экспоненциальному закону длительности обслуживания заявок в различных системах сети. В установившемся режиме вероятность передачи заявки из системы S_i в систему S_j равна доле потока, поступающего из системы S_i в систему S_j . Если система без потерь, то на входе системы S_i имеется поток с интенсивностью

$$\lambda_i = \sum_{j=0}^n p_{ij} \lambda_0, \quad i = 0, 1, \dots, n. \quad (1)$$

Из этой системы уравнений находятся соотношения интенсивностей потоков λ_j и λ_0 в виде

$$\lambda_j = \alpha_j \lambda_0, \quad j = 1, \dots, n,$$

где α_j — коэффициент передачи, который определяет среднее число этапов обслуживания одной заявки в системе S_j .

Для замкнутой сети принимается $\lambda_0 = 1$.

Определение вероятности состояний. В стационарном установившемся режиме вероятность того, что в системе S_j находится M_j заявок, определяется из выражения

$$\pi_{M_j} = \beta_j^{M_j} R_j(M_j) \pi_{0j},$$

где

$$\beta_j = \lambda_j \vartheta_j = \begin{cases} \rho_j & \text{при } m_j = 1; \\ k_j & \text{при } m_j > 1. \end{cases}$$

$$R_j(M_j) = \begin{cases} 1/M_j! & \text{при } M_j \leq m_j; \\ 1/m_j! m_j^{M_j - m_j} & \text{при } M_j > m_j; \end{cases}$$

$$\pi_{0j} = \left(\sum_{M_j=0}^{m_j-1} \frac{\beta_j^{M_j}}{M_j!} + \frac{\beta_j^{m_j}}{m_j!(1-\beta_j/m_j)} \right)^{-1}.$$

Вероятность состояния замкнутой сети $P(M_1, \dots, M_n)$, характеризующая вероятность того, что в системе S_1 находится M_1 заявок, и т. д., вычисляется по формуле

$$P(M_1, \dots, M_n) = \frac{\prod_{j=1}^n R_j(M_j) (\alpha_j \vartheta_j)^{M_j}}{\sum_{A(M,n)} \prod_{j=1}^n R_j(M_j) (\alpha_j \vartheta_j)^{M_j}}, \quad (7)$$

где $\sum_{A(M,n)}$ — символ суммирования по всем возможным состояниям, для которых

$$\sum_{j=1}^n M_j = M.$$

По вероятностям состояний определяются характеристики сети.

Вычисление характеристик разомкнутой сети. В разомкнутой экспоненциальной сети в стационарном режиме согласно теореме Джексона можно считать, что сеть состоит из совокупности независимых СМО с простейшими входными потоками. Для каждой СМО характеристики определяются отдельно. Для каждой S_j СМО сети средняя длина очереди

$$l_j = \frac{\beta_j^{m_j+1}}{m_j! m_j (1-\beta_j/m_j)^2} \pi_{0j},$$

среднее число в системе

$$n_j = l_j + \beta_j,$$

а среднее время ожидания в очереди ω_j , и пребывания в системе u_j определяется по формулам Литтла из (9) и (10). Тогда характеристики сети в целом: среднее число заявок во всех очередях

$$l = \sum_{j=1}^n l_j,$$

среднее число заявок в сети

$$N = \sum_{j=1}^n n_j,$$

среднее время ожидания в очередях

$$\omega = \sum_{j=1}^n \alpha_j \omega_j$$

среднее время реакции сети

$$u = \sum_{j=1}^n \alpha_j u_j$$

Вычисление характеристик замкнутой сети. Для любой S_j СМО сети средняя длина очереди

$$l_j = \sum_{r=m_j+1}^M (r - m_j) P(M_j = r), \quad (1)$$

среднее число заявок в СМО

$$n_j = \sum_{r=0}^M r P(M_j = r),$$

средние времена ожидания ω_j и пребывания u_j вычисляются по формулам Литтла. Среднее время цикла сети, т. е. интервал времени между двумя последовательными выходами одной и той же заявки из СМО S_i составляет $U_j = M/\lambda_j$.

В заключение отметим, что простые аналитические выражения для определения характеристик ВС как сетевых моделей имеются лишь для узкого круга систем. В ряде случаев приходится прибегать к численным моделям. В работе изложены некоторые вопросы численных методов исследования систем, и в частности, здесь рассмотрен пакет программ, который позволяет автоматизировать составление и решение уравнений для ВС, описываемых марковскими процессами.

Существенный недостаток сетевых моделей — трудность учета таких ситуаций, когда заявке требуется одновременно несколько ресурсов. В теории массового обслуживания слабо учитывается использование накопителей, неоднородность потока заявок. Считается, что заявка обслуженная одной СМО, поступает на другую, в то время как в ВС в передаче одновременно участвуют передающее устройство, коммуникатор и принимающее устройство. Применение аналитического моделирования целесообразно для предварительной оценки характеристик ВС.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Основные подходы исследования характеристик ВС в нестационарном режиме.
2. Исследование характеристик в режиме перегрузок.
3. Описание стохастической сети применительно к ВС.
4. Определение вероятности состояний.
5. Определение характеристик разомкнутой и замкнутой сети.

ЛИТЕРАТУРА

- 50.
- 51.
- 52.
- 53.
- 54.

ЛЕКЦИЯ 12. ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 ЧАСА)

План

1. Процедура имитационного моделирования
2. Обобщенные алгоритмы имитационного моделирования

1. ПРОЦЕДУРА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Метод имитационного моделирования заключается в создании логико-аналитической (математической) модели системы и внешних воздействий, в имитации функционирования системы, т. е. в определении временных изменений состояния системы под влиянием внешних воздействий, и в получении выборок значений выходных параметров, по которым определяются их основные вероятностные характеристики. Данное определение справедливо для стохастических систем. При исследовании детерминированных систем отпадает необходимость в получении выборок значений выходных параметров.

Модель системы со структурным принципом управления представляет собой совокупность моделей элементов и их функциональные взаимосвязи. Модель элемента (агрегата, обслуживающего прибора) — это, в первую очередь, набор правил (алгоритмов) поведения устройства по отношению к входным воздействиям (заявкам) и правил изменения состояний элемента. При моделировании ВС на системном уровне элемент отображает функциональное устройство на том или ином уровне детализации.

В простейшем случае устройство может находиться в работоспособном состоянии или в состоянии отказа. В работоспособном состоянии устройство может быть занято, например, выполнением операции по обслуживанию заявки или свободно. К правилам поведения устройства относятся правила выборки заявок из очереди; реакция устройства на поступление заявки, когда устройство занято или к нему имеется очередь заявок; реакция устройства на возникновение отказа в процессе обслуживания заявки и некоторые другие.

Функциональные взаимосвязи устройств определяют возможные пути продвижения заявок по системе от входных устройств к выходным. Они формируют функциональную структуру ВС.

Модель внешних воздействий — это правила определения моментов поступления входных сигналов (заявок) в систему, маршрута заявок в системе по каждому из потоков в соответствии с алгоритмами обработки, приоритетов обслуживания заявок одних потоков по отношению к другим, трудоемкости обслуживания заявок устройствами, допустимого времени пребывания заявок в системе и др.

В процессе имитации функционирования системы измеряются те выходные характеристики, которые интересуют исследователя. При изучении стохастических систем измерения производятся многократно с тем, чтобы можно было с достаточной точностью определить вероятностные характеристики системы.

Модель системы с программным принципом управления представляет собой, в основном, формализованное описание параллельно протекающих процессов с указанием используемых ресурсов и алгоритмов управления процессами.

Имитационное моделирование — это метод исследования, который основан на том, что анализируемая динамическая система заменяется имитатором и с ним проводятся эксперименты для получения информации об изучаемой системе. Роль имитатора зачастую выполняет специальная программа ВС.

Основная идея метода имитационного моделирования стохастических систем во многом исследована методом вычисления случайных величин, который называется методом статистических испытаний или методом Монте-Карло. Заключается он в следующем. Пусть необходимо определить функцию распределения случайной величины y . Допустим, что искомая величина y может быть представлена в виде зависимости

$$y = y(\alpha, \beta, \gamma, \dots, \varpi),$$

где $\alpha, \beta, \gamma, \dots, \varpi$ — случайные величины с известными функциями распределения.

Для решения задач такого вида применяется следующий алгоритм:

1) по каждой из величин $\alpha, \beta, \gamma, \dots, \varpi$ производится случайное испытание, в результате которого определяется некоторое конкретное значение случайной величины $\alpha_i, \beta_i, \gamma_i, \dots, \varpi_i$ (способы проведения случайного испытания описаны ниже);

2) используя найденные величины, определяется одно частное значение y_i по вышеприведенной зависимости;

3) предыдущие операции повторяются N раз, в результате чего определяется N значений случайной величины y ;

4) на основании N значений величины y находится ее эмпирическая функция распределения.

Имитация функционирования системы. Предположим, что ВС состоит из процессора 1 с основной памятью, устройства ввода 4, печатающего устройства 2 и монитора 3. Через устройство ввода поступает поток заданий X_1 . Процессор обрабатывает задания и результаты обработки выдает на печатающее устройство (принтер). Одновременно с этим ВС используется, например, как информационно-справочная система. Оператор-пользователь, работающий за монитором, посыпает в систему запросы X_2 , которые обрабатываются процессором, и ответы выводятся на монитор. Процессор работает в двух программном режиме: в одном разделе обрабатываются задания X_1 , в другом, с более высоким относительным приоритетом, — запросы X_2 .

Представим данную ВС в упрощенном варианте в виде стохастической сети из четырех СМО. Потоки заданий и запросов будем называть потоками заявок. Считаем потоки X_1 и X_2 независимыми. Известны функции распределения периодов следования заявок τ_1 и τ_2 и длительностей обслуживания T_{1k} и T_{2k} заявок k -м устройством. Требуется определить времена загрузки каждого устройства и времена реакции по каждому из потоков.

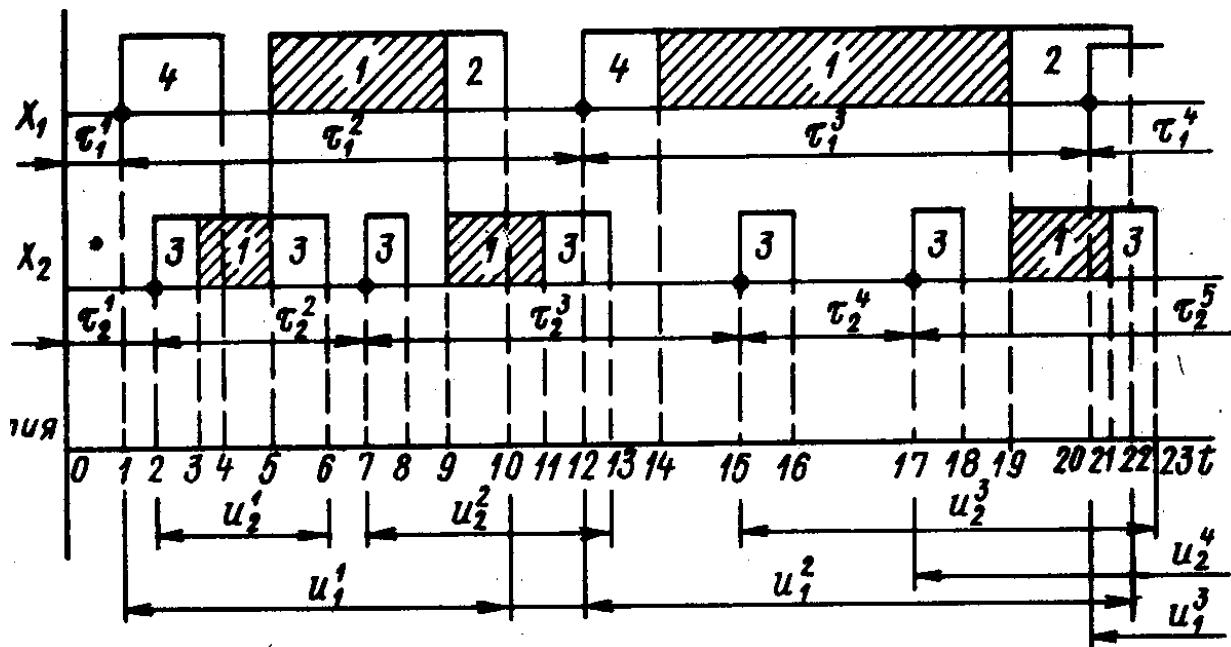


Рис. 1. Временная диаграмма функционирования ВС

В начале определяется момент поступления в систему первой заявки потока X_1 по результатам случайного испытания в соответствии с функцией распределения периода следования заявок. На рис. 1 это момент времени $t_1 = 0 + \tau_1^1$ (здесь и далее верхний индекс обозначает порядковый номер заявки данного потока). То же самое делается для потока X_2 . На рис. 2 момент поступления первой заявки потока $X_2 - t_2 = 0 + \tau_2^1$. Затем находится минимальное время, т.е. наиболее раннее событие. В примере— это время t_1 . Для первой заявки потока X_1 определяется путем случайного испытания время обслуживания устройством ввода T_{14}^1 и отмечается момент окончания обслуживания $t_4 = t_1 + T_{14}^1$. На рисунке показан ступенькой переход устройства 4 в состояние «занято». Одновременно определяется момент поступления следующей заявки потока $X_1 : t_{12} = t_1 + \tau_1^2$.

Следующее минимальное время — это момент поступления заявки потока $X_2 - t_2$. Для этой заявки находится время обслуживания на мониторе T_{23}^1 и отмечается время окончания обслуживания $t_3 = t_2 - T_{23}^1$. Определяется момент поступления второй заявки потока $X_2 - t_7 = t_2 + \tau_2^2$. Снова выбирается минимальное время— это t_3 . В этот момент заявка потока X_2 начинает обрабатываться процессором. По результату случайного испытания определяется время ее обслуживания T_{23}^1 и отмечается момент $t_5 = t_3 + T_{23}^1$ окончания обслуживания. Следующее минимальное время t_4 — момент завершения обслуживания заявки потока X_1 устройством 4. С этого момента заявка может начать обрабатываться процессором, но он занят обслуживанием заявки потока X_2 . Тогда заявка потока X_1 переходит в состояние ожидания, становится в очередь.

В следующий минимальный момент времени t_5 освобождается процессор. С этого момента процессор начинает обрабатывать заявку потока X_1 , а заявка потока

X_2 переходит на обслуживание монитором, т. е. ответ на запрос пользователя передается из основной памяти в буферный накопитель монитора. Далее определяются соответствующие времена обслуживания T_{11}^1 и T_{23}^1 и отмечаются моменты времени $t_9 = t_5 + T_{11}^1$ и $t_6 = t_5 + T_{23}^1$. В момент t_6 полностью завершается обработка первой заявки потока X_2 . По разности времен t_6 и t_2 вычисляется время реакции по этой заявке:

$$u_2^1 = t_6 - t_2$$

Следующий минимальный момент времени t_7 — это поступление второй заявки потока X_2 . Определяется время поступления очередной заявки этого потока $t_{15} = t_7 + \tau_2^3$. Затем вычисляется время обслуживания второй заявки на мониторе T_{23}^2 и отмечается момент $t_8 = t_7 + T_{23}^2$ после чего заявка становится в очередь, так как процессор занят. Эта заявка поступает на обслуживание в процессор только после его освобождения в момент времени t_9 . В этот же момент заявка потока X_1 начинает обслуживаться принтером. Определяются времена обслуживания T_{21}^2 и T_{12}^1 по результатам случайных испытаний и отмечаются моменты окончания обслуживания $t_{11} = t_9 + T_{21}^2$ и $t_{10} = t_9 + T_{12}^1$. В момент времени t_{10} завершается полное обслуживание первой заявки потока X_1 . Разность между этим моментом и моментом времени t_1 дает первое значение времени реакции по потоку $X_1 - u_1^1$.

Вторая заявка потока X_2 в момент t_{11} поступает с процессора на монитор и обслуживается им в течение времени T_{23}^2 , которое завершается в момент $t_{13} = t_{11} + T_{23}^2$. Снова определяется очередное минимальное время. Это время — t_{12} когда в систему поступает вторая заявка из потока X_1 . Тогда вычисляется время поступления третьей заявки потока $X_1 - t_{20} = t_{12} + \tau_1^3$. Вторая заявка обслуживается устройством ввода в течение времени T_{14}^2 (момент завершения — $t_{14} = t_{12} + T_{14}^2$) и процессором — T_{11}^2 (момент завершения — $t_{19} = t_{14} + T_{11}^2$). В момент t_{13} состояние системы не изменяется, но вычисляется второе значение времени реакции по потоку X_2 :

$$u_2^2 = t_{13} - t_7$$

В момент времени t_{15} систему поступает третья заявка потока X_2 . Определяется момент поступления четвертой заявки потока $X_2 - t_{17} = t_{15} + \tau_2^4$ (предполагается, что пользователь может посылать запросы, не дожидаясь ответов на предыдущие запросы). Третья заявка обслуживается монитором в течение времени T_{23}^3 , но с момента окончания обслуживания ($t_{16} = t_{15} + T_{23}^3$) переходит в состояние ожидания, так как занят процессор.

Следующее минимальное время t_{17} , — это время поступления четвертой заявки потока X_2 . После ее обслуживания монитором (момент завершения $t_{18} = t_{17} + T_{23}^4$) она также переходит в ожидание, т. е. образуется очередь из двух заявок. После освобождения процессора в момент t_{19} начнется обслуживание процессором третьей заявки потока X_2 , а затем с момента $t_{21} = t_{19} + T_{21}^3$ — монитором. По завершении этого обслуживания ($t_{23} = t_{21} + T_{23}^3$) можно будет вычислить третье значение времени реакции по потоку X_2 :

$$u_2^3 = t_{23} - t_{15}$$

С момента времени t_{19} принтер начнет обслуживание второй заявки потока X_1 и завершит его к моменту $t_{22} = t_{19} + T_{12}^2$, после чего определяется второе значение времени реакции по потоку X_1 :

$$u_1^2 = t_{22} - t_{12}$$

Указанные процедуры выполняются до истечения времени моделирования. В результате получается некоторое количество (выборка) случайных значений времен реакции $\{u_1\}$ и $\{u_2\}$ по первому и второму потокам. По этим значениям могут быть определены эмпирические функции распределения и вычислены количественные вероятностные характеристики времен реакции. В процессе моделирования можно суммировать продолжительности занятости каждого устройства обслуживанием всех потоков. Например, на рис. 1 занятость процессора 1 выделена заштрихованными ступеньками. Если результаты суммирования разделить на время моделирования, то получатся коэффициенты загрузки устройств.

Одновременно появляется возможность определения таких характеристик системы, как время ожидания заявок в очереди, число заявок, обслуженных системой, средняя и максимальная длина очереди заявок к каждому из устройств, требуемая емкость памяти и некоторые другие характеристики.

Имитационное моделирование дает возможность учесть надежностные характеристики ВС. В частности, если известны времена наработки на отказ и восстановления всех входящих в систему устройств, определяются моменты возникновения отказов устройств в период моделирования и моменты восстановления. Если в моменты возникновения отказа устройство занято обслуживанием заявки, то может приниматься разное решение в зависимости от типа устройства и режима его работы: заявка снимается и больше не обслуживается (выбывает из системы) или заявка помещается в очередь, а после восстановления устройства дообслуживается или поступает на повторное обслуживание.

2. ОБОБЩЕННЫЕ АЛГОРИТМЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Алгоритм моделирования по принципу особых состояний. В изложенном выше примере моделирование проводилось по принципу особых состояний (событий). В качестве особых событий выделены поступление заявки в систему, освобождение элемента после обслуживания заявки, завершение моделирования. В общем случае в системе могут быть выделены события и других типов, например возникновение отказа устройства в процессе обслуживания заявки и завершение восстановления устройства после отказа.

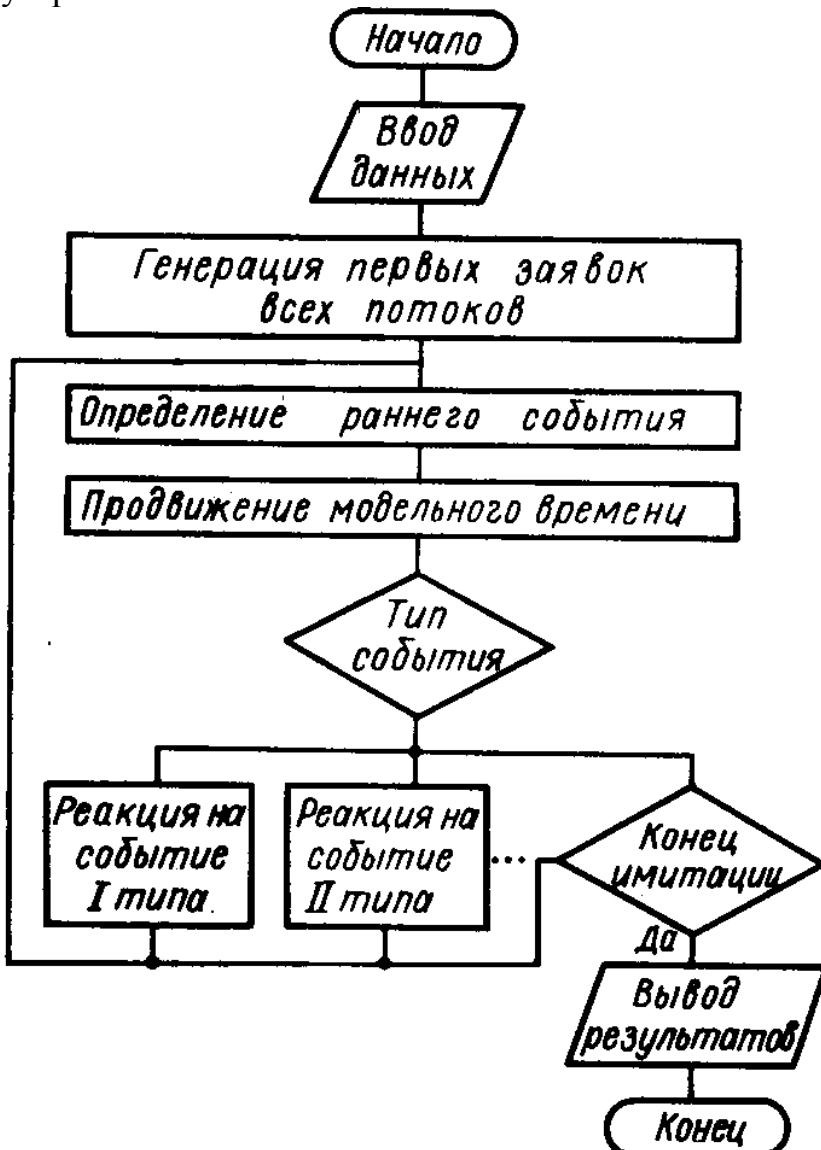


Рис. 3. Алгоритм моделирования по принципу особых состояний

Процесс имитации функционирования системы развивался во времени с использованием управляющих последовательностей, определяемых по функциям распределения вероятностей исходных данных путем проведения случайных испытаний. В качестве управляющих последовательностей использовались в примере последовательности значений периодов следования заявок по каждому i -му потоку

$\{\tau_i\}$ и длительностей обслуживания заявок i -го потока k -м устройством $\{T_{ik}\}$. Моменты наступления будущих событий определялись по простым рекуррентным соотношениям. Эта особенность дает возможность построить простой циклический алгоритм моделирования, который сводится к следующим действиям:

- 1) определяется событие с минимальным временем — наиболее раннее событие;
- 2) модельному времени присваивается значение времени наступления наиболее раннего события;
- 3) определяется тип события;
- 4) в зависимости от типа события предпринимаются действия, направленные на загрузку устройств и продвижение заявок в соответствии с алгоритмами их обработки, и вычисляются моменты наступления будущих событий; эти действия называют реакцией модели на события;
- 5) перечисленные действия повторяются до истечения времени моделирования.

В процессе моделирования производится измерение и статистическая обработка значений выходных характеристик. Обобщенная схема алгоритма моделирования по принципу особых состояний (принцип δz) изображена на рис. 2. Вначале производится инициализация моделирующей программы — подготавливаются массивы, вводятся и размещаются в оперативной памяти входные данные, настраиваются датчики случайных чисел. Затем генерируются первые заявки по каждому потоку — определяются моменты их поступления в систему и конкретизируются другие параметры.

Остальные операции выполняются в цикле. Определяется момент наступления наиболее раннего события и до этого момента смещается модельное время. После определения типа события реализуется соответствующая реакция на событие. В процессе этих действий могут возникать те или иные события в будущем, что фиксируется в массиве событий. Реакция на определенное событие, кроме всего прочего, приводит, как правило, к исключению его из массива событий. Кончается каждая реакция возвратом к определению события с минимальным временем.

Такой цикл повторяется до достижения конца моделирования, после чего завершается обработка статистики и выводятся результаты. На этом заканчивается модельный эксперимент.

В системе-оригинале могут протекать одновременно несколько процессов. Это приводит к тому, что в один и тот же момент времени в системе может возникнуть несколько особых событий. При моделировании на однопроцессорной ВС эти ситуации разрешаются таким образом: последовательно в установленном порядке реализуются реакции на все одновременно возникшие ситуации без продвижения модельного времени. Это называется псевдопараллельной имитацией нескольких процессов.

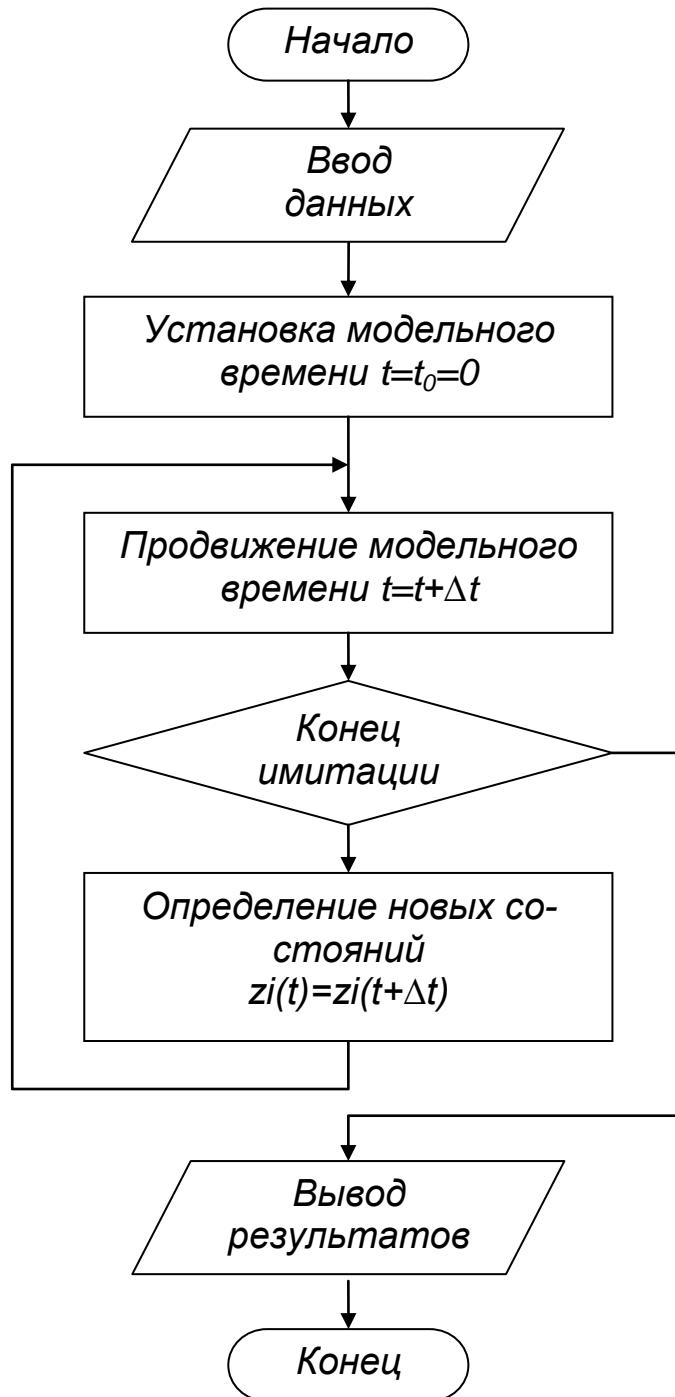


Рис. 3. Алгоритм моделирования по принципу временных приращений

Алгоритм моделирования по принципу Δt . Укрупненная схема моделирующего алгоритма, который реализует принцип постоянного приращения модельного времени (принцип Δt), представлена на рис. 3. Как и в предыдущем алгоритме, вначале инициализируется программа, в частности, вводятся значения $z_i(t_0)$, $i = 1, \dots, n$, которые характеризуют состояние системы в n -мерном фазовом пространстве состояний в начальный момент времени t_0 . Модельное время устанавливается $t = t_0 = 0$.

Основные операции, с помощью которых имитируется функционирование системы, выполняются в цикле. Функционирование системы отслеживается по последовательной смене состояний $Z_i(t)$. Для этого модельному времени дается некоторое приращение Δt . Затем по вектору текущих состояний определяются новые состояния $z_i(t + \Delta t)$, которые становятся текущими. Для определения новых состояний по текущим в формализованном описании системы должны существовать необходимые математические зависимости. Такой цикл продолжается до тех пор, пока текущее модельное время меньше заданного времени моделирования T_m .

По ходу имитации измеряются, фиксируются и обрабатываются требуемые выходные характеристики. При $t \geq T_m$ завершается обработка измерений и выводятся результаты моделирования.

При моделировании стохастических систем вместо новых состояний вычисляются распределения вероятностей для возможных состояний. Конкретные значения вектора текущего состояния определяются по результатам случайных испытаний. В результате проведения имитационного эксперимента получается одна из возможных реализаций случайного многомерного процесса в заданном интервале времени (t_0, T_m) .

Моделирующий алгоритм, основанный на принципе Δt , применим для более широкого круга систем, чем алгоритм, построенный по принципу особых состояний. Однако при его реализации возникают проблемы с определением величины Δt . Для моделирования ВС на системном уровне в основном применяется принцип особых состояний.

Контрольные вопросы

1. Общее определение имитационного моделирования.
2. Имитация функционирования ВС.
3. Алгоритм моделирование по принципу особых состояний.
4. Алгоритм моделирования по принципу временных приращений.

Литература

- 55.
- 56.
- 57.
- 58.
- 59.

ЛЕКЦИЯ 13. МЕТОДЫ ИССЛЕДОВАНИЯ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ (2 ЧАСА)

План

1. Методы определения характеристик вычислительных систем
2. Метод повторных экспериментов
3. Методы генерации случайных величин и последовательностей

1. МЕТОДЫ ОПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

Измеряемые характеристики ВС. При имитационном моделировании можно измерять значения любых характеристик, интересующих исследователя. Обычно по результатам измерений вычисляют характеристики всей системы, каждого потока и устройства.

Для всей ВС производится подсчет поступивших в систему заявок, полностью обслуженных и покинувших систему заявок без обслуживания по тем или иным причинам. Соотношения этих величин характеризуют производительность ВС при определенной рабочей нагрузке.

По каждому потоку заявок могут вычисляться времена реакции и ожидания, количества обслуженных и потерянных заявок. По каждому устройству зачастую определяются время загрузки при обслуживании одной заявки и число обслуженных устройствам заявок, время простоя устройства в результате отказов и количество отказов, возникших в процессе моделирования, длины очередей и занимаемые емкости памяти.

При статистическом моделировании большая часть характеристик — это случайные величины. По каждой такой характеристике у определяется N значений, по которым строится гистограмма относительных частот, вычисляются математическое ожидание, дисперсия и моменты более высокого порядка, определяются средние по времени и максимальные значения. Коэффициенты загрузки устройств вычисляются по формуле

$$\rho_k = \vartheta_k N_{ok} / T_m,$$

где ρ_k — коэффициент загрузки k -го устройства; ϑ_k — среднее время обслуживания одной заявки k -м устройством; N_{ok} — количество обслуженных устройством заявок за время моделирования T_m

Определение условий удовлетворения стохастических ограничений при имитационном моделировании производится путем простого подсчета количеств измерений, вышедших и не вышедших за допустимые пределы.

Расчет математического ожидания и дисперсии выходной характеристики. В случае анализа стационарного эргодического процесса функционирования системы вычисление математического ожидания и дисперсии характеристики у

производится усреднением не по времени, а по множеству N значений, измеренных по одной реализации процесса достаточной продолжительности. В целях экономии основной памяти ВС, на которой проводится моделирование, математическое ожидание и дисперсия вычисляются в ходе моделирования путем наращивания итогов при появлении очередного измерения случайной характеристики по рекуррентным формулам. Математическое ожидание случайной величины y для ее n -го измерения y_n :

$$m_n = m_{n-1} \frac{n-1}{n} + \frac{y_n}{n},$$

где m_{n-1} — математическое ожидание, вычисленное по предыдущим $n - 1$ измерениям.

Дисперсия для n -го измерения:

$$d_n = d_{n-1} \frac{n-2}{n-1} + \frac{1}{n} (y_n - m_{n-1})^2,$$

где d_{n-1} — дисперсия, вычисленная по предыдущим $n - 1$ измерениям. Вначале дисперсия принимается равной нулю.

При большом количестве измерений эти оценки являются состоятельными и несмещенными.

Расчет среднего по времени значения выходной характеристики. В процессе моделирования постоянно ведется подсчет длины очереди к каждому устройству и занимаемой емкости накопителей. При этом отслеживается максимальное значение и вычисляется среднее по времени значение. Например, средняя длина очереди вычисляется по формуле

$$l_N = \sum_{i=1}^N l_i \tau_i / T_m,$$

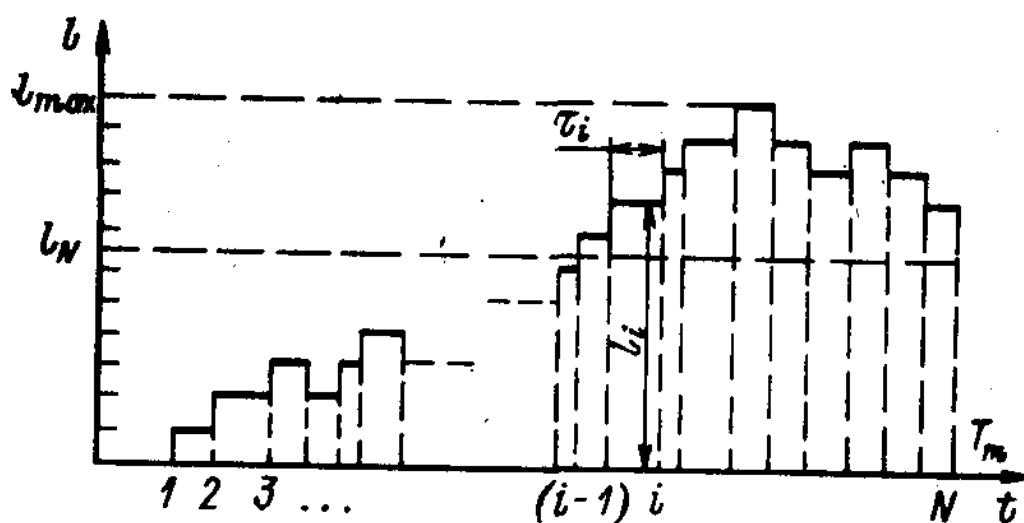


Рис. 1. Временная диаграмма изменения длины очереди

где i — номер очередного изменения состояния очереди (занесения заявки в очередь или исключения из очереди) (рис. 1); N —количество изменений состояния очереди; τ_i ; — интервал времени от ($i-1$)-го до 1-го изменения состояния; l_i — число заявок в очереди в интервале

По аналогичной формуле вычисляется средняя по времени используемая емкость накопителя:

$$q_n = \sum_{i=1}^N q_i \tau_i / T_m,$$

где q_i — емкость накопителя, занятая в интервале времени между двумя последовательными обращениями к накопителю.

Формулы (4) и (5) приводят к виду, удобному для вычисления путем наращивания итогов.

Построение гистограммы. Основное достоинство имитационного моделирования заключается в том, что по любой выходной характеристике может быть построена гистограмма относительных частот — эмпирическая плотность распределения вероятностей, вне зависимости от сочетаний распределений параметров системы и внешних воздействий. При исследовании стационарной системы гистограмма строится по следующей методике.

Перед началом моделирования задаются предположительные границы изменения интересующей выходной характеристики y , т. е. нижний y_u и верхний y_v пределы, и указывается число интервалов гистограммы N_g . По этим данным вычисляется интервал

$$\Delta = (y_v - y_u) / N_g.$$

Затем в процессе моделирования по мере появления измерений характеристики y , определяется число попаданий этой случайной величины в i -й интервал гистограммы R_i ; и подсчитывается общее число измерений N . По полученным данным вычисляется относительная частота по каждому интервалу:

$$G_i = R_i / (N\Delta).$$

Этих данных достаточно для построения гистограммы относительных частот: на оси абсцисс откладываются пределы изменения анализируемой характеристики y ; весь диапазон изменения подразделяется на заданное число интервалов; над каждым t -м интервалом проводится отрезок, параллельный оси абсцисс, на расстоянии, равном G_i от оси абсцисс (рис. 2).

Рис. 2. Гистограмма относительных частот

Отметим, что площадь гистограммы относительных частот равна единице, так же, как интеграл от плотности вероятности в пределах от минус до плюс бесконечности равен единице. Площадь гистограммы равняется сумме площадей прямоугольников, построенных на каждом i -ом интервале:

$$S_g = \sum_{i=1}^{N_g} G_i \Delta = \sum_{i=1}^{N_g} \frac{R_i}{N\Delta} \Delta = \sum_{i=1}^{N_g} \frac{R_i}{N} = 1,$$

поскольку общее число измерений характеристики y равно сумме чисел попаданий в каждый из интервалов:

$$N = \sum_{i=1}^{N_g} R_i.$$

При необходимости выдвигается гипотеза о том, что полученное эмпирическое распределение согласуется с некоторым теоретическим распределением, имеющим аналитическое выражение для функции или плотности распределения. Эта гипотеза проверяется по тому или другому критерию. Например, при использовании критерия хи-квадрат в качестве меры расхождения используется выражение

$$\chi^2 = \frac{\sum_{i=1}^{N_g} (R_i - N_{p_i})^2}{N_{p_i}},$$

где p_i — определенная из выбранного теоретического распределения вероятность попадания случайной величины в i -й интервал.

Из теоремы Пирсона следует, что для любой функции распределения $F(y)$ случайной величины y при $N \rightarrow \infty$ распределение величины χ^2 имеет вид

$$H_k(z) = P(\chi^2 < z) = \frac{1}{2^{k/2} \Gamma(k/2)} \int_0^z t^{k/2-1} e^{-t/2} dt \quad (7)$$

где z — значение случайной величины χ^2 ; $k = Ng - (r + 1)$ — число степеней свободы распределения хи-квадрат; r — количество параметров теоретического закона распределения; $\Gamma(k/2)$ — гамма-функция.

Функция распределения хи-квадрат табулирована по вычисленному значению и числу степеней свободы с помощью таблиц определяется вероятность (7). Если она превышает заданный уровень значимости C , то выдвинутая гипотеза принимается.

2. МЕТОД ПОВТОРНЫХ ЭКСПЕРИМЕНТОВ

Характеристики нестационарных процессов. Для системы, процесс функционирования которой отличается от стационарного эргодического, нельзя вычислять вероятностные характеристики по одной реализации процесса, поскольку оценки могут получиться смещенными или несостоительными. Следует ожидать, что если на ВС поступает поток заявок, интенсивность которого изменяется во времени, как это показано, например, на рис. 3. выходные характеристики такой системы относятся к нестационарным случайным функциям (процессам).

Известно, что в основе всех формальных методов лежит представление случайного процесса с помощью ансамбля реализации и описание его посредством характеристик, получаемых усреднением по ансамблю. Предположим, что случайный процесс $Y(t)$ задан ансамблем реализации $\{y_i(t)\} (i=1,2,\dots)$, а интересующая исследователя вероятностная характеристика в $\Theta[Y(t)]$ определяется предельным соотношением

$$\Theta[Y(t)] = \lim_{N_i \rightarrow \infty} \frac{1}{N_i} \sum_{i=1}^{N_i} \theta[y_i(t)],$$

где θ — оператор преобразования, лежащий в основе определения характеристики Θ ; N_i — количество реализаций, по которым производится усреднение.

При усреднении по ограниченной совокупности реализаций прямые измерения значений вероятностной характеристики в $\Theta[Y(t)]$ могут быть выполнены в соответствии с формулой

$$\Theta^*[Y(t)] = \frac{1}{N_i} \sum_{i=1}^{N_i} \theta[y_i(t)].$$

Полной вероятностной характеристикой случайного процесса может служить многомерная функция распределения вероятности мгновенных значений процесса. Случайный процесс $Y(t)$ считается исчерпывающе описанным в вероятностном смысле на интервале $(t_1, \dots, t_{N_r}) \in (0, T)$, если задана его N_r -мерная функция распределения:

$$F_{N_r}(Y_1, t_1, \dots, Y_{N_r}, t_{N_r}) = P[Y(t_1) \leq Y_1, \dots, Y(t_{N_r}) \leq Y_{N_r}] \quad (10)$$

которая соответствует любому сочетанию моментов времени t_r на интервале $(0, T)$ при произвольном N_r , в том числе при $N_r \rightarrow \infty$. Исходя из этих предпосылок, при исследовании стохастических систем необходимо получать для каждой выходной характеристики совокупности N_i реализаций, причем по каждой реализации следует измерять значения по N_r сечениям (отсчетам), как это показано на рис. 3. Таким способом могут быть получены вероятностные оценки выходных характеристик, если они являются нестационарными или стационарными неэргодическими процессами.

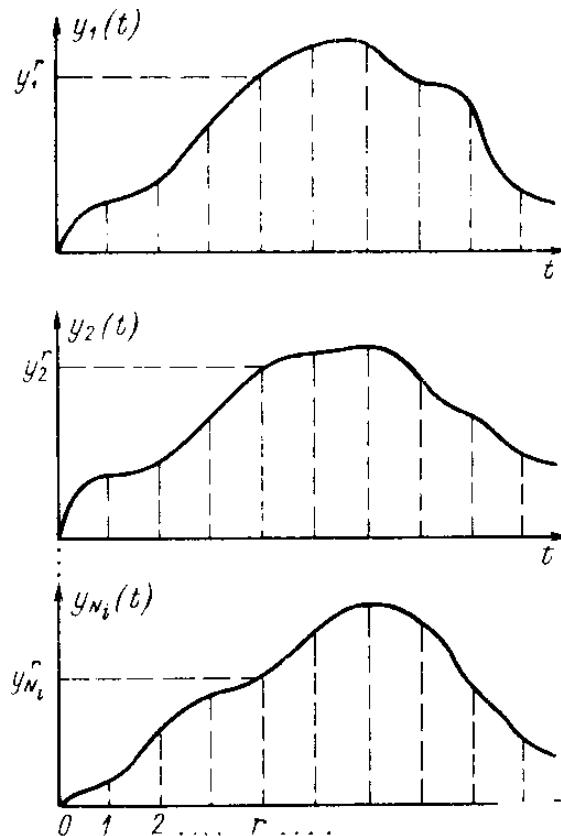


Рис. 7. Реализации нестационарного случайного процесса

Алгоритм повторных экспериментов. При имитационном моделировании нестационарного режима функционирования ВС требуется получить N_i реализаций случайных процессов по всем выходным характеристикам. С этой целью необходимо проводить N_i имитационных экспериментов в интересующей исследователя области определения случайных процессов (O, T).

Изменение характера сочетаний случайных событий в каждом последующем эксперименте может быть достигнуто заменой начальных значений датчиков случайных чисел, используемых в процессе моделирования, в частности, для генерации управляющих последовательностей. Это можно реализовать, например, так: при каждом последующем эксперименте в качестве начальных значений датчиков случайных чисел используют последние значения предыдущего эксперимента.

Исследование систем с нестационарным режимом путем имитационного моделирования с применением метода повторных экспериментов сводится к следующему. Проводится N_i имитационных экспериментов. При каждом последующем эксперименте параметры системы и нагрузки устанавливаются на исходные значения и в процессе каждого эксперимента остаются неизменными или изменяются по одним и тем же зависимостям. Датчики случайных чисел устанавливаются в начальное состояние только один раз — перед моделированием и до конца моделирования вырабатывают последовательности случайных чисел.

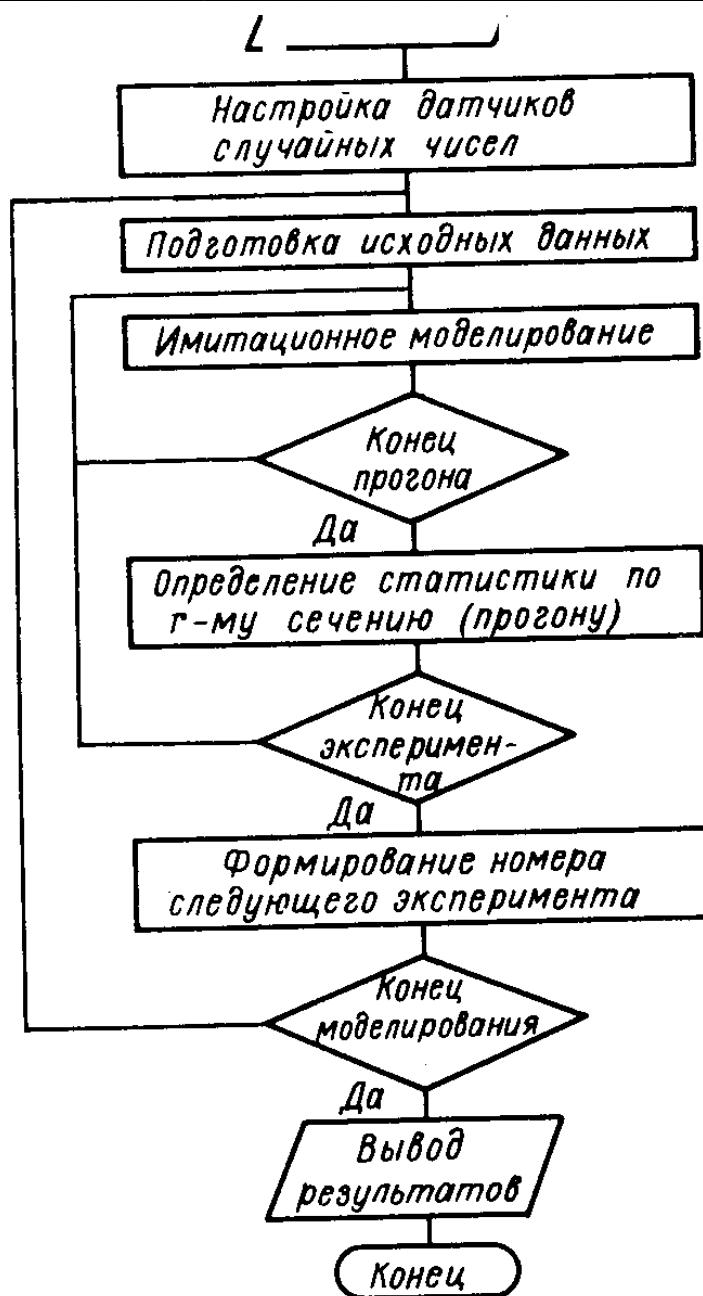


Рис. 4. Алгоритм моделирования по методу повторных экспериментов

Период моделирования T_m по каждому эксперименту разделяется на N_r сечений. Интервал времени моделирования между двумя соседними сечениями будем называть прогоном. В каждом эксперименте для фиксированных моментов времени $t_r (1, 2, \dots, N_r)$ определяются численные значения выходных характеристик.

По каждому i-му сечению для всех выходных характеристик может определяться эмпирическая многомерная функция $K_y(t_r, t'_r)$ или плотность распределения, оцениваться математическое ожидание $m_y(t_r)$, корреляционная функция или дисперсия $D_y(t_r)$ по всей совокупности N_i реализаций.

Для обеспечения статистической достоверности результатов обычно требуется проведение нескольких сотен или тысяч экспериментов с измерением и обработкой вектора выходных характеристик по нескольким десяткам сечений. Очевидно, что

выполнение этой работы «вручную» не представляется возможным в приемлемые сроки. Это означает, что на ВС должно быть возложено не только проведение имитации, но и реализация метода повторных экспериментов.

Машинный алгоритм повторных экспериментов представлен на рис. 4. В первом блоке выполняется ввод данных о моделируемой системе и нагрузке, а также производится инициализация программы. Отдельно выделен блок настройки датчиков случайных чисел, чтобы подчеркнуть, что датчикам задаются начальные значения до основных циклов моделирования. В дальнейшем датчики вырабатывают неповторяющиеся последовательности случайных чисел. Затем вводятся и размещаются в соответствующих массивах и переменных исходные данные. В частности, задаются количества прогонов N_r , экспериментов N_i и период моделирования T_m .

В следующем блоке выполняется имитационное моделирование процесса функционирования так, как это описано в п. 2, по тому или другому алгоритму. В ходе имитации изменяются значения тех переменных параметров, которые заданы как функции времени, и постоянно отслеживается достижение конца прогона, т. е. событие, когда текущее модельное время станет равным t_r . При выполнении этого условия определяются, вычисляются и запоминаются статистические данные по г-му сечению, после чего имитация продолжается.

Если выполнено N_r прогонов, т. е. завершен очередной эксперимент, формируется номер следующего эксперимента и управление передается блоку подготовки исходных данных. После проведения N_i экспериментов завершается обработка и вывод результатов моделирования.

Расчет характеристик по методу повторных экспериментов. При анализе систем с нестационарным процессом функционирования путем имитационного моделирования с использованием метода повторных экспериментов для каждой выходной характеристики, которая является случайным процессом, зачастую оценивается математическое ожидание, корреляционная функция или дисперсия, а также могут быть построены по каждому сечению гистограммы, чем определяется многомерная плотность распределения вероятностей.

Математическое ожидание случайного процесса $Y(t)$ — это неслучайная функция $m_y(t)$, которая при каждом значении аргумента t_r представляет собой математическое ожидание соответствующего сечения случайной функции:

$$m_y(t_r) = M[Y(t_r)].$$

Корреляционная функция случайного процесса — это тоже неслучайная функция двух аргументов $K_y(t, t')$, которая при каждой паре значений аргументов t, t' равна корреляционному моменту соответствующих сечений случайной функции:

$$K_y(t_r, t'_r) = M[Y(t_r) - m_y(t_r)(Y(t'_r) - m_y(t'_r))].$$

При равенстве $t=t'$ корреляционная функция превращается в дисперсию случайной функции:

$$\mathbf{K}_y(t_r, t_r) = \mathbf{D}_y(t_r) = \mathbf{D}[y(t_r)].$$

В соответствии с формулой (9) N_r -мерная плотность распределения вероятностей оценивается по формуле:

$$f_{N_r}(Y_1, t_1, \dots, Y_{N_r}, t_{N_r}) = \frac{1}{N_i} \sum_{i=1}^{N_i} \psi_i[Y_1 | t_1, \dots, Y_{N_r} | t_{N_r}] \quad (11)$$

где значения функции $\psi_i[\dots]$ равны 1 при $Y_1 < y_i(t_r) \leq Y_1 + dY, \dots, Y_{N_r} < y_i(t_{N_r}) \leq Y_{N_r} + dY$ и равны 0, если хотя бы одно из этих неравенств для i -й реализации не выполняется.

В результате вычисления по формуле (11) получается многомерная плотность распределения вероятностей, подобная той, которая изображена на рис. Важно отметить, что для проверки выполнения стохастических ограничительных условий типа (1.5) нет необходимости вычислять плотность распределения по формуле (11), а достаточно подсчитать общее количество N_y и количество N_p значений измерения случайной величины, попадающих между ограничивающими пределами, по всей совокупности экспериментов, а затем определить вероятность выполнения ограничительного условия:

$$P_y = N_p / N_y \geq \beta.$$

При моделировании вычисления результатов производятся в конце каждого прогона путем наращивания итогов. В частности, количественные вероятностные значения таких выходных характеристик, как длины очередей к каждому устройству, времена реакции по каждому потоку заявок и времена загрузки каждого устройства, определяются по r -му сечению для i -го эксперимента по следующим формулам. Оценка математического ожидания длины очереди к устройству:

$$L_i = L_{i-1}(i-1)/i + l_i/i,$$

где L_{i-1} — математическое ожидание длины очереди за предыдущие ($i-1$) экспериментов; l_i — длина очереди по r -му сечению для i -го эксперимента.

Дисперсия длины очереди

$$D[L_i] = \frac{1}{i-1} (D[L_{i-1}](i-2) + (l_i - L_i)^2), \quad (14)$$

где $D[L_{i-1}]$ — дисперсия длины очереди за предыдущие ($i-1$) экспериментов; в первом эксперименте дисперсия принимается равной нулю.

По временам реакции и загрузки при достаточно большом количестве сечений, когда процесс можно считать стационарным на протяжении одного прогона, теку-

щие значения математических ожиданий и дисперсии вычисляются по формулам (2) и (3) соответственно. Тогда в r -м сечении i -го эксперимента оценка математического ожидания времени загрузки каждого устройства при обслуживании одной заявки:

$$V_i = V_{i-1} \frac{N_{z(i-1)}}{N_{z(i-1)} + n_{zi}} + \vartheta_i \frac{n_{zi}}{N_{z(i-1)} + n_{zi}} \quad (15)$$

где V_{i-1} — математическое ожидание времени загрузки за предыдущие ($i-1$) экспериментов; $N_{z(i-1)}$ — число загрузок устройства на r -м прогоне по предыдущим ($i-1$) экспериментам; ϑ_i — математическое ожидание времени загрузки на r -м прогоне в i -м эксперименте; n_{zi} — число загрузок устройства на r -м прогоне в i -м эксперименте.

Дисперсия времени загрузки каждого устройства на r -м прогоне

$$D[V_i] = \frac{1}{N_{z(i-1)} + n_{zi}} (D[V_{i-1}])(N_{z(i-1)} - 1) + d_i (n_{zi} - 1) + \frac{N_{z(i-1)} n_{zi}}{N_{z(i-1)} + n_{zi}} (\vartheta_i - V_{i-1})^2,$$

где $D[V_{i-1}]$ — дисперсия, вычисленная по предыдущим ($i-1$) экспериментам; d_i — дисперсия, вычисленная в i -м эксперименте.

Коэффициенты загрузки устройств вычисляются по формуле (1).

Оценки математического ожидания и дисперсии времени реакции U_i по каждому потоку в r -м сечении i -го эксперимента выполняются по формулам, аналогичным (15) и (16). Для экономии места в памяти моделирующей ВС обновленные статистические характеристики по r -му сечению записываются на место прежних, вычисленных в ($i-1$) эксперименте.

3. МЕТОДЫ ГЕНЕРАЦИИ СЛУЧАЙНЫХ ВЕЛИЧИН И ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Датчики равномерно распределенных случайных чисел. При статистическом моделировании стохастических систем возникает необходимость в определении случайных событий, величин и последовательностей по заданным статистическим характеристикам. В основе их определения лежит использование последовательности чисел, равномерно распределенных в интервале (0,1). Программы ВС, формирующие такие последовательности, называют датчиками или генераторами случайных чисел.

Для получения последовательности равномерно распределенных случайных чисел с помощью ВС часто используется мультиплексионный способ. Случайные числа получаются из рекуррентного соотношения

$$\alpha_i = (A\alpha_{i-1} + C) \bmod M, \quad (16)$$

где A, C — константы; M — достаточно большое положительное целое число.

При соответствующем выборе констант и задании некоторого начального значения α_0 эта формула позволяет получать последовательность целых чисел, равномерно распределенных в интервале $(0, M - 1)$. Последовательность имеет период повторения, равный M . Поэтому точнее называть числа псевдослучайными. Случайные числа, равномерно распределенные в интервале $(0,1)$, находятся масштабным преобразованием полученных целых чисел.

Моделирование случайных событий и дискретных величин. Для моделирования случайного события X , наступающего с вероятностью P , берется значение α случайного числа, равномерно распределенного на интервале $(0,1)$, и сравнивается с P . Если $\alpha \leq P$, то считается, что произошло событие X .

Предположим, что дискретная случайная величина Y может принимать значения y_1, \dots, y_n вероятностями p_1, \dots, p_n . При этом

$$y_1 < y_2 < \dots < y_n; \sum_{i=1}^n p_i = 1.$$

Тогда берется значение α случайного числа, распределенного равномерно на интервале $(0, 1)$, и определяется такое k , принадлежащее совокупности $(1, n)$, при котором удовлетворяется неравенство

$$\sum_{i=1}^{k-1} p_i \leq \alpha \leq \sum_{i=1}^k p_i, \quad k \in (1, n).$$

Тогда случайная величина Y принимает значение y_k .

Моделирование случайных непрерывных величин. Пусть непрерывная случайная величина Y имеет произвольный закон распределения. Предположим, что она задается эмпирической плотностью распределения $f^*(y)$ — гистограммой (рис. 5, а). Из гистограммы определяется эмпирическая функция распределения $F^*(y)$ — дискретная кумулятивная функция (рис. 5, б) для середин интервалов группирования случайной величины в пределах (y_{\min}, y_{\max}) .

Для определения одного конкретного значения случайной величины Y берется значение α случайного числа, равномерно распределенного на интервале $(0, 1)$. Затем находится такое k , при котором

$$F_{k-1} < \alpha \leq F_k$$

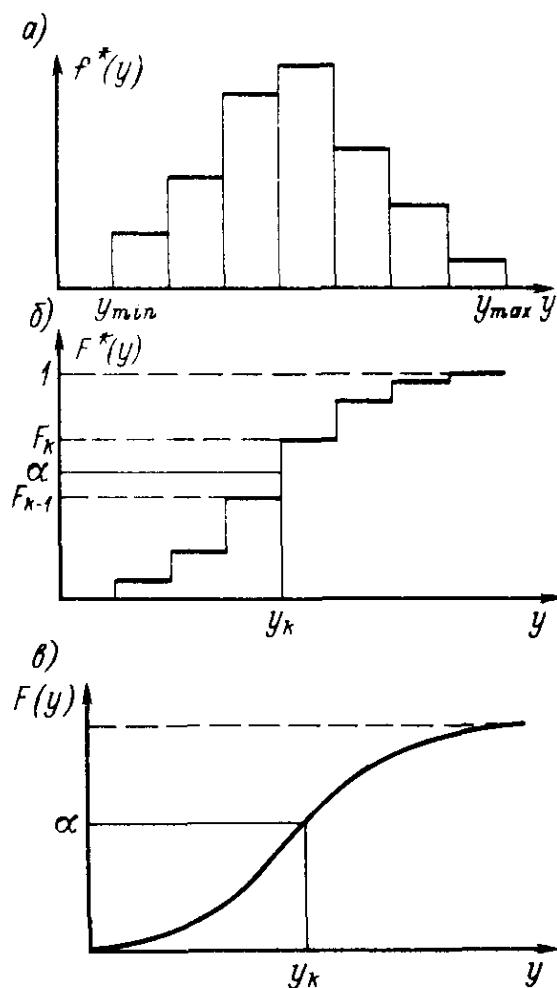


Рис. 5. Графики для определения значения случайного числа по дискретной и интегральной функции распределения

Тогда искомое случайное число равное (рис. 5, б). Это же правило применимо и при задании случайной непрерывной величины интегральной функцией распределения $F(y)$, как показано на рис. 5, б. Оно вытекает из теоремы: если случайная величина Y имеет плотность распределения $F(y)$, то ее распределение

$$F(y) = \int_{-\infty}^{+\infty} f(y) dy$$

является равномерным на интервале $(0, 1)$.

Для некоторых законов распределения (экспоненциальный, Эрланга) имеются простые аналитические зависимости $y = \Phi(\alpha)$. Например, пусть требуется получить конкретное значение случайного числа Y с экспоненциальным законом распределения. Подставим в формулу (18) a и плотность распределения:

$$\alpha = \int_0^{y_k} \lambda e^{-\lambda y} dy, \quad y > 0.$$

После интегрирования получается

$$\alpha = 1 - e^{-\lambda y_k}.$$

Разрешая это уравнение относительно y_k , имеем

$$y_k = -\ln(1-\alpha) / \lambda.$$

Учитывая, что если случайная величина ξ подчинена равномерному закону распределения в интервале $(0, 1)$, то случайная величина $\psi = 1 - \xi$ также имеет равномерный закон распределения в интервале $(0, 1)$, последнее соотношение можно заменить следующим:

$$y_k = -\ln \alpha / \lambda$$

Процедура определения конкретного значения случайного числа с заданным законом распределения называется случайным испытанием или «бросанием жребия».

Моделирование случайных процессов сводится на практике к определению последовательностей случайных величин. Исходными данными являются функции распределения, определенные в требуемые моменты времени, и последовательность случайных чисел $\alpha_1, \alpha_2, \dots$, подчиняющаяся равномерному закону распределения в интервале $(0, 1)$. Конкретные значения случайных процессов в нужные моменты времени находят по изложенным выше правилам.

В большом числе публикаций, рассматриваются вопросы алгоритмизации, программирования и исследования качества датчиков случайных чисел.

В процессе статистического моделирования существует необходимость в частом обращении к датчикам случайных чисел. С их помощью формируются конкретные значения случайных параметров каждой заявки, параметров обслуживания заявки каждым устройством; определяются пути продвижения заявки по тому или иному маршруту при вероятностной дисциплине маршрутизации и т. д. Имитационное моделирование систем по принципу особых состояний проводится с постоянным использованием датчиков случайных чисел для формирования всех управляющих последовательностей.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Измеряные характеристики ВС.
2. Основные формулы для расчета выходных характеристик ВС.
3. Методика пострения гистограммы и ее использование для исследования ВС.
4. Основная идея метода повторных экспериментов.
5. Что понимается под датчиком случайных величин?
6. Какие методы (алгоритмы, программы) генерации последовательностей случайных величин вы знаете?
7. Примеры использования датчиков случайных чисел.

ЛИТЕРАТУРА

- 60.
- 61.
- 62.
- 63.
- 64.

ЛЕКЦИЯ 14. МОДЕЛИ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 ЧАСА)

План

1. Модели линейной дискретной оптимизации с булевыми переменными
2. Преобразование задачи с дискретными переменными к задаче с булевыми переменными
3. Преобразование задачи линейного булева программирования к задаче нелинейного булева программирования

1. МОДЕЛИ ЛИНЕЙНОЙ ДИСКРЕТНОЙ ОПТИМИЗАЦИИ С БУЛЕВЫМИ ПЕРЕМЕННЫМИ

Задача оптимизации линейной целевой функции с булевыми переменными и линейными ограничениями является одной из самых распространенных моделей дискретного программирования. Комбинаторные задачи с булевыми переменными, принимающими значения 0 или 1, встречаются при решении многих практических проблем из экономики, проектирования, управления и других областей.

Задача оптимизации (минимизации или максимизации) линейной целевой функции с булевыми переменными и линейными ограничениями в общем виде описывается с помощью одной из следующих моделей линейного булева программирования:

I. Модель А - для задачи минимизации

$$q(x) = (c, x) = \sum_{j=1}^n c_j x_j \longrightarrow \min_{x_j \in \{0;1\}} \quad (1)$$

$$h_i(x) = (a_i, x) = \sum_{j=1}^n a_{ij} x_j \geq b_i, \quad (i = 1, 2, \dots, m) \quad (2)$$

II. Модель В - для задачи максимизации

$$q(x) = (c, x) = \sum_{j=1}^n c_j x_j \longrightarrow \max_{x_j \in \{0;1\}} \quad (3)$$

$$h_i(x) = (a_i, x) = \sum_{j=1}^n a_{ij} x_j \leq b_i, \quad (i = 1, 2, \dots, m) \quad (4)$$

т.е. требуется минимизировать или максимизировать линейную целевую функцию $q(x)$ по булевым переменным $x_j \in \{0,1\}$ при выполнении условия, задаваемого системой линейных неравенств вида (2) или (4).

Проблема отыскания решения задачи (1), (2) или (3), (4) может в принципе решаться с применением метода полного перебора, суть которого заключается в переборе всех булевых векторов заданной длины, проверке для каждого вектора выполнения линейных ограничений, вычислении значений целевой функции для допустимых векторов и выборе из них минимального или максимального значения целевой функции. Однако решение, полученное методом полного перебора, связа-

но с большим объемом вычислений, который неосуществим при больших размерностях задачи даже на сверхпроизводительных ЭВМ. Так как каждая из n компонент независимо от других может принимать два значения 0 или 1, поэтому общее число булевых векторов длины n равно 2^n .

Это величина и характеризует сложность алгоритма полного перебора. В связи с тем, что для большинства задач дискретной оптимизации полный перебор неосуществим, были разработаны различные методы неявного перебора, которые обеспечивают нахождение точного решения без пересмотра всех булевых векторов. Такими являются различные варианты метода отсечений, метода ветвей и границ, метода динамического программирования. Но опыт применения этих методов при решении реальных задач с достаточно большим числом переменных показал, что многие задачи линейного программирования с булевыми переменными еще являются нерешаемыми на ЭВМ из-за нехватки машинного времени. Следовательно, с ростом размера задачи n она становится "труднорешаемой", т.е. практически неразрешимой.

Задача (1), (2) или (3), (4) в числе многих задач комбинаторной оптимизации отнесена к классу "труднорешаемых" или, NP (no polynomial) - полных задач. Причина заключается в том, что алгоритма, решающего поставленную задачу за время, ограниченное полиномом от "размера задачи", в настоящее время нет.

В связи с этим исследования, направленные на разработку новых эффективных или полиномиальных методов решения задач линейного программирования с булевыми переменными, являются, несомненно, актуальными.

Сначала рассмотрим широко распространенные практические задачи, описываемые моделями линейного программирования с булевыми переменными.

2. ПРЕОБРАЗОВАНИЕ ЗАДАЧИ С ДИСКРЕТНЫМИ ПЕРЕМЕННЫМИ К ЗАДАЧЕ С БУЛЕВЫМИ ПЕРЕМЕННЫМИ

Пусть x дискретная переменная, принимающая только целые значения $0, 1, 2, \dots, k$. Тогда эту переменную можно представить как линейную комбинацию $(p+1)$ булевых переменных y_0, y_1, \dots, y_p , т.е.:

$$x = y_0 + 2y_1 + 2^2 y_2 + \dots + 2^p y_p , \quad (1)$$

где $y_i \in \{0;1\}$, ($i = 0, 1, 2, \dots, p$); p -наименьшее целое число, удовлетворяющее условию

$$k \leq 2^{p+1} - 1.$$

Пример. Рассмотрим следующую задачу целочисленного программирования

$$q(x_1, x_2) = 2x_1 + x_2 \rightarrow \max_{\{x_1, x_2\}}$$

$$3x_1 + 2x_2 \leq 15$$

$$0 \leq x_1 \leq 5$$

$$0 \leq x_2 \leq 3$$

x_1, x_2 – целые.

Переменная x_1 принимает шесть целых значений: 0, 1, 2, 3, 4, 5. Для этой переменной $k_1=5$. Возьмем для переменной x_1 наименьшее целое число p_1 , оно определяется из условия (2)

$$5 \leq 2^{p_1+1} - 1, \text{ откуда } p_1 = 2$$

Исходя из (1), можно произвести замену переменных:

$$x_1 = y_0 + 2y_1 + 4y_2, \quad y_0, y_1, y_2 \in \{0;1\}.$$

Переменная x_2 принимает четыре целых значения 0, 1, 2, 3.

Следовательно, для x_2 : $k_2 = 3$, $p_2 = 1$;

$$x_2 = y_3 + 2y_4, \quad y_3, y_4 \in \{0;1\}.$$

Исходная задача дискретного программирования преобразована к следующей задаче булева программирования :

$$q(y) = 2y_0 + 4y_1 + 8y_2 + y_3 + 2y_4 \rightarrow \max_{y_i \in \{0;1\}}$$

$$3y_0 + 6y_1 + 12y_2 + 2y_3 + 4y_4 \leq 15$$

$$y_0 + 2y_1 + 4y_2 \leq 5$$

$$y_3 + 2y_4 \leq 3$$

$$y_i \in \{0;1\}, \quad (i = 0, 1, 2, 3, 4).$$

Если дискретная переменная x принимает произвольные целые дискретные значения $c_1, c_2, c_3, \dots, c_k$, то в этом случае соотношения (1) и (2) соответственно преобразуются в следующие соотношения:

$$x = c_0 + y_0 + 2y_1 + 2^2 y_2 + \dots 2^p y_p,$$

$$c_k \leq 2^{p+1} - 1.$$

Таким образом, всегда можно ограничиться рассмотрением задачи булева программирования вместо задачи дискретного программирования.

3. ПРЕОБРАЗОВАНИЕ ЗАДАЧИ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ К ЗАДАЧЕ НЕЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ

Задача вида

$$\begin{aligned} q(x) = (c, x) &= \sum_{j=1}^n c_j x_j \rightarrow \max_{x_j \in [0;1]} \\ h(x) = (a, x) &= \sum_{j=1}^n a_j x_j \leq b \end{aligned} \quad (2)$$

в числе многих задач комбинаторной оптимизации отнесена к классу “труднорешаемых”. Следовательно, для эффективного решения на вычислительных машинах задач большого размера нужно искать алгоритмические принципы, позволяющие определять оптимальное решение без необходимости явно перечислять элементы множества булева вектора $x = (x_1, x_2, \dots, x_n) \in 2^n$.

Именно эта идея неявного (в противоположность явному) перечисления решений и лежит в основе методов, предлагаемых и исследуемых в данной книге. Идейная основа последних заключается в преобразовании исходной задачи (1), (2) к следующей задаче булева программирования:

$$\Phi(x) = \frac{\sum_{j=1}^n c_j x_j}{\sum_{j=1}^n a_j x_j} \rightarrow \max_{x_j \in [0;1]} . \quad (3)$$

При переходе от задачи (1), (2) к задаче (3) возникает вопрос: не теряется ли смысловое содержание исходной задачи при переходе к новой форме. Оказывается, не теряется, это объясняется следующим образом: Максимизируя функцию $\Phi(x)$ по булевым переменным $x = (x_1, x_2, \dots, x_n)$, практически мы, во-первых, максимизируем числитель выражения (3), т.е. линейную функцию булевых переменных

$$q(x) = (c, x) = \sum_{j=1}^n c_j x_j ,$$

что соответствует задаче (1). Во-вторых, при максимизации выражения (3), минимизируется знаменатель этого выражения, т.е.

$$(a, x) = \sum_{j=1}^n a_j x_j ,$$

Следовательно, минимизация выражения $(a, x) = \sum_{j=1}^n a_j x_j$, хотя это процедура

не обеспечивает точного выполнения условия (2), а способствует выполнению этого же условия.

В дальнейшем исследовании задачи о рюкзаке (см. п.п. 4.1) будет изложена вычислительная схема получения решения задачи (3) в виде упорядоченного ряда булевых переменных

$$x_{j_1}, x_{j_2}, \dots, x_{j_n} , \quad (4)$$

но этот ряд булевых переменных может не удовлетворять выполнению ограничения вида (2). После получения упорядоченного ряда (4) каждый очередной элемент этого ряда поочередно слева направо будет вставлен в выражения (2) и таким образом проверено выполнение условия

$$(a, x) = \sum_{j=1}^n a_j x_j \leq b$$

до максимального числа первых n' переменных упорядоченного ряда (4).

Таким образом, переход от исходной задачи линейного булева программирования вида (1), (2) к новой задаче – максимизации фишерского типа функционала вида (3) логичен.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Роль и место моделей булева программирования в проблеме оптимизации.
2. Основные виды моделей линейного булева программирования.
3. О полиномиальных и не полиномиальных задачах в дискретной оптимизации.
4. Преобразование задачи с дискретными переменными к задаче с булевыми переменными.
5. Преобразование задачи *** к задаче нелинейного булева программирования.
6. Как записывается функционал Фишерского типа.

ЛИТЕРАТУРА

- 65.
- 66.
- 67.
- 68.

ЛЕКЦИЯ 15. НОВЫЕ МОДЕЛИ ЗАДАЧ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 ЧАСА)

ПЛАН

1. Задача и модель оптимизации работы насосной станции
2. Модель задачи автоматической классификации
3. Задача об оптимизации размещения букв алфавита на клавиатуре ЭВМ

1. ЗАДАЧА И МОДЕЛЬ ОПТИМИЗАЦИИ РАБОТЫ НАСОСНОЙ СТАНЦИИ

Исследование и моделирование крупных насосных станций является одной из ключевых задач в системе машинного водоподъема (СМВ). В настоящее время в СМВ в основном используется диспетчерское управление, основанное на простых методах принятия решений, исходя из личного опыта и интуиции лица, принимающего решения (диспетчера) и решающего задачу управления для текущего момента времени. Такое управление приводит к перерасходу электроэнергии на водоподъем, непроизводительным сбросам и потерям воды, невыполнения графика водоподачи. Поэтому задачи, связанные с исследованием, моделированием и разработкой оптимальных алгоритмов управления работой насосной станции, особенно становятся актуальными в связи с переходом к рыночным условиям хозяйствования.

Исследуем работу насосной станции и определим ее основные управляющие параметры.

В крупных насосных станциях обычно устанавливается несколько насосных агрегатов, предназначенных для подъема воды на высоту определенного диапазона. В большинстве случаев насосная станция работает в режимах, при которых недоиспользуются полные возможности, заложенные в насосных агрегатах. Следовательно, возникает необходимость создания таких методов управления, которые позволяют максимально использовать все потенциальные возможности насосной станции и создать оптимальную систему управления по заданному критерию.

В насосных станциях используются крупные осевые насосные агрегаты типа "Р" и "ОП", центробежные типа "В". Для подъема воды на высоту до 25 м используются осевые насосные агрегаты, а на высоту более 25 м -центробежные. Для моделирования процесса водоподачи основными являются гидроэнергетические и расходные характеристики. Гидроэнергетические характеристики можно найти в каталогах. Расходная характеристика Q насосного агрегата зависит от высоты подъема H и от угла ψ разворота лопасти насосного агрегата:

$$Q = Q(H, \psi).$$

В каталогах насосных агрегатов расходная характеристика осевого насосного агрегата задается в виде семейства кривых при различных углах разворота лопастей:

$$Q_j(H) = Q(H, \psi_j), (j=1,2,\dots,n),$$

где ψ_j - угол разворота лопастей, соответствующий j -ой кривой; n - количество кривых.

Таким образом, расходная характеристика насосного агрегата полностью определяется двойкой

$$\langle Q_j(H), \psi_j \rangle, (j=1,2,\dots,n).$$

Состояние насосной станции определяется количеством работающих насосных агрегатов m_p из общего числа насосных агрегатов m и последовательностью углов разворота лопастей работающих насосных агрегатов

$$\psi_1, \psi_2, \dots, \psi_{m_p}.$$

Например, i -й насосный агрегат может работать в n положениях

$$\psi_{i1}, \psi_{i2}, \dots, \psi_{in}, (i=1,2,\dots,m),$$

т.е. положение насосного агрегата определяется положением угла разворота лопастей.

Для работающих насосных агрегатов насосной системы введем следующие обозначения:

$$M_p = \{i_1, i_2, \dots, i_{m_p}\}, \\ \Psi = \{\psi_1, \psi_2, \dots, \psi_{m_p}\},$$

где M_p - множество номеров работающих насосных агрегатов;

Ψ_p - множество углов разворота лопастей работающих насосных агрегатов.

Следовательно, состояние насосной станции в каждый момент времени определяется тройкой $\langle M_p, \Psi_p \rangle$.

Общая расходная характеристика насосной станции, соответствующая ее состоянию, определяется как алгебраическая сумма расходов каждого работающего насосного агрегата:

$$Q_{HC} = \sum_{i \in M_p} Q_i(H, \psi_i), \psi_i \in \Psi_p,$$

где $Q_i(H, \psi_i)$ - расходная характеристика i -го насосного агрегата;

H - высота подъема воды;

ψ_i - угол разворота лопастей i -го насосного агрегата.

Потребляемая мощность насосной станции также определяется как алгебраическая сумма мощностей каждого работающего насосного агрегата:

$$C_{HC} = \sum_{i \in M_p} c_i,$$

где $c_i = \frac{9.81 * H_i * Q_i}{\eta_i}$ /кВт/ - мощность i -го насосного агрегата;

H_i - напор, м;

Q_i - расход i -го насосного агрегата, м³/с;

η_i - КПД i -го насосного агрегата.

Оптимизация управления заключается в определении количества и номеров работающих насосных агрегатов, а также углов разворота их лопастей, обеспечивающих минимум потребляемой насосной станцией мощности для реализации заданного графика водоподачи.

Приводим постановку задачи оптимизации, которая является общепринятой в СМВ.

Пусть управляемый процесс в области

$$D = \begin{cases} \psi_{\min} \leq \psi \leq \psi_{\max}, \\ \underline{H}_{kp} \leq H \leq \bar{H}_{kp}, \end{cases}$$

характеризуется определением регулирующей тройки

$$\langle m_p, M_p, \Psi_p \rangle, \quad (1)$$

где ψ_{\min} и ψ_{\max} - минимальное и максимальное допустимые значения углов разворота лопастей насосных агрегатов;

\bar{H}_{kp} и \underline{H}_{kp} - критические значения уровней верхнего и нижнего бьефов насосных станций, при которых требуется минимизировать функционал

$$C_{HC} = \sum_{i \in M_p} c_i \rightarrow \min \quad (2)$$

с выполнением ограничения следующего вида

$$\left| \sum_{i \in M_p} Q_i - Q_n \right| \leq \varepsilon, \quad (3)$$

здесь $\varepsilon = 0.05 * Q_n$ - допустимая погрешность управления.

Задача оптимизации (2), (3) с оптимизируемой тройкой параметров (1) не подлежит эффективному решению существующими методами. В связи с этим возникшую задачу сформулируем как задачу линейного булева программирования в обобщенной постановке.

Предполагается, что в насосной станции имеются m насосных агрегатов:

$$P_1, P_2, \dots, P_m.$$

Насосный агрегат P_i может работать в n_i положениях

$$\psi_{i1}, \psi_{i2}, \dots, \psi_{in_i}, \quad (i = 1, 2, \dots, m) \quad (4)$$

где n_i - количество углов разворота лопастей i -го насосного агрегата; причем рассматривается случай, когда количество углов разворота лопастей для разных насосных агрегатов различно; кроме того, насосный агрегат P_i может находиться только в одном из перечисленных положений (4).

Известна производительность (производительность насосных агрегатов обычно регулируется углом разворота лопастей рабочего колеса работающих насосных агрегатов) каждого положения i -го насосного агрегата, т.е.

$$q_{i1}, q_{i2}, \dots, q_{in_i}, \quad (i = 1, 2, \dots, m)$$

где q_{ij} - расходная характеристика i -го насосного агрегата при j -м положении угла разворота лопастей.

Также считаются известными значения следующих параметров

$$c_{i1}, c_{i2}, \dots, c_{in_i}, \quad (i = 1, 2, \dots, m),$$

где c_{ij} - потребляемая мощность i -го насосного агрегата при j -м положении угла разворота лопастей.

Введем булевые переменные x_{ij} по следующему правилу: $x_{ij}=1$, если i -й насосный агрегат работает в j -м положении; $x_{ij}=0$ - в противном случае.

Задача оптимизации работы насосной станции может быть сформулирована как задача линейного булева программирования следующего вида:

$$(C, X) = \sum_{i=1}^m \sum_{j=1}^{n_i} c_{ij} x_{ij} \rightarrow \min_{x_{ij} \in \{0;1\}} \quad (5)$$

$$\underline{Q} \leq \sum_{i=1}^m \sum_{j=1}^{n_i} q_{ij} x_{ij} \leq \bar{Q}, \quad (6)$$

$$\sum_{j=1}^{n_i} x_{ij} \leq 1, \quad (i = 1, 2, \dots, m) \quad (7)$$

где \underline{Q} и \bar{Q} нижний и верхний пределы общей расходной характеристики насосной станции.

По полученному решению задачи (5)-(7) можно устанавливать количество и номера работающих насосных агрегатов. Количество работающих насосных агрегатов определяется из соотношения

$$m_p = \sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}^*,$$

где x_{ij}^* - ($i=1,2,\dots,m$; $j=1,2,\dots,n_i$) - решение задачи (5)-(7). Номера работающих насосных агрегатов определяются из следующего условия: если для произвольного значения i $x_{ij}^* = 1$ ($j=1,2,\dots,n_i$), то i -й насосный агрегат работает.

Здесь путем минимизации целевой функции (5) уменьшается общая потребляемая мощность насосной станции. Выполнения ограничения вида (6) обеспечивает подъем воды в допустимых пределах $[\underline{Q}, \bar{Q}]$. Выполнением ограничения вида (7) обеспечивается работа каждого насосного агрегата только в одном из возможных положений (4).

2. МОДЕЛЬ ЗАДАЧИ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ

Задача автоматической классификации возникает во многих прикладных вопросах, когда требуется разбиение множества из конечного числа объектов на определенное (может быть и заранее неопределенное) число классов таким образом, чтобы минимизировать некоторый критерий их взаимной несогласованности. Вводим необходимые обозначения и понятия.

Предполагается, что заданы n объектов, и для каждой пары объектов i и j предполагается заданным число d_{ij} ($d_{ij} \geq 0$), f называемое расстоянием между объектами i и j (если эти объекты могут рассматриваться как элементы n мерного евклидова пространства).

Задача состоит в разбиении множества всех объектов на p классов (предполагается, что целое число p задано) и выборе в каждом классе специального объекта, называемого представителем этого класса, так чтобы сумма расстояний от объектов до их представителей была минимальна.

Вводится матрица булевых переменных по следующему правилу

$$x_{ij} = \begin{cases} 1, & \text{если объект с номером } i \text{ отнесен к классу, представителем которого является объект с номером } j; \\ 0, & \text{в противном случае.} \end{cases}$$

Здесь для идентификации объектов через булевые переменные x_{ij} использованы обозначения с двумя индексами, первый из которых i -указывает номер объекта в исходной выборке; второй j -номер объекта, который определен в качестве представителя одного из классов.

Приведенные обозначения и понятия разъясним на следующем примере. Пусть задано 10 объектов и требуется их разбиения на 3 класса, т.е. $n=10$ и $p=3$.

Предположим, что в результате решения задачи автоматической классификации получены следующие данные:

$$x_{17}=1; x_{22}=1; \quad x_{32}=1; \quad x_{49}=1; \quad x_{59}=1;$$

$$x_{62}=1; \quad x_{77}=1; \quad x_{87}=1; \quad x_{99}=1; \quad x_{102}=1;$$

Остальные элементы матрицы булевых переменных X равны нулю. В качестве представителей классов определены следующие объекты:

$x_{22}=1$ - представитель (условно первого) класса, куда входят следующие объекты:

$$x_{22}, x_{32}, x_{62} \text{ и } x_{102};$$

$x_{77}=1$ - представитель (условно второго) класса, куда входят следующие объекты:

$$x_{17}, x_{77} \text{ и } x_{87},$$

$x_{99}=1$ - представитель (условно третьего) класса, куда входят следующие объекты:

$$x_{49}, x_{59} \text{ и } x_{99}.$$

Как стало ясно из примера, представителями классов могут быть определены только диагональные элементы булевой матрицы X .

Теперь мы можем описать задачу автоматической классификации в виде следующей модели линейного булева программирования:

$$q(x) = Dx = \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij} \rightarrow \min_{x_{ij} \in \{0;1\}} \quad (8)$$

$$\sum_{j=1}^n x_{ij} = 1, \quad (i = 1, 2, \dots, n), \quad (9)$$

$$\sum_{j=1}^n x_{jj} = p, \quad (10)$$

$$x_{ij} \in \{0;1\}, \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, n).$$

Задача минимизации суммы расстояний от объектов до их представителей реализуется решением задачи (8). Ограничения (9) выражают тот факт, что каждый объект должен быть привязан к одному и только одному классу. Ограничение (10) обеспечивает определения p числа объектов, которые являются представителями образованных p классов.

3. ЗАДАЧА ОБ ОПТИМИЗАЦИИ РАЗМЕЩЕНИЯ БУКВ АЛФАВИТА НА КЛАВИАТУРЕ ЭВМ

С момента появления пишущих машин, особенно средств вычислительной техники, исследования, связанные с изучением проблемы оптимального размещения букв любого алфавита на клавиатуре ЭВМ и пишущей машинки являются актуальными по сей день и требует своего разрешения эффективными методами.

Из литературных источников известно, что также задача оптимального размещения букв алфавита на клавиатуре ЭВМ как и любая задача размещения, относится к задачам комбинаторной оптимизации.

В данной работе впервые предлагается новая модель для задачи оптимального размещения букв алфавита любого языка на клавиатуре ЭВМ в виде линейной модели булева программирования. Для формализации задачи вводится булева матрица

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix}$$

по следующему правилу

$$x_{ij} = \begin{cases} 1, & \text{если буква с номером } i \text{ размещена на клавише} \\ & \text{с номером } j; \\ 0, & \text{в противном случае.} \end{cases}$$

Первый индекс i используется для определения порядкового номера буквы в исходном алфавите, второй индекс j - для идентификации порядкового номера клавиши на клавиатуре.

Вводятся следующие понятия:

a_i - частота появления буквы с порядковым номером i в генеральной выборке слов рассматриваемого алфавита (в данном случае необходимо исследовать достаточно большое число слов);

c_j - расстояние от центра клавиатуры до клавиши с порядковым номером j .

После введения необходимых обозначений и понятий задачу об оптимизации размещения букв алфавита на клавиатуре ЭВМ можно описать с помощью следующей модели линейного булева программирования:

$$q(x) = \sum_{i=1}^n \sum_{j=1}^n a_i c_j x_{ij} \rightarrow \min_{x_{ij} \in \{0,1\}} \quad (11)$$

$$\sum_{i=1}^n x_{ij} = 1, \quad (j = 1, 2, \dots, n) \quad (12)$$

$$\sum_{j=1}^n x_{ij} = 1, \quad (i = 1, 2, \dots, n) \quad (13)$$

$$x_{ij} \in \{0,1\}, \quad (i, j = 1, 2, \dots, n).$$

При математической формулировке задачи основными критериями выступают суммарные передвижения пальцев по клавиатуре ЭВМ. При минимизации этого критерия, как известно, снижаются расходы пользователя по времени, что приводит к естественному уменьшению утомляемости.

Ограничения видов (12) и (13) использованы для обеспечения закрепления каждой буквы только за одной клавишей и обратно.

Относительно модели (11) - (13) можно сделать следующее замечание: в модели рассматривается случай с исследованием частоты появления отдельных букв, при котором и реализуется общий случай, когда исследуется комбинация букв в тексте. Потому что комбинация (сочетание) букв в тексте состоит из отдельных букв.

Для решения задачи (11) - (13) можно использовать эффективные методы линейного булева программирования.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Общее описание работы насосной станции.
2. Линейная булевая модель работы насосной станции.
3. Что понимается под задачей автоматической классификации?
4. Модель задачи оптимального размещения букв алфавита на клавиатуре ЭВМ.

ЛИТЕРАТУРА

- 69.
- 70.
- 71.
- 72.
- 73.

ЛЕКЦИЯ 16. РЕШЕНИЯ ПРАКТИЧЕСКИХ ЗАДАЧ ЛИНЕЙНОГО БУЛЕВА ПРОГРАММИРОВАНИЯ (2 ЧАСА)

План

1. Моделирование и оптимизация работы насосной станции
2. Проверка адекватности математической модели
3. Алгоритм оптимального управления работы насосной станции

1. МОДЕЛИРОВАНИЕ И ОПТИМИЗАЦИЯ РАБОТЫ НАСОСНОЙ СТАНЦИИ

Эффективное управление внутристанционным режимом работы насосной станции предполагает определение количества и номеров работающих насосных агрегатов, а также положения углов разворота лопастей, которые обеспечивают минимум потребляемой мощности энергии для реализации заданного графика водоподачи. В настоящее время на многих магистральных каналах с каскадом насосных станций управление процессом водоподачи осуществляется центральной диспетчерской службой, то есть диспетчер осуществляет управление водозабором, транспортированием и распределением воды. Процесс принятия решения диспетчером сводится к сравнению фактического состояния процесса водоподачи с запланированным и исходя из этого, выработки приемлемых в данный момент времени мероприятий на основе личного опыта и интуиции.

При реализации стратегии управления диспетчер опрашивает дежурных инженеров о параметрах гидравлических режимов участков канала и насосных станций, состоянии основного технологического процесса. Стратегия управления, выработанная диспетчером, реализуется на каждой насосной станции и гидротехническом сооружении дежурными инженерами. При нормальной эксплуатации диспетчер получает информацию от насосных станций о значениях технологических параметров через каждые шесть часов, а о параметрах потребителей через каждый час. Диспетчер, проводя анализ обстановки в каскаде, принимает решение для управления процессом водоподачи, которое по диспетчерской связи сообщается дежурным инженерам насосных станций. По этим распоряжениям они осуществляют пуск или остановку насосных агрегатов, изменяют производительность насосных станций разворотом лопастей на определенный градус, включая или выключая насосные агрегаты. То есть управление процессом водоподачи производится в режиме ручного диспетчерского управления.

Диспетчер, управляя системой, полагается на свой личный опыт и интуицию, основанные на простых методах принятия решений. Управление оборудованием и объектами осуществляется вручную, а для связи между диспетчером и объектами управления используется телефон или факс. Как уже говорилось такое управление приводит к перерасходу электроэнергии на водоподъем, непроизводительным сбросам и потерям воды, невыполнению графика водоподачи.

Целью разработки является минимизация расхода электроэнергии и относительная погрешность уровня водоподачи заданного графика насосной станции путём исследования различных режимов работы насосной станции.

Считается, что известны плановый объём водоподачи, количество насосных агрегатов, количество положений каждого насосного агрегата в насосной станции, а также гидротехнические и расходные характеристики в каждом положении. Требуется определить номера работающих насосных агрегатов и их положения, при которых обеспечивается требуемый объём водоподачи с минимальными потерями и расходом электроэнергии.

Пусть управляемый процесс в области

$$D = \begin{cases} j_{\min} < j < j_{\max}, \\ \underline{H}_{kp} < H < \overline{H}_{kp} \end{cases}$$

где j_{\min} и j_{\max} - минимально и максимально допустимые значения углов разворота лопастей насосных агрегатов;

\underline{H}_{kp} , \overline{H}_{kp} - критические значения уровня верхнего и нижнего бьефов насосных станций;

при которой требуется минимизировать функционал

$$C_{hc} = \sum_{i \in M_p} c_i \rightarrow \min ,$$

здесь C_{hc} - общий расход электроэнергии насосной станцией;

c_i - расходная характеристика i -го насосного агрегата с выполнением ограничения следующего вида

$$|\sum_{i \in M_p} Q_i - Q_n| \leq \varepsilon ,$$

здесь Q_i - расход i -го насосного агрегата;

$\varepsilon = 0.05 * Q_n$ - допустимая погрешность управления.

В приведенной постановке задача о насосной станции относится к классу "трудно решаемых задач", и для ее решения не существует эффективного метода.

Но, переходя от исходной постановки к другой (5-7), задача становится разрешимой с применением эффективного метода.

Здесь также была поставлена и реализована проблема минимизации относительной погрешности уровня водоподачи. Погрешность (ε) обозначим как разницу между плановым объёмом водоподачи (Q_{plan}) и реальным объемом водоподачи

$$\sum_{i=1}^m \sum_{j=1}^{n_i} q_{ij} x_{ij} .$$

Тогда задача оптимизации работы насосной станции может быть сформулирована как задача линейного булева программирования следующего вида:

$$(C, X) = \sum_{i=1}^m \sum_{j=1}^{n_i} c_{ij} x_{ij} \rightarrow \min_{x_{ij} \in \{0;1\}} \quad (1)$$

$$|\sum_{i=1}^m \sum_{j=1}^{n_i} q_{ij} x_{ij} - Q_{plan}| \leq \varepsilon, \quad (2)$$

$$\sum_{j=1}^{n_i} x_{ij} \leq 1, \quad (i = 1, 2, 3, \dots, m) \quad (3)$$

Для решения поставленной задачи можно использовать эффективный алгоритм метода обобщенных неравенств, для чего переходим к другой форме задачи булева программирования:

$$f(x) = \frac{(a, x)}{(b, x)} = \frac{\sum_{k=1}^l a_k x_k}{\sum_{k=1}^l b_k x_k} \rightarrow \max_{x_k \in \{0;1\}}, \quad (4)$$

где матрицы q_{ij}, c_{ij}, x_{ij} ($i=1,2,\dots,m; j=1,2,\dots,n_i$), используемые в (1)-(3) преобразованы в вектора a_k, b_k, x_k ($k=1,2,\dots,l$), соответственно, построчно.

Как известно, при решении задачи (1)-(3) методом обобщенных неравенств сначала получим решение задачи (4), для которой проверяется выполнение условия (3) в первую очередь, а затем условия (2).

При нарушении условия (3.) сразу же переходят к очередному насосному агрегату путем наращивания на единицу индексной переменной i в формуле (3), не проверяя работы остальных положений текущего насосного агрегата, так как насосный агрегат может работать только в одном положении. Решением задачи (1)-(3.) будет такое решение задачи (4.) с минимальной мощностью, при котором выполняются условия (2) и (3.) одновременно.

2. ПРОВЕРКА АДЕКВАТНОСТИ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

С целью проверки адекватности предложенных в этой работе математической модели и алгоритма управления процессом водоподъема насосной станции было проведено тестирование написанного программного обеспечения. В качестве тестовых данных были использованы данные насосных станций, расположенных на территориях Бухарской и Кашкадарьинской областей. Технические характеристики насосных агрегатов одной из насосных станций приведены в табл. 1.

Результаты численного моделирования с использованием метода обобщенных неравенств для оптимального управления работой насосной станции проведены в табл. 2.

Полученные результаты показывают адекватность математической модели и алгоритма оптимизации работы насосной станции. Отклонение от заданного гра-

фика водоподачи не превышает 4%, при этом затрачиваемая электроэнергия минимальна.

Проведен трехэтапный вычислительный эксперимент с использованием данных о насосной станции. На первом этапе мы получили результаты в виде булевых переменных, дающие оптимальные решения о режиме работы насосной станции. На втором этапе были получены данные в виде таблиц, в которых отражены номера положений лопастей, число работающих насосных агрегатов, данные о потребляемой мощности и относительной погрешности водоподачи. На третьем этапе была минимизирована относительная погрешность уровня водоподачи.

Таблица 1. Технические характеристики насосных агрегатов одной из насосных станций, расположенных на территориях Бухарской и Кашкадарьинской областей.

Марка насоса	Подача, $Q, \text{м}^3/\text{ч}$	Подача, $Q, \text{м}^3/\text{с}$	Напор, $H, \text{м}$	КПД $\eta, \%$	Мощность на валу насоса N, kBt	Угол установки лопастей $V, \text{град}$
ОП5-87	8784	2,44	8,8	80	263	-6°30'
	9288	2,58	7,8	80	246	-3°20'
	10080	2,8	11,7	80	400	0
	12060	3,35	11	85,5	423	+2°30'
	14220	3,95	7,15	80	345	+6°30'
ОП3-110	14400	4	22	80	1080	-7°
	17982	4,98	15,4	83	905	-4°
	19152	5,32	22,8	87	1368	-2°
	20062	5,57	21,5	87,5	1350	-0°
	22500	6,25	14,6	77	1160	+1°30'
ОП5-145	24120	6.7	10.4	78	886	-6°30'
	26820	7.45	8.3	80	747	-6°30'
	27720	7.7	12.8	78	1380	-3°20'
	36360	10.1	11	85,5	1275	0
	41040	11.4	7.7	80	1075	+2°30'
ОП11-185	52920	14.7	15.5	80	2785	-8°
	54900	15.25	14.3	80	2680	-6°
	61200	17.0	20.4	82	4040	-4°
	69840	19.4	18	88	3890	-2°
	79920	22.2	12.7	84	3290	-0°
ОП10-260	102240	28.4	23.1	80	8050	-9°
	105840	29.4	22.4	81	7960	-6°
	129240	35.9	27.8	87	11250	-3°
	136800	38	26	87,5	11130	-1°
	152640	42.4	21	84,5	10330	-0°
ОП10-145	25920	7,2	15,3	80	1350	-9°
	27720	7,7	14,1	82	1300	-6°
	32400	9	18	87	1985	-3°
	34920	9,7	17	87,5	1855	-1°
	39960	11,1	12,9	83	1680	0

где: $V, \text{град}$ – углы разворота лопастей рабочего колеса;

$Q, \text{м}^3/\text{с}$ – водоподача;

N, kBt – затрачиваемая энергия;

$\eta, \%$ - КПД насосного агрегата.

Таблица 2. Результаты численного моделирования

Насосные агрегаты	1-й насосный агрегат	2-й насосный агрегат	3-й насосный агрегат	4-й насосный агрегат	5-й насосный агрегат	6-й насосный агрегат	Плановый объем водоподачи $Q, \text{ м}^3/\text{с}$	Плановый объем водоподачи $Q, \text{ м}^3/\text{с}$	Потребляемая электроэнергия $C, \text{ кВт}$	Отклонения от графика $DQ, \%$	Число работающих насосных агрегатов N
№ эксперимента											
1.	4пол	2пол	1пол	-	-	-	15.000	15.030	2214.00	0.200	3
2.	-	2пол	-	-	-	5пол	16.000	16.080	2585.00	0.500	2
3.	-	-	-	3пол	-	-	17.000	17.000	4040.00	0.000	1
4.	-	5пол	-	-	-	5пол	18.000	17.350	2840.00	3.611	2
5.	2пол	2пол	5пол	-	-	-	19.000	18.960	2226.00	0.211	3
6.	-	3пол	-	1пол	-	-	20.000	20.020	4153.00	0.100	2
7.	5пол	2пол	5пол	-	-	-	21.000	20.330	2325.00	3.190	3
8.	-	-	-	5пол	-	-	22.000	22.200	3290.00	0.909	1
9.	-	2пол	1пол	-	-	5пол	23.000	22.780	3471.00	0.957	3
10.	-	-	-	2пол	-	4пол	24.000	24.950	4535.00	3.958	2
11.	-	-	-	2пол	-	4пол	25.000	24.950	4535.00	0.200	2
12.	1пол	2пол	2пол	-	-	5пол	26.000	25.970	3595.00	0.115	4
13.	4пол	2пол	2пол	-	-	5пол	27.000	26.880	3755.00	0.444	4
14.	-	1пол	-	2пол	-	3пол	28.000	28.250	5745.00	0.893	3
15.	-	3пол	-	1пол	-	3пол	29.000	29.020	6138.00	0.069	3
16.	2пол	2пол	5пол	-	-	5пол	30.000	30.060	3906.00	0.200	4
17.	5пол	2пол	5пол	-	-	5пол	31.000	31.430	4005.00	1.387	4
18.	5пол	2пол	5пол	-	-	5пол	32.000	31.430	4005.00	1.781	4
19.	-	3пол	-	4пол	-	3пол	33.000	33.200	7131.30	0.602	3
20.	-	1пол	3пол	2пол	-	1пол	34.000	34.150	6490.00	0.441	4
21.	-	1пол	3пол	2пол	-	1пол	35.000	34.150	6490.00	2.429	4
22.	-	5пол	-	5пол	-	2пол	36.000	36.150	5750.00	0.417	3
23.	-	-	3пол	5пол	-	1пол	37.000	37.100	6020.00	0.270	3
24.	-	-	-	-	4пол	-	38.000	38.000	11130.0	0.000	1

Результаты вычислительного эксперимента показали адекватность математической модели и алгоритма оптимизации работы насосной станции.

3. АЛГОРИТМ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ РАБОТОЙ НАСОСНОЙ СТАНЦИИ

Ниже приведен алгоритм оптимального управления работой насосной станции с применением метода обобщенных неравенств.

1. Ввести $m, n_i, c_{ij}, q_{ij}, Q_{plan}$.

2. Перевести c_{ij} и q_{ij} в одномерные массивы a_k и b_k ($k=1,2,\dots,l$; $l = \sum_{i=1}^m n_i$).

3. Вычисление значений функции

$f_k^{(1)} = \frac{a_k}{b_k}$, ($k=1,2,\dots,l$), для каждой переменной x_k и определение первого решения

$x_{k_1}^* = 1$ (при этом $x_k=0$, $k=1,2,\dots,l$; $k \neq k_1$), которому соответствует максимальное значение функций среди вычисленных значений функции.

4. Вычисление значений функции последовательным присоединением новых элементов,

$f_k^{(2)} = \frac{a_{k_1}^* + a_k}{b_{k_1}^* + b_k}$, ($k=1,2,\dots,l$; $k \neq k_1$) и определение второго решения, $x_{k_2}^* = 1$, которому

соответствует максимальное значение функции среди вычисленных значений функции.

5. Вычисление значений функции последовательным присоединением новых элементов, т.е.

$$f_k^{(l)} = \frac{\sum_{k=1}^{l-1} a_{k_i}^* + a_k}{\sum_{k=1}^{l-1} b_{k_i}^* + b_k}, \quad (k = 1,2,\dots,l; k \neq k_i, i = 1,2,\dots,l-1).$$

6. Последовательно повторяя этот процесс и присоединяя остальные элементы, получаем упорядоченный ряд $x_1^*, x_2^*, \dots, x_l^*$.

7. Возвращаемся от одномерных массивов a_k и b_k ($k=1,2,\dots,l$) к двухмерным c_{ij} , q_{ij} массивам ($i=1,2,\dots,m$; $j=1,2,\dots,n_i$).

8. Решению задачи будет соответствовать минимальное число n' первых элементов полученного упорядоченного ряда, которое будет удовлетворять условиям (2) и (3).

9. Результаты вывести на экран, в файл или на принтер.

На рис.1 приведена блок - схема алгоритма оптимального управления работой насосной станции.

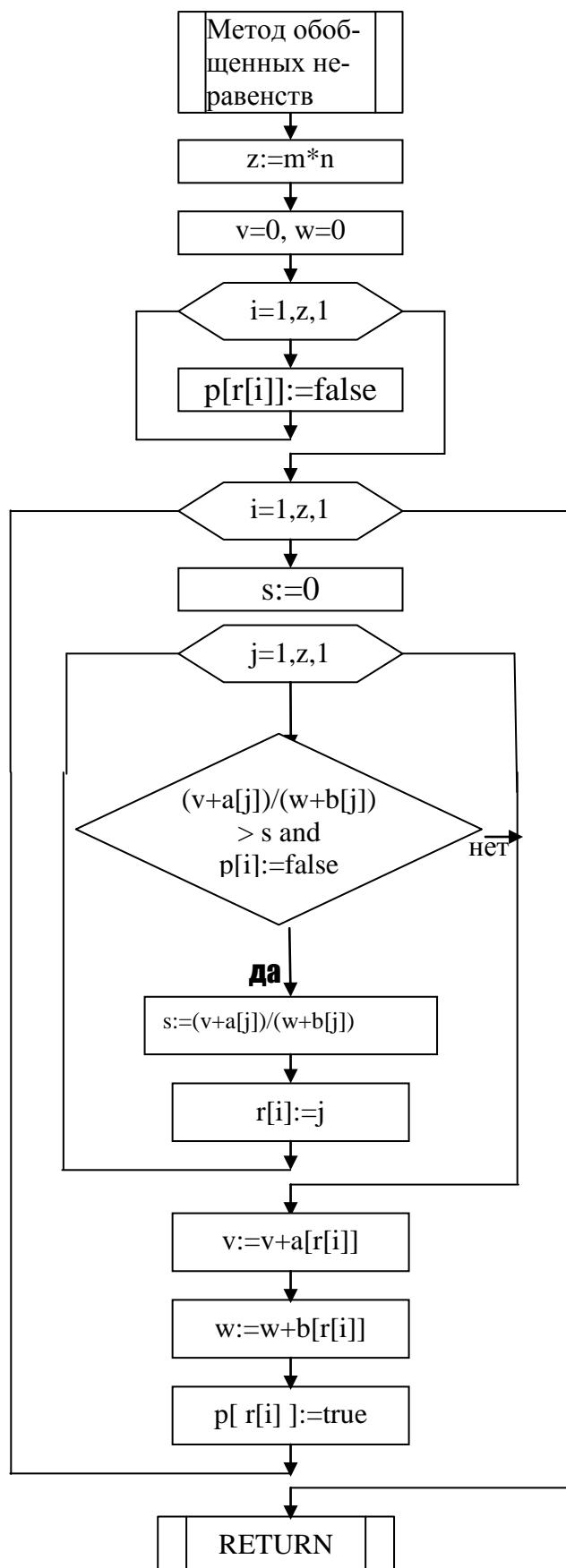


Рис. 1. Блок-схема алгоритма оптимального управления работой насосной станции с применением метода обобщенных неравенств

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Почему задача оптимизации работы насосной станции преобразовалась к задаче нелейного булева программирования?
2. Для чего ведется проверка адекватности полученной модели оптимизации работы насосной станции?
3. Какая сложность использованного алгоритма оптимального управления работы насосной станции.
4. Обоснование выбора языка программирования для реализации поставленной задачи.

ЛИТЕРАТУРА

74.
75.
76.
77.
78.

ЛЕКЦИЯ 17. ИССЛЕДОВАНИЕ МЕТОДОВ ПОСТРОЕНИЯ ИНФОРМАТИВНЫХ НАБОРОВ ПРИЗНАКОВ (2 ЧАСА)

План

1. Сравнительный анализ методов построения информативных наборов признаков
2. Аналитический подход к формированию информативной подсистемы признаков в задаче распознавания
3. Упрощенный метод классификации с использованием аналитического подхода формирования информативной подсистемы признаков при наличии обучающей выборки

1. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПОСТРОЕНИЯ ИНФОРМАТИВНЫХ НАБОРОВ ПРИЗНАКОВ

Построение информативного набора признаков по обучающей выборке является одной из центральных проблем классической теории распознавания образов. Рассмотрим основное содержание этой проблемы в классической постановке. Назначение указанной проблемы - определение наиболее полезных для классификации набора признаков. Предпосылкой к реализации данной проблемы является следующее важное обстоятельство.

Когда исходная система признаков определена, то она, как правило, оказывается весьма избыточной. Существуют разные точки зрения за и против относительно сохранения такой избыточности: аргумент "за" объясняется тем, что увеличение числа признаков позволяет более полно описать объекты: аргумент "против" - увеличение числа признаков увеличивает "шум" в данных, затрудняет обработку и требует больших вычислительных затрат.

В связи с поиском ответа на вопрос: сколько признаков следует оставлять в информативном наборе признаков и сколько объектов нужно брать при соблюдении условия статического ансамбля (с точки зрения статистики, а не предметной области) для получения результата с заданной точностью, целесообразно обратиться к результатам исследований по оцениванию ошибки распознаванию при различном объеме обучающей выборки и числу признаков. На основании анализа этих результатов можно сделать следующие выводы:

- ошибка быстро увеличивается с ростом числа признаков и медленно уменьшается с ростом числа объектов;
- увеличение числа признаков требует значительного повышения объема обучающей выборки для достижения той же ошибки.

Следовательно, первостепенное значение для выбора числа признаков должны иметь нестатистические соображения, вытекающие из существа решаемой задачи и особенностей предметной области. При решении многих прикладных задач, где для определения значения признаков используются дорогостоящие измерительные приборы, снижение размерности признакового пространства приобретает решаю-

шую роль. Как правило, такое преобразование признакового пространства сводится к определению относительно небольшого числа признаков, которые обладают свойством наибольшей информативности в смысле, определяемом выбранным критерием информативности. Причем, говоря о преобразовании признакового пространства и выборе критерия информативности, следует иметь в виду, что преобразование признаков, осуществляющее вне связи с качеством классификации, приводит к задаче представления исходных данных в пространстве меньшей размерности. Полученный при этом информативный набор признаков определяется путем максимизации (или минимизации) некоторого функционала - критерия информативности, не учитывающего классовую структуру объектов. Если же признаки выбираются с целью улучшения характеристик классифицируемой системы, то критерий такого выбора связан с разделимостью классов. В соответствии с этими задачами в прикладных исследованиях обычно используются два подхода к снижению размерности исходного пространства признаков. При первом подходе новые признаки определяются без учета качества классификации (задача представления данных), а при втором - поиск признаков связан с оценкой качества классификации (задача распознавания объектов). Очевидно, что реализация рассмотренного выше этапа формирования информативного описания распознаваемых объектов связана с использованием методов снижения размерности, основанных на втором подходе.

Проблеме преобразования признакового пространства посвящена обширная литература. В этих работах с различной степенью глубины и обобщенности рассматриваются методы, алгоритмы и критерии минимизации исходного описания применительно к определенным классам задач обработки данных: распознавания образов, регрессионного анализа, кластеризации и т.д.

Рассмотрим наиболее известные методы снижения размерности признакового пространства, которые нашли широкое применение в различных прикладных исследованиях.

Снижение размерности исходного пространства признаков представляет по существу переход от исходной системы признаков $x=(x^1, x^2, \dots, x^n)$ к новой системе $\tilde{x}=(x^{i_1}, x^{i_2}, \dots, x^{i_{n'}})$, включающей меньшее число признаков ($n' < n$), чем исходная система. При этом новые признаки x^{i_j} ($j=1, 2, \dots, n'$) формируются из исходных признаков x^j ($j=1, 2, \dots, n$). Снижение размерности исходного пространства осуществляется путем решения оптимизационной задачи следующего вида:

$$\Phi(x) \rightarrow \max_x \Rightarrow \tilde{x},$$

где $\Phi(x)$ - заданная мера информативности n -мерной системы признаков x . Согласно, конкретный выбор меры информативности Φ приводит к конкретному методу снижения размерности. Для задачи распознавания образов определение нового набора признаков \tilde{x} обычно связано с оценкой качества классификации. Как отмечено в, при малых выборках поиск такого набора можно производить, основываясь на минимизации верхней границы или математического ожидания ошибок классификации при контроле. Известно также, что данные ошибки могут характеризо-

ваться величиной критерия качества разделимости классов в отобранном наборе признаков или величиной ошибочно расклассифицированных объектов, если выбор признаков осуществляется для конкретного правила классификации. При рассмотрении методов выборов признаков далее будем предполагать, что качество (информативность) признаков и их наборов оценивается по одному из этих критериев.

Задача определения оптимального набора признаков имеет относительно простое решение в условиях независимости исходных признаков x^1, x^2, \dots, x^n . В этом случае последние ранжируются по значениям критерия информативности каждого из этих признаков и для заданного n' определяется наилучший набор $x^{i_1}, x^{i_2}, \dots, x^{i_{n'}}$ включающий первые n' наиболее информативных признаков ранжированного ряда. Подробное описание этого метода приведена в.

Ситуация существенно усложняется при коррелированности исходных признаков x^1, x^2, \dots, x^n : для определения наилучшего набора из n' признаков необходимо сравнить все возможные $C_n^{n'}$ наборов из n признаков. Если n' не зафиксировано, то количество вариантов достигает величины

$$P = \sum_{n'=1}^n C_n^{n'} = 2^n - 1 .$$

Отсюда следует, что метод полного перебора хотя и гарантирует определение оптимального набора признаков $(x^{i_1}, x^{i_2}, \dots, x^{i_{n'}})$, однако его можно использовать лишь при небольшом числе исходных признаков.

В связи с этим обстоятельством в прикладных исследованиях обычно используются методы частичного перебора, позволяющие получить субоптимальный набор признаков при сравнительно малых вычислительных затратах. Среди них наиболее распространеными считаются методы, использующие процедуры последовательного отбрасывания признаков (алгоритм "последовательной селекции назад"), последовательного присоединения признаков (алгоритм "последовательной селекции вперед"), а также комбинации этих процедур. Результаты сравнительного анализа этих методов по вычислительным затратам приведены в. Автором показано, что в условиях представительной выборки процедура последовательного присоединения обеспечивает результаты более близкие к оптимальным, чем процедура последовательного отбрасывания. Каждому из этих методов присущи свои недостатки. Например, при реализации процедуры последовательного присоединения нельзя отбросить признак, включенный в оптимальный набор на предыдущих шагах, а при использовании процедуры последовательного отбрасывания не учитывается статистическое влияние исключенных ранее признаков.

Другим широко используемым методом выбора информативных признаков является случайный поиск адаптации (метод СПА). Суть этого метода состоит в случайному поиске субоптимального или оптимального набора из n' признаков $x^{i_1}, x^{i_2}, \dots, x^{i_{n'}}$ с "поощрением" и "наказанием" отдельных признаков x^1, x^2, \dots, x^n . В то же время этому методу присущи свои недостатки. Например, неизвестно как его использовать при конкурирующих группах признаков. Более того, в условиях ма-

лой обучающей выборки, решения, полученные с помощью этого метода, могут быть статистически неустойчивыми.

Рассмотренные методы выбора признаков отличаются друг от друга способом организации перебора различных сочетаний исходных признаков. В то же время они обладают общей особенностью: каждый из этих методов для оценки качества признаков и их наборов может использовать различные критерии информативности. Однако такая универсальность порождает проблему выбора эффективного критерия, т.к. результат выбора признаков существенно зависит от используемого критерия информативности.

В настоящее время разработан ряд методов выбора признаков, которые ориентированы на использование конкретного критерия информативности. Один из таких методов предложен и использован в. Суть его заключается в использовании меры важности исходного признака, представляющего собой обработанную надлежащим образом степень уменьшения так называемых "голосов" при удалении этого признака. Авторами метода эта мера названа информационным весом признака. Для определения информационных весов используются специальные вычислительные алгоритмы, которые позволяют учитывать взаимосвязи признаков. Основное достоинство этого метода состоит в том, что вычислительные процедуры определения информационных весов признаков и распознавания построены на единой основе - алгоритмах вычисления оценок.

Интересный подход к выбору информативных признаков рассмотрен в. В рамках этого подхода предложен ряд методов определения информативных наборов признаков с учетом ограничений, связанных с затратами на создание технических средств измерения этих признаков. Мера информативности последних задается критериями, основанными на использовании евклидового расстояния. В случае, когда признаки детерминированные, определение информативных наборов сводится к решению оптимизационных задач, которые можно описать следующим образом.

Пусть заданы обучающие выборки $X_1 = \{x_{11}, x_{12}, \dots, x_{1k_1}\}$,

$$X_2 = \{x_{21}, x_{22}, \dots, x_{2k_2}\}, \dots, X_m = \{x_{m1}, x_{m2}, \dots, x_{mk_m}\},$$

каждая из которых представляет соответствующий класс и содержит объекты

$$x_{pi} = (x_{pi}^1, x_{pi}^2, \dots, x_{pi}^n), \quad p=1,2,\dots,m; \quad i=1,2,\dots,k_p \text{ где}$$

x_{pi}^j - значение j -го признака, замеренного на i -ом объекте p -го класса, m -количество заданных классов, k_p -количество объектов в p -ом классе.

Для формирования наборов признаков в исходной системе $x=(x^1, x^2, \dots, x^n)$ используется n' - мерный вектор $\lambda=(\lambda_1, \lambda_2, \dots, \lambda_n)$, компоненты которого принимают значения 0 или 1, и указывают на отсутствие ($\lambda_j=0$) или наличие ($\lambda_j=1$) соответствующего признака (x^j) в рассматриваемом наборе. Если предположить, что n -мерное векторное пространство является евклидовым, то расстояние между двумя объектами x_{pt} и x_{ql} относительно набора признаков, заданного вектором λ , будет выражаться через

$$d(x_{pt}, x_{ql}) = \sqrt{\sum_{j=1}^n \lambda_j (x_{pt}^j - x_{ql}^j)^2}$$

$$(p, q = 1, 2, \dots, m; t = 1, 2, \dots, k_p; l = 1, 2, \dots, k_q).$$

Кроме того, для формирования критерия эффективности системы распознавания авторами вводятся следующие меры близости между объектами:

$$S(X_p) = \sqrt{\frac{2}{k_p} \frac{1}{k_p - 1} \sum_{t=1}^{k_p} \sum_{l=1}^{k_p} d^2(x_{pt}, x_{pl})}, \quad (1)$$

характеризующая среднеквадратичный разброс объектов внутри данного класса X_p , ($p=1, 2, \dots, m$);

$$R(X_p, X_q) = \sqrt{\frac{1}{k_p} \frac{1}{k_q} \sum_{t=1}^{k_p} \sum_{l=1}^{k_q} d^2(x_{pt}, x_{ql})}, \quad (2)$$

характеризующая среднеквадратичный разброс данной пары классов

$$X_p \text{ и } X_q, (p, q = 1, 2, \dots, m; p \neq q).$$

В качестве алгоритма распознавания выбирается классификатор евклидового расстояния, определяющий величину

$$L(x, X_p) = \sqrt{\frac{1}{k_p} \sum_{t=1}^{k_p} d^2(x, x_{pt})}, \quad (3)$$

которая является среднеквадратичным расстоянием между распознаваемым объектом x и объектами класса X_p , ($p=1, 2, \dots, m$) по набору признаков, определяемому вектором λ . Решающее правило этого классификатора состоит в следующем: $x \in X_p$ если

$$L(x, X_p) = \min_{q=1, 2, \dots, m} L(x, X_q).$$

Рассматривая в качестве критерия эффективности системы распознавания функционал, зависящий в общем случае от функций (1), (2) и выбранного алгоритма распознавания (3), задачу определения информативного набора признаков можно свести к оптимизационной

$$\Phi(\lambda) = \Phi[S(\lambda); R(\lambda); L(\lambda)] \rightarrow \underset{\lambda}{\operatorname{extr}} \quad (4)$$

при

$$C(\lambda) = \sum_{j=1}^n c_j \lambda_j \leq c_0, \quad (5)$$

где c_j - затраты на создание технического средства, предназначенного для определения признака x_j ; c_0 - ресурсы, выделенные на создание технических средств наблюдений.

В прикладных исследованиях обычно можно ограничиться решением оптимизационных задач для частных видов функционала Φ , например:

$$\Phi_1(\lambda) = \max_{\gamma=1,2,\dots,m} \{S(X_\gamma)/R(X_p, X_q) \geq R^0(X_p, X_q)\} \rightarrow \min_{\lambda}; \quad (6)$$

$$\Phi_2(\lambda) = \min_{p,q=1,2,\dots,m} \{R(X_p, X_q)/S(X_\gamma) \leq S^0(X_\gamma)\} \rightarrow \max_{\lambda}; \quad (7)$$

$$\Phi_3(\lambda) = \min_{p,q=1,2,\dots,m} \left[\frac{R^2(X_p, X_q)}{S(X_p)S(X_q)} \right] \rightarrow \max_{\lambda}; \quad (8)$$

при ограничении (5).

Решение каждой из задач (6), (7), (8) направлено на определение информативных наборов признаков, улучшающих разделимость заданных классов, что, в конечном итоге, повышает эффективность системы распознавания, при соблюдении ограничения, на общую сумму стоимостей технических средств наблюдений. Причем допустимой областью решений этих задач является подмножество вершин единичного гиперкуба, число которых равно $C_n^{n'}$.

Один из предлагаемых в методов решения задач (6), (7), (8) основан на построении непрерывного аналога для задачи дискретной оптимизации. В этом случае вместо дискретной задачи

$$\begin{cases} \Phi(\lambda) \rightarrow \text{extr}, \\ \sum_{j=1}^n \lambda_j = n', \\ \lambda_j \in \{0; 1\} \end{cases} \quad (9)$$

решается задача непрерывной оптимизации

$$\begin{cases} \Phi(\lambda) \rightarrow \text{extr}, \\ \sum_{j=1}^n \lambda_j = n', \\ \lambda_j \in [0; 1], \end{cases} \quad (10)$$

в которой параметры λ_j могут принимать произвольные значения на отрезке $[0,1]$.

Очевидно, что такая замена имеет смысл в том случае, если функционал $\Phi(\lambda)$, определенный для дискретных значений λ_j , имеет свое очевидное распространение на непрерывный случай.

Если решением задачи (10) является некоторый вектор λ^* , то решением задачи (9) будет считаться такой вектор λ^0 , у которого компоненты, соответствующие n' максимальным компонентам λ^* , равны 1, а остальные - 0.

Задача (10) является классической задачей нелинейного программирования и ее решение можно получить, привлекая различные численные методы. В для решения задачи (10) предварительно вводятся соответствующие штрафные функции, а затем используется градиентный метод оптимизации. Преимущество данного подхода к определению информативного набора признаков по сравнении с рассмотренными выше методами частичного перебора заключается в надежности получаемых результатов, которая обеспечивается использованием хорошо развитого "математического аппарата численных методов оптимизации". В то время как методы частичного перебора базируются лишь на том или ином эвристическом принципе выбора информативных признаков, который при всей своей "разумности" не исключает возможности пропуска оптимального набора признаков, так как при таком переборе не учитывается большая часть возможных вариантов составления комбинаций признаков.

Завершая рассмотрение наиболее распространенных методов выбора признаков, следует отметить, что, несмотря на достигнутые результаты, решение задачи определения информативных наборов признаков и поныне не потеряло своей актуальности в связи с расширением сферы применения методов распознавания образов на новые прикладные задачи со специфическими особенностями. Эффективному использованию известных методов выбора информативных наборов признаков при решении таких задач препятствует ряд проблем. Одна из них связана с выбором наиболее подходящей для конкретной задачи комбинации метода и критерия, т.е. фиксированием для этой задачи такой пары метода и критерия, при использовании которых минимизируются вычислительные затраты и улучшается качество распознавания. Поскольку по приблизительным подсчетам к настоящему времени разработано более 20 методов выбора информативных наборов признаков и почти 30 критериев информативности, то при решении каждой новой задачи необходимо просмотреть порядка 600 пар методов и критериев, что несомненно вызовет большие трудности. Поэтому в дальнейших исследованиях на первый план должна выступать проблема создания для наиболее распространенных критериев информативности соответствующих методов, которые в отличие от существующих методов частичного перебора были бы более обоснованными и строгими.

Решение этой проблемы позволит значительно упростить процедуру выбора для конкретной задачи наиболее подходящей пары метода и критерия: поиск ее будет связан лишь с определением подходящего критерия.

2. АНАЛИТИЧЕСКИЙ ПОДХОД К ФОРМИРОВАНИЮ ИНФОРМАТИВНОЙ ПОДСИСТЕМЫ ПРИЗНАКОВ В ЗАДАЧЕ РАСПОЗНАВАНИЯ

Рассмотрим задачи формирования информативной подсистемы признаков в классической постановке. Пусть задана априорная информация в виде множества объектов $X=\{x\}$, каждый из которых описан своим набором признаков $x=(x^1, x^2, \dots, x^n)$. Предполагаем, что произведено разбиение объектов на непересекающиеся классы X_1, X_2, \dots, X_m и каждый класс X_p содержит k_p объектов, т.е.

$$X_p = \{x_{p_i} : x_{p_i} = (x_{p_i}^1, x_{p_i}^2, \dots, x_{p_i}^n), i=1,2,\dots,k_p\}$$

Ставится следующая задача. Из исходной системы n признаков необходимо выделить подсистему из n' признаков (где $n' < n$), которая является наиболее информативной среди всех подсистем мощности n' в смысле наилучшего качества разбиения объектов на классы.

Предлагаемый ниже метод позволяет аналитически построить информативную подсистему признаков на примере одного вида функционала на основе максимизации которого в исходной системе из n признаков выбираются n' наиболее информативных признаков. Метод легко распространяется и на другие виды функционалов, например, на функционал, предложенный в. Следуя, введем необходимые понятия, которые будут использованы ниже. За меру близости, характеризующую среднеквадратичный разброс объектов классов X_p и X_q , будем использовать величину

$$R(X_p, X_q) = \sqrt{\frac{1}{k_p} \frac{1}{k_q} \sum_{l=1}^{k_p} \sum_{t=1}^{k_q} \sum_{j=1}^n \lambda_j (x_{pl}^j - x_{qt}^j)^2}.$$

В качестве меры близости между объектами данного класса X_p , $p=1,2,\dots,m$, примем величину

$$S(X_p) = \sqrt{\frac{2}{k_p} \cdot \frac{1}{k_p-1} \sum_{l=1}^{k_p} \sum_{t=1}^{k_p} \sum_{j=1}^n \lambda_j (x_{pl}^j - x_{pt}^j)^2},$$

которая характеризует среднеквадратичный разброс объектов внутри класса X_p .

Совокупность признаков объектов может быть описана n -мерным булевским вектором $\lambda=(\lambda_1, \lambda_2, \dots, \lambda_n)$, компоненты которого принимают значения 1 или 0 в зависимости от того, будет включен данный признак в информативную подсистему или нет.

Решим задачу определения информативной подсистемы признаков исходя из необходимости максимизации отношения расстояний между классами к среднеквадратичным разбросам объектов внутри классов. Формально это будет следующая задача максимина:

$$\max_{\lambda} \min_{p,q=1,2,\dots,m} \Phi = \max_{\lambda} \min_{p,q=1,2,\dots,m} \frac{\frac{1}{k_p} \frac{1}{k_q} \sum_{l=1}^{k_p} \sum_{t=1}^{k_q} \sum_{j=1}^n \lambda_j (x_{pl}^j - x_{qt}^j)^2}{\sqrt{\frac{2}{k_p} \frac{1}{k_q-1} \sum_{l=1}^{k_p} \sum_{t=1}^{k_q} \sum_{j=1}^n \lambda_j (x_{pl}^j - x_{pt}^j)^2}} \mathbf{x}$$

$$\mathbf{x} \frac{1}{\sqrt{\frac{2}{k_q} \frac{1}{k_q-1} \sum_{l=1}^{k_q} \sum_{t=1}^{k_q} \sum_{j=1}^n \lambda_j (x_{ql}^j - x_{qt}^j)^2}}. \quad (1)$$

Введем обозначения:

$$a_j = \sum_{l=1}^{k_p} \sum_{t=1}^{k_q} (x_{pl}^j - x_{qt}^j)^2; \quad (2)$$

$$c_j = \sum_{l=1}^{k_q} \sum_{t=1}^{k_q} (x_{ql}^j - x_{qt}^j)^2;$$

$$\sigma = \frac{1}{2} \sqrt{(1 - \frac{1}{k_p})(1 - \frac{1}{k_q})}.$$

Согласно (2) a_j, b_j, c_j - положительные постоянные величины.

Для σ имеем $0 \leq \sigma < \frac{1}{2}$.

Критерий информативности признаков, заданный в виде функционала задачи (1), является эвристическим и базируется на основополагающей в распознавании образов гипотезе компактности: с увеличением расстояния между классами улучшается их разделимость. Предпочтительными считаются те признаки, которые максимизируют это расстояние. В настоящем разделе для оценивания информативности признаков далее используются эвристические критерии, а именно - критерии, представленные функционалами вида

$$\Phi_1(\lambda) = \frac{\sum_{j=1}^n \lambda_j a_j}{\sqrt{\sum_{j=1}^n \lambda_j b_j} \sqrt{\sum_{j=1}^n \lambda_j c_j}}$$

и

$$\Phi_2(\lambda) = \frac{\sum_{j=1}^n \lambda_j a_j}{\sum_{j=1}^n \lambda_j b_j}$$

или функционалами сводимыми к этим видам, которые поддаются легко реализуемой максимизации. Выбор этих критериев обусловлен следующими обстоятельствами.

Известно, что наиболее обоснованным и естественным из всех критериев информативности признаков является оценка ошибки классификации, поставленная в соответствии всем наборам признаков, сформулированным на основе исходной системы переменных $x=(x^1, x^2, \dots, x^n)$. Преимущество такого определения критерия информативности состоит в том, что основанием служит задача распознавания: в результате выбора определенного набора признаков дается непосредственная количественная оценка информативности данного набора, связанная с качеством классификации. Однако реализация такого подхода к оцениванию информативности признаков требует выполнения значительного объема сложных вычислений.

Другая группа критериев, используемых для оценки качества признаков или их наборов, основана на статистике и теории информации. Эта группа включает критерии, основанные на мерах вероятностной зависимости. Результаты подробного анализа этих критериев приведены в, поэтому здесь отметим лишь, что все они, за исключением расстояния Бхаттачари, дивергенции, имеют сложный аналитический вид. Как отмечено в, большинство критериев этой группы в параметрическом виде использовать неудобно (их расчеты требуют многократного интегрирования), а в параметрическом – оценки вероятности ошибочной классификации получаются смещенными. Ввиду этого использование их для выбора информативных признаков весьма затруднено. Следовательно, возвращаясь к рассмотрению эвристических критериев, следует отметить, что они в отличие от критериев, основанных на статистике и теории информации, являются более простыми. Причем, несмотря на простоту, эвристические критерии в ряде случаев являются оптимальными. Именно эти особенности обусловили широкое применение их при решении практических задач классификации.

Справедлива следующая

Теорема 3.1. Пусть все множество объектов подразделено на классы X_1, X_2, \dots, X_m , все классы априорно описаны на языке признаков $X=(x^1, x^2, \dots, x^n)$. Если априорная информация (2) ранжирована в виде

$$\frac{a_{j_1}}{\sqrt{b_{j_1} c_{j_1}}} \geq \frac{a_{j_2}}{\sqrt{b_{j_2} c_{j_2}}} \geq \dots \geq \frac{a_{j_n}}{\sqrt{b_{j_n} c_{j_n}}}, \quad (3)$$

то каждый набор информативной подсистемы из n' признаков формируется из первых n' признаков, соответствующих этой ранжированной последовательности.

Доказательство. Ограничим рассмотрение теоремы двумя классами, т.е. $p=q=1$ и далее, не нарушая общности рассуждений. Будем полагать $\lambda_j=1$ ($j=1, 2, \dots, n$), так как аналитическим способом выводится формула, на основе которой может быть определена информативная подсистема, состоящая из произвольного числа признаков n' , где $1 \leq n' \leq n$. Полагая $\lambda_j=1$ ($j=1, 2, \dots, n$) и используя обозначения (2), функционал (1) перепишем следующим образом:

$$\Phi = \sigma \frac{\sum_{j=1}^n a_j}{\sqrt{(\sum_{j=1}^n b_j)(\sum_{j=1}^n c_j)}}. \quad (4)$$

Для знаменателя функционала (4) используем неравенство Гельдера. Неравенство Гельдера в общем случае имеет вид

$$\sum_{k=1}^n x_k y_k \leq \left(\sum_{k=1}^n x_k^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^n y_k^q \right)^{\frac{1}{q}},$$

где $p > 1$; $q = p/(p-1)$; $x_k, y_k \geq 0$, ($k = 1, 2, \dots, n$).

Полагая $p=2$ и $q=2/(2-1)=2$, знаменатель функционала (1) представляем в виде

$$\left[\sum_{j=1}^n (b_j^{\frac{1}{2}})^2 \right]^{\frac{1}{2}} \left[\sum_{j=1}^n (c_j^{\frac{1}{2}})^2 \right]^{\frac{1}{2}} \geq \sum_{j=1}^n (b_j^{\frac{1}{2}} c_j^{\frac{1}{2}}) = \sum_{j=1}^n (b_j c_j)^{\frac{1}{2}}. \quad (5)$$

Учитывая (5), выражение (4) записываем в виде:

$$\Phi = \sigma \frac{\sum_{j=1}^n a_j}{\sqrt{(\sum_{j=1}^n b_j)(\sum_{j=1}^n c_j)}} \leq \sigma \frac{\sum_{j=1}^n a_j}{\sum_{j=1}^n (b_j c_j)^{\frac{1}{2}}}. \quad (6)$$

Учитывая, что $0 \leq \sigma < 1/2$, из (6) получаем

$$\sigma \frac{\sum_{j=1}^n a_j}{\sum_{j=1}^n \sqrt{b_j c_j}} < \frac{\sum_{j=1}^n a_j}{\sum_{j=1}^n \sqrt{b_j c_j}}. \quad (7)$$

Последовательность величин a_j, b_j, c_j в (7) ранжируем следующим образом:

$$\frac{a_{j_1}}{\sqrt{b_{j_1} c_{j_1}}} \geq \frac{a_{j_2}}{\sqrt{b_{j_2} c_{j_2}}} \geq \dots \geq \frac{a_{j_n}}{\sqrt{b_{j_n} c_{j_n}}}.$$

Тогда для положительных величин $a_{j_i}, b_{j_i}, c_{j_i}$, ($j=1, 2, 3, \dots, n$) можно доказать справедливость неравенства

$$\frac{\sum_{i=l_1}^{l_2} a_{j_i}}{\sum_{i=l_1}^{l_2} \sqrt{b_{j_i} c_{j_i}}} \geq \frac{\sum_{i=l_3}^{l_4} a_{j_i}}{\sum_{i=l_3}^{l_4} \sqrt{b_{j_i} c_{j_i}}}, \quad (8)$$

где $1 \leq l_1 \leq l_2 \leq l_3 \leq l_4 \leq n$.

Доказательство неравенства (8) приведено в разделе 3.3. Используя неравенство (8) для выражения (7), приходим к следующему результату:

$$\frac{\sum_{i=1}^n a_{j_i}}{\sum_{i=1}^n \sqrt{b_{j_i} c_{j_i}}} \leq \frac{\sum_{i=1}^{n-1} a_{j_i}}{\sum_{i=1}^{n-1} \sqrt{b_{j_i} c_{j_i}}} \leq \dots \leq \frac{\sum_{i=1}^2 a_{j_i}}{\sum_{i=1}^2 \sqrt{b_{j_i} c_{j_i}}} \leq \frac{a_{i_1}}{\sqrt{b_{i_1} c_{i_1}}} \quad (9)$$

С помощью соотношений (9) можно определить информативную подсистему, состоящую из произвольного числа признаков n' в пределах от 1 до n ,

Теорема 2. Пусть имеется два набора информативных признаков и соответственно n' и n'' - числа признаков в каждом наборе, причем $n' < n''$. Тогда информативный признак, включенный в набор с числом признаков n' , входит в набор с числом признаков n'' .

Если функционал (1) представим как функцию от числа выделенных информативных признаков n' , т.е. $\Phi(n')$, то справедлива следующая

Теорема 3. Если априорная информация (2) ранжирована в виде последовательности (3), то в зависимости от числа информативных признаков для функционала справедливо выражение

$$\Phi(n'=1) \geq \Phi(n'=2) \geq \dots \geq \Phi(n'=n)$$

Доказательства последних двух теорем непосредственно следуют из неравенств (9). Результаты, полученные для двух классов, обобщим для m классов.

Все пары из m классов можно перенумеровать, вводя индекс $v=1, 2, \dots, \theta$, где

$$\theta = C_m^2 = \frac{m!}{(m-2)! 2!} = \frac{(m-1)m}{2}.$$

Следовательно, для каждой пары из m классов можно записать свое выражение (9). При фиксированном n' выбираем тот набор признаков, для которого выражение

$$\frac{\sum_{i=1}^{n'} a_{j_i}}{\sum_{i=1}^{n'} \sqrt{b_{j_i} c_{j_i}}}$$

принимает наименьшее значение.

Сформулируем описанный здесь метод формирования информативной подсистемы признаков в виде алгоритма, обозначаемого далее как A_1 .

Алгоритм A_1

Шаг 1. Вычисление значения функционала

$$\Phi_j = \frac{a_j}{\sqrt{b_j c_j}}, \quad (j=1,2, \dots, n)$$

для каждого исходного признака в отдельности.

Шаг 2. Ранжирование значений функционала, полученных на шаге 1, в виде последовательности неравенств (3).

Шаг 3. Формирование наиболее информативной подсистемы признаков заданной мощности n' из тех исходных признаков, которым соответствуют первые n' элементов ранжированной последовательности (3).

3. УПРОЩЕННЫЙ МЕТОД КЛАССИФИКАЦИИ С ИСПОЛЬЗОВАНИЕМ АНАЛИТИЧЕСКОГО ПОДХОДА ФОРМИРОВАНИЯ ИНФОРМАТИВНОЙ ПОДСИСТЕМЫ ПРИЗНАКОВ ПРИ НАЛИЧИИ ОБУЧАЮЩЕЙ ВЫБОРКИ

На основе разработанного аналитического подхода формирования информативной подсистемы признаков предложен упрощенный метод. Он состоит в выделении в исходной информативной системе признаков наименьшей подсистемы признаков, являющейся максимально информативной в смысле заданного критерия качества распознавания контрольных объектов. Следуя, введем необходимые понятия и обозначения.

Пусть задана обучающая выборка в виде множества объектов $\{x\}$, каждый из которых описан своим набором признаков $x=(x^1, x^2, \dots, x^n)$. Предполагается, что произведено разбиение объектов на непересекающиеся классы X_1, X_2, \dots, X_m и каждый класс X_p содержит k_p объектов

$$X_p = \{x_{pi} : x_{pi} = (x_{pi}^1, x_{pi}^2, \dots, x_{pi}^n), \quad i=1,2, \dots, k_p\}$$

В качестве меры близости между объектами одного класса X_p ($p=1,2, \dots, m$) используем величину

$$S(X_p) = \sqrt{\frac{1}{k_p - 1} \sum_{l=1}^{k_p} \sum_{j=1}^n \lambda_j (x_{pl}^j - \bar{x}_p^j)^2}, \quad (p=1,2, \dots, m), \quad (1)$$

которая характеризует среднеквадратичный разброс объектов внутри класса; здесь $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ - булевский вектор, причем

$$\sum_{j=1}^n \lambda_j = n'; \quad \bar{x}_p^j = \frac{1}{k_p} \sum_{l=1}^{k_p} x_{pl}^j, \quad (j=1,2,\dots,n),$$

$\bar{x}_p = (\bar{x}_p^1, \bar{x}_p^2, \dots, \bar{x}_p^n)$ - усредненный объект класса .

Пусть $x = (x^1, x^2, \dots, x^n)$ - некоторый контрольный объект, который необходимо распознавать. За меру близости объекта x к классу X_p будем использовать величину

$$\mu_p(x) = \frac{\|x - \bar{x}_p\|}{S(X_p)} = \frac{\sqrt{\sum_{j=1}^n \lambda_j (x^j - \bar{x}_p^j)^2}}{\sqrt{\frac{1}{k_p-1} \sum_{l=1}^{k_p} \sum_{j=1}^n \lambda_j (x_{pl}^j - \bar{x}_p^j)^2}}, \quad (2)$$

$(p=1,2,\dots,m).$

Принцип распознавания определим следующим образом: будем считать, что объект x на основе априорной информации вероятнее всего принадлежит классу X_q , если

$$\mu_q(x) = \min_{p=1,2,\dots,m} \mu_p(x). \quad (3)$$

Теперь введем меру достоверности распознавания. Если каждый класс X_p содержит k_p объектов, то общее количество объектов будет равно $\sum_{p=1}^m k_p$. Пусть m^* - число правильно распознанных объектов из общего числа объектов $\sum_{p=1}^m k_p$ по решающему правилу (3). Тогда в качестве меры достоверности распознавания можно использовать величину

$$P = \frac{m^*}{\sum_{p=1}^m k_p} \quad (4)$$

Как следует из (4), величина параметра P колеблется в пределах от 0 до 1 . Очевидно, что чем больше значение P , тем больше правильно распознанных объектов по алгоритму (3), т.е. лучше решена задача распознавания.

В (1) в качестве $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ будем использовать результаты аналитического подхода формирования информативной подсистемы признаков. Сущность задачи на данном этапе заключается в том, что в начале используются все информативные подсистемы признаков $n' = 1, 2, \dots, n$, удовлетворяющие решению задачи

$$\Phi(\lambda_1, \lambda_2, \dots, \lambda_n) \rightarrow \max_{\substack{\lambda_j \in \{0,1\} \\ \sum_{j=1}^n \lambda_j = n'}} ,$$

где n' - число информативных признаков.

Затем среди информативных подсистем признаков $n'=1,2,\dots,n$ требуется выбрать такую подсистему, которая обеспечивает наилучшее качество распознавания. Формально это выражается с помощью следующей задачи оптимизации

$$P(n^*) = \max_{1 \leq n' \leq n} P(n').$$

Здесь приведены результаты по оценке работоспособности разработанных алгоритмов построения информативного набора признаков.

Экспериментальным исследованиям подвергались алгоритмы, имеющие условные обозначения A_1, A_2 и A_3 (см. разделы 3.2 и 3.4). Первый из указанных алгоритмов основан на аналитическом методе формирования информативной подсистемы признаков, предложенном в настоящей работе (см. п. 3.2); второй - реализует метод, основанный на максимизации критерия информативности, представленный с помощью функционала Фишерского типа (см. п. 3.4); третий реализует известный метод "случайный поиск с адаптацией", являющийся в настоящее время одним из наиболее эффективных методов построения информативного набора признаков.

При формировании обучающей выборки использовались 33 модельных сигналов двух классов (первый класс содержал 9 объектов, второй - 24), сгенерированных для проведения экспериментов. Каждый из этих сигналов был представлен значениями 22 признаков.

В качестве единого критерия информативности наборов признаков при реализации алгоритмов A_1, A_2 и A_3 использовался функционал (см. п.п. 3.2 и 3.3) вида

$$\Phi(\lambda_1, \lambda_2, \dots, \lambda_n) = \frac{\sum_{j=1}^n \lambda_j a_j}{\sum_{j=1}^n \lambda_j \sqrt{b_j c_j}}$$

В таблице 1, 2 и 3 приведены результаты экспериментов по определению информативных наборов признаков с использованием алгоритмов A_1, A_2 и A_3 . Представленные в таблице 1, 2 и 3 данные показывают об эффективности алгоритма A_2 в смысле максимизации заданного критерия информативности. В то же время информативные подсистемы признаков, найденные с помощью алгоритма A_1 , несколько отличаются по составу от соответствующих подсистем признаков, найденных с применением алгоритма A_3 .

Таким образом, на конкретной экспериментальной обучающей выборке подтвердился факт о том, что среди представленных алгоритмов на экспериментальное

исследование алгоритм A_2 обеспечивает лучшие результаты по сравнению с алгоритмами A_1 и A_3 , для рассматриваемого критерия информативности наборов признаков.

В результате применения алгоритма A_2 получена наименьшая подсистема информативных признаков ($n'=4$), являющаяся максимально информативной в смысле наилучшего качества классификации.

При этом информативный набор включал в себя следующие признаки

$$x^2, x^3, x^{21}, x^{22}.$$

Таблица 1. Результаты реализации алгоритма A_1

Число информативных признаков, n'	Номер признака, составляющего информативную подсистему	Максимальное знач. критер. информ., Φ
1	21	2.574857
2	2	1.470701
3	3	1.173123
4	19	0.810150
5	16	0.735473
6	15	0.677316
7	5	0.07и401
8	4	0.674766
9	13	0.673679
10	11	0,671148
11	1	0.668976
12	7	0.664957
13	6	0.661200
14	12	0.659206
15	22	0.658930
16	14	0.655815
17	10	0.653644
18	8	0.650639
19	20	0.637749
20	18	0.623271
21	17	0.618469
22	9	0.616698

Таблица 2. Результаты реализации алгоритма A_2 .

Число информативных признаков, n'	Номер признака, составляющего информативную подсистему	Максимальное знач. критер. информ., Φ
1	21	2.574857
2	22	2.021782
3	2	1.389256
4	3	1.141824
5	13	1.002369
6	5	0.921154
7	1	0.847427
8	4	0.800245
9	19	0.783590
10	12	0.776468
11	10	0.769044
12	11	0.761983
13	9	0.754082
14	14	0.745229
15	8	0.736916
16	6	0.728415
17	7	0.720398
18	17	0.706514
19	20	0.683260
20	18	0.658761
21	16	0.638858
22	15	0.616698

Таблица 3. Результаты реализации алгоритма A_3

Число информативных признаков, n'	Номера признаков, составляющих информативную подсистему	Максимальное знач. критер. информ., Φ
1	21	2,574857
2	21;3;	1,268297
3	21;3;8;	0,879507
4	21;3;8;12;	0,796589
5	2;5;10;14;19;	0,760506
6	22;3;8;13;19;1;	0,754368
7	1;5;9;14;19;10;2;	0,745188
8	21;3;8;13;19;1;5;9	0,764260
9	3;8;12;19;22;4;9;14	0,728429
10	22;4;9;13;19;5;14;1;10;2	0,736259
11	4;8;13;19;22;9;14;1;5;2;10;	0,727484
12	21;3;8;13;18;22;4;9;19;14;1;15	0,704212
13	21;3;8;12;18;22;4;9;13;15;14;19	0,700529
14	21;4;8;13;18;22;9;19;1;5;14;10;2;6	0,697358
15	12;17;21;4;8;13;18;22;9;14;19;1;5;10;2	0,687816
16	21;3;7;12;17;22;4;8;13;19;1;5;9;14;10;2	0,717666
17	21;3;8;12;18;22;4;9;13;19;1;5;14;10;2;6;11;	0,692491
18	12;17;22;4;8;13;18;1;5;9;14;19;10;2;6;21;3;11;	0,681318
19	11;17;21;3;7;12;8;13;18;4;9;19;1;5;14;2;10;6;	0,676344
20	21;3;7;12;17;22;4;8;13;18;5;9;19;1;10;14;20;2;6;11	0,658761
21	6;11;16;20;3;7;12;17;21;8;18;22;4;9;13;19;1;5;14;10;2;	0,638858
22	1;5;10;14;19;6;15;20;2;7;11;16;3;12;17;21;4;8;13;22;9;18;	0,616698

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Сущность задачи построения информативного набора приказов и основные подходы к её решению.
2. Каким образом формировались критерий информативности признаков?
3. Оценка сложности алгоритма А1.
4. Сущность упрощенного метода классификации объектов в чем заключается.

ЛИТЕРАТУРА

79.
80.
81.
82.
83.

ЛИТЕРАТУРА

1. Альянах И.Н. – Моделирование вычислительных систем, Ленинград: Машиностроение, 1988 г. — 223 стр.
2. Растигин Л.А. – Современные принципы управления сложными объектами, Москва: Советское радио, 1980 г. — 232 стр.
3. Растигин Л.А., Маджаров Н.Е. – Введение в идентификацию объектов управления, Москва: Энергия, 1977 г. — 216 стр.