

**ГОСУДАРСТВЕННЫЙ КОМИТЕТ СВЯЗИ, ИНФОРМАТИЗАЦИИ И
ТЕЛЕКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ РЕСПУБЛИКИ УЗБЕКИСТАН**

ТАШКЕТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

На правах рукописи

УДК 004.421

БЕГИМОВ ЎКТАМ ИБРОГИМОВИЧ

**Разработка алгоритма и программного обеспечения
для установления системных связей на основе интервального анализа**

ДИССЕРТАЦИЯ

на соискание академической степени магистра
5A330201 – Компьютерные системы и их программное
обеспечение (по отраслям)

Научный руководитель:

к.ф-м.н. Кабулов Р.В.

СОДЕРЖАНИЕ

| | | |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| | Введение | 3 |
| I глава. | ВИДЫ НЕОПРЕДЕЛЕННОСТЕЙ В ИСХОДНЫХ ДАННЫХ СИСТЕМ И ПРОЦЕССОВ УСТАНОВЛЕНИЯ СИСТЕМНЫХ СВЯЗЕЙ..... | 13 |
| 1. | Понятие системы и процесса установления системных связей | 13 |
| 2. | Полнота и достоверность исходных данных..... | 18 |
| | Выводы по главе | 20 |
| II глава. | МЕТОДЫ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ И СТРУКТУРИРОВАНИЕ НЕОПРЕДЕЛЕННОСТЕЙ В ИСХОДНЫХ ДАННЫХ..... | 21 |
| 1. | Исследование методов восстановления пропусков в массивах данных | 21 |
| 2. | Обзор методов структурирования данных в условиях неопределенности | 30 |
| 3. | Структурирование данных при моделировании эволюционных дискретных процессов | 35 |
| | Выводы по главе | 68 |
| III глава. | МОДЕЛИРОВАНИЕ СИСТЕМ И ПРОЦЕССОВ УСТАНОВЛЕНИЯ СИСТЕМНЫХ СВЯЗЕЙ ЗАЩИТЫ ИНФОРМАЦИИ В УСЛОВИЯХ НЕПОЛНОТЫ И НЕДОСТОВЕРНОСТИ ИСХОДНЫХ ДАННЫХ..... | 69 |
| 1. | Метод интервального анализа оценки уровня риска информационной системы..... | 69 |
| 2. | Задачи защиты информации, формализуемые как экстремальные задачи на графах с интервальными весами | 78 |
| 3. | Описания программного обеспечения..... | 80 |
| | Выводы по главе | 85 |
| | Заключение | 86 |
| | Список использованных источников | 87 |

АННОТАЦИЯ МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

В данной магистерской работе рассматривается тема «Разработка алгоритма и программного обеспечения для установления системных связей на основе интервального анализа». При решении многих прикладных задач приходится сталкиваться с неопределенностью в исходных данных. Учет таких неопределенностей в практических задачах и ее внедрение в виде автоматизации какого-либо процесса является актуальной проблемой сегодняшних дней.

Объект исследования – системы и процессы установления системных связей. Предмет исследования – условия неполноты и недостоверности исходных данных в показателях процессов и систем установления системных связей.

Диссертация состоит из введения, трех глав, заключения, списка литературы. В ходе выполнения диссертации разработаны интервальные аналоги решения задач линейного программирования симплекс методом, алгоритмы реализации машинной интервальной арифметики для компьютера, разработаны алгоритм и программное обеспечение для установления системных связей на основе интервального анализа.

МАГИСТРЛИК ДИССЕРТАЦИЯСИ АННОТАЦИЯСИ

Мазкур диссертация ишида “интервал тахлил асосида алоқа тизимларини ўрнатиш учун алгоритм ва дастурий таъминот ишлаб чиқиш” мавзусига бағишланган. Кўп амалий масалаларни ечиш жараенида биз бошланғич маълумотларни ноаниқлилиги билан келганига дуч келамиз. Шундай ноаниқликларни амалий масаларда амалда ишлатиш ва тадбиқ қилиш хозирги кунга келиб долзарб ҳисобланади.

Тадқиқот объекти – алоқа тизимларни ўрнатиш тизим ва жараенлари. Тадқиқот предмети эса алоқа тизимларни ўрнатиш жараенларининг бошланғич маълумотларнинг ноаниқлилиги.

Мазкур диссертация кириш, уч бобдан, хулоса ва адабиётлар рўйхатидан иборат. Диссертацияни бажариш давомида чизиқли дастурлаш симплекс усули масалаларни ечиш интервал ўхшаш усул яратилган, компьютерлар учун машина интервал арифметикасини бажариш алгоритмлари, интервал тахлил асосида алоқа тизимларини ўрнатиш компьютер тизимларини ҳимоялаш учун алгоритм ва дастурий таъминот ишлаб чиқилган.

MASTER'S DISSERTATION ANNOTATION

This master's thesis deals with the topic "Development of algorithms and software system for establishing relations based on interval analysis." When dealing with many applications have to deal with the uncertainty in the source data. Accounting for these uncertainties in practical problems and its implementation as the automation of a process is an urgent problem today.

The object of study - systems and processes of establishing systemic linkages. The subject of the study - the conditions of incompleteness and unreliability of the source data in terms of processes and systems to establish systemic connections.

The thesis consists of an introduction, three chapters, conclusion, bibliography. During the interval of the thesis developed counterparts for solving linear programming simplex method, the algorithms of the machine interval arithmetic for the computer algorithm and software system for establishing relations based on interval analysis.

Введение

Актуальность работы. При решении многих прикладных задач приходится сталкиваться с неопределенностью в исходных данных. Как правило, она вызвана ошибками измерений, округления, приближенным представлением и распределенным характером самого исследуемого процесса и т.д. Учёт таких неопределенностей всегда был важен в тех разделах научно-прикладных исследований, которые в качестве исходных данных используют измерения, и в настоящее время их учёт остается актуальным.

В последние годы широкое распространение в вычислительной математике получили, методы интервального анализа.

Интенсивное развитие и проникновение в различные области математики интервальных методов привело к созыву в 1975 г. Первого международного симпозиума по интервальной математике, второй симпозиум проведен в 1980г. Первоначально интервальные методы возникли, как средство автоматического контроля ошибок округления на ЭВМ в последствии превратились в один из разделов современной прикладной математики. При этом в основе лежала двухсторонняя аппроксимация, контроля переучетом погрешности приводит к необходимости обобщения понятия вещественного числа, а именно, к понятию интервального числа. В монографии Мура, по существу впервые были изложены последовательные основы нового направления в вычислительной математике. Последующие исследования показали, что методы интервального анализа могут служить не только для учета ошибок округления на ЭВМ, но и являются новым аналитическими методами для теоретических исследований.

Широкое внедрение в вычислительную практику высокопроизводительных компьютеров вызвало необходимость не только в априорном анализе ошибок округлений, но и в методах, позволяющих

проводить апостериорный анализ ошибок округлений. В течение последнего десятилетия интенсивное развитие получили интервальные методы. Появившиеся как средство автоматического учета ошибок округлений, интервальные методы превратились в способ аналитических исследований и позволяют учитывать как погрешности в начальных данных, так и погрешности метода, тем самым интервальные методы дают возможность определить в процессе вычислений гарантированные двусторонние приближения к искомым точным решениям.

В связи с бурным ростом потребности в информации наблюдается тенденция увеличения обмена трафиками. В тоже время предъявляются достаточно жесткие требования к надежности и устойчивости функционирования сетей телекоммуникации. Надежность сетей телекоммуникации зависит от надежности элементов сети и при их высокой степени надежности устойчивость функционирования сети определяется наличием обходных и резервных путей. В условиях перехода к полностью оптическим сетям телекоммуникации, позволяющих обслуживать очень большие трафики, задача обеспечения высокой надежности и устойчивости функционирования сетей телекоммуникации Узбекистана является актуальной.

В работе И.Б. Шибинского отмечается, что решению проблемы построения отказоустойчивых вычислительных систем уделяется большое внимание ведущими фирмами США, Японии и других развитых государств. Для решения этой проблемы применяются разные способы. Основными из них являются: четырехкратное резервирование всех основных компонент вычислительных систем, мажоритарное резервирование процессоров по методу “два из трех”, дублирование ЭВМ, скользящее резервирование процессоров кратности $1/t$ с обеспечением возможности самостоятельно находить себе работу.

В работе со ссылкой на первоисточники отмечается, что четырехкратное резервирование основных компонент на уровне мультимикропроцессоров

систем связано с большими экономическими затратами и вместе с тем не защищает от отказов компараторов, число которых вдвое больше числа основных компонент. Мажоритарное резервирование менее эффективно, чем дублирование микропроцессоров. Скользящее резервирование кратности $1/t$ мало эффективно и, вдобавок требует дополнительных коммутирующих устройств. Дублирование компьютер вдвое увеличивает финансовые затраты.

В работе приведены аналитические методы получения вероятностных функций для расчета безотказности, отказоустойчивости, живучести и безопасности технических систем.

В работе рассматриваются вопросы защиты аппаратуры связи от перенапряжений и помех. Это связано с тем, что современная элементная база (интегральные микросхемы и др.) характеризуются высокой плотностью размещения элементов на кристалле и чрезвычайно малыми энергиями полезных сигналов, поэтому защита от электромагнитных воздействий становится обязательным условием бесперебойной работы телекоммуникационного оборудования.

Более вероятными путями проникновения перенапряжений в аппаратуру являются сигнальные и питающие цепи, поэтому предлагается, в первую очередь, защищать их путем установки специальных устройств.

В работе предлагается комплексный подход к решению проблемы защиты средств радио и проводной связи от импульсных грозовых перенапряжений. При этом отмечается, что основными техническими мероприятиями для защиты от импульсных перенапряжений, возникающих между различными элементами и составными частями аппаратуры или объекта в целом при прямом или близком ударе молнии, являются:

В работе рассмотрены способы минимизации последствий радиальной диффузии воды в конструкцию оптического кабеля (ОК). Дело в том, что стойкость к воздействию воды означает не только защиту ОК от продольного распространения воды во избежание выхода из строя подключенной к кабелю аппаратуры (аварийный режим при обрыве ОК или оболочки), но и

защиту в штатном режиме элементов ОК от воды, диффундирующей через его оболочки. При длительной эксплуатации вода, проникающая через оболочки кабеля к его "влаг чувствительным" элементам (поверхности оптического волокна, изоляции, например, из облученного полиэтилена и т.д.), может привести к ухудшению его оптических, электрических, механических характеристик. На основе результатов исследований процесса водопоглощения в различных кабельных конструкциях авторы предлагают следующие способы защиты кабелей и их конструктивных элементов от радиального воздействия воды (паров воды):

- уменьшение концентрации влаги в кабеле за счет выбора материала внешней оболочки с минимально возможным предельным поглощением;

- использование полых трубок, обеспечивающих эффект "непроницаемости" сердечника кабеля для воды;

- увеличение времени насыщения конструкций влагой.

Построению «само залечивающихся» сетей на основы СЦИ способствуют следующие факторы:

- наличие встроенных средств контроля и управления;

- деление сети на независимые функциональные слои;

- «интеллектуальные» возможности мультиплексоров и аппаратуры оперативного переключения (АОП).

Процедура реконфигурации может быть централизованной или децентрализованной (распределенной). В первом случае должен быть центр управления сетью, который собирает информацию о состоянии всех ее элементов, при необходимости принимает решение о реконфигурации и рассылает соответствующие команды на переключение. Основное преимущество централизованного метода – меньшая сложность его реализации, а главный недостаток – критичность к отказам самого центра управления, к потере или искажению информации, как поступающей в него, так и направленной из него.

Распределенные процедуры реконфигурации не требуют наличия

центра управления. При возникновении отказов на сети комплекты АОП различных узлов, обмениваясь между собой сообщениями, сами определяют состояние сети, вырабатывают согласованное решение по ее реконфигурации и реализуют его. Основным недостатком распределенных процедур – гораздо большая сложность их реализации и, как следствие, увеличение времени выполнения реконфигурации.

Организация «самозалечивания» на основе АОП предполагает наличие весьма развитой системы сетевого управления.

Описанные выше основные конфигурации сетей и методы организации в них «самозалечивания» можно применять не только в чистом виде, но и комбинировать их. Если сеть СЦИ состоит из нескольких колец, то в целях обеспечения сопряжения и взаимодействия их используются МВВ и АОП. Заметим, что использование АОП дает большие возможности. Это можно увидеть на схеме объединения колец. В этом случае могут быть организованы различные логические кольца, охватывающие и связывающие несколько цепочек МВВ между собой.

С помощью смешанной архитектуры, использующей как кольцевые структуры, так и АОП, можно строить более эффективные сети с таким уровнем отказоустойчивости, как и у чисто кольцевой сети, но с меньшей суммарной пропускной способностью всех линий.

Разумеется, что более простой и дешевый вариант объединения колец – наличие для двух смежных колец только одного общего узла. Недостаток его – прерывание связи между кольцами при выходе этого узла из строя. Поэтому рекомендуется сопряжение колец в двух узлах, что обеспечивает устойчивость сети к одиночным отказам её элементов.

Таким образом, выбор структуры сети телекоммуникации предполагает детальный технико-экономический анализ, в котором необходимо учитывать, в частности, размеры сети, требования к её надежности и живучести, распределение трафика между узлами и другие факторы.

В отчается, что одним из основных направлений обеспечения

живучести сетей телекоммуникации является поиск структур, устойчивых к возмущениям и вредным воздействиям. Предлагается способ обеспечения живучести сети путём отыскания структур с минимальной средней длиной линий передачи.

В предлагается при проектировании сети телекоммуникации учитывать результаты геофизических исследований температурных зависимостей и грозодеятельности районов, от которых зависит коэффициент готовности участков сети. Это позволяет осуществить мероприятия, повышающие структурную надежность участков сети, а также надежность отдельных ребер сети, что сократит неоправданные затраты на проведение работ по повышению структурной надежности элементов и всей сети, а также экономические потери, возникающие при простое систем передачи из-за низкой защищенности сетей от внешних воздействий.

В рассмотрены вопросы повышения надежности ВОСП на этапе их эксплуатации. Предлагается автоматизировать процессы технической эксплуатации, что позволит повысить оперативность обнаружения, локализации и устранения возникающих неисправностей и в целом приводит к уменьшению коэффициента простоя. Отмечается, что структурное и поэлементное резервирование, наряду с автоматизацией и централизацией функций технической диагностики в процессе эксплуатации ВОСП, широко используется при создании и внедрении систем и сетей связи в зарубежных развитых странах. Техническая диагностика высокоскоростных цифровых систем телекоммуникации включает в себя следующие функции:

обнаружение неисправности в трактах передачи;

переключение вышедшего из строя тракта передачи (или компонента аппаратуры) на резерв;

поиск и локализация неисправности в вышедшем из строя тракте передачи;

сбор, передача и обработка служебной информации (сигналов телеуправления, телесигнализации и служебной связи).

В отмечается также, что при разработке программно-технических средств автоматизации важная роль, с точки зрения повышения качества функционирования ВОСП, отводится выбору оптимальной стратегии восстановления функционирования, с учетом условий эксплуатации, резервирования, размещения ремонтно-восстановительных бригад на обслуживаемом участке магистрали. В этой работе рекомендуется также в стратегии восстановления учесть использование предотказового состояния аппаратуры ВОСП.

В автор предлагает для повышения надежности ВОСП использовать структурное резервирование (общее или отдельное). Для оптимизации рекомендует схему отдельного резервирования, так как надежность, с ростом кратности резерва возрастает быстрее, чем в схеме общего резервирования. Кроме того, в такой схеме можно применить различную кратность резервирования по каждому из элементов объекта технической эксплуатации.

Из анализа методов устойчивого функционирования систем и сетей телекоммуникаций следует, что разработка алгоритма и программного обеспечения для установления системных связей на основе интервального анализа является актуальной задачей. Всё еще остается актуальной проблема поиска эффективных мер обеспечения устойчивого функционирования систем и сетей телекоммуникаций. Дополнительными мерами повышения надежности, живучести, информационной и физической безопасности систем телекоммуникаций могут быть использование методов и средств прогнозирующего контроля, что в значительной мере позволит улучшить показатели их устойчивого функционирования.

Основной целью исследований является разработка алгоритма и программного обеспечения для установления системных связей на основе интервального анализа, а также исследование свойств интервальных отображений для получения оптимальных интервальных расширений и установление связей между вещественными и интервальными

отображениями; изучения задачи интерполяции интервального анализа; построения интервальных методов решения оптимизационных задач линейного программирования; разработка алгоритмов реализации машинной интервальной арифметики с учетом особенностей устройства компьютера; исследование существующих подходов к моделированию систем и процессов в условиях неполноты и недостоверности исходных данных для возможности обоснованного выбора математического аппарата для решения конкретных задач защиты информации.

Постановка задачи

Для установления системных связей решаются следующие задачи:

- привести обзор возможных неопределенностей в исходных данных систем и процессов установления системных связей защиты информации;
- исследовать методы восстановления пропущенных значений и структурирования неопределенностей в исходных данных;
- решить задачи оценки уровня риска информационной системы и многокритериальной оценки средств защиты информации с использованием методов структурирования неопределенностей в исходных данных.

Чтобы «максимизировать» целевые функции, можно использовать понятия «оптимального решения», аналогичные 1) идею удовлетворяющего решения, 2) идею α - эффективного решения.

Для каждого критерия мы допускаем существование заданной дополнительной цели - некоторого нечеткого множества на вещественной оси. Определение цели может быть существенным для качества «оптимального» решения. Его смысл зависит, однако, от природы критериев, будучи в некотором смысле идеальными значениями соответствующих критериев. Нечеткое значение функции критерия сравнивается с целью с помощью некоторого нечеткого отношения, также заданного извне. Тогда нечеткие критерии обрабатываются как ограничения.

В рамках интервального анализа данную проблему можно интерпретировать как максимизацию модель принятия решения.

Из анализа методов устойчивого функционирования систем и сетей телекоммуникаций следует, что:

в настоящее время используется большое количество мер для обеспечения отказоустойчивости и проектного качества функционирования систем телекоммуникаций;

основным методом обеспечения устойчивого функционирования систем и сетей телекоммуникаций является метод структурного резервирования;

мало используются методы прогнозирующего контроля за состоянием отдельных элементов и в целом систем и сетей телекоммуникаций;

несмотря на принятые многочисленные меры обеспечения устойчивости функционирования аппаратуры и линейно-кабельных сооружений в процессе эксплуатации имеют место отказы аппаратуры и повреждении линий передачи.

Введенные, понятия эквивалентных интервальных отображений, интервальных расширений могут быть использованы как при применении интервальных методов, так и при проведении интервально-аналитических исследований. Предложенные интервальные методы решения задач линейного программирования симплекс методом, могут применяться для решения задач с неточно заданной начальной информацией.

Объект исследования – системы и процессы установления системных связей.

Предмет исследования – условия неполноты и недостоверности исходных данных в показателях процессов и систем установления системных связей.

Общая методика исследований

В работе используются методы интервального анализа, получившие развитие в работах Шокина, Назаренко, Юлдашева и других авторов.

Основные результаты исследований сводятся к следующему:

- Разработаны интервальные аналоги решения задач линейного программирования симплекс методом; алгоритмы реализации машинной интервальной арифметики для компьютера;
- разработаны алгоритм и программное обеспечение для установления системных связей на основе интервального анализа.

Опубликованность результатов. Основные результаты диссертации опубликованы в виде научных статей, материалов конференций. Список публикаций приведен в списке литературы.

Структура и объем диссертации. Диссертация состоит из введения, трех глав, заключения, списка литературы и приложений. Объем диссертации 89 страниц. Список литературы содержит 20 наименований.

Глава 1. ВИДЫ НЕОПРЕДЕЛЕННОСТЕЙ В ИСХОДНЫХ ДАННЫХ СИСТЕМ И ПРОЦЕССОВ УСТАНОВЛЕНИЯ СИСТЕМНЫХ СВЯЗЕЙ

1.1 Понятие системы и процесса установления системных связей

Задачи установления системных связей характеризуются следующими основными особенностями:

- наличием большого числа факторов, влияющих на эффективность решения задачи;
- отсутствием количественных достоверных исходных данных об этих факторах;
- отсутствием формальных (математических) методов получения оптимальных результатов решения по совокупности исходных данных.

Решение произвольных задач производится на основе моделей исследуемых объектов или процессов. Универсальной моделью любого объекта является представление его в виде системы.

Система – совокупность элементов и связей между ними, обладающая определенной целостностью.

Процесс – совокупность состояний системы, упорядоченных по изменению какого-либо параметра, определяющего свойства системы.

Система установления системных связей – совокупность взаимосвязанных элементов, функционирование которых направлено на обеспечение безопасности информации.

Элементами безопасности информации являются люди, инженерные конструкции и технические средства, обеспечивающие защиту информации независимо от их принадлежности к другим системам.

Система безопасности информации описывается следующими параметрами:

- 1) цели и задачи (конкретизированные в пространстве и во времени цели);
- 2) входы и выходы системы;
- 3) процессы внутри системы, обеспечивающие преобразование входов в выходы.

Цели представляют собой ожидаемые результаты функционирования системы защиты информации.

Задачи – необходимые действия для достижения системой поставленных целей.

Возможность решения задач зависит от ресурса, выделяемого на установления системных связей. Ресурс включает в себя людей, решающих задачи защиты информации, финансовые, технические и другие средства, расходуемые на защиту информации.

Входами установления системных связей защиты информации являются угрозы информации.

Выходами – меры, которые нужно применить для предотвращения угроз или снижение их до требуемого уровня.

Процесс образуют мероприятия, действия и технологии, определяющие меры защиты информации. Для выбора оптимальных мер защиты из множества известных необходимо иметь показатели эффективности.

На рисунке 1. приведена схема этих параметров.

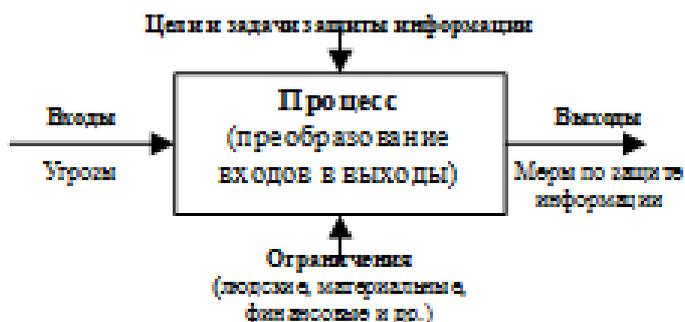


Рисунок 1 – Параметры установления системных связей защиты информации

Решение задачи установления системных связей защиты информации с точки зрения системного подхода можно сформулировать как трансформацию существующей системы, не обеспечивающей требуемый уровень защищенности, в систему с заданным уровнем безопасности информации.

Основным методом исследования установления системных связей защиты информации является моделирование. Моделирование предусматривает создание модели и ее исследование (анализ). Описание или физический аналог любого объекта, в том числе установления системных связей защиты информации и ее элементов, создаваемые для определения и исследования свойств объекта, представляют собой его модель. В модели учитываются существенные для решаемой задачи элементы, связи и свойства изучаемого объекта.

Различают вербальные, физические и математические модели и соответствующее моделирование.

Вербальная модель описывает объект на национальном и профессиональных языках. Человек постоянно создает вербальные модели его окружающей среды и руководствуется ими при принятии решений. Чем точнее модель отображает мир, тем эффективнее при прочих равных условиях деятельность человека. На естественном или профессиональном языке можно описать любой объект или явление. Сложные модели прошлой, настоящей, будущей жизни людей создают писатели. Но вербальные модели позволяют анализировать связи между ее элементами лишь на качественном уровне.

Физическая модель представляет материальный аналог реального объекта, который можно подвергать в ходе анализа различным воздействиям и получать количественные соотношения между этими воздействиями и результатами. Часто в качестве физических моделей исследуют уменьшенные копии крупных объектов, для изучения которых отсутствует

инструментарий. Модели самолетов и автомобилей продувают в аэродинамических трубах, макеты домов для сейсмических районов испытывают на вибростендах и т.д. Но возможности физического моделирования объектов защиты и угроз ограничены, так как трудно и дорого создать физические аналоги реальных объектов. Действительно, для того, чтобы получить физическую модель канала утечки, необходимо воспроизвести его элементы, в том числе среду, а также априори неизвестные средства и действия злоумышленника.

По мере развития вычислительной математики и техники расширяется сфера применения математического моделирования. Математическое моделирование предусматривает создание и исследование математических моделей реальных объектов и процессов. Математические модели могут разрабатываться в виде аналитических зависимостей выходов системы от входов, уравнений для моделирования динамических процессов в системе, статистических характеристик реакций системы на воздействия случайных факторов. Математическое моделирование позволяет наиболее экономно и глубоко исследовать сложные объекты, чего нельзя добиться с помощью вербального моделирования или что чрезмерно дорого при физическом моделировании. Возможности математического моделирования ограничиваются уровнем формализации описания объекта и степенью адекватности.

Для моделирования сложных систем все шире и шире применяется метод математического моделирования, называемый имитационным моделированием. Оно предполагает определение реакции модели системы на внешние воздействия, которые генерирует ЭВМ в виде случайных чисел. Статистические характеристики (математическое ожидание, дисперсия, вид и параметры распределения) этих случайных чисел должны с приемлемой точностью соответствовать характеристикам реальных воздействий. Функционирование системы при случайных внешних воздействиях описывается в виде алгоритма действий системы и их характеристик в ответ

на каждое воздействие на входе. Таким образом, имитируется работа сложной системы в реальных условиях. Путем статистической обработки выходных результатов при достаточно большой выборке входных воздействий получаются достоверные оценки работы системы. Например, достаточно объективная оценка эффективности системы защиты информации при многообразии действий злоумышленников, которые с точки зрения службы безопасности носят случайный характер, возможна на основе имитационного моделирования системы защиты.

Другое перспективное направление математического моделирования, которое представляет интерес для моделирования объектов защиты и угроз информации – компьютерные деловые игры. Компьютерные деловые игры – аналог деловых игр людей, применяемый для решения проблем в организационных структурах. Деловая игра имитирует процесс принятия решения в сложных условиях недостаточности достоверной информации людьми, играющими роль определенных должностных лиц. Участниками компьютерной игры являются два человека или компьютер и человек. Причем за сотрудника службы выступает человек, а злоумышленника – компьютер или человек. Например, злоумышленник – компьютер устанавливает в случайном месте закладное устройство, а другой игрок – человек производит поиск закладного устройства с помощью различных выбранных средств по показаниям виртуальных приборов моделей этих средств.

В чистом виде каждый вид моделирования используется редко. Как правило, применяются комбинации вербального, физического и математического моделирования. С вербального моделирования начинается сам процесс моделирования, так как нельзя создать физические или математические модели, не имея образного представления об объекте и его словесного описания. Если есть возможность исследовать свойства объекта на физической модели, то наиболее точные результаты обеспечиваются при физическом моделировании.

Моделирование объектов защиты предусматривает определение источников с защищаемой информацией и разработку моделей материальных объектов защиты. К объектам защиты относятся источники защищаемой информации и контролируемые зоны, в которых находятся эти источники.

В результате этого этапа определяются:

- 1) модели объектов защиты с указанием всех источников информации с описанием факторов, влияющих на их безопасность;
- 2) цена C_{ii} защищаемой информации каждого i -го источника.

На основе полученных результатов на этапе моделирования угроз выявляются угрозы безопасности информации, производится оценка ожидаемого от их реализации потенциального ущерба и ранжирование угроз по потенциальному ущербу. При моделировании угроз определяются риск (вероятность) угрозы P_y и ущерб C_y в случае ее реализации.

1.2 Полнота и достоверность исходных данных

Полнота представляет собой показатель, характеризующий достаточность информации для решения соответствующих задач. Поэтому, чтобы иметь возможность определять данный показатель, необходимо для каждой задачи или группы задач заблаговременно составить перечень сведений, которые требуются для их решения. Для представления таких сведений удобно воспользоваться так называемыми объектно-характеристическими таблицами (ОХТ). ОХТ – это двухмерная матрица, по строкам которой приведен перечень наименований объектов, процессов или явлений, входящих в круг интересов соответствующей задачи, а по столбцам – наименования их характеристик (параметров), необходимых для решения задачи.

Значения характеристик при этом будут располагаться на пересечении соответствующих строк и столбцов. Совокупность всех ОХТ, необходимых для обеспечения решения всех задач объекта, называется его

информационным кадастром. Для оценки полноты информации часто используют следующую методику.

Обозначим через $d_{\mu\nu}$ элемент, находящийся в μ -й строке и ν -м столбце интересующего нас компонента соответствующей ОХТ, причем:

$$d_{\mu\nu} = \begin{cases} 1, & \text{если по данному элементу информация имеется;} \\ 0, & \text{в противном случае.} \end{cases}$$

Тогда в качестве коэффициента полноты информации в данной ОХТ можно принять величину:

$$K_{\Pi} = \frac{\sum_{\mu} \sum_{\nu} d_{\mu\nu}}{mn},$$

где m – число строк, а n – число столбцов ОХТ.

Однако при этом не учитывается важность (значимость) различных элементов. В целях устранения этого недостатка положим, что $K_{\mu\nu}^{(g)}$ есть коэффициент важности элемента μ -й строки и ν -го столбца. Тогда очевидно, в качестве меры взвешенной полноты информации в рассматриваемой ОХТ можно принять величину:

$$K_{\Pi}^{(g)} = \frac{\sum_{\mu} \sum_{\nu} d_{\mu\nu}}{mn \sum_{\mu} \sum_{\nu} K_{\mu\nu}^{(g)}}.$$

Достоверность определяется как «уровень разумной уверенности в истинности некоего высказывания, который удовлетворяет некоторым правилам непротиворечивости и в соответствии с этими правилами формально может быть выражен числом».

Известные подходы к решению проблемы оценки достоверности связаны с применением теоремы Байеса и теории нечетких множеств, на основе которых в настоящее время разработаны и применяются в экспертных системах практические способы объединения свидетельств, регистрирующих качественные и логико-семантические связи между фрагментами базы данных.

Используя идею байесовского подхода, можно поставить вопрос о

достоверности фрагментов информационной базы данных в более общем плане, рассматривая любой ее фрагмент как гипотезу, а фрагменты, с которыми он связан, как свидетельства относительно фрагмента-гипотезы. Под фрагментом информационной базы данных понимают часть содержания или структуры, обладающую свойством дискретности и независимости, т.е. некоторую совокупность данных или высказываний, имеющую самостоятельный смысл. Достоверность (D) фрагмента поступающего в информационную базу данных, зависит от достоверности источника информации и методики ее получения. Каждый вновь поступающий в информационную базу данных фрагмент (НФЗ – новый фрагмент знаний) есть пара:

$$НФЗ = \langle Z, D \rangle,$$

где Z – значение фрагмента; D – достоверность информации.

1.3 Выводы по главе

В главе приведены понятия процесса и установления системных связей защиты информации. Процесс – совокупность состояний системы, упорядоченных по изменению какого-либо параметра, определяющего свойства системы. Система защиты информации – совокупность взаимосвязанных элементов, функционирование которых направлено на обеспечение безопасности информации. При этом элементами системы защиты информации являются люди, инженерные конструкции и технические средства, обеспечивающие защиту информации независимо от их принадлежности к другим системам. Кроме того приведены понятия полноты и достоверности исходных данных.

Глава 2. МЕТОДЫ ВОССТАНОВЛЕНИЯ ПРОПУЩЕННЫХ ЗНАЧЕНИЙ И СТРУКТУРИРОВАНИЕ НЕОПРЕДЕЛЕННОСТЕЙ В ИСХОДНЫХ ДАННЫХ

2.1 Исследование методов восстановления пропусков в массивах данных

С проблемой обработки пропусков в массивах данных приходится сталкиваться при проведении разнообразных социологических, экономических и статистических исследований. Традиционными причинами, приводящими к появлению пропусков, являются невозможность получения или обработки, искажение или сокрытие информации. В результате на вход программ анализа собранных данных поступают неполные сведения.

Самым простым решением обработки данных является исключение некомплектных наблюдений, содержащих пропуски, и дальнейший анализ полученных таким образом "полных" данных. Понятно, что такой подход приводит к сильному различию статистических выводов, сделанных при наличии в данных пропусков и при их отсутствии.

Поэтому более перспективным является иной путь – заполнение пропусков перед анализом фактических данных. Можно выделить следующие преимущества данного подхода: ясное представление структуры данных; вычисление необходимых итоговых значений; уверенная интерпретация результатов анализа, так как можно опираться на традиционные характеристики и суммарные значения.

Сегодня создано множество методов восстановления пропусков, однако единая методология обработки подобных данных отсутствует, несмотря на ее необходимость.

Основной задачей данной работы является сравнительный анализ существующих методов восстановления пропущенных значений в массивах (рядах, таблицах) данных, в том числе с практической проверкой восстанавливающей способности наиболее известных алгоритмов.

Для решения поставленной задачи использовалось изучение

литературных источников и компьютерное моделирование ряда алгоритмов (замена пропуска общим средним, замена пропуска средним из ближайших, метод сплайн-интерполяции, Zet-алгоритм) с анализом их эффективности для реальных массивов данных.

Поясним постановку задачи предсказания (восстановления) значений пропущенных элементов на примере обработки таблицы размером $m \times n$ не содержащей пропусков.

Пусть в нашем распоряжении имеется набор различных методов (алгоритмов):

$$F = \langle f_1, f_2, \dots, f_k, \dots, f_n \rangle,$$

предназначенных для предсказания значений пропущенных элементов. Закроем в таблице известный элемент x_{ij} , стоящий на пересечении строки i и столбца j , и предскажем его с помощью всех методов F поочередно. Каждый метод f_k предскажет свое значение x_{ij}^* , при этом относительная ошибка предсказания будет определена как:

$$d_{ij}^* = \frac{|x_{ij}^* - x_{ij}|}{x_{ij}} \cdot 100\%.$$

Восстановим в таблице элемент x_{11} , уберем элемент x_{12} и повторим процедуру, получив при этом относительную ошибку d_{12}^* . Прделавав это по очереди со всеми элементами таблицы и просуммировав обнаруженные ошибки, получаем величину относительных ошибок D_k для каждого метода. Наилучшим из них естественно считать такой метод f_k^* , который дает минимальную сумму ошибок.

Методы из набора F могут отличаться друг от друга лежащими в их основании эвристическими предположениями (гипотезами). Возможность использования методов разной степени сложности связана с тем, насколько простым или сложным является механизм, согласно которому данные оказываются пропущенными. Используя терминологию, предложенную в,

будем называть пропуски в данных полностью случайными (data are missing completely at random – MCAR), если условная вероятность $P(X_j \text{ пропущено/прочие } X)$ не зависит ни от X_j , ни от прочих X (то есть эта вероятность постоянна для всех наблюдений, и наблюдаемые X_j являются случайной под выборкой тех X_j , которые должны были получиться в эксперименте). Пропуски в данных называются случайными (missing at random – MAR), если вероятность $P(X_j \text{ пропущено/прочие } X)$ не зависит от X_j , но может зависеть от других X . Оказывается, что в этих случаях механизм пропусков несущественен (ignorable), и к данным применимы вариации методов восстановления пропусков. Наконец, если $P(X_j \text{ пропущено/прочие } X)$ зависит от самого X_j , то механизм пропусков является существенным (non-ignorable), и для корректного анализа данных необходимо знать этот механизм. Введенные понятия относятся к отдельным переменным, и в пределах одной и той же базы данных можно, в принципе, наблюдать все приведенные варианты. Можно построить тесты, отличающие MAR от MCAR, однако по данным невозможно отличить, являются ли они MAR, или же механизм пропусков существенен.

Характеристика известных методов восстановления пропусков по литературным данным. Первый возможный подход к обработке данных с пропусками – это просто исключение некомплектных объектов. Данный метод легко реализуется, но необходимым условием его применения является следование данных требованию MCAR. Кроме того, необходимо, чтобы количество пропусков было небольшим, иначе происходят сильные смещения, кроме того, как показывает практика, данный метод не очень эффективен.

Следующим подходом, реализованным в большом числе алгоритмов, является подход с заполнением пропусков. Наряду с очевидными преимуществами данного подхода ему присущи недостатки: «...Опасность

этого подхода в том, что он не позволяет отличать ситуации, где задача не очень трудна и может быть корректно решена таким способом, от ситуаций, где обычные оценки по реальным и подставленным данным сильно смещены».

Выделяют следующие два принципиальных недостатка данного подхода.

1. Как правило, параметры для алгоритма заполнения пропусков вычисляются по присутствующим данным, что вносит зависимость между наблюдениями. Конечно, такой искусственной зависимости не возникает, если проводится заполнение константой или случайными значениями, не зависящими от присутствующих наблюдений в выборке, или методом подстановки без подбора. Но на практике эти методы представляют малую ценность. Зависимости можно также избежать, разделяя исходную выборку на две под выборки и вычисляя подстановки (например, средне выборочные значения) для анализируемой под выборки по значениям наблюдений во второй под выборке. При таком подходе приходится жертвовать частью информации, чтобы заполнить пропущенные значения.

2. Распределение данных после заполнения будет отличаться от истинного, даже если пренебречь зависимостью, указанной выше. Этот факт особенно нагляден для простых методов заполнения (средне выборочных, по регрессии и т.п.). Различные варианты данных методов заполнения приведут к смеси истинного и вырожденных распределений с вырождением на гиперплоскостях, на которых располагаются предсказываемые значения.

Отметим, что анализ подобных полных данных стандартными методами неправилен и приводит к таким недостаткам, как несостоятельность и смущённость оценок параметров. Однако методы данной группы широко применяются. Приведем особенности наиболее известных методов глобального заполнения (то есть заполнения с использованием данных всей выборки).

Заполнение средними (средним по всей выборке или средними по

группам) – применение имеет смысл только в случае следования данных условию MAR, данная группа методов легко реализуема; недостатки – искажение распределения данных, уменьшение дисперсии.

Методы заполнения с подбором – подобны методам заполнения средними, легко реализуемы, кроме того, при применении данных методов больше разброс дисперсии предсказанных значений пропусков, не искажается распределение. Недостаток – выявление распределения обрабатываемой генеральной совокупности данных по полученной выборке комплектных данных.

Заполнение по регрессии. В основу данной группы методов положены хорошо известные алгоритмы регрессионного анализа. Из условий применения данного метода можно выделить требование о следовании данных условию MAR (хотя для частных случаев возможно применение более слабых требований) и требования, относящиеся к выполнению предпосылок регрессионного анализа. Недостатки подобных методов очевидны: качество предсказания (восстановления пропусков) напрямую зависит от успешного выбора взятой за основу регрессионной модели.

Метод сплайн-интерполяции – обоснованный математически метод интерполяции, показывающий хорошие результаты. Для успешного применения необходимо, чтобы данные следовали условию MAR. Недостатки метода следуют из самой его идеи. Например, в случае восстановления группы пропусков, следующих подряд друг за другом, результат аппроксимации сплайном данной группы не всегда может дать оценки, приближающиеся с достаточной точностью к значениям, которые могли бы быть на месте пропусков.

Методы многократного заполнения. Основное их преимущество в том, что они преодолевают недостаток методов однократного заполнения в смысле большего разброса дисперсии оценки; посылки применимости данного метода полностью определяются используемыми методами формирования, множества вариантов восстановления пропуска.

МП-оценивание (EM-алгоритм) – относится к категории методов моделирования. Особенность данных методов – построение модели порождения пропусков с последующим получением выводов на основании функции правдоподобия, построенной при условии справедливости данной модели, с оцениванием параметров методами типа максимального правдоподобия. Отметим, что если другие методы восстановления пропусков требуют, чтобы данные отвечали условию MAR (или MCAR как более жесткому), то для данных методов возможно построение моделей, учитывающих конкретную специфику области, как следствие, возможна постановка более слабых условий к данным. Недостаток – необходимость построения модели порождения пропусков.

Использование методов факторного анализа. Особенности метода: отсутствие требования априорного заполнения пропусков, необходимость в предварительной нормировке данных, наличие требований факторного анализа. В случае использования нелинейных моделей данных метод имеет очевидное преимущество по сравнению с регрессионными методами. В связи с большим количеством шагов алгоритма данному методу присуща некоторая трудоемкость реализации.

Использование методов кластерного анализа. Особенность метода – его применение не опирается на какую-либо вероятностную модель, но при этом оценить его свойства в статистических терминах не представляется возможным. Однако данный метод обладает существенным достоинством, а именно, он позволяет указать предпочтительный порядок восстановления данных и выявить случаи, когда пропуски не могут быть восстановлены по имеющимся данным.

Локальные алгоритмы восстановления пропусков. Алгоритмы семейства Zet (Wanga), по сути, являются детально проработанной и апробированной технологией верификации экспериментальных данных, основанной на гипотезе их избыточности. Внешне они сходны с методом локального заполнения. Данные алгоритмы хорошо показали себя, но

необходимость задания ряда важных параметров приводит к необходимости убедиться в правдоподобности восстановленных значений.

Существуют другие альтернативные подходы к восстановлению данных. В качестве примера отметим использование нейросетей. Несмотря на некоторую эквивалентность моделей статистики и нейросетевых парадигм, трудно выявить явные преимущества и недостатки применения нейросетевых технологий для восстановления данных.

Практическое исследование восстанавливающей способности выбранных методов заполнения пропущенных значений было программно реализовано в системе Mathcad и частично в Microsoft Excel с тестированием на большом множестве реальных временных рядов с различной структурой пропусков (создаваемых в процессе исследования искусственно).

Итоговые результаты по исследовавшимся методам таковы.

1. Метод замены пропущенного значения общим средним из присутствующих элементов.

Данный метод был выбран для исследования в связи с тем, что это наиболее известный, простой способ восстановления пропусков, включенный как средство борьбы с ними в большинство статистических пакетов. Эксперименты показали полную несостоятельность данного метода даже на простых рядах данных.

2. Метод замены пропущенного значения средним из ближайших присутствующих элементов переменной.

Данный метод является эффективным развитием метода замены пропусков общим средним, и эксперименты показали хорошую точность метода в случае одиночных пропусков на достаточно гладких рядах данных. Благодаря простоте реализации можно даже рекомендовать использование данного метода в приведенных выше условиях, но только в них. Наличие в данных групповых пропусков или сильные флуктуации значений ряда сводят эффективность метода к нулю.

Таким образом, этот метод можно использовать только для

восстановления одиночных пропусков в рядах данных.

3. Метод восстановления пропущенного значения сплайн-интерполяцией по присутствующим элементам.

Все эксперименты показали, что в случае наличия в данных одиночных пропусков настоящий метод показывает лучшие результаты восстановления среди всех методов независимо от сложности поведения ряда, за исключением, конечно, вырожденных случаев, в которых количество пропусков намного превышает количество существующих наблюдений.

Однако в случае групповых пропусков результаты применения данного метода оказались неожиданно сильно зависящими от структуры пропусков, особенно в случае большой протяженности группы. Причины подобной особенности этого алгоритма понятны: по сути, метод просто строит наилучшую аппроксимирующую поверхность для существующих наблюдений и не более того. Такие же особенности ряда, как периодичность, наличие линии тренда и т.п., никак не принимаются в расчет при восстановлении группового пропуска. Поэтому при применении данного метода возрастает степень участия исследователя в процессе восстановления пропусков, который должен не просто выполнить программу алгоритма, но и проконтролировать полученные результаты и, если необходимо, даже отменить их.

4. Метод восстановления пропущенного значения на основе использования Zet-алгоритма.

Данный алгоритм интересен для исследования тем, что при восстановлении учитывает закономерности ряда, может работать как с одномерными рядами данных, так и с таблицами данных, состоящих из множества взаимосвязанных рядов, указывает случаи, в которых имеющихся данных недостаточно для восстановления пропусков. И действительно, эксперименты показали целесообразность применения Zet-алгоритма.

Так, в случае одиночных пропусков данный метод имеет несомненное преимущество перед простыми методами восстановления пропусков (общего

среднего, среднего из ближайших) и несколько уступает методу сплайн-интерполяции. Однако в случае наличия в данных групповых пропусков качество восстановления пропущенных значений при помощи Zet-алгоритма лучше остальных рассматриваемых методов, причем, результаты применения метода стабильны, учитывают закономерности исследуемого ряда, достаточно слабо зависят от структуры пропусков и, иногда являются единственно достоверными по сравнению с результатами применения других методов. Особо следует отметить возможности этого метода при восстановлении групповых пропусков в таблицах данных. Ни один из остальных исследуемых методов не способен выполнить такую задачу с приемлемой точностью, в то время как Zet-алгоритм показывает хорошие результаты. Конечно, у данного метода есть ограничения. Например, между данными должна прослеживаться причинно-следственная (вероятностная) связь, а количество существующих наблюдений, по которым восстанавливаются пропуски, не должно быть малым. Если данные сильно зашумлены и искажены, обладают большой долей пропусков, то результат восстановления, естественно, будет некорректен: здесь как нельзя более ясно работает правило "мусор на входе – мусор на выходе". Однако даже в такой ситуации описываемый алгоритм будет искать закономерности в присутствующих данных, и осуществлять восстановление пропусков, в отличие от других методов, для которых в подобной ситуации пропуск заполняется совершенно фантастическими значениями.

Приведенные результаты отражают поставленную задачу исследований и, более того, позволяют предложить следующую, как представляется, новую и достаточно эффективную методику восстановления пропусков в массивах данных: сначала к одиночным пропускам применяется метод сплайн-интерполяции (хотя иногда достаточно применить метод заполнения средним из ближайших), затем к результирующему набору данных с восстановленными одиночными пропусками и незатронутыми групповыми пропусками применяется Zet-алгоритм.

2.2 Обзор методов структурирования данных в условиях неопределенности

Многообразие, а зачастую и противоречивость различных требований к проектируемой системе или оптимизируемому объекту, неполнота информации, неточность исходных данных для используемых моделей неизбежно приводят к тому, что реальную задачу оптимизации приходится решать в условиях неопределенности. В настоящей монографии осуществлена попытка систематизировать существующие (классические и новейшие) методы структурирования данных, т.е. методы адекватного отражения, прогнозирования и принятия конкретных оценок или значений параметров для управленческих решений, относящихся к слабо формализованным эволюционным дискретным процессам и системам, а также предложить новый инструментарий для структурирования данных в условиях неопределенности.

Проблема неточности данных. В фундаментальном научном издании отмечено, что даже для установившихся режимов той или иной системы соотношение между источниками ошибки в типичном случае выглядит следующим образом:

- 1) 82-84 % – из-за неточности исходных данных,
- 2) 14-15 % – из-за неточности математической модели,
- 3) 2-3 % – из-за неточности применяемого метода.

Ввиду такой большой доли погрешности исходных данных неизбежно возникает и погрешность в расчете целевой функции моделируемого процесса, что в реальной ситуации приводит к значительной зоне неопределенности при выборе оптимального режима работы системы. Отсюда появляется необходимость разработки методов, учитывающих неопределенность исходных данных при решении задач многоуровневого управления эволюционными процессами.

В классических работах, посвященных этой проблеме, предлагаются

различные методы принятия решений в условиях больших ошибок во входных данных. Эти методы можно разделить на две основные группы:

- подавление влияния неточной информации с дальнейшим использованием обычных детерминированных алгоритмов;
- переход при наличии неточной информации на специальные алгоритмы (стохастические, нечеткие, интервальные).

Для первой группы характерным является применение различных методов фильтрации и сглаживания исходной информации, усреднения и взвешивания данных. Применяются также методы восстановления отсутствующих данных, интерполирования и экстраполирования, робастные алгоритмы.

К настоящему времени в адрес методов, основанных на методах первой группы, высказано немало обоснованных критических замечаний. Предварительная фильтрация данных, отсеечение выбросов и сглаживание оставшихся данных приводят к неадекватности этих данных наблюдаемому процессу по следующим причинам:

- применение процедур сглаживания и отсеечения может быть обоснованным только при непосредственном учете специфики наблюдаемого процесса;
- алгоритмы, используемые для реализации методов первой группы, являются достаточно сложными;
- говоря об адекватности допущений, положенных в основу методов первой группы, можно говорить об их эвристическом характере, т.е. отсутствии достаточно строгой обоснованности.

При использовании стохастических моделей возникает ряд принципиальных трудностей, связанных со сложностью получения плотностей распределения вероятностей для параметров модели. Чаще всего эти трудности порождаются известной проблемой, называемой «проблема малых выборок». Действительно, в реальных ситуациях чаще всего удается обеспечить лишь несколько десятков наблюдений оцениваемого параметра, в

то время как необходимым является количество порядка тысячи и более.

Судя по растущему количеству публикаций, посвященных обсуждаемой проблеме, все большее число исследователей склоняется к тому, что в реальном математическом моделировании наиболее целесообразным подходом можно считать представление исходных данных в виде нечетких множеств или интервальных значений.

Проблема неопределенности цели. Можно считать несомненным тот факт, что в процессе моделирования сложной системы исследователь должен принимать во внимание не одну цель, а две или больше целей, которые в некотором смысле «равноправны», т.е. в каждой паре целей нет доминируемой и доминирующей. Описать их одним показателем (критерием) невозможно. Конструктору самолета, например, необходимо обеспечить не только безопасность пассажиров, но и минимальную стоимость перелета. Экономисту нужно построить такой план, чтобы «при минимуме затрат добиться максимума выпуска продукции» и т.п., причем, эти требования, как мы видим, часто противоречат друг другу.

Легко понять, что свести подобные многокритериальные задачи к точно поставленным задачам классического математического программирования нельзя в принципе. Этот вопрос выходит за рамки области деятельности исследователя в процессе построения математической модели и разработки метода нахождения наилучшего решения.

В условиях многокритериальности в выборе и принятии одного решения заложено противоречие, т. к. обязательно найдутся решения, которые по некоторым критериям являются лучше выбранного. Следовательно, указанных противоречий можно избежать лишь в случае, если говорить о нахождении множества «подходящих» решений. Этот тезис первым достаточно четко сформулировал итальянский экономист Вильфредо Парето еще в 1904 году в форме так называемого принципа Парето. Согласно Парето, возможные решения следует искать среди не улучшаемых альтернатив, т.е. альтернатив, улучшение которых по одним критериям

приводит к ухудшению по другим критериям. Принцип этот достаточно очевидный и очень важный с чисто прикладной точки зрения: он позволяет, во-первых, сжать множество альтернатив, во-вторых, он демонстрирует те потери, которые имеет оперирующая сторона по тем или иным показателям, стремясь улучшить какой-то определенный показатель.

К настоящему времени в рамках дискретной оптимизации можно говорить о самостоятельном направлении дискретного программирования для многокритериальных задач. Однако, пока не известно ни одного эффективного алгоритма решения какой-либо многокритериальной задачи с нечеткими данными.

В самом общем виде математическая постановка дискретной многокритериальной задачи состоит из описания условий, определяющих конечное или счетное множество допустимых решений $X = \{x\}$ и заданной на X векторной целевой функции (ВЦФ)

$$F(x) = (F_1(x), F_2(x), \dots, F_N(x)) \quad (2.1)$$

критерии, которой, могут принимать вид:

$$F_v(x) \rightarrow \text{extr}, \quad v = \overline{1, N}, \quad \text{extr} \in \{\min, \max\} \quad (2.2)$$

Если фиксированы все параметры ВЦФ (2.1) и система условий, определяющих МДР X , то принято говорить об индивидуальной задаче.

Под математическим решением индивидуальной задачи дискретной многокритериальной оптимизации следует понимать нахождение того или иного множества альтернатив (МА). Из найденного МА впоследствии с помощью методов многокритериального выбора осуществляется выбор и принятие решения.

Перечислим наиболее известные типы МА.

- множество допустимых решений X , которое рассматривается в качестве МА в случае, когда критерий выбора и принятия решения является очень сложным.

- паретовское множество \tilde{X} , состоящее из всех паретовских

оптимумов. Для данной индивидуальной задачи с ВЦФ (2.1) – (2.2), состоящей из максимизируем критериев: (в (2.2) $extr = \max$),

решение $\tilde{x} \in X$ называется паретовским оптимумом, если не существует такого элемента $x^* \in X$, который удовлетворяет неравенствам:

$$F_\nu(x^*) \geq F_\nu(\tilde{x}), \nu = 1, \dots, N,$$

среди которых хотя бы одно является строгим. Эти неравенства заменяем на обратные, если в (2.2) значение $extr = \min$.

– полное множество альтернатив X^0 , которое определяется как подмножество $X^0 \subseteq \tilde{X}$ минимальной мощности $|X^0|$ и такое, что:

$$F(X^0) = F(\tilde{X}),$$

$$F(X^*) = \{F(x) : x \in X^*\} \quad \forall X^* \subseteq X.$$

ПМА является обобщением понятия «оптимум», определенного для 1-критериальных, т.е. оптимизационных задач.

Для всякой индивидуальной задачи, представленные, выше МА образуют иерархически упорядоченную цепочку включений $X^0 \subseteq \tilde{X} \subseteq X$. При исследовании какой-либо дискретной многокритериальной задачи в качестве основной математической проблемы обычно рассматривается вопрос построения достаточно эффективного алгоритма нахождения требуемого МА этой задачи.

2.3 Структурирование данных при моделировании эволюционных дискретных процессов

2.3.1 Дискретные задачи с интервальными данными

Применение интервального анализа и различных минимаксных (гарантированных) подходов обладает целым рядом преимуществ:

- не требуется знание вероятностных характеристик неопределенных факторов, которые редко бывают точно известны на практике;
- при минимаксном подходе можно получить строгие оценки для самих искомых величин, а не для вероятностей или математических ожиданий, что имеет большое значение при наличии малого числа замеров параметров и одной или нескольких реализаций;
- статистические характеристики не могут гарантировать определенный исход одного конкретного опыта;
- во всех случаях даются гарантированные двусторонние аппроксимации искомых решений.

Однако процесс моделирования задач с интервальными данными требует осуществления алгебраических операций над областями, в результате чего на выходе используемых алгоритмов получаются области весьма сложной формы, что и порождает известные трудности решения интервальных задач. Главная проблема в решении реальных задач с интервальными данными – это проблема точности получаемого результата. В общем случае точность интервального результата полностью определяется следующими четырьмя факторами.

1. Неопределенность в задании исходных данных.
2. Округления при выполнении операций, изменяющих или порождающих интервальные объекты.
3. Приближенный характер используемого численного метода.
4. Степень учета зависимостей между участвующими в вычислении

интервальными объектами (переменными и константами).

Численное решение интервальных задач базируется на интервальной арифметике.

Приведем интервальную арифметику.

Пусть R – множество всех вещественных чисел. Под интервалом:

$$A = [a_1, a_2], \quad a_1 \leq a_2,$$

понимается замкнутый ограниченный отрезок на числовой оси.

Множество всех интервалов обозначим через IR . Элементы IR записываются прописными буквами. Если A – элемент:

$$IR (A \in IR),$$

то его левый и правый концы обозначим a_1 и a_2 , а также:

$$A = [a_1, a_2].$$

Символы \in, \cap, \subset и т.п. понимаются в обычном теоретико-множественном смысле, причем \subset обозначает не обязательно строгое включение, т.е. соотношение $A \subset B$ допускает равенство интервалов.

Два интервала:

$$A = [a_1, a_2], \quad B = [b_1, b_2],$$

равны тогда и только тогда, когда:

$$a_1 = b_1, \quad a_2 = b_2.$$

Отношение строгого порядка на множестве IR определяется следующим образом: $A < B$ тогда и только тогда, когда $a_2 < b_1$. Возможно также упорядочение по включению: A не превосходит B , если $A \subset B$. Пересечение $A \cap B$ интервалов A и B пусто, если $A < B$ или $B < A$, в противном случае:

$$A \cap B = [\max\{a_1, b_1\}, \min\{a_2, b_2\}], \text{ является снова интервалом.}$$

Для интервала $[a_1, a_2]$ симметричным, по определению, является интервал $[b_1, b_2]$, у которого:

$$-b_1 = a_2, \quad -b_2 = a_1.$$

Шириной $\omega(A)$ интервала A называется величина:

$$\omega(A) = a_2 - a_1.$$

Середина $m(A)$ есть полу сумма его концов:

$$m(A) = (a_1 + a_2)/2.$$

Абсолютная величина $|A|$ определяется как:

$$|A| = \max \{ |a_1|, |a_2| \}.$$

Нетрудно заметить, что $|A| \leq |B|$, когда $A \subset B$, причем $\omega(A) \leq \omega(B)$, если $A \cap B$ и $A \neq B$.

Расстояние $r(A, B)$ между элементами $A, B \in IR$ вводится равенством:

$$r(A, B) = \max \{ |a_1 - b_1|, |a_2 - b_2| \}.$$

Вырожденный интервал, т.е. интервал с совпадающими концами $a_1 = a_2 = a$ отождествляется с вещественным числом a . Таким образом, имеем включение $R \subset IR$.

Арифметические операции над интервальными числами определяются следующим образом. Пусть $* \in \{+, -, \cdot, / \}$, $A, B \in IR$, тогда:

$$A * B = \{ a * b \mid a \in A, b \in B \} \quad (2.3)$$

В случае деления $0 \notin B$.

Определение (2.3) эквивалентно соотношениям:

$$A + B = [a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2] \quad (2.4)$$

$$A - B = [a_1, a_2] - [b_1, b_2] = [a_1 - b_2, a_2 - b_1] \quad (2.5)$$

$$A \cdot B = [a_1, a_2] \cdot [b_1, b_2] = [\min \{ a_1 b_1, a_2 b_1, a_1 b_2, a_2 b_2 \}, \max \{ a_1 b_1, a_2 b_1, a_1 b_2, a_2 b_2 \}] \quad (2.6)$$

$$A / B = [a_1, a_2] / [b_1, b_2] = [a_1, a_2] \cdot [1/b_2, 1/b_1] \quad (2.7)$$

Операцию вычитания можно выразить через сложение и умножение, положив:

$$-B = (-1)B = [-1, -1]B \text{ и } A - B = A + (-B).$$

Если A и B – вырожденные интервалы, то равенства (2.4) – (2.7) совпадают с обычными арифметическими операциями над вещественными числами. Из определения (2.3) непосредственно видно, что интервальные сложения и умножения ассоциативны и коммутативны, т.е. для $A, B, C \in IR$

имеют место равенства:

$$A + (B + C) = (A + B) + C, \quad A + B = B + A,$$

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C, \quad A \cdot B = B \cdot A.$$

Роль нуля и единицы играют обычные 0 и 1, которые отождествляются с вырожденными интервалами $[0, 0]$ и $[1, 1]$, т.е.:

$$A + 0 = 0 + A = A, \quad A \cdot 1 = 1 \cdot A = A.$$

Равенства (2.3) – (2.7) показывают, что если один из операндов является невырожденным интервалом, то и результат арифметической операции также невырожденный интервал. Исключение составляет умножение на $0 = [0, 0]$. Отсюда следует, что для невырожденного интервала A не существует обратных по сложению и умножению элементов, т.к. если $A + B = 0$, $A \cdot C = 1$, то A, B, C должны быть вырожденными, т.е. вычитание не является обратным сложению, деление не обратное умножению. Значит, $A - A \neq \emptyset$, $A / A \neq 1$, когда $\omega(A) > 0$. Всегда $0 \in A - A$ и $1 \in A / A$.

Основная теорема интервальной арифметики отражает такое важное транзитивное свойство, как монотонность по включению. Это значит, что если $A \subset C$, $B \subset D$, то выполняются соотношения:

$$A + B \subset C + D, \quad A - B \subset C - D, \quad AB \subset CD, \quad A / B \subset C / D \quad (\text{если } 0 \notin D).$$

2.3.2 Нечеткие интервальные данные, возникающие в процессе математического моделирования

В процессе реального моделирования сложных систем значительная часть необходимой для их математического описания информации существует в виде нечетких представлений или пожеланий экспертов, параметры системы оказываются неопределенными (хотя и не случайными) и, в то же время, сильно влияющими на ход решения. Общепринятые количественные методы анализа по своей сути мало пригодны и неэффективны для систем такого рода. Неточно заданные параметры либо не

принимаются во внимание, либо с учетом определенных предположений и допущений заменяются средними оценками. Именно в этом смысле традиционные методы точного количественного анализа не имеют требуемого практического значения в реальных экономических, социальных и других системах. Кроме того, при моделировании процессов, связанных с участием человека, классические подходы не в состоянии отразить нечеткость человеческого мышления и поведения. Все указанное выше приводит к мысли о том, что для моделирования процессов управления больше подошли бы «нечеткие математические методы», нежели классические.

Согласно работе М. Блэка, неопределенность имеет место, когда универсальное множество состоит более, чем из одной точки. Если для этих элементов множества заданы соответствующие вероятности или другие вероятностные характеристики, то имеет место вероятностная неопределенность. Если известны только граничные элементы множества – интервальная неопределенность. При задании для каждого элемента множества соответствующей степени – нечеткость. Неопределенность можно классифицировать по степени неопределенности (полная определенность, вероятностная, лингвистическая, интервальная, полная неопределенность), по характеру неопределенности (параметрическая, структурная, ситуационная) и по использованию получаемой в ходе управления информации (устраняемая и неустраняемая).

Для преодоления трудностей представления неточных понятий, анализа и моделирования систем, в которых участвует человек, американским математиком Лотфи Заде в 1965г. была предложена теория нечетких (размытых) множеств. Подход на основе теории нечетких множеств является, по сути дела, альтернативой общепринятым количественным методам анализа систем. Он имеет три основные отличительные черты:

– вместо или в дополнение к числовым переменным используются нечеткие величины и так называемые «лингвистические» переменные;

– простые отношения между переменными описываются с помощью нечетких высказываний;

– сложные отношения описываются нечеткими алгоритмами.

Такой подход дает приближенные, но в тоже время эффективные способы описания поведения систем, настолько сложных и плохо определенных, что они не поддаются точному математическому анализу. До работ Л. Заде подобная качественная информация, по существу, просто терялась – было непонятно, как ее использовать в формальных схемах анализа альтернатив. Теоретические же основания данного подхода вполне точны и строги в математическом смысле и не являются сами по себе источником неопределенности. В каждом конкретном случае степень точности решения может быть согласована с требованиями задачи и точностью имеющихся данных. Подобная гибкость составляет одну из важных черт методов нечетких множеств и их дальнейшего развития – методов нечетких систем. Основные приложения данного подхода находятся в таких областях, как искусственный интеллект, лингвистика, поиск информации, процессы принятия решений, распознавание образов, медицинская диагностика, психология, право, экономика и другие отрасли человеческой деятельности.

Для реальных сложных систем характерно наличие одновременно разнородной информации:

– точечных замеров и значений параметров;

– допустимых интервалов измерения;

– лингвистических критериев и ограничений, полученных от специалистов-экспертов и т.д.

Реальные задачи содержат в себе нечеткие условия и некоторую нечеткость цели в связи с тем, что их постановку осуществляет человек. Учет фактора неопределенности при решении задач во многом изменяет методы принятия решения: меняется принцип представления исходных данных и параметров модели, становятся неоднозначными понятия решения задачи и

оптимальности решения. Чаще всего конкретное содержание задачи требует обеспечения заданного уровня нечеткости решения. Наличие неопределенности может быть учтено непосредственно в моделях соответствующего типа представлением недетерминированных параметров как случайных величин с известными вероятностными характеристиками, как нечетких с заданными функциями принадлежности или как интервальных величин с фиксированными интервалами изменения. Попытки применения какого-либо конкретного математического аппарата (интервального анализа, статистических методов, теории игр, детерминированных моделей и т.д.) для принятия решений в условиях неопределенности позволяет отразить в модели лишь отдельные виды данных и приводит к безвозвратной потере информации других типов.

Обычно на практике всегда имеется возможность наряду с точечной оценкой параметра (наиболее допустимым его значением) указать минимальное и максимальное значение (интервал), которые может принимать нечеткая величина. Кроме того, иногда удается построить и функцию, характеризующую допустимость каждого значения внутри заданного интервала на основе статистического материала или опроса группы экспертов. Теория нечетких множеств дает возможность проводить вычисления не с одним точечным значением, а с характеристической функцией и получать в результате вычислений нечеткую величину, для которой с помощью одного из методов процедуры денацификации может быть получена точечная (четкая) оценка. Применение теории нечетких множеств позволяет провести также согласование различных нечетких решений при наличии нечетких целей, ограничений, коэффициентов, начальных и граничных условий. Даже в тех случаях, когда неопределенность в процессе принятия решений может быть представлена вероятностной моделью, обычно удобнее оперировать с ней методами теории нечетких множеств без привлечения аппарата теории вероятностей.

2.3.3 Структурирование данных при моделировании эволюционных дискретных процессов

Пусть эволюционный процесс определяется векторным итерационным уравнением

$$\vec{X}_{t+1} = F(\vec{X}_t), \quad t = 1, 2, \dots \quad (2.8)$$

Здесь \vec{X}_t – это вектор из n компонент, где n может быть очень большим числом и обычно включает много переменных, о которых мы ничего не знаем. Функция F в (1) переводит систему из одного момента времени в следующий, вид ее тоже неизвестен. Исследователь наблюдает временной ряд скалярных величин $x_t, t = 1, 2, \dots, T$. Наблюдения генерируются в соответствии с некоторой функцией

$$x_t = h(X_t) \quad (2.9)$$

Будем называть функцию h «функцией наблюдателя». Временной ряд образует траекторию, которая является плотной на аттракторе [V.1].

Для получения сведений об исходной системе нужен некоторый способ, с помощью которого мы сможем возвращаться от наблюдаемой к исследуемой системе. Этот способ осуществляется путем построения фазовой траектории размерности ρ :

$$\Phi_\rho(\vec{X}) = \{(x_t, x_{t+1}, \dots, x_{t+\rho-1})\}, \quad t = 1, 2, \dots, T \quad (2.10)$$

Термин «фазовая траектория» обычно подразумевает, что соседние точки множества (2.9) для наглядности соединены отрезками прямой или кривой линии. Объективную информацию о характере поведения эволюционного процесса (2.10) исследователь может получить через наблюдения (2.9), опираясь на теорему Такенса: если система, которая порождает временной ряд, является n - размерной, и обеспечено выполнение неравенства $\rho \geq 2n + 1$, тогда в общем случае фазовые траектории воссоздадут динамику исследуемой системы. Существует диффеоморфизм между фазовыми траекториями и истинными данными, порождаемыми системой.

Этот результат позволяет делать выводы о поведении системы, опираясь на данные наблюдений, и, более того, получать информацию для прогнозирования этого поведения.

В отличие от наиболее изученных дифференцируемых динамических систем в настоящей работе рассматриваем эволюционные процессы, которым присуще дискретное изменение наблюдаемых показателей во времени, т.е. изменения, происходящие в определенные промежутки времени (скачки). В этом случае соответствующее фазовое пространство является дискретным, а упорядоченная во времени последовательность значений наблюдаемого процесса называется временным рядом. Если эволюционный процесс, а точнее, изменение во времени его состояний подчиняется некоторым вероятностным закономерностям, то его принято называть стохастическим процессом.

Термин «структурирование данных» подразумевает эффективное представление данных в некотором классе задач, подлежащих решению на ЭВМ. Структурирование данных обычно происходит с помощью фиксированных в используемом языке способов построения и представления, исходя из конечного числа базовых структур. Например, деревья удобны для представления часто встречающихся иерархических структур. К настоящему времени методологически завершенной является теория структурирования целей в системном анализе, а также структурирование компонент временных рядов с точки зрения их прогнозирования.

Применительно к эволюционным процессам структурирование данных направлено на то, чтобы исследователь имел четкое представление о моделях, отражающих динамику эволюционировано, т.е. поведения учитываемых данных, а также о методах, с помощью которых эти данные можно оценить.

Рассматривая стохастический процесс, значения наблюдаемого показателя представляем в виде числовой функции $X(t)$ (дискретного) времени t . В общем случае характеристики рассматриваемого

стохастического процесса в любой момент времени t являются случайными величинами с определенными распределениями вероятности.

Приведем сложившуюся к настоящему времени классификацию стохастического процесса по зависимости между значениями $X(t)$ в различные моменты времени t :

– стохастический процесс с независимыми значениями: при любых t_1 и t_2 , $t_1 \neq t_2$, случайные величины $X(t_1)$ и $X(t_2)$ независимы;

– стохастический процесс с независимыми приращениями: для любых непересекающихся интервалов $[t'_1, t'_2]$ и $[t''_1, t''_2]$, $t'_2 < t''_1$, случайные величины: $X(t'_2) - X(t'_1)$ и $X(t''_2) - X(t''_1)$ независимы;

– стохастический процесс с Марковским свойством, т.е. случайный процесс без последствия. Случайный процесс $X(t)$ называется Марковским процессом, если для любых двух моментов времени t_0 и t_1 , $t_0 < t_1$ условное распределение $X(t_1)$ при условии, что заданы все значения $X(t)$ при $t \leq t_0$, зависит только от $X(t_0)$;

– стационарные случайные процессы: вероятностные характеристики стохастического процесса неизменны во времени, в частности при любых t и s случайные величины:

$X(t)$ и $X(t+s)$ имеют одинаковое распределение;

- случайные величины $(X(t_1), X(t_2))$ и $(X(t_1+s), X(t_2+s))$ имеют одинаковые совместные распределения и т.д.

«Теории неточных вероятностей» позволили моделировать кроме случайностей и другие виды неопределенностей, включающие непротиворечивость, неточность и неполноту имеющей информации.

В этих математических моделях вероятности событий задаются неточным образом, например, интервально, т.е. путем определения верхней и нижней границ вероятностей событий. При этом важно отметить, что нечеткие меры в отличие от вероятностных мер не обладают свойством аддитивной в общем случае. Вместе с тем, в теориях нечетких вероятностей

постулируется гипотеза о том, что вероятностный закон нам неизвестен, но он имеет место быть и достаточно адекватно отражается одной из вероятностных мер. Последнее определяется с помощью некоторой функции множеств, которая дает точные нижние оценки $\underline{\mu}(A)$ вероятности событий $A \in \mathfrak{F}$, где $\mathfrak{F} = 2^X$ – алгебра всех подмножеств базового множества:

$$X = \{x_1, x_2, \dots, x_n\}.$$

Верхние вероятности получаются с помощью отношения двойственности $\underline{\mu}(A) = 1 - \underline{\mu}(\bar{A})$.

В рамках указанных выше теорий выбор и принятие решений осуществляется на базе различных моделей логического вывода без вероятностного обоснования, но при этом существенным образом используя «интуитивное» понимание нечетких множеств.

Возможные модели логического вывода обобщают классическую модель пропозициональной логики в том смысле, что высказывания задаются нечеткими множествами. При этом важно отметить, что в эти множества вкладывается вероятностная интерпретация – каждое нечеткое множество задает распределение возможностей на базовом измеримом пространстве, а значение меры возможности интерпретируется как верхняя оценка вероятности.

Ключевое различие между случайностью и нечеткостью состоит в том, как пересекаются взаимоисключающие события. В классической теории вероятности их пересечение – это пустое множество, а в теории нечетких множеств – нет. В теории вероятности сумма вероятностей двух взаимно исключающих друг друга событий всегда равна единице, в то время как в теории возможностей она может быть и больше единицы.

Таким образом, в теории возможностей используется функция принадлежности. Ее значения представляют собой не вероятность наступления события, а степень возможности реализации данного исхода. При этом не предполагается, что сумма степеней появления двух

взаимоисключающих исходов равна 1.

Задачи статистического анализа обычно разделяют на три обширных класса:

- статистическое оценивание;
- проверка статистических гипотез;
- построение статистических зависимостей.

В круг задач статистического оценивания входит задача вычисления оценки функции распределения. Оценивание распределения можно считать несущественной задачей, но только в случае неограниченных возможностей в получении статистического материала. Действительно, в таком случае с помощью широко известных методов можно очень точно оценивать параметры распределений и иметь небольшие ошибки при проверке гипотез. Если же необходимо оценить распределение, то это несложно выполняется одним из распространенных методов, например с помощью построения полигона частот или гистограммы.

На практике часто приходится работать в условиях ограниченных объемов выборок. При анализе статистического материала ограниченного объема задача оценивания функции распределения принимает проблематичный характер. Единственным методом ее решения в таких условиях, если есть в этом необходимость, является построение эмпирической функции распределения, имеющей вид ступенчатого графика. Эта задача приобретает главенствующее значение для выборок очень малого объема, содержащих менее десяти наблюдений.

Выборку следует считать малой, если при ее обработке методами, основанными на группировке наблюдений, нельзя достичь заданных точности и достоверности. Из этого определения следует, что выборку можно считать большой, если при ее обработке имеется возможность перейти к группировке наблюдений без ощутимой потери информации. При этом должны достигаться заданные точность и достоверность. Границу, разделяющую большие и малые выборки (достаточная выборка), нельзя,

конечно, понимать как точку в ряду натуральных чисел. Достаточные выборки, естественно, в силу случайности выбора, образуют некоторое конечное множество. Но самым существенным, что следует из определения малой выборки, является необходимость при обработке малой выборки индивидуального подхода к каждой отдельной реализации.

Проблема малых выборок приобретает важное значение в случае, когда возникает вопрос: имеет ли рассматриваемый экономический временной ряд тяжелые хвосты и если да, то как соотносится вес «головой» этого распределения и вес его «хвоста» («хвостов»)?

Определение термина «тяжелый хвост» относится к случаю, когда объем выборки (длина временного ряда) является вполне достаточным для достоверного определения имеющегося распределения. Случайная величина W имеет распределение с тяжелым хвостом, если вероятность:

$$P(W > x) \sim cx^{-\alpha}, \quad x \rightarrow \infty,$$

где параметр $\alpha \in (0, 2)$ и c – положительная константа. Указанное определение означает, что хвост распределения затухает по гиперболическому закону, т.е. медленно, в силу чего дисперсия этого распределения стремится к бесконечности. В противоположность этому распределение с легким хвостом, например, нормальное имеет экспоненциально спадающий хвост. В этом случае уже за пределами $M \pm 3\sigma$ (т.е. за пределами радиуса, равного трем стандартным отклонениям окрестности математического ожидания M) вес хвоста составляет, как правило, несколько процентов от веса головы.

Таким образом, возникает проблема установления наличия или отсутствия тяжелого хвоста для реальных временных рядов, длина которых составляет несколько десятков или порядка 100. В этом случае просто не представляется возможность сколь-нибудь достоверно установить характер функции распределения для вероятности $P(W > x)$. Для установления факта наличия тяжелого хвоста существует инструментарий, идея которого связана с понятием лепт эксцесса.

Определение термина «лептоэксцесс» базируется на том факте, что для нормального распределения, обладающего легким хвостом, значение коэффициента эксцесса равно одной и той же константе, равной числу 3 при любых значениях параметров этого распределения. Явление лепт эксцесса подразумевает, что вес хвоста растет с ростом значения коэффициента эксцесса. При этом с учетом известного «правила трех сигм» в рассматриваемом эмпирическом распределении к области хвоста (хвостов) относятся такие точки этого распределения, которые находятся за пределами области «головы» $[M - 3\sigma, M + 3\sigma]$. Эмпирический вывод, полученный на основании анализа многочисленных реальных временных рядов сводится к следующему утверждению: рассматриваемое распределение обладает тяжелым хвостом, если вклад точек, находящихся за пределами «головы», вносит в значение коэффициента эксцесса вес, больший числа 3.

Структурирование данных для эволюционных процессов и систем чаще всего включает в себя этап прогнозирования, которое, вообще говоря, может относиться к одному из измерений для более качественного предсказания другого измерения. Последнее в рамках эконометрики обычно реализуется двумя базовыми инструментами: корреляционный анализ, позволяющий оценить степень взаимосвязи между двумя факторами (если такая взаимосвязь вообще существует) и регрессионный анализ, показывающий, как можно предсказать, или управлять одной из двух переменных с помощью другой.

Временные ряды отличаются от данных об одном временном срезе в том отношении, что в случае временных рядов сама последовательность наблюдений несет в себе важную информацию. В частности, чтобы охарактеризовать какую-либо совокупность данных в целом, нам уже недостаточно знать лишь типичное значение этих данных (например, среднее значение) или даже изменчивость этой совокупности данных (описываемую, например, стандартным отклонением). В этом случае желательно знать, что скорее всего произойдет дальше. Подобный прогноз должен по возможности

точнее экстраполировать ближайшее поведение системы с точки зрения моделей поведения этой системы в прошлом.

Временной ряд не является случайной выборкой из некоторой генеральной совокупности (исключением является процесс чистого случайного шума). Гораздо вероятнее, например, что завтрашняя цена окажется ближе к сегодняшней, чем к прошлогодней цене; последовательные наблюдения не являются независимыми друг от друга.

Современное прогнозирование предполагает научно обоснованное суждение о возможных состояниях некоторой системы в будущем, об альтернативных путях и сроках его осуществления, оно предпочитает получение количественных оценок этих состояний при помощи математических и инструментальных средств реализации. Методологии прогнозов присущи общие черты. Все они в той или иной мере используют экстраполяцию прошлых тенденций в отношении как общенациональных, так и частичных показателей производства, народонаселения, технического прогресс и т.д.

Методы, составляющие инструментарий классического прогнозирования, согласно, могут быть разделены на 3 большие группы: статистические, причинно-следственные и комбинированные. Использование этого инструментария предполагает, что исходные данные обычно представляют собой результаты выборочных наблюдений. При этом не каждую совокупность зарегистрированных реальных данных нужно считать подходящим временным рядом, на основании, которого можно составлять прогноз. Из такой совокупности должны быть исключены выбросы, т.е. такие результаты наблюдений, которые не являются характерными или типичными для рассматриваемого эволюционного процесса. К выбросам могут быть отнесены такие уровни ряда, для которых отклонение от среднего превосходит 4σ , σ – стандартное отклонение. Прогнозирование временного ряда осуществляется в предположении, что для него выполняются условия стационарности и независимости уровней, составляющих этот ряд. Уместно

отметить, что указанное свойство независимости обуславливает подчинение уровней ВР нормальному закону.

К основным вопросам анализа временных рядов относится выявление сезонности, а также тренда, имеющего, как правило, полиномиальное представление. Выявление трендовой составляющей базируется, как правило, на различных подходах к построению скользящей средней.

Процесс построения прогнозной модели обычно подразумевает использование тех или иных методов уточнения прогноза (в современном представлении – методов «обучения» прогнозной модели). К их числу относится метод адаптивного сглаживания, метод Бокса-Дженкинса или их обобщение, получившее название «байесовские прогнозы».

Классика прогнозирования явно или «по умолчанию» подразумевает, что для предлагаемой прогнозной модели осуществляется ее верификация и валидация. Термин «верификация» подразумевает содержательный и логический анализ гипотезы, на которой базируется предлагаемая прогнозная модель, а также эмпирическое подтверждение или отрицание правильности этой гипотезы. Валидация – оценка точности прогнозирования.

Представляется естественным вычислить ошибки прогноза в виде разницы между фактическими значениями ВР и прогнозными значениями рассматриваемого отрезка ВР. Однако, эта разница, строго говоря, представляет собой оценку точности аппроксимации. Для последней существуют вполне надежные критерии оценки точности – дисперсия, критерии Фишера и Стьюдента и др. Применение этих критериев для оценки точности прогноза вызывает серьезное возражение, т.к. не дает гарантии, что в будущем эти оценки будут правомерны. Существует лишь предположение, что в будущем эволюционный процесс сохранит свои характеристики. Это условие означает стабильность или устойчивость динамики временного ряда, которая является необходимой для получения прогноза вообще.

Наряду с верификацией прогнозной модели существует проблема верификации исходных данных. Эта проблема возникает в связи с тем, что

классические методы прогнозирования базируются на предположении, что выполняется требование нормальности распределения ошибок в исходной информации.

Развитие экономико-математического моделирования, анализа и прогнозирования в современных условиях связано с последовательным ростом уровня их формализации. Основу этого развития заложил прогресс в области прикладной математики, математической статистики, методов оптимизации, теории приближений, в эконометрике, прогностике и пр.

Для процессов управления в промышленности, экономике, финансовом бизнесе характерна определенная стабильность, инертность, сложившаяся структура взаимосвязей. При выборе подходящей аппроксимирующей функции нужно знать класс экономического поведения показателя. Это поведение может представлять собой сезонные (периодические) процессы, процессы накопления или распада экспоненциального типа, достаточно гладкое трендовое поведение и др. Менеджер, как правило, нуждается в уточнении этого временного класса.

К настоящему времени в экономико-математических публикациях уже сложилось классическое представление о структурировании временного ряда при построении для него прогнозной модели. Эту структуру составляют следующие 4 компоненты:

- 1) долгосрочная тенденция (тренд);
- 2) точное повторение сезонных моделей поведения;
- 3) циклы, т.е. непериодические («блуждающие») циклические подъемы и падения;
- 4) остаточная компонента в виде случайного, нерегулярного шума.

Прогноз получается путем наложения прогнозных моделей, соответствующих каждой из перечисленных четырех компонент.

Трендом временного ряда:

$$Z = \langle z_t \rangle, z_t = z(t), t = 1, 2, \dots, n,$$

называют плавно изменяющуюся нециклическую (непериодическую)

компоненту, описывающую чистое влияние долговременных факторов. В простейших случаях тренд представляется в виде следующих функций от t :

- линейная модель вида $z(t) = a_0 + a_1 t$;
- степенная модель $z(t) = a_0 + a_1 t^{a_2}$, $a_2 < 1$;
- полиномиальная модель $z(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n$;
- показательная модель $a_0 + a_1 b^t$;
- логарифмическая модель вида $z(t) = a_0 + a_1 \ln t$;
- экспоненциальная модель $z(t) = a_0 + a_1 e^{-t}$;
- логистическая модель вида $z(t) = a / (1 + b e^{-ct})$;
- модель Гомперца $\log(z(t)) = a - br^t$, где $0 < r < 1$;
- гиперболическая $z(t) = a_0 + a_2 / t$.

Для последней существует несколько разновидностей гиперболы, обладающих разными свойствами:

- равносторонняя гипербола, асимптотическая к линии, параллельной оси t :

$$x = \frac{1}{a_0 + a_1 t};$$

- равносторонняя гипербола, асимптотическая к линиям, параллельным обеим осям:

$$\frac{1}{Z(t)} = a_0 + a_1 \left(\frac{1}{t} \right);$$

- логистическая функция:

$$Z(t) = \frac{y_{\max} - y_{\min}}{1 + e^{a_0 + a_1 t}} + y_{\min},$$

где y_{\max} , y_{\min} – максимальное и минимальное значения уровней ряда динамики (положительные или отрицательные); a_0 , a_1 – неизвестные параметры уравнения тренда.

Рассмотрим тренды и циклы в техническом анализе. Термин «технический анализ» означает исследование динамики любых показателей

(рынка) при помощи графических методов с целью прогнозирования будущего направления их эволюции. С точки зрения менеджера технический анализ ВР базируется на трех краеугольных камнях:

- временные ряды учитывают всё;
- временные ряды подчиняются тенденциям;
- временные ряды содержат закономерности.

По видимому, правомерность первого из этих утверждений можно оценивать в терминах нечетких множеств, у которых не совсем близки к 1 значения функции принадлежности для лингвистических переменных, составляющих понятие «всё». Второе из этих утверждений закладывает основу для трендового анализа в том смысле, что наметившийся тренд с большой вероятностью продлится, нежели изменит направление. Вместе с тем, рыночные отношения объективно обуславливают непредсказуемость появления и окончания трендов. Третье из приведенных утверждений означает, что правила рынка, действовавшие в прошлом, будут действовать в настоящем и будущем.

Важнейшим инструментом технического анализа являются тренды. В этой области тренд, он же тенденция определяется как однонаправленное движение анализируемого показателя, действующее в течение определенного непрерывного периода времени. Главная причина трендовости или трендоустойчивости процессов заключается в объективно существующей инерционности экономических и природных явлений.

В техническом анализе направление тренда определяют при помощи различных процедур усреднения наблюдаемого временного ряда. Для рассматриваемого временного ряда:

$$Z : z_i, i = 1, 2, \dots, n,$$

при выбранной длине m интервала сглаживания элемент \bar{z}_t , $t > m/2$ простого скользящего среднего (Moving Average - MA) вычисляется по формуле:

$$\bar{z}_t = \frac{1}{m} \cdot \sum_{i=t-p}^{t+p} z_i, \quad p = \left\lceil \frac{m}{2} \right\rceil, \quad t = \overline{p+1, n-p} \quad (2.11)$$

В техническом анализе уже сложилась классификация трендов, относящаяся к рыночным показателям. Тренд бывает не только растущим («бычьим»), убывающим («медвежьим») или нейтральным, но и:

- сильным, слабым или нормальным;
- молодым, зрелым или старым;
- долгосрочным, среднесрочным или краткосрочным.

Перечислим основные принципиальные особенности прогнозного процесса. Базовой компонентой при этом является процесс авторегрессии скользящего среднего, который состоит из линейной функции от предыдущих наблюдений Y_{t-1} , Y_{t-2} . Прогнозируемые приращения вычисляются по формуле

$$\Delta t = Y_t - Y_{t-1} = \delta + \varphi(Y_{t-1} - Y_{t-2}) + \varepsilon_t + \theta\varepsilon_{t-1} \quad (2.12)$$

где δ - const – коэффициент авторегрессии, ε_t – случайный шум, θ – некоторая доля предыдущего случайного шума ε_{t-1} . Пользователю прогнозного процесса полезно учитывать следующие его особенности:

1. Компонента авторегрессии в комбинации со скользящей средней обладает памятью о своем прошлом. Однако эта память ограничена двумя предшествующими наблюдениями Y_{t-1} и Y_{t-2} .

2. Компонента чистого интегрированного процесса состоит из компоненты δ , определяющей состояние дрейфа, плюс случайный шум ε_t , подчиняющийся нормальному распределению.

Таким образом, прогнозная модель является полезной в тех ситуациях, в которых нет тенденции возврата к долгосрочному среднему значению (например, индекс потребительских цен) или, наоборот, в которых ряд стремится оставаться вблизи долгосрочного среднего значения (например, уровень безработицы или процентные ставки). При этом в первом из указанных случаев (не стационарности) долгосрочные прогнозы проявляют

тенденции к бесконечному нарастанию или наоборот, монотонному снижению.

К настоящему времени наибольшее распространение и применение в реальных расчетах получили методы третьей группы из перечисленных выше. Чаще всего в реальном экономико-математическом моделировании основное внимание уделяется анализу трендов и сезонности. При этом построение прогнозной модели рассматриваемого временного ряда реализуется через преобразование его в базовую модель временного ряда. Точно так же каждый элемент, т.е. каждое число в этой базовой модели ВР получается путем перемножения четырех компонент:

«Данные = тренд × сезонность × цикличность × регулярность».

Содержательное определение этих четырех компонент в случае экономического прогнозирования состоит в следующем.

Эти четыре базовые компоненты временного ряда (тренд, сезонность, циклическая и случайная компоненты) можно оценивать различными способами. Ниже приведено краткое изложение метода, который базируется на скользящей средней. В основе этого метода лежит деление элементов ряда на значения ординат скользящей средней следующим образом:

– скользящая средняя используется для устранения сезонных эффектов усреднения по всему году, а также для уменьшения нерегулярной компоненты и получения комбинации тренда и циклической компоненты;

– деление элементов исходного ряда на значения соответствующих ординат сглаженного ряда скользящей средней дает нам «отношение к скользящей средней», которое представляет нам как сезонные, так и нерегулярные значения. Выполняя группирование по сезонным периодам, например, по времени года, а затем усреднение в полученных группах, находим сезонный индекс для каждого времени года. Выполняя деление каждого значения ряда на соответствующий сезонный индекс для соответствующего времени года, находим значения с сезонной поправкой;

– регрессия ряда с сезонной поправкой по времени служит для

оценивания долгосрочного тренда в виде прямой линии как функции от времени, т.е. эта переменная времени может состоять из чисел 1, 2, 3,... . Этот тренд (тенденция) не отражает сезонных колебаний и дает возможность получить прогноз с сезонной поправкой;

– прогнозирование можно выполнять с помощью сезонности тренда. Получая из уравнения регрессии прогнозируемые значения (тренд) для будущих периодов времени и затем, умножая их на соответствующий сезонный индекс, мы получаем прогнозы, которые отражают как долгосрочную тенденцию, так и сезонное поведение.

Анализ публикаций, посвященных методам и моделям прогнозирования, позволяет утверждать о существовании двух параллельных направлений в этой теории. Упомянутые объекты первого из этих направлений имеют социально-экономическое содержание. Объектами второго направления являются сложные системы техногенного происхождения из различных областей жизнедеятельности.

2.3.4 Структурирование данных на базе интеллектуальных инструментальных средств

Структурирование данных в основном заключается в реализации процедуры исключения из дальнейшего рассмотрения малозначимых составляющих. Для реального осуществления процесса структурирования могут быть использованы различные методы экспертной оценки – ранжирование, нормирование, упорядочения разного рода.

В течение последнего десятилетия пришло осознание того, что реальным эволюционным социально-экономическим процессам и системам присущи высокая степень неопределенности и чрезмерная сложность динамики их эволюционировано, что усложняет применение классических моделей. Недостаточность классических подходов особенно отчетливо проявляется при попытках структурирования зашумленных цифровых

данных. Наряду с проблемой зашумленности существуют еще две проблемы: проблема выбора из неограниченно расширяющегося множества допустимых решений, мощность которого чаще всего растет экспоненциально с ростом размерности задач; и проблема адекватного описания таких явлений и понятий, которые имеют многозначный и нечеткий характер, в особенности, когда достигнутые значения представляют в виде правил, основанных на нечетких множествах и лингвистических переменных.

Факт существования вышеупомянутых проблем объясняет основные причины, которые вызывают рост интереса у исследователей-теоретиков и менеджеров-практиков к новым интеллектуальным технологиям поддержки принятия управленческих решений. Эти интеллектуальные технологии реализуются на базе таких методов, как нейронные сети, генетические алгоритмы, клеточные автоматы и нечеткие системы.

Термин «нечеткие системы» подразумевает соединение инструментария нечетких множеств с реализацией новых информационных технологий. К настоящему времени теория нечетких систем – это единственная теория, которая математически оперирует со смысловым содержанием слов человека. Иными словами, это математический метод, созданный для того, чтобы представлять смысловые нечеткости слов человека и математически обрабатывать субъективные данные.

Наряду с определенной самодостаточностью нечеткие системы способны усилить возможности нейронных сетей и генетических алгоритмов. Использование в этом случае нечетких систем не только ослабляет требование к точности данных в процессе построения моделей, но и позволяет описать сложные системы с помощью переменных, значения которых определяются на интуитивном уровне. Возникающая в результате этого парадигма моделирования, управления, выработки решений и т.п. ведет к формированию лингвистических аргументов логических функций. Такие функции, описывающие реальные объекты, могут уточняться в процессе обучения по имеющимся данным. Более того, появляется возможность

формирования нечетких правил вывода в процессе обучения.

В конечном счете, приходим к новой парадигме искусственного интеллекта, которая подразумевает способность извлекать новые знания из имеющихся данных, накапливать эти знания в процессе обучения, генерировать новые правила вывода и, как следствие, обобщать информацию.

Инструментарий искусственных нейронных сетей. Нейронные сети были созданы в результате наблюдения за естественными процессами, происходящими в нервной системе живых существ, и попыток воспроизведения этих процессов. Термин нейрон, обозначающий основной исполнительный элемент искусственных нейронных сетей, был непосредственно заимствован из теории природных нервных систем.

Схема одно нейронной сети представлена на рисунке 2, где нейрон n – го слоя выполняет функцию адаптивного сумматора с регулируемыми уровнями входных сигналов:

$$x_r(n), r = \overline{1, p},$$

который осуществляет дополнительную линейную или нелинейную обработку вычисленной суммы с целью получения результата. Нейрон получает входные сигналы либо от сенсоров (которыми могут являться нейроны предшествующего слоя сети) либо в форме центростремительных сигналов с выходов других формальных ячеек.

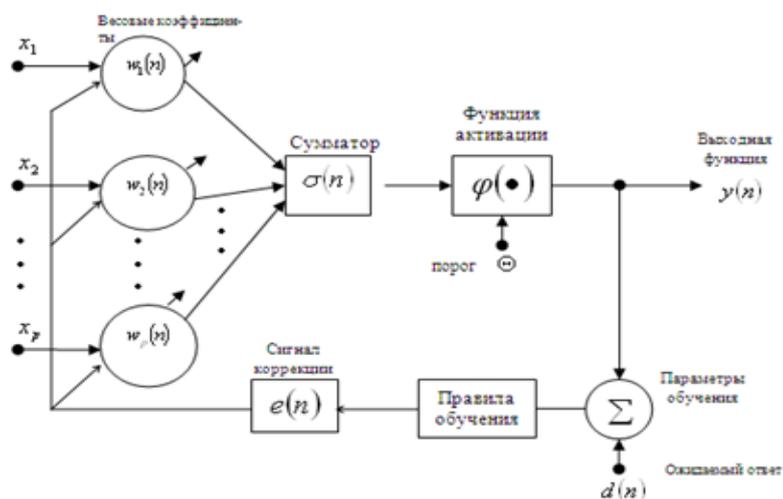


Рисунок 2. – Однослойная нейронная сеть

Алгоритм, используемый для обучения данного нейрона, получил название «Алгоритм наименьших квадратов» (Least-Mean-Square Algorithm, или LMS).

Искусственная нейронная сеть (ИНС), показанная на рисунке 2, имеет один недостаток – пороговое значение устанавливается жестко и корректируется только вручную. Этого можно избежать, если добавить к входу нейрона один синапс с постоянно установленным на нем потенциалом -1, а вес, на который умножается этот потенциал, подбирать так же, как и остальные веса – в процессе обучения.

Общая проблема искусственных нейронных сетей заключается в построении искусственной системы с заданным функциональным поведением; в контексте нейронных сетей она понимается, как задача синтеза требуемой искусственной сети. Она может включать в себя следующие подзадачи:

- 1) выбор существенных для решаемой задачи признаков и формирование признаковых пространств;
- 2) выбор или разработка архитектуры нейронной сети, адекватной решаемой задаче;
- 3) получение обучающей выборки из наиболее представительных, по мнению эксперта, векторов признаковых пространств;
- 4) обучение нейронной сети на обучающей выборке.

Отметим, что первые три подзадачи во многом требуют экспертного опыта работы с нейронными сетями, и здесь нет исчерпывающих формальных рекомендаций.

Под процессом обучения персептрона понимается алгоритмическая корректировка весовых коэффициентов синаптических связей каждого участвующего в процессе обучения нейрона, направленная на достижение минимальной ошибки в определении параметров выходного вектора для каждого из входных «образцов».

На этапе обучения на вход сети последовательно подаются входные

сигналы из заранее подготовленного для тренировки сети набора. Каждому из входных сигналов (данным) соответствуют заранее известные параметры выходного вектора, определение которых для произвольного набора данных, в том числе не использованных в процессе обучения, является целью задачи. Такими параметрами могут быть, например, логические утверждения принадлежности входного вектора тому или иному классу решений или его соответствия одному из тестовых образов, коэффициенты разложения входной функции относительно некоторого базиса и т.д.

Существует очень большой класс задач, когда желаемые значения выходов вообще неизвестны, а нейронной сети предлагается найти скрытые закономерности в массиве данных. Например, от нейронной сети может потребоваться нахождение компактного представления большого массива избыточных данных, т.е. произвести его оптимальное кодирование. Такое обучение называют обучением без учителя.

Обучение без учителя чаще всего используют в задачах предобработки большого количества данных. В отсутствие желаемых значений выходов, «учителем» сети становятся сами данные, а именно, имеющиеся в них зависимости. Значительная избыточность входных данных позволяет нейронной сети выделять из них закономерности и находить их более компактное представление. Практически можно сказать, что самообучающиеся сети «кодируют» входную информацию наиболее компактным, при заданных архитектурой ограничениях, кодом.

Подход к обучению ИНС учитывает следующие обстоятельства:

а) модифицируются только компоненты матрицы весов, отвечающие ненулевым значениям входов;

б) знак приращения веса соответствует знаку ошибки, т.е. положительная ошибка ($\delta > 0$, значение выхода меньше требуемого) приводит к усилению связи;

в) обучение каждого нейрона происходит независимо от обучения остальных нейронов, что соответствует важному с биологической точки

зрения принципу локальности обучения.

Инструментарий клеточных автоматов. Клеточным автоматом называют сеть из элементов, меняющих свое состояние в дискретные моменты времени в зависимости от состояния самого элемента и его ближайших соседей в предшествующий момент времени.

В более общем представлении клеточный автомат – это определенная динамическая система, состоящая из множества A идентичных, имеющих предел машин или ячеек, которые повторно меняют «цвет» или состояние, следуя заранее определенным правилам, и эти правила одинаково действуют по отношению ко всем элементам множества A в дискретном временном отрезке.

В общем случае ячейки в клеточном автомате могут принимать один из k различных цветов, очень часто обращаются к двум цветам (белому и черному).

Различные клеточные автоматы могут демонстрировать весьма разнообразное поведение, которое может быть адаптировано для целей обработки информации за счет выбора:

- закона изменения состояния элемента, т.е. другой функции перехода;
- конкретного определения понятия «ближайшие соседи».

В публикациях, посвященных нейронным сетям, отмечено, что существует определенное сходство работы алгоритмов для нейронных сетей и для клеточных автоматов. Например, рассматриваемая без обучения нейронная сеть Хопфилда вполне может рассматриваться, как клеточный автомат, элементами которого являются формальные нейроны. В качестве закона изменения состояния нейроавтомата используется пороговое преобразование взвешенной суммы входов нейронов, а ближайшими соседями каждого элемента являются все прочие элементы автомата.

В теории клеточных автоматов имеется классификация, согласно которой все автоматы делятся на четыре класса, в зависимости от типа динамики изменяющихся состояний. Автоматы первого класса по истечении

конечного времени достигают однородного состояния, в котором значения всех элементов одинаковы и не меняются со временем. Ко второму классу автоматов относятся системы, приводящие к локализованным структурам стационарных или периодических во времени состояний элементов. Третий класс составляют «блуждающие» автоматы, которые с течением времени посещают произвольным (непериодическим) образом все возможные состояния элементов, не задерживаясь ни в одном из них. И, наконец, четвертый класс составляют «странные» автоматы, характер динамики которых зависит от особенностей начального состояния элементов. Некоторые начальные состояния приводят к однородному вырождению автомата, другие – к возникновению циклической последовательности состояний, третьи – к непрерывно меняющимся (как «по системе», так и без видимой системы) картинам активности элементов.

Инструментарий генетических алгоритмов. Генетические (эволюционные) алгоритмы возникли в результате наблюдения и попыток копирования естественных процессов, происходящих в мире живых организмов, в частности, эволюции и связанной с ней селекции (естественного отбора) популяций живых существ. Иными словами, генетические алгоритмы – это процедуры поиска, основанные на механизмах естественного отбора и наследования.

Механизм селекции заключается в выборе хромосом с наивысшей оценкой (т.е. наиболее приспособленных), которые репродуцируют чаще, чем хромосомы с более низкой оценкой (хуже приспособленные). Репродукция означает создание новых хромосом в результате рекомбинирования генов родительских хромосом. Рекомбинация – это процесс, в результате которого возникают новые комбинации генов. Для этого используются две операции: скрещивание, позволяющее создать две совершенно новые хромосомы потомков путем комбинирования генетического материала пары родителей, а также мутация, которая может вызывать изменения в отдельных хромосомах.

В генетических алгоритмах применяется ряд терминов, заимствованных из генетики, прежде всего, хромосомы, а также популяция, особь, аллель, генотип, фенотип. От традиционных методов оптимизации генетические алгоритмы отличаются следующими базовыми элементами:

- 1) обрабатывают не значения параметров самой задачи, а их закодированную форму;
- 2) осуществляют поиск решения, исходя не из точки, а из некоторой популяции;
- 3) используют только целевую функцию, а не ее иную дополнительную информацию;
- 4) применяют вероятностные, а не детерминированные правила выбора.

Перечисленные четыре свойства, которые можно сформулировать как кодирование параметров, а также использование минимума информации о задаче и рандомизация операций приводят в результате к устойчивости генетических алгоритмов и к их превосходству над другими широко применяемыми технологиями.

С помощью генетических алгоритмов эффективно решается проблема выбора из неограниченно расширяющегося множества допустимых решений. В работе этих алгоритмов приближение или выбор наиболее целесообразного решения достигается за счет генерации серии последовательных приближений аргументов с лучшими значениями функций приспособленности, генерируемой в результате мутации и скрещивания «хромосом». Можно утверждать, что применение генетических алгоритмов является целесообразным для сжатия необозримых объемов информации, подаваемой на вход используемой информационно-вычислительной системы.

Поток огромного количества информации, справляться с которой становится все труднее и труднее, является одной из отрицательных черт современного информационного общества. Причем, во многих случаях эта информация является явно избыточной, что лишь затрудняет ее обработку и

восприятие. Стремлением избавиться от этого отчасти объясняется наметившийся рост интереса к сжатию информации с целью отбрасывания несущественных данных и выделения только наиболее значимых.

Экспертные системы. Эксперт – это человек, который является профессионалом достаточно высокого уровня в каком-то вопросе, чьи оценки и суждения по поводу объекта экспертизы учитываются при принятии решений.

Под экспертизой обычно понимают проведение группой компетентных специалистов измерения некоторых характеристик для подготовки принятия решений. Типичными проблемами, требующими проведения экспертизы, являются, например: определение целей, стоящих перед объектом управления; прогнозирование; разработка сценариев; генерирование альтернативных вариантов решений; определение рейтингов и другие.

Выделяют следующие основные этапы экспертизы:

- формулировка цели экспертизы;
- построение объектов оценивания или перечисление их характеристик;
- формирование экспертной группы;
- определение способа экспертного оценивания и способа выражения экспертами своих оценок;
- проведение экспертизы;
- обработка и анализ результатов экспертизы;
- повторные туры экспертизы, если есть необходимость уточнить или сблизить мнения экспертов;
- формирование вариантов рекомендаций.

Экспертиза, т.е. измерение или сравнение объектов по определению связана с каким-либо оцениванием объектов. Оценки бывают разных видов. Количественные оценки представляются с помощью абсолютной шкалы. Такие шкалы используются, когда результат измерений определяется однозначно (например, число студентов в аудитории).

В отличие от количественных оценок, соответствующих, как правило, объективным измерениям объективных показателей, балльные оценки обычно характеризуют субъективные мнения. Значения балльной шкалы представляют собой ограниченный ряд равноудаленных друг от друга чисел.

Балльные оценки бывают двух видов. Оценки первого вида производятся по объективному критерию, по общепринятому эталону, в соответствии с градациями этого эталона. Как правило, таковы оценки в спортивном судействе или правила присвоения рабочих разрядов. Эти оценки – оценки по балльной шкале.

Балльная оценка второго вида – это оценка, производимая, когда не только нет общепринятых эталонов, но и сомнительно даже наличие некоего единственного объективного критерия, субъективными отражениями которого являются оценки. В таком случае говорят о порядковой (или ранговой) шкале. Оценки, произведенные по ранговой шкале, имеет смысл сравнивать только по отношению «больше - меньше».

Следующий вид оценки – ранжирование. Под ранжированием понимают упорядочение объектов в соответствии с убыванием их предпочтительности (допускается указание на равноценность некоторых объектов). Примером ранжирования является определение призеров какого-либо конкурса. Отметим, что ранжирование представимо как оценка в ранговой шкале: рангом объекта (т.е. значением оценки) можно считать номер места, которое этот объект занимает в ранжировании при обратной нумерации мест. Этот способ оценки состоит в указании предпочтительного объекта в каждой паре объектов (иногда допускается объявление обоих объектов равноценными и несравнимыми).

Метод по парного сравнения применяется потому, что считается более легким качественное сравнение двух объектов, чем отражение предпочтения в балльной (или ранговой) шкале. Иногда при по парных сравнениях эксперту предлагается баллами оценить интенсивность своего предпочтения для каждой пары объектов. Для упорядочения объектов на основании

качественного критерия иногда удобен метод средней точки: выбирается лучший и худший объекты; потом объект, который может быть расположен посередине между ними; потом объекты, которые могут быть расположены посередине между худшими и ранее найденным средним, а также посередине между лучшим и средним, и т.д.

Центральная парадигма интеллектуальных технологий сегодня – это обработка знаний. Системы, ядром которых является база знаний или модель предметной области, описанная на языке сверхвысокого уровня, приближенном к естественному, называют интеллектуальными.

Чаще всего интеллектуальные системы применяются для решения сложных задач, где основная сложность решения связана с использованием слабоформализованных знаний специалистов-практиков и где логическая (или смысловая), обработка информации превалирует над вычислительной. Например, понимание естественного языка, поддержка принятия решения в сложных ситуациях, постановка диагноза и рекомендации по методам лечения, анализ визуальной информации, управление диспетчерскими пультами и др.

Экспертные системы – наиболее распространенный класс интеллектуальных систем, ориентированный на тиражирование опыта высококвалифицированных специалистов в областях, где качество принятия решений традиционно зависит от уровня экспертизы, например, таких, как медицина, юриспруденция, геология, экономика, военное дело и др. Экспертные системы эффективны лишь в специфических «экспертных» областях, где важен эмпирический опыт специалистов.

Основные факторы, влияющие на целесообразность и эффективность разработки экспертных систем:

- нехватка специалистов, затрачивающих значительное время для оказания помощи другим;
- выполнение небольшой задачи требует многочисленного коллектива специалистов, поскольку ни один из них не обладает достаточным знанием;

- сниженная производительность, поскольку задача требует полного анализа сложного набора условий, а обычный специалист не в состоянии просмотреть (за отведенное время) все эти условия;
- большое расхождение между решениями самых хороших и самых плохих исполнителей;
- наличие экспертов, готовых поделиться своим опытом.

Главное отличие интеллектуальных систем и экспертных систем от других программных средств – это наличие базы знаний, в которой знания хранятся в форме, понятной специалистам в предметной области, и могут быть изменены и дополнены также в понятной форме. Это и есть языки представления знаний, создание которых до последнего времени являлось центральной проблемой при разработке экспертных систем.

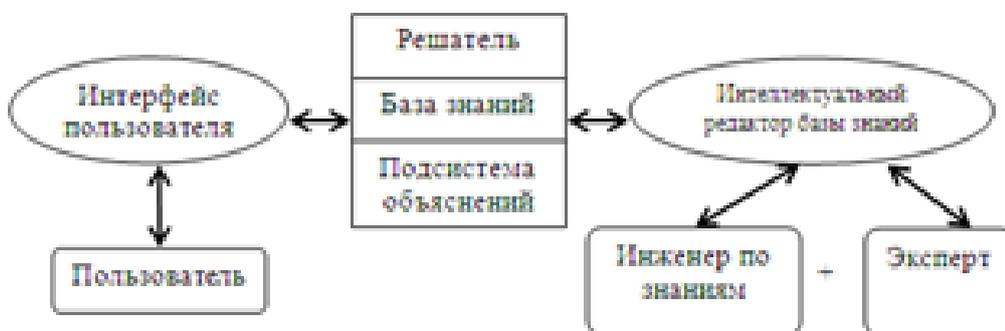


Рисунок 3 – Структура экспертной системы

Обобщенная структура минимальной экспертной системы представлена на рисунке 3. Следует учесть, что реальные экспертные системы могут иметь более сложную структуру, однако блоки, изображенные на этом рисунке, непременно присутствуют в любой действительно экспертной системе, поскольку представляют собой стандарт структуры современной экспертной системы.

2.4 Выводы по главе

Глава посвящена исследованию методов восстановления пропущенных значений и структурирования неопределенностей во временных рядах. Рассмотрены следующие методы восстановления пропущенных значений: замена пропуска общим средним, замена пропуска средним из ближайших, метод сплайн-интерполяции, Zet-алгоритм. Под неопределенностью понимается: нечеткие множества, интервалы, временные ряды. В главе приведен обзор существующих методов структурирования неопределенностей.

Глава 3. МОДЕЛИРОВАНИЕ СИСТЕМ И ПРОЦЕССОВ УСТАНОВЛЕНИЯ СИСТЕМНЫХ СВЯЗЕЙ ЗАЩИТЫ ИНФОРМАЦИИ В УСЛОВИЯХ НЕПОЛНОТЫ И НЕДОСТОВЕРНОСТИ ИСХОДНЫХ ДАННЫХ

3.1 Метод интервального анализа оценки уровня риска информационной системы

В настоящее время управление информационными рисками представляет собой одно из наиболее актуальных и динамично развивающихся направлений стратегического и оперативного менеджмента в области защиты информации. Его основная задача – объективно идентифицировать и оценить наиболее значимые для бизнеса информационные риски компании, а также адекватность используемых средств контроля рисков для увеличения эффективности и рентабельности экономической деятельности компании. Фактически риск представляет собой интегральную оценку того, насколько эффективно существующие средства защиты способны противостоять информационным атакам. Под термином «управление информационными рисками» обычно понимается системный процесс идентификации, контроля и уменьшения информационных рисков компаний в соответствии с определенными ограничениями российской нормативно-правовой базы в области защиты информации и собственной корпоративной политики безопасности. Считается, что качественное управление рисками позволяет использовать оптимальные по эффективности и затратам средства контроля рисков и средства защиты информации, адекватные текущим целям и задачам бизнеса компании.

Обычно выделяют две основные группы методов расчёта рисков информационной безопасности. Первая из них позволяет установить уровень риска путём оценки степени соответствия определённому набору требований по обеспечению информационной безопасности. В качестве источников таких требований могут выступать:

- нормативно-правовые документы предприятия, касающиеся вопросов информационной безопасности;

- рекомендации международных стандартов (ISO 17799, OSTAVE, CoBIT и др.);

- рекомендации компаний-производителей программного и аппаратного обеспечения (Microsoft, Oracle, Cisco и др.).

Вторая группа методов оценки рисков информационной безопасности базируется на определении вероятности реализации атак, а также уровней их ущерба. В данном случае значение риска вычисляется отдельно для каждой атаки и в общем случае представляется как произведение вероятности проведения атаки на величину возможного ущерба от этой атаки. Значение ущерба определяется собственником информационного ресурса, а вероятность атаки вычисляется группой экспертов, проводящих процедуру аудита.

В настоящей работе предлагается метод интервального анализа оценки уровня риска информационной системе. Под уровнем риска при этом подразумевается уровень возможного ущерба, наносимого информационной системе.

Формирование модели угроз, определение взаимосвязи между угрозами и рисками информационной безопасности. Формирование модели угроз информационной безопасности состоит в выборе адекватной решаемой задаче классификации угроз и выделении наиболее распространенных классов из них.

В публикации классификацию угроз выполняют по 2 базовым признакам: по действию на характеристики безопасности информации и по природе источника.

По признаку «действие на характеристики безопасности информации» классификация угроз имеет вид:

- К - тип (угроза конфиденциальности);
- Ц - тип (угроза целостности);

- Д - тип (угроза доступности);
- КЦ - тип (угроза конфиденциальности и целостности);
- КД - тип (угроза конфиденциальности и доступности);
- ЦД - тип (угроза целостности и доступности);
- КЦД - тип (угроза конфиденциальности, целостности и доступности).

По признаку «природа источника» классификация угроз имеет вид:

- объективная (угроза, возникновение которой не зависит от прямой деятельности человека и связанная с разными стихийными природными явлениями);
- субъективная (угроза, возникновение которой зависит от деятельности человека).

Основным недостатком этих двух классификаций является зависимость угрозы от ресурса, на который она воздействует. При этом не отражаются возможные альтернативные сценарии развития угрозы.

В настоящей работе предлагается классифицировать угрозы информационной безопасности по признаку «способ распространения»:

- атаки с использованием вредоносного кода;
- сетевые атаки;
- атаки на получение несанкционированного доступа;
- злоупотребления полномочиями;
- сбои в работе аппаратуры;
- кражи и чрезвычайные ситуации;
- чрезмерное использование систем защиты, ухудшающие работу автоматизированной системы.

Уровень риска информационной безопасности предприятия определяется, как сказано ранее, уровнем ущерба, наносимого предприятию при реализации возможных видов угроз. Уровень ущерба представляет собой качественную характеристику. В таблице 3.1 приведена качественная шкала уровня ущерба компании.

Таблица 3.1 – Качественная шкала оценки уровня ущерба

| № п/п | Уровень ущерба | Описание |
|----------|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Малый | Приводит к незначительным потерям материальных активов, которые быстро восстанавливаются, или к незначительному влиянию на репутацию компании |
| 2 | Умеренный | Вызывает заметные потери материальных активов или к умеренному влиянию на репутацию компании |
| 3 | Ущерб средней тяжести | Приводит к существенным потерям материальных активов или значительному урону репутации компании |
| 4 | Большой | Вызывает большие потери материальных активов и наносит большой урон репутации компании |
| 5 | Критический | Приводит к критическим потерям материальных активов или к полной потере репутации компании на рынке, что делает невозможным дальнейшую деятельность организации |

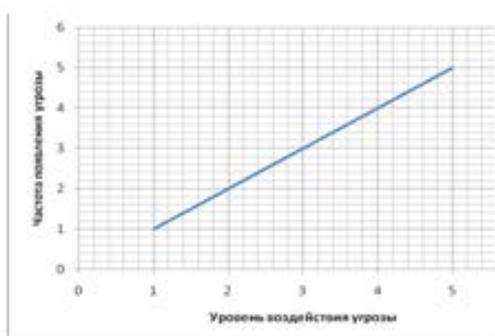
В настоящей работе предлагается метод получения оценки уровня ущерба информационной системе в зависимости от частоты проявления той или иной угрозы. Такая оценка представляется в виде интервала, у которого носитель – уровень ущерба, функция принадлежности – степень проявления угрозы (частость).

Метод интервального анализа оценки уровня ущерба информационной

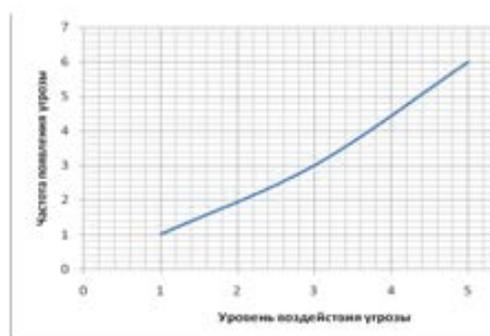
системе состоит в следующем. Группе экспертов предлагается оценить зависимость частоты появления выделенных видов угроз и соответствующего уровня ущерба предприятия. Такая зависимость представляет собой аналитическую функцию. Существуют 4 основных вида трендов (линейный, экспоненциальный, логарифмический, полиномиальный) функций такой зависимости, которые в общем случае можно представить в виде 3 видов линий:

- 1) линейно возрастающая;
- 2) нелинейная с монотонным возрастанием;
- 3) нелинейная с периодом возрастания и убывания.

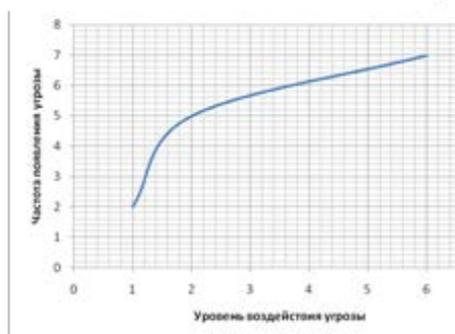
В качестве иллюстрации на рисунке 4 приведены эти 3 вида линий.



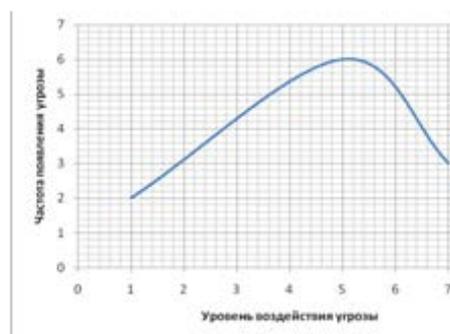
а) линейно возрастающий тренд



б) нелинейный тренд с монотонным возрастанием



в) нелинейный тренд с монотонным возрастанием



г) нелинейный тренд с периодом возрастания и убывания

Рисунок 4 – Виды функции уровня ущерба информационной системе

Уточнение выбранного тренда функции поведения информационной состоит в задании ключевых точек:

- 1) начальной $(x_{нач}, \mu_{нач})$ и конечной $(x_{кон}, \mu_{кон})$ точек для линейного тренда;
- 2) начальной точки $(x_{нач}, \mu_{нач})$, конечной точки $(x_{кон}, \mu_{кон})$, точки перегиба $(x_{пр}, \mu_{пр})$ для экспоненциального, логарифмического и полиномиального трендов.

Пример заполненной анкеты эксперта приведен в таблице 3.2.

Таблица 3.2 – Пример заполненной анкеты эксперта

| № п/п | Класс угрозы | Способ распространения | Вид тренда | | | |
|-------|------------------------------------------|---------------------------------------------------------|----------------------------------------------|------------------------------------------------------------------|-----------------|----------------|
| | | | Линейный | Экспоненциальный | Логарифмический | Полиномиальный |
| 1 | Атаки с использованием вредоносного кода | через файл, прикрепленный к сообщению электронной почты | $(x_{нач}, \mu_{нач}), (x_{кон}, \mu_{кон})$ | - | - | - |
| | | через посторонние дискеты и CD диски | $(x_{нач}, \mu_{нач}), (x_{кон}, \mu_{кон})$ | - | - | - |
| | | через скаченный из Интернета файл | $(x_{нач}, \mu_{нач}), (x_{кон}, \mu_{кон})$ | - | - | - |
| | | с пиратскими программами | $(x_{нач}, \mu_{нач}), (x_{кон}, \mu_{кон})$ | - | - | - |
| | | со СПАМом | $(x_{нач}, \mu_{нач}), (x_{кон}, \mu_{кон})$ | - | - | - |
| 2 | Сетевые атаки | на переполнение буфера | - | $(x_{нач}, \mu_{нач}), (x_{пр}, \mu_{пр}), (x_{кон}, \mu_{кон})$ | - | - |
| | | IP-spoofing | - | $(x_{нач}, \mu_{нач}), (x_{пр}, \mu_{пр}), (x_{кон}, \mu_{кон})$ | - | - |

| | | | | | | |
|---|-------------------------------------------------|---------------------------------------------------|---------------------------------------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------------|---|
| | | на систему НСД | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - | - |
| | | cracking Web-серверов | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - | - |
| 3 | Атаки на получение несанкционированного доступа | установка и использование посторонних программ | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | сканирование IP адресов и портов сети | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | загрузка с дискеты | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | подбор паролей | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | атаки на переполнение буфера | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | подключение модемов и других аппаратных устройств | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ | - |
| | | | | | | |
| 4 | Злоупотребления полномочиями и | использование компьютера в личных целях | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | ошибки персонала | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | продажа корпоративных данных | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | раскрытие конфиденциальных данных | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |

| | | | | | | |
|---|-------------------------------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------|---|---|----------------------------------------------------------------------------|
| | | использование компьютеров для непроизводственной деятельности | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| 5 | Сбои в работе аппаратуры | отказ связи | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | аппаратный сбой | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | потеря питания | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | зависание компьютера | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| 6 | Кражи и чрезвычайные ситуации | воровство активов | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | похищение персонала | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | пожар | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | землетрясение | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| | | наводнение | $(x_{нач}, \mu_{нач}),$ $(x_{кон}, \mu_{кон})$ | - | - | - |
| 7 | Чрезмерное использование систем защиты, ухудшающие работу автоматизированных систем | антивирусная защита | - | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ |
| | | криптографическая защита | - | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ |
| | | защита точек доступа, сетевых служб и сетевых | - | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ |

| | | | | | | |
|--|----------------|-------------------------------------------------------------|---|---|---|----------------------------------------------------------------------------|
| | ванной системы | коммуникаций (МЭ, DNS, DHCP и др.) | | | | |
| | | защита от НСД (встроенные средства и внешние устройства) | - | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ |
| | | разграничение прав доступа, групповая политика и мониторинг | - | - | - | $(x_{нач}, \mu_{нач}),$ $(x_{пр}, \mu_{пр}),$ $(x_{кон}, \mu_{кон})$ |

Для получения обобщенных по всем экспертам функций, принадлежности нечетких множеств уровня ущерба информационной системе предлагается брать взвешенную среднюю по формуле:

$$\mu_{cp} = \frac{1}{n} \sum_{i=1}^n w_i \mu_i,$$

где w_i – весовой коэффициент i -го эксперта (степень его компетентности), n – количество экспертов, μ_i – выбранная степень принадлежности i -го эксперта для заданной степени принадлежности.

Предложенный метод построения, функций принадлежности оценки уровня риска информационной системы отличается от существующих тем, что:

- во-первых, он не использует аппарат теории вероятностей в силу отсутствия реальной статистики воздействия угроз;
- во-вторых, не применяет процедуру оценки степени соответствия информационной системы определённому набору требований по обеспечению информационной безопасности, что может являться весьма дорогой процедурой для предприятия.

3.2 Задачи защиты информации, формализуемые как экстремальные задачи на графах с интервальными весами

Задачу многокритериальной оценки средств защиты информации можно свести к интервальной экстремальной задаче на графах. В заданном n -вершинном графе:

$$G = (V, E),$$

каждое ребро $e \in E$ взвешено интервалом $w(e)$, т.е. отрезком:

$$w(e) = [w_1(e), w_2(e)],$$

где $w_1(e) \leq w_2(e)$. Допустимое решение рассматриваемой задачи представляем в виде подграфа:

$$x = (V_x, E_x), V_x \subseteq V, E_x \subseteq E.$$

Обозначим через $X = \{x\}$ МДР рассматриваемой задачи, на котором определена интервальная целевая функция (ИЦФ) вида *MAXSUM* :

$$w(x) = \sum_{e \in E_x} w(e) \rightarrow \max \quad (3.1)$$

или ИЦФ вида *MAXMIN*

$$w(x) = \min_{e \in E_x} w(e) \rightarrow \max \quad (3.2)$$

Значение этих ИЦФ можно получить из свойств представленных выше операций сложения интервалов и сравнения интервалов, откуда значение ИЦФ (3.1) или (3.2) также есть интервал:

$$w(x) = [w_1(x), w_2(x)],$$

где $w_i(x) = \sum_{e \in E_x} w_i(e)$, $i = 1, 2$. Под решением интервальной задачи понимается такой элемент $x^0 \in X$, на котором значение ИЦФ (3.1) или (3.2) достигает требуемого экстремума.

В случае интервальных весов нахождение оптимума наталкивается на проблему выбора наиболее целесообразного решения из множества несравнимых альтернатив. В связи с этим необходимо ввести отношения предпочтения, эквивалентности и несравнимости.

Бинарное отношение (БО) предпочтительности условимся обозначать символом \prec , БО несравнимости – символом \leftrightarrow , БО эквивалентности – символом \sim . БО предпочтительности определяется с учетом того, является ИЦФ максимизируемой или, наоборот, она является минимизируемой. Из двух элементов x_1 и x_2 , $x_1, x_2 \in X$, решение x_1 предпочтительнее решения:

$$x_2 \quad (x_1 \prec x_2),$$

$$\text{если } w_i(x_1) \leq w_i(x_2), \quad i=1,2,$$

в случае минимизируемой ИЦФ.

$$w_i(x_1) \geq w_i(x_2), \quad i=1,2,$$

в случае максимизируемой ИЦФ.

Решения x_1 и x_2 несравнимы ($x_1 \leftrightarrow x_2$), когда имеет место строгое вложение интервалов $w(x_1) \subset w(x_2)$, либо $w(x_2) \subset w(x_1)$. Эти решения эквивалентны ($x_1 \sim x_2$), если совпадают соответствующие им интервалы $w(x_1) = w(x_2)$.

Отношения предпочтения и несравнимости порождают на МДР X паретовское множество (ПМ) $\tilde{X} \subseteq X$, состоящее из паретовских оптимумов (ПО).

Весьма важным является принципиальная возможность сведения задач с интервальными данными к многокритериальным задачам.

3.3. Описания программного обеспечения

3.3.1. Структура программного обеспечения

Программой обеспечения на среде Delphi-6 . Как известно файлы Delphi называются проектами. Проекты Delphi состоит из формы и модули. Формы Delphi имеют такой же вид, как окна Windows позволяют создавать удобный для пользователя графический интерфейс. Модули являются библиотеками подпрограмм, которой можно использовать для решения различных задач. Формы Delphi состоят из отдельных элементов управления которая служат для ввода данных в удобный для пользователя виде и отображения результатов. В основе программирование Delphi лежит понятие событийно управляемого программирования. Программный под окна создаются как реакция на отдельных событие вызываемого действиями пользователя. Такими событиями могут быть ввода текста, нажатие клавиши, мыши и т.д.

Программой обеспечения состоит из двух отдельных проектов:

- 1) Проект, служат для решения задач простым симплекс методом.
Основная форма этого проекта переведена на рисунке 5.
- 2) Проект, , служат для решения задач интервальным симплекс методом .
Основная форма этого проекта переведена на рисунке 6.

Элементы использованные в этих формах и назначения, о также модули описаны в разделе 3.3.2.

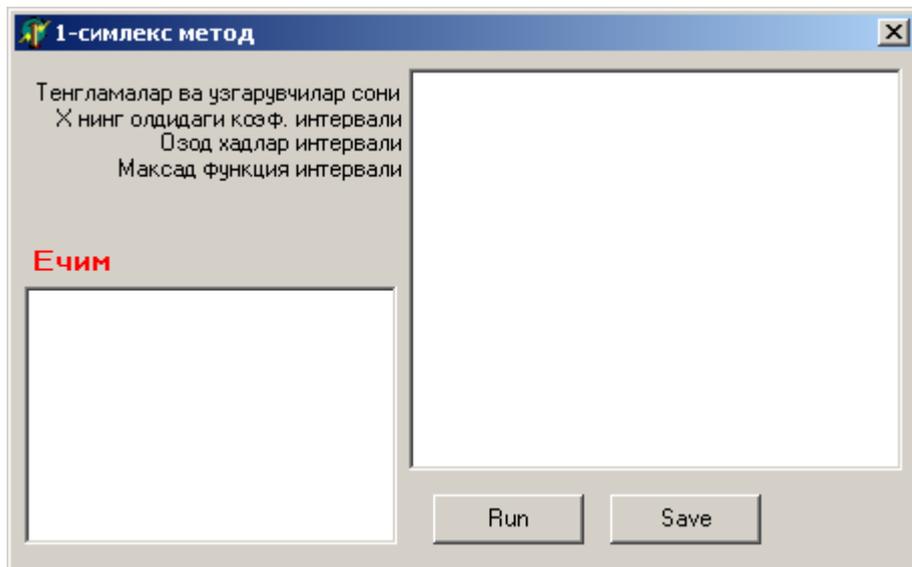


Рисунок 5 – Форма программы обычного симплекс метода.

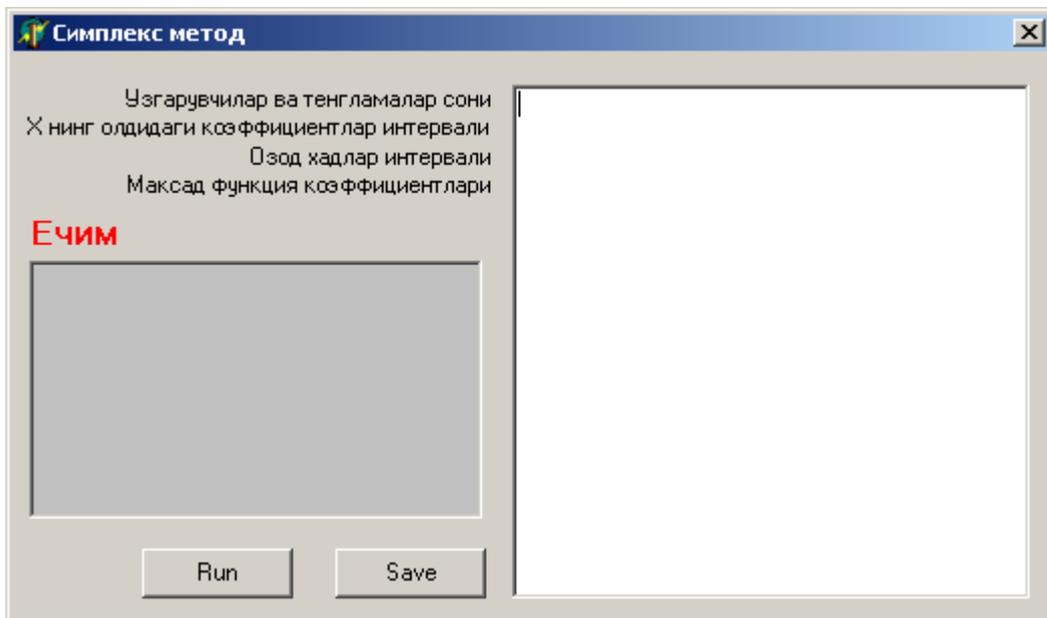


Рисунок 6 – Форма программы симплекс метода на основе интервального анализа.

3.3.2. Основные модули программного обеспечения

Модули программы симплекс метода

TFORM1- Форма предназначена для реализации обычного симплекс метода. С помощью этой формы вводятся исходные данные, а также отображаются результаты которые могут быть сохранены в виде файла.

Button 1- данная кнопка запускает программу симплекс метода после ввода данных.

- Button 2- данная кнопка предназначена для сохранения результатов в файле.
- Label 1
- Label 2
- Label 3
- Label 4
- Label 5
- Memo 1
- Memo 2
- Save dialog 1

В программе используются функции:

- Min Bnis- функция вычисления минимальных « B_{io} » коэффициентов.
- Min Delta- функция вычисления минимального значения дельта для каждого столбца в каждой таблице.

Модули программы симплекс метода на основе интервального анализа

TFORM2- Форма предназначена для реализации интервального симплекс метода. С помощью этой формы вводятся исходные интервальные данные и производится вычисления, а также отображаются интервальные результаты которые могут быть сохранены в виде файла.

В программе используется функции:

- Min Bnis- функция вычисления минимальных интервальных « B_{i0} » коэффициентов.
- Min Delta- функция вычисления минимального значения дельта для каждого столбца в каждой таблицы.

Button 1- данная кнопка запускает программу интервального симплекс метода после ввода данных.

Button 2- данная кнопка предназначена для сохранения результатов в файле.

3.3.3. Инструкция по использованию программного обеспечения

Прежде чем, привести программу в рабочее состояние надо обеспечить следующие операции:

1. Загрузить компьютер
2. Проверить, в данном компьютере имеется среда Delphi 6, если она есть, то программа будет работать как exe файл, а в противном случае необходимо установить программу.
3. Если сама программа находится в гибком диске A:, то компьютер будет работать только с этим диском.
4. После загрузки компьютера, на рабочем столе появиться иконка, мой компьютер, щелкнут на него дважды.
5. После того как окно откроется, на экране появятся разные папки из них выбрать A: и щелкнут в него дважды.
6. Когда окно откроется, нужно выбрать папку данной программы, и запустить её.
7. После того как программа откроется, на экране появиться форма программы, которая состоит из двух частей. На верху правого угла даются переменные по соответствующим надписям рядом, а внизу окно с готовыми результатами.

8. После завершения ввода переменных надо сохранить значения , для этого надо нажать кнопку **SAVE**
9. После нажатия кнопки **save** , на экране компьютера выйдет диалоговое окно, из которого надо будет выбрать диск или папку с которым мы имеем дело, в данном случае это диск A:.
10. После выбора диска A:, внизу надо будет показать куда и в какое место (в какой файл) в данном диске надо сохранить данные.
11. После того как все необходимые операции выполнены, связанные с сохранением нужно будет нажать на кнопку сохранить диалоговом окне
12. Потом на экране компьютера выйдет прежняя форма программы, и после всего, в конечном этапе , следует нажать на кнопку **RUN** и появится результат .

Если необходимо закрыть программу, то надо нажать на кнопку **X**, на верхнем правом углу экрана.

3.4 Выводы по главе

В главе приведены методы решения двух задач защиты информации: оценки уровня риска информационной системы и многокритериальной оценки средств защиты информации. Для решения первой из этих задач применяется инструментарий теории нечетких множеств. Для решения второй задачи – интервальная арифметика и методы многокритериального выбора.

Заключение

1. Приведен обзор возможных неопределенностей в исходных данных систем и процессов защиты информации. Отмечено, что основными видами неопределенности являются: неполнота информации, недостоверность информации. Неполнота информации – это пропущенные значения в исходных данных, недостоверность – представление исходных данных в виде нечетких множеств и интервалов.

2. Проведен анализ существующих методов структурирования неопределенностей, а именно: методы восстановления пропусков в массивах данных, арифметических операций над интервальными и нечеткими множествами, методы прогнозирования временных рядов.

3. Предложены подходы к решению двух задач: построения функций, принадлежности оценки уровня риска информационной системы и многокритериальной оценки средств защиты информации с использованием методов структурирования неопределенностей в исходных данных.

Список используемых источников

I. Законы Республики Узбекистан

1. Закон «о телекоммуникациях» Республики Узбекистан. 20 августа 1999 г.

II. Указы и постановления Президента Республики Узбекистан, Кабинета министров

1. Указ Президента Республики Узбекистан за №УП-3080 от 30.05. 2002 г. «О дальнейшем развитии компьютеризации и внедрении информационно-коммуникационных технологий».

2. Постановление Президента Республики Узбекистан «О дополнительных мерах по дальнейшему развитию информационно-коммуникационных технологий», г. Ташкент. 8 июля 2005 г.

3. Постановление Президента Республики Узбекистан «О мерах по дальнейшему внедрению и развитию современных информационно-коммуникационных технологий», г. Ташкент. 21 марта 2012 г.

III. Произведения Президента Республики Узбекистан И.А.Каримова

1. И.А. Каримов. «Узбекистан на пороге XXI века. Угрозы безопасности условия и гарантии прогресса.» Ташкент: «Шарқ», 2006г.

IV. Основная литература

1. Малюк А.А. Информационная безопасность: концептуальные и методологические основы защиты информации. – М: Горячая линия-Телеком, 2004. – 280с.

2. Калмыков С.А, Шокин Ю.И, Юлдашев З.Х. “Методы интервального анализа”. 1987 г.

3. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во ин-та математики, 2003

4. Литтл Р. Дж. А., Рубин Д.Б. Статистический анализ данных с пропусками. – М.: Финансы и статистика, 1990

5. И.Б. Шубинский. Активная защита от отказов микропроцессорных вычислительных систем – М.: Знание, 1987.– 116с.

6. Надежность, живучесть и безопасность корабельных электроэнергетических систем // Рябинин И.А., Парфенов Ю.М. – Санкт-Петербург; ВМА, 1997. – 430 с.

7. Г.Н. Черкесов. Методы и модели оценки живучести сложных систем. М.: Знание, 1997.– 118с.

V. Дополнительная литература

1. В.А. Ершов, Н.А. Кузнецов. Мультисервисные телекоммуникационные сети. – М.: Изд-во МГТУ им- Н.Э. Баумана, 2003.- 432 с.

2. Е.Б. Алексеев. Особенности технической эксплуатации волоконно-оптических систем и сетей синхронной цифровой иерархии // Учебное пособие, ИПК М ТУСИ. – М.: ЗАО «Информсвязь издат»,1999 г.

VI. Периодические издания, статистические сборники и отчеты

1. Н.Л. Сторожук. Устройства защиты от перенапряжений и эффективность их применения // Электросвязь. – 2006. - №2.

2. А.Л. Зоричев, В.Г. Лищинский. Молниезащита объектов связи // Электросвязь -2006. - №2.

3. Э.Я. Геча, В.А. Нестерко. Способы защиты кабельных конструкций от радиальной диффузии воды // Электросвязь - 2005. - №8.

4. А. Х. Султанов, В.Н. Акульшин, И.Л. Виноградова. Подход к повышению надежности и быстродействия волоконно- оптических систем передачи // Электросвязь -2005. - №12.

5. В.А. Нетес. Основные принципы организации самозалечивающихся сетей на основе синхронной цифровой иерархии // Электросвязь. – 1995. - №1

6. Г.А. Птицын. Живучесть сетей связи // Электросвязь. -2001. - №2.

7. В.Е. Митрохин. Структурная надёжность развивающейся сети при грозовых разрядах в условиях неоднородной структуры земли // Электросвязь. – 2002. - № 6.

8. Е.Б. Алексеев. Особенности эксплуатации ВОСП и пути повышения качества их функционирования // Электросвязь.- 1997. - №6.

9. Е.Б. Алексеев. Оптимизация решений при проектировании и организации

технической эксплуатации ВОСП по критерию надежности // Электросвязь.- 2002.-№6.

10. РФ 45.047-99 «Линии передачи волоконно-оптические на магистральной и внутризоновых первичных сетях ВСС России. Техническая эксплуатация». Руководящий технический материал. – М.: ООО «Резонанс». 200г.

11. Р.И. Исаев. Статистический анализ надежности и живучести кабельных систем передачи // Aloqa dunyosi. – 2006. - № 1/6.

VII. Интернет ресурсы