

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖА БЕРУВЧИ
DSc.27.06.2017.Т.07.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АСОСИДАГИ БИР МАРТАЛИК ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ

МАДРАХИМОВ ШАВКАТ ФАЙЗУЛЛАЕВИЧ

УМУМЛАШГАН БАҲОЛАРНИ ҲИСОБЛАШ УСУЛЛАРИ АСОСИДА
ЯШИРИН ҚОНУНИЯТЛАРНИ АНИҚЛАШ ТИЗИМЛАРИ

05.01.03 – Информатиканинг назарий асослари

ТЕХНИКА ФАНЛАРИ ДОКТОРИ (DSc)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент – 2020

Докторлик (DSc) диссертацияси автореферати мундарижаси
Оглавление автореферата докторской (DSc) диссертации
Content of the Doctoral (DSc) Dissertation Abstract

Мадрахимов Шавкат Файзуллаевич

Умумлашган баҳоларни ҳисоблаш усуллари асосида яширин қонуниятларни аниқлаш тизимлари..... 3

Мадрахимов Шавкат Файзуллаевич

Системы обнаружения скрытых закономерностей на базе методов вычисления обобщенных оценок..... 25

Madrakhimov Shavkat Fayzullayivich

Systems for detecting hidden regularities on the basis of methods for calculating generalized estimates..... 47

Эълон қилинган ишлар рўйхати

Список опубликованных работ
List of published works..... 51

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖА БЕРУВЧИ
DSc.27.06.2017.Т.07.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АСОСИДАГИ БИР МАРТАЛИК ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ

МАДРАХИМОВ ШАВКАТ ФАЙЗУЛЛАЕВИЧ

УМУМЛАШГАН БАҲОЛАРНИ ҲИСОБЛАШ УСУЛЛАРИ АСОСИДА
ЯШИРИН ҚОНУНИЯТЛАРНИ АНИҚЛАШ ТИЗИМЛАРИ

05.01.03 – Информатиканинг назарий асослари

ТЕХНИКА ФАНЛАРИ ДОКТОРИ (DSc)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент – 2020

Техника фанлари бўйича фан доктори (DSc) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2017.1.DSc/T27 рақами билан рўйхатга олинган.

Докторлик диссертацияси Ўзбекистон Миллий университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгашнинг веб саҳифасида (www.tuit.uz) ва «Ziyonet» Ахборот таълим порталида (www.ziyonet.uz) жойлаштирилди.

Илмий маслаҳатчи: **Игнатъев Николай Александрович**
физика-математика фанлари доктори, профессор

Расмий оппонентлар: **Шарипбай Алтинбек Амирули**
техника фанлари доктори, профессор

Гулямов Шухрат Маннопович
техника фанлари доктори, профессор

Маматов Нарзилло Солиджонович
техника фанлари доктори, катта илмий ходим

Етакчи ташкилот: **Қозоғистон Миллий университети**

Диссертация химояси Тошкент ахборот технологиялари университети ҳузуридаги DSc.27.06.2017.Г.07.01 рақамли Илмий кенгашнинг 2020 йил «___» _____ соат ___ даги мажлисида бўлиб ўтади. (Манзил: 100202, Тошкент ш., Амир Темур кўчаси, 108. Тел.: (+99871) 238-64-43; факс: (+99871) 238-65-52, e-mail: tuit@tuit.uz).

Диссертация билан Тошкент ахборот технологиялари университетининг Ахборот-ресурс марказида танишиш мумкин (_____ рақами билан рўйхатга олинган). Манзил: 100202, Тошкент ш., Амир Темур кўчаси, 108. Тел.: (+99871) 238-64-43.

Диссертация автореферати 2020 йил «___» _____ кун тарқатилди.
(2019 йил «___» _____ даги _____ рақамли реестр баённомаси).

Р.Х.Ҳамдамов

Илмий даражалар берувчи илмий
кенгаш раиси, т.ф.д., профессор

Ф. М. Нуралиев

Илмий даражалар берувчи илмий
кенгаш илмий котиби, т.ф.д., доцент

Ш.Ҳ.Фозилов

Илмий даражалар берувчи Илмий
кенгаш қошидаги бир марталик илмий
семинар раиси, т.ф.д., профессор

КИРИШ (фан доктори (DSc) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳонда компьютер технологияларини жадал суратларда ривожланиши ва уларни инсон фаолиятининг турли соҳаларида қўлланилиши билан боғлиқ ҳолда берилганларга сифатли (интеллектуал) ишлов беришга эҳтиёж кузатилмоқда. Жамиятнинг инновацион ривожланишидаги асосий омил – бу билимдир. Мавжуд билимларни ўзлаштириш билан бир қаторда, янги қонунлар ва қонуниятларни очиш ҳисобига улар ҳажмларини тўлдириш долзарб масалалардан биридир. Дунёнинг ривожланган мамлакатлари, шу жумладан, АҚШ, Буюк Британия, Германия, Франция, Россия Федерацияси, Япония, Хитой, Жанубий Корея ҳамда Amazon, Alibaba, Google, Yandex, Microsoft ва SAN каби халқаро компанияларда прогноз қилиш, регрессия ва классификация масалаларини ечишга йўналтирилган сунъий интеллект тизимларини яратиш доирасида берилганларга интеллектуал ишлов бериш усулларини ишлаб чиқиш ва қўллаш бўйича фаол илмий тадқиқотлар олиб борилмоқда.

Жаҳонда берилганларни интеллектуал таҳлил қилиш усуллари асосида аналитик тавсифига эга бўлмаган ва фақат тажриба берилганларининг тўплами билан қаноатланадиган катта кўламдаги амалий масалалар ечишга йўналтирилган интеллектуал тизимларни яратиш бўйича кенг қамровли илмий тадқиқотлар олиб борилмоқда. Бу борада, жумладан, берилганларнинг ўтказиб юборилган (тўлдирилмаган) қийматларини тиклаш, танланманинг «шовқин» объектларини аниқлаш ва ўчириш, берилганлар базасидаги яширин қонуниятларни аниқлаш, берилганларга дастлабки ишлов бериш, информатив аломатларни аниқлаш, объектлар тавсифидаги янги аломатлар фазосига ўтиш, объектларнинг жойлашув тузилмасини визуаллаштириш, қатъиймас мантик аппаратида фойдаланган ҳолда интеллектуал интерфейсларни яратиш усул ва алгоритмларини ишлаб чиқиш муҳим тадқиқот мавзулари сифатида қаралади.

Республикада ахборот технологияларини ижтимоий ва ишлаб-чиқариш соҳаларига кенг қамровли тадбиқ қилиш кузатилмоқда, давлат бошқаруви тизимининг самарадорлигини ошириш учун ижтимоий ва ишлаб чиқариш соҳаларида ахборот технологияларини кенг жорий этиш тадбирлари, электрон ҳукумат дастурлари ишлаб чиқилмоқда. 2017-2021 йилларда Ўзбекистон Республикасини янада ривожлантириш бўйича ҳаракатлар стратегиясида, жумладан «... иқтисодиёт, ижтимоий соҳага, бошқариш тизимига инфор­мацион-коммуникацион технологияларни жорий этиш, ... муҳандислик-коммуникация ва ижтимоий инфратузилмани ривожлантириш ҳамда модернизация қилиш»¹ вазифалари белгиланган. Мазкур вазифаларни амалга оширишда, интеллектуал ахборот тизимлари учун янги билимлар

¹ Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги Фармони

манбаи бўлиб хизмат қилувчи, берилганлар базасидан яширин қонуниятларни ажратиб олиш муҳим вазифалардан бири ҳисобланади.

Ўзбекистон Республикасининг «Электрон ҳукумат тўғрисида»ги Қонуни (2015), Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги, 2017 йил 29 ноябрдаги ПФ-5264-сон «Ўзбекистон Республикаси инновацион ривожлантириш вазирлигини ташкил этиш тўғрисидаги» ва 2018 йил 19 февралдаги ПФ-5349-сон «Ахборот технологиялари ва коммуникациялари соҳасини янада такомиллаштириш чора-тадбирлари тўғрисида» Фармонлари, 2012 йил 21 мартдаги ПҚ-1730-сон «Замонавий ахборот-коммуникация технологияларини жорий этиш ва янада ривожлантириш чора-тадбирлари тўғрисидаги», 2014 йил 3 апрелдаги ПҚ-2158-сон «Иқтисодиётнинг реал секторида ахборот коммуникация технологияларини янада жорий қилиш тўғрисида» Қарорлари ҳамда ахборот технологиялари соҳасига тегишли бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Мазкур тадқиқот республика фан ва технологиялар ривожланишининг IV. «Ахборотлаштириш ва ахборот коммуникация технологияларини ривожлантириш» устувор йўналиши доирасида бажарилган.

Диссертация мавзуси бўйича хорижий илмий тадқиқотлар шарҳи². Берилганларни интеллектуал таҳлил қилиш усуллари ва алгоритмларини ишлаб чиқишга йўналтирилган илмий тадқиқотларда янги усулларни яратиш ва асослаш билан бирга мавжудларини предмет соҳаларнинг математик моделлари доирасида амалий қўллаш бўйича кенг қамровли ишлар жаҳоннинг етакчи илмий марказлари ва олий таълим муассасаларида олиб борилмоқда, жумладан, Европа, Америка ва Осиё мамлакатларининг Халқаро сунъий интеллект ассоциациялари, Россия Сунъий интеллект илмий-тадқиқот институти, Москва давлат университети, Томск давлат университети, Санкт-Петербург информатика ва автоматлаштириш институти (Россия Федерацияси), NYU Center for Data Science, National Center for Data Mining of the University of Illinois at Chicago, University of California, Department of Electrical and Computer Engineering University of Texas (АҚШ), Darmstadt University of Applied Sciences (Германия), University of Lions (АҚШ, Канада), Centre for Data Mining and Database Technology (Ҳиндистон), Data mining and Big data institute (Исроил) ва бошқаларни мисол келтириш мумкин.

² Диссертация мавзуси бўйича илмий тадқиқотлар шарҳи www.braude.ac.il, sonatech.ac.in, www.msu.ru, cds.nyu.edu, www.ncdm.uic.edu, www.ece.utexas.edu, www.h-da.de/, www.ifmo.ru, www.universite-paris-saclay.fr, www.en.uni.lu, medicine.utah.edu, www.icyb.kiev.ua, www.titech.ac.jp, www.keele.ac.uk, www.cam.ac.uk, www.ox.ac.uk, www.conf.ling.cornell.edu, www.compling.stonybrook.edu, gtnlp.wordpress.com, ling.uni-konstanz.de, www.biominingu.org, mi.eng.cam.ac.uk/mi/Main/Speech ва бошқа манбалар асосида ишлаб чиқилган.

Интеллектуал таҳлил усуллари асосида берилганлар базасидан яширинган қонуниятларни аниқлаш усуллари ва алгоритмларини яратиш ҳамда машинали ўргатиш, образларни англаш ва тавсифлаш тизимларини такомиллаштиришга оид жаҳонда олиб борилган тадқиқотлар натижасида қатор, жумладан қуйидаги илмий натижалар олинган: диабетик касалликларни ташҳис қилиш масалаларига «чуқур ўргатиш» усули яратилган (Université Paris-Saclay, Франция); табиий тилдаги матнларга ишлов бериш масаласини ечиш учун text mining усуллари ишлаб чиқилган (Speech Research Group, Буюк Британия, University of Luxembourg, Люксембург, Bharathiar Universiteti, Ҳиндистон, University of Oxford, Буюк Британия, Cornell Computational Linguistics Lab, Stony Brook University, АҚШ, Konstanz universiteti, Германия, Georgia Tech Computational Linguistics Lab, Грузия, АБҲҲ, Россия Федерацияси); машинали ўргатиш тизимлари учун тажрибавий валидация усуллари яратилган (Center for Data Science and Public Policy at the University of Chicago, АҚШ); аломатлар фазосини танлаш усуллари ва алгоритмларининг математик асоси ишлаб чиқилган (Бошқарув муаммолари институти, Россия Федерацияси); танланма селекцияси учун «шовқин» объектларини аниқлаш ва ўчириш усуллари ишлаб чиқилган (Томск давлат университети, Россия Федерацияси); берилганларни интеллектуал таҳлил ва машинали ўрганиш усуллари яратилган ва нанозаррачаларнинг биомедицина хоссаларини прогноз қилиш масаласи ечилган (Department of Biomedical Informatics, University of Utah, АҚШ); кўп поғонали бошқарувни амалга оширувчи интеллектуал тизимни қуриш усул ва воситалари яратилган (В.М.Глушков номидаги Кибернетика институти, Украина); матндан маълумотларни топиш масаласини ечишда ҳавола ва ассоциацияларни таҳлил қилиш, визуаллаштириш ва интеллектуал аналитика усуллари ишлаб чиқилган (Tokyo Institute of Technology, Япония); рандомлашган дарахтлар ўрмони асосида алгоритмлар ансамбли аниқланган (Keele university, Буюк Британия); кўп параметрли образларни англаш масалаларини моделлаштириш ва ечиш, ҳамда берилганлар таҳлили учун тадбиқий статистика алгоритмлари яратилган (Cambridge university, Буюк Британия).

Дунёда сунъий интеллектнинг берилганларни интеллектуал таҳлили соҳасидаги усулларни такомиллаштириш ва янгиларини яратиш бўйича бир қатор, жумладан, қуйидаги устувор йўналишларда тадқиқотлар олиб борилмоқда: образларни англаш алгоритмларининг умумлаштириш қобилиятини баҳолаш учун мезон ва усулларини яратиш; мантиқий қонуниятларни излашдаги комбинатор мураккабликни камайтириш усул ва алгоритмларини ишлаб чиқиш; предмет соҳанинг ўзига хослигини инобатга олган ҳолда табиий тиллардан машинали таржима қилиш усулларини ишлаб чиқиш.

Муаммони ўрганилганлик даражаси. Берилганларни таҳлил қилиш, хусусан, берилганлар базасидан яширинган билимларни ажратиб олиш, информатив аломатларни аниқлаш ва танланма объектлари баҳоларини ҳисоблаш масалаларининг илмий асослари Ю.И.Журавлев, Н.Г.Загоруйко,

В.А.Дюк, В.П.Гладун, А.С.Потапов, Ф.И.Андон, В.Н.Вапник, А.Ю.Зиновьев, К.В.Воронцов, Г.С.Лбов, В.Г.Царегродцев, П.Флах, Г.Пятецкий-Шапиро, Х.Эдельштайн, А.А.Барсегян, Т.А.Гаврилова, В.Ф.Хорошевский ва бошқа олимлар ишларида баён этилган. Ю.И.Журавлев образларнинг англаш масалаларига алгебраик ёндошувни амалга оширган, В.А.Дюк тадқиқотлари берилганлардаги мантикий қонуниятларни излаш алгоритмларни ишлаб чиқишдаги комбинатор мураккабликни қисқартиришга бағишланган, Ю.А.Зиновьев ва В.Г.Царегродцев ишларида қўйилган масалага сунъий нейрон тўри моделларини қўллаш асослари келтирилган.

Республикамизда образларни англаш ва сунъий интеллект соҳасида илмий тадқиқотларнинг назарий асосларини такомиллаштириш ва ҳисоблаш усулларини ишлаб чиқишда муваффақиятга эришган бир қатор олимларни кўрсатиш мумкин, жумладан, академик М.М.Камилов, Т.Ф.Бекмуратов, Ф.Т.Адилова, Ш.Ҳ.Фозилов, Д.Т.Мухамадиева, Р.Н.Усманов, Р.Ҳ.Ҳамдамов, А.Х.Нишановлар тадқиқотларида информатив аломатларни саралашнинг аналитик усуллари, баҳоларни ҳисоблаш алгоритмлари, нейро-қатъиймас мантиқ моделлари, танланмага дастлабки ишлов бериш усуллари қаралган.

Берилганлар базасидан ички яширинган қонуниятларни аниқлашда эвристик алгоритмлардан фойдаланилиши сабабли, одатда заиф формаллашувчи предмет соҳалар учун информацион моделларни қуриш қатъий математик асосга эга эмас. Шу сабабли, предмет соҳалардаги жараёнлар ва ҳодисаларни адекват равишда тавсифловчи моделларни шакллантириш мезонлари ва алгоритмларини танлаш услубиятини такомиллаштириш талаб қилинади.

Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқотлари Ўзбекистон Миллий университети илмий тадқиқот ишлар режасининг Ф-4-64 «Берилганлар интеллектуал таҳлилида умумлашган баҳолар ва объектларнинг индивидуал метрикасини ҳисоблаш усулларини яратиш ва асослаш» (2012-2016), ОТ-Атех-2018-224 «Йўғон ичак яллиғланиши ва колоректал саратонининг ривожланиши, кечиши ва оқибатларини прогноз қилиш, ҳамда этиопатогенез ва таваккалчилик факторларини инобатга олган ҳолда диагностикани амалга ошириш ва даволашнинг ахборот моделини ишлаб чиқиш» (2018-2019) мавзусидаги лойиҳалар доирасида бажарилган.

Тадқиқотнинг мақсади тажриба берилганлар базаси бўйича яширинган қонуниятларни излаш учун берилганларнинг интеллектуал таҳлили мезонлари ва усулларини яратиш, ҳамда асослашдан иборат.

Тадқиқотнинг вазифалари:

миқдорий аломатлар қийматларини, чегараларида ўзаро кесишмайдиган синфлар вакиллари устунлик қиладиган интервалларга бўлиш масаласи ечимининг мавжудлигини асослаш;

умумлашган баҳолар асосида мумкин бўлган объектлар тавсифидаги янги аломатлар фазосини шакллантириш;

турли тоифадаги аломатларнинг кўрсатилган ва аниқланувчи тўплам остилари бўйича объектларни сон ўқиға акслантиришнинг стохастик ва детерминистик усуллари ишлаб чиқиш;

объектларнинг умумлашган баҳолари асосида ўргатувчи танланма селекциясини амалға ошириш;

дарахт атрибутлари сифатида, аниқланувчи объектлар тўплами бўйича ҳисобланган умумлашган баҳолар қийматлари билан ҳисобланувчи латент аломатларни танлаш орқали қарор қилиш дарахтини қуриш;

берилганларнинг интеллектуал таҳлилининг сонли натижалари асосида хулоса қоидаларини шакллантириш.

Тадқиқотнинг объекти сифатида берилганларнинг интеллектуал таҳлил қилиш усуллари асосида турли предмет соҳаларнинг берилганлар базасидан яширин қонуниятларни изловчи информаион моделлар олинган.

Тадқиқотнинг предмети умумлашган баҳоларни ҳисоблашға асосланган берилганларнинг интеллектуал таҳлили усулларида иборат.

Тадқиқотнинг усуллари. Ишдаги назарий тадқиқотлар берилганларнинг интеллектуал таҳлили назарияси ва алгоритмлари, берилганларға ишлов бериш ҳамда қатъиймас мантиқ усуллариға таянилган, ҳисоблаш экспериментини ўтказиш учун берилганлар базасидан яширин қонуниятларни ажратиб олишға асосланган компьютер воситасида моделлаштиришдан фойдаланилган.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

мумкин бўлган объектларни тавсифловчи миқдорий аломатларнинг иккита синфдан бирининг вакиллари устунлик қиладиган интервалларға бўлишнинг турғунлигини баҳолаш усули (формуласи) ишлаб чиқилган;

умумлашган баҳолар асосида бошланғич аломатлар боғлиқмаслиги инобатға олинган ҳолда аниқланувчи объектлар тўплам остиси бўйича ҳисобланган латент аломатларни қарор қилиш дарахти тугунларининг атрибути сифатида танлаш усули ишлаб чиқилган.

латент аломатлар бўйича янги аломатлар фазосини шакллантириш алгоритми ишлаб чиқилган;

танланмаларни селекция қилиш учун умумлашган баҳолардан фойдаланиш асосида шовқин объектларни ажратиб олиш мезони аниқланган;

интервал усуллари асосида экспертлар томонидан кўрсатилган ёки аниқланган аломатлар тўплам остиси бўйича қарор қабул қилиш жараёнини қатъиймас мантиқ терминларида изоҳлаш усули ишлаб чиқилган.

Тадқиқотнинг амалий натижаси қуйидагилардан иборат:

танланма объектларининг ўз синфида турғунлигини ҳисоблаш асосида танланма сифатини (компактлик гипотезасининг ростлик даражасини) аниқлаш усули яратилган;

танланма объектларини тавсифловчи миқдорий аломатлар қийматларини интервалларға бўлиш орқали объектларнинг синфларға тегишлилик даражасини аниқлаб берувчи қатъиймас мантиқ қоидаларини шакллантириш усуллари яратилган;

умумлашган баҳолар асосида қарор қилиш дарахтини қуриш алгоритми ишлаб чиқилган;

стохастик ва детерминистик усулларда ҳисобланган умумлашган баҳолар (латент аломатлар) асосида танланма объектларининг тузилмавий жойлашувини визуаллаштиришдан фойдаланган ҳолда классификация масаласини ечиш технологияси ишлаб чиқилган;

объектларни тавсифловчи бошланғич аломатлар тўпламини иерархик агломератив гуруҳлаш алгоритми асосида қисқартирилган, янги аломатлар фазосига ўтиш орқали «ўлчамлар ланъати» муаммосини ечиш усули таклиф қилинган;

предмет соҳаларнинг инфор­мацион моделларини қуриш учун берилганларнинг интеллектуал таҳлили алгоритмларини танлаш усули ишлаб чиқилган.

Тадқиқот натижаларининг ишончлилиги. Тадқиқот натижаларининг ишончлилиги предмет соҳалар танланмаларида, ҳамда Internet глобал тармоғининг махсус сайтларида жойлаштирилган репозиторийлардан олинган берилганлар базаларида ўтказилган ҳисоблаш тажрибалари, синфлар объектлари ўртасидаги муносабатлар тузилмасини таҳлил қилишдаги компактлик гипотезасининг ростлик даражасини ҳисоблайдиган объектлар турғунлиги мезонларидан фойдаланиш, яширинган қонуниятларни ажратиб олиш учун усуллар кетма-кетлигини танлаш технологияларини қўллаш билан изоҳланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти умумлашган баҳолардан фойдаланган ҳолда берилганлар базаси ва сақлагичларидан янги билимларни ажратиб олиш усуллари ва технологияларини яратиш орқали берилганларни интеллектуал таҳлили назариясини ривожлантиришга хизмат қилади. Ишлаб чиқилган усул ва алгоритмлар образларни англаш, ижтимоий ва техника соҳасидаги ташхис тизимларини яратишда, технологик жараёнлар, саноат объектларини бошқариш каби техник ва саноат иловаларини яратишда кенг қўламда қўлланиши билан изоҳланади.

Тадқиқотнинг амалий аҳамияти, ишлаб чиқилган алгоритмлар ва дастурий таъминотни турли заиф формаллашган предмет соҳалар – тиббиёт, фармакология, социология масалаларини ечишда ҳамда таълим жараёнига қўлланиши билан изоҳланади. Умумлашган баҳоларни ҳисоблаш усулларини қўллаш объект, ҳодиса ва жараёнлар бўйича яширинган қонуниятларни аниқлашга сарфланадиган ресурсларни қисқартириш имконини беради.

Тадқиқот натижаларининг жорий қилиниши. Умумлашган баҳоларни ҳисоблаш усулларига асосланган берилганлар базасидан яширинган қонуниятларни ажратиб олиш усуллари ва алгоритмлари амалга оширилган дастурий таъминоти бўйича олинган натижалар:

микдорий аломатлар қийматларини интервалларга бўлиш ва умумлашган баҳоларни ҳисоблаш усуллари, умумлашган баҳолар асосида шовқин объектларни ажратиб олиш мезонлари, ҳамда экспертлар томонидан

кўрсатилган ёки аниқланган аломатлар тўпلام остиси бўйича қарор қабул қилиш жараёнини қатъиймас мантқиқ терминларида изоҳлаш услубияти «Ўзбекинвест» компаниясига жорий қилинган («Ўзбекинвест» экспорт-импорт миллий суғурта компаниясининг 2018 йил 8 июндаги 132-сон маълумотномаси). Илмий тадқиқот натижалари суғуртага олинадиган таваккалчиликни баҳолаш, қайтадан суғурталашга бериладиган таваккалчилик қисмини кейинги баҳоси ва суғурта тарифлари ўлчамини аниқлаш ҳамда жисмоний ва юридик шахслар билан суғурта шартномасини тузишдаги санаб ўтилган ҳолатларни рўй беришини прогноз қилишда ишлатилган берилганларга ишлов бериш вақтини 15%га камайтириш имконини берган;

бошланғич аломатларнинг иерархик агрегатив гуруҳлаш қоидаси ва умумлашган баҳолаш асосида шовқин объектларни ажратиш усуллари Тошкент давлат стоматология институтига жорий қилинган (Ўзбекистон Республикаси Соғлиқни сақлаш вазирлигининг 2019 йил 2 октябрдаги 8н-4.323-сон маълумотномаси). Натижада клиник берилганлар базасига ишлов беришда аниқланган ички қонуниятлар асосида экспертлар таҳлилидан ечимга олиб келмайдиган йўллари қирқиб ташлаш орқали қарор қабул қилиш жараёнини 10-15%га қисқартириш имконини берган;

мумкин бўлган объектларни тавсифловчи миқдорий аломатларнинг иккита синфдан бирининг вакиллари устунлик қиладиган интервалларга бўлишнинг турғунлигини баҳолаш усули Ўзбекистон Республикаси Монополияга қарши кураш қўмитасига жорий қилинган (Ўзбекистон Республикаси монополияга қарши кураш қўмитасининг 2019 йил 25 октябрдаги 1443/04-22-сон маълумотномаси). Натижада ўтказилган биржа битимлари базасидан аниқланган ички қонуниятлар экспертлар томонидан қарор қабул қилиш жараёнини 10-15%га қисқаришига олиб келган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқотнинг назарий ва амалий натижалари 13 та халқаро ва 7 та республика илмий-амалий анжуманларида маъруза қилинган ва муҳокамадан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Диссертация мавзуси бўйича жами 41 та илмий иш эълон қилинган бўлиб, жумладан, 1 та монография, Ўзбекистон Республикаси Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижалари чоп этишга тавсия этилган илмий нашрларда 16 илмий мақолалар, 10 таси республика ва 6 таси хорижий журналларда нашр этилган ҳамда Ўзбекистон Республикаси интеллектуал мулк агентлигининг 2 та ихтиро учун патент ва 2 та ЭҲМ учун яратилган дастурий воситаларни қайд қилиш гувоҳномалари олинган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, бешта боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловалардан иборат. Диссертация ҳажми 171 саҳифадан ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида диссертация мавзусининг долзарблиги ва зарурлиги асосланган, мақсад ва вазифалар шакллантирилган, тадқиқот объекти ва предмети белгилаб берилган, тадқиқотнинг Ўзбекистон Республикаси фан ва технологияларининг устувор йўналишларига мослиги кўрсатилган, олинган натижаларнинг назарий ва амалий аҳамияти, ишончлилиги асосланган, тадқиқот натижаларининг амалиётга тадбиқларининг рўйхати, ишнинг апробация натижалари, нашр этилган ишлар ва диссертация тузилмаси тўғрисидаги маълумотлар келтирилган.

Диссертациянинг **«Берилганларни интеллектуал таҳлиliga асосланган инфор­мацион моделлар»** деб номланган биринчи бобда заиф формаллашган предмет соҳалар масалаларини ечишга мўлжалланган билимларга асосланган инфор­мацион моделларида, берилганлар базалари ва сақлагичларидан маълумотлар, билим кўринишидаги эксперт маълумотлари, жараёнлар ва ҳодисаларни математик белгилашлар терминларида ёзиш талаб қилиниши таъкидланади. Ушбу модел доирасида билим манбаи ҳисобланган яширинган қонуниятларни излаш амалга оширилади. Эксперт қарорларни қабул қилиш учун баҳолаш усуллари тавсифланган. Умумлашган баҳо­лар ошкор равишда ўлчанмайдиган латент аломатлар ҳисобланади ва объектлар тавсифидаги янги аломатлар фазосини танлашда фойдаланилади. Тавсифланишида латент аломатлар ишлатилган объектлар учун мумкин бўлган объект тушунчасининг ўринсизлиги исбот қилинган. Латент аломатлар қийматлари сифатида кўрсатилган ёки аниқланувчи аломатлар тўплам ости бўйича объектлар тавсифини сон ўқи­га аксланишидаги қийматларни олиш тавсия қилинади. Акслантиришнинг чизиқли ва чизиқли бўлмаган усулларини танлашга боғлиқ равишда латент аломатлар фазосидаги синфлар объектлари ўртасидаги муносабатлар тузилмаларининг ҳам ўзгариб туриши кўрсатилган. Чизиқли ва чизиқсиз усуллар билан ҳисобланадиган умумлашган баҳо­лар билан баҳо­ларни ҳисоблаш алгоритмлар моделлари ўртасида қиёсий таҳлили ўтказилган. Ҳисоблаш натижаларини изоҳлаш мақсадида билимларни вербаллаштириш учун тегишлилик функцияси қийматларини ва умумлашган баҳо­ларни лингвистик ўзгарувчиларга бириктириш технологияси таклиф қилинади.

Диссертациянинг **«Берилганларни таҳлил қилишнинг интервал усуллари»** деб номланган иккинчи бобда мантиқий қонуниятлар тақдим этиш шакллари ва уларни аниқлаш усуллари тавсифлари келтирилади. Берилганлар базасида қонуниятларни излаш учун миқдорий аломатлар қийматларини интервалларга бўлишдан фойдаланиш таклиф қилинган.

Образларни англашнинг стандарт қўйилган масаласи қаралади. Объектлар тўплами $E_0 = \{S_1, \dots, S_m\}$ берилган бўлиб, улар ўзаро кесишмайдиган $K_1, \dots, K_l, l > 1$ синфларга бўлинган. Объектлар n та турли тоифадаги $X(x_1, \dots, x_n)$ аломатлар билан тавсифланган бўлиб, уларнинг ξ таси интервал, $n - \xi$ таси номинал шкалаларда ўлчанади. Турли тоифадаги

аломатлар тўплам остисидан миқдорий латент аломатларни шакллантиришга имкон берувчи алмаштиришлар мавжуд деб ҳисобланади.

Миқдорий аломатлар қийматларини ўзаро кесишмайдиган интервалларга бўлишнинг иккита шарти берилган:

1. Интерваллар сони синфлар сонига тенг;
2. Интерваллар сони синфлар сонидан катта ёки тенг.

Биринчи ва иккинчи шартлар бўйича интерваллар чегараларининг қийматларини ҳисоблаш талаб қилинади.

Айтайлик, $u_1^1, u_1^2 (u_2^1, u_2^2)$ – мос равишда $[c_0, c_1]$ ва $(c_1, c_2]$ интерваллардаги $K_i, i=1,2$ синф вакилларининг x_j аломат бўйича қийматларининг миқдори, $p - E_0$ объектларининг x_j аломат бўйича қийматларининг камаймайдиган кўринишда тартибланган

$$r_{j_1}, \dots, r_{j_p}, \dots, r_{j_m} \quad (1)$$

кетма-кетлигидаги $c_0 = r_{j_1}, c_1 = r_{j_p}, c_2 = r_{j_m}$ кўринишидаги интерваллар чегараларини аниқловчи элементнинг тартиб номери.

Қуйидаги

$$\left(\frac{\sum_{p=1}^2 \sum_{i=1}^2 u_i^p (u_i^p - 1)}{\sum_{t=1}^2 |K_t| (|K_t| - 1)} \right) \left(\frac{\sum_{p=1}^2 u_1^p (|K_2| - u_2^p) + u_2^p (|K_1| - u_1^p)}{2|K_1||K_2|} \right) \rightarrow \max_{c_0 < c_1 < c_2} \quad (2)$$

мезон, $[c_0, c_1], (c_1, c_2]$ интерваллар оралиғидаги чегаранинг оптимал қийматини ҳисоблаш ва миқдорий аломатнинг номинал шкала ўлчамидаги градацияларини аниқлашда қўллаш имкониятини беради. Мезоннинг чап қавс ичидаги ифода синф ичидаги ўхшашликни, ўнг қавсдагиси – синфлар ўртасидаги фарқланишни билдиради.

Иккинчи шарт учун (1) бўйича

$$\left| \frac{d_t^i(u, v)}{|K_t|} - \frac{d_{3-t}^i(u, v)}{|K_{3-t}|} \right| \rightarrow \max \quad (3)$$

мезон асосида шундай кесишмайдиган интерваллар қуриш талаб этиладики, уларнинг чегараларида $K_t, t \in \{1, 2\}$ синф вакиллари қийматларининг миқдори устунлик қилсин. Бу ерда $d_t^i(u, v) - [r_{c_u}, r_{c_v}]^i$ интервалдаги K_t синф вакилларининг миқдори.

Миқдорий c – аломатнинг $[r_{c_u}, r_{c_v}]^i$ интервал бўйича K_1 синфга тегишлилик функцияси қийматини

$$f_{ci} = \frac{\eta_{1i}}{\eta_{1i} + \eta_{2i}} \quad (4)$$

кўринишида аниқлаймиз. Бу ерда $\eta_{ti} = \frac{d_t^i(u, v)}{|K_t|}, t \in \{1, 2\}$.

Барча интерваллар бўйича турғунлик $(0.5,1]$ ораликда ётади ва қуйидаги

$$U(c) = \frac{1}{m} \sum_{\{[r_{c_u}, r_{c_v}]\}} \begin{cases} f_{c_i}(v-u+1), & f_{c_i} > 0.5, \\ (1-f_{c_i})(v-u+1), & f_{c_i} < 0.5, \end{cases} \quad (5)$$

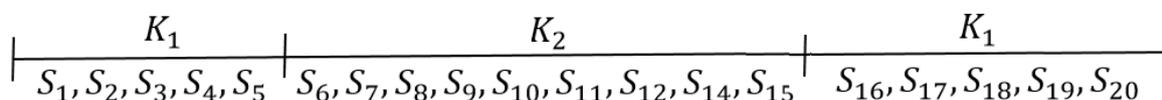
формула билан ҳисобланади.

Аломатлар қийматлари тўпламида (3) бўйича интервалларга бўлиш мумкин бўлмаслик шартлари аниқланган.

Фараз қилайлик E_0 ўргатувчи танланмада $|K_1| = \alpha|K_2|$, $\alpha = \frac{c}{d}$, $c, d \in \{1, 2, \dots\}$ тенглик бажарилсин. E_0 танланма объектларининг тавсифланишида шундай миқдорий аломат мумжуд бўлиб, у p та турли қийматлар билан берилган бўлсин $p(2 \leq p \leq m)$. Аломатнинг K_1 ва K_2 синфдаги бир хил қийматлар сонларини мос равишда a_1, \dots, a_p ва b_1, \dots, b_p

орқали белгилайлик. Агар $\frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_p}{b_p} = \alpha$ шарт бажарилса, у ҳолда мумкин

бўлган интерваллар сони 2 дан p гача ва ихтиёрий интерваллар бўйича объектнинг K_1 ёки K_2 синфларга тегишлилик функциясининг қиймати 0.5 тенг бўлади, яъни интервалда бирорта синф вакилларининг устунлигига эришиб бўлмайди. 1-расмда аломат қийматларининг биринчи ва иккинчи шарт бўйича интервалларга бўлиниши тест мисолида кўрсатилган.



1-расм. Тест мисол.

Аломат қийматларини $[1,5]$, $(5,20]$ интервалларга бўлишдаги (1) мезон оптимал қиймати 0.3611 тенг бўлиб, аломат бўйича синфлар объектларининг бўлиниши (компактлиги) нисбатан ёмон эканлигини кўрсатади. Бунга қарама-қарши равишда (3) мезоннинг ҳар бир интервал бўйича қиймати 1 тенг бўлган 3 та интервалга бўлиниши аломат қийматлари бўйича яхши шартланган кластерлар борлигини намоён этади.

Диссертациянинг «Умумлашган баҳоларни ҳисоблаш усуллари» деб номланган учинчи бобида объектлар ўртасидаги муносабатларни сон ўқиға акслантириш учун умумлашган баҳоларни ҳисоблаш усуллари қаралади. Объектларнинг умумлашган баҳоларининг қийматлари икки синфли англаш масаласи доирасида ҳисобланади. Синф объектлари қуйидаги оппозициялар шаклида тақдим этилади – “*бемор-деярли соғ*”, “*меҳнатсевар-дангаса*”, “*бой-камбагал*”, “*аълочи-қарздор*”, “*қўли очиқ-зиқна*” ва ш.к. Умумлашган баҳоларни ҳисоблашнинг иккита усули таклиф этилган: стохастик ва детерминистик.

Стохастик усулда $S_a \in E_0, S_a = (x_{a1}, \dots, x_{an})$ объектнинг умумлашган баҳоси

$$R(S_a) = \sum_{i \in I} w_i t_i (x_i - c_1^i) / (c_2^i - c_0^i) + \sum_{i \in I} \mu_i(x_{ai})$$

кўринишида ҳисобланади. Бу ерда $[c_0^i, c_1^i][c_1^i, c_2^i], w_i$ қийматлари (2) бўйича аниқланади, $T = (t_1, \dots, t_n)$ векторнинг $t_i \in \{-1, 1\}$ қийматлари эса

$$\min_{S_p \in K_1} R(S_p) - \max_{S_p \in K_2} R(S_p) \rightarrow \max$$

шарти билан ҳисобланади.

Номинал $r \in J$ аломат градациялари сонини p орқали, K_d синф объектлари тавсифидаги r -аломатнинг t -градациясининг ($1 \leq t \leq p$) қийматлари миқдорини g_{dr}^t орқали, K_d синфдаги r -аломат градациялари миқдорини l_{dr} орқали белгилайлик. K_1 ва K_2 синфлар ўртасидаги r -аломат бўйича фарқланиш

$$\lambda_r = 1 - \frac{\sum_{t=1}^p g_{1r}^t g_{2r}^t}{|K_1| |K_2|}. \quad (6)$$

катталиқ сифатида аниқланади.

Номинал r -аломатнинг K_1 ва K_2 синфлар бўйича бир жинслик даражаси (синф ичидаги ўхшашлик ўлчови) β_r қуйидаги формулалар билан ҳисобланади

$$D_{dr} = \begin{cases} (|K_d| - l_{dr} + 1)(|K_d| - l_{dr}), & p > 2, \\ |K_d|(|K_d| - 1), & p \leq 2; \end{cases}$$

$$\beta_r = \begin{cases} \frac{\sum_{t=1}^p g_{1r}^t (g_{1r}^t - 1) + g_{2r}^t (g_{2r}^t - 1)}{D_{1r} + D_{2r}}, & D_{1r} + D_{2r} > 0, \\ 0, & D_{1r} + D_{2r} = 0. \end{cases} \quad (7)$$

Юқорида келтирилган (6) ва (7) ёрдамида номинал $r \in J$ аломат вази

$$v_r = \lambda_r \beta_r \quad (8)$$

кўринишида ҳисобланади.

Кўришиб турибдики, номинал аломатнинг p та градацияларини идентификация қилувчи сонлар тўпламини $\{1, \dots, p\}$ тўпламига акслантириш мумкин. Шундай акслантиришни ҳисобга олган ҳолда $S = (x_1, \dots, x_n)$ объект учун $x_i = j, i \in J, j \in \{1, \dots, p\}$ аломатнинг умумлашган баҳога қўшган ҳиссаси

$$\mu_i(j) = v_i \left(\frac{\alpha_{ij}^1}{|K_1|} - \frac{\alpha_{ij}^2}{|K_2|} \right), \quad (9)$$

катталиқ билан аниқланади. Бу ерда $\alpha_{ij}^1, \alpha_{ij}^2$ – мос равишда K_1 ва K_2 синфлардаги i -аломатнинг j -градацияси қийматларининг сони, v_i – i -аломатнинг (8) бўйича аниқланадиган вазни.

Детерминистик усулда $S \in E_0 \cap K_d$, $S = (b_1, \dots, b_n)$ объектнинг умумлашган баҳоси

$$R(S) = \frac{1}{|K_{3-d}|} \sum_{S_j \in K_{3-d}} \left(\sum_{c \in I} \left\{ \begin{array}{l} f_{ci}, b_c \in [r_{cu}, r_{cv}]^i \text{ ва } x_{jc} \notin [r_{cu}, r_{cv}]^i, \\ \frac{f_{ci} |b_c - x_{jc}|}{|r_{cu} - r_{cv}|}, r_{cu} \neq r_{cv} \\ 0, \quad r_{cu} = r_{cv} \end{array} \right\}, b_c, x_{jc} \in [r_{cu}, r_{cv}]^i + \sum_{c \in J} \left\{ \begin{array}{l} f_{ci}, b_c \neq x_{jc}, \\ 0, b_c = x_{jc}, \end{array} \right\} \right) \quad (10)$$

формула билан ҳисобланади. Бу ерда ҳар бир номинал $c \in J$ аломатнинг τ_c градациялари $\{1, 2, \dots, \tau_c\}$ тўпламга тегишли, $[r_{cu}, r_{cv}]^i$ – $c (c \in I)$ -аломатнинг i -устунлик интервали, f_{ci} – c -аломатнинг (4) бўйича ҳисобланадиган тегишлилик функциясининг қиймати. Агар $c \in J$ бўлса, η_{1i}, η_{2i} қийматлари i -градациянинг мос равишда K_1 ва K_2 синфлардаги қийматлари миқдори деб қаралади.

Умумлашган баҳолар асосида қарор қабул қилиш дараҳтини қуриш усули таклиф этилган бўлиб, унда информатив аломатлар жуфтлиги (латент аломат) бўйича умумлашган баҳо қийматлари лексографик тартибдаги *if-then* қарор қабул қоидалар кетма-кетлиги шаклланишида қўлланиладиган дарахт тугунларидаги классификаторларни шакллантиришда ишлатилади. Ушбу усул тадқиқ қилинаётган предмет соҳа ҳақидаги янги, олдин маълум бўлмаган билимларни ажратиб олишдаги берилганларнинг интеллектуал таҳлил қилиш воситаси сифатида зарурдир.

Диссертациянинг «**Берилганлар базасидан яширинган қонуниятларни аниқлаш**» деб номланган тўртинчи бобида мантиқий қонуниятлар ёрдамида берилганлар базасидан билимларни ажратиб олиш усуллари қаралган. Мантиқий қонуниятлар таҳлили, улар турли тоифадаги аломатлар фазосидаги турғунлиги асосида амалга оширилади.

Синфлар объектларининг турғунлигини ҳисоблаш жараёни миқдорий аломатлар ўлчов масштабларига нисбатан инвариант бўлиши кўрсатилган.

Фармакологик воситаларни экспертиза қилиш мисолида турли тоифадаги аломатлар кўрсатилган ёки аниқланувчи тўплам остилари бўйича ҳисобланган умумлашган баҳоларни қатъиймас мантиқ терминларида изоҳлаш таклиф этилган.

Иккита K_1 ва K_2 врачлар гуруҳларининг, мос равишда аллергологлар ($|K_1| = 23$) ҳамда пульмонолог ва терапевтларнинг ($|K_2| = 68$) дори

воситаларининг кўрсатилган 3 та – χ^1 (Глюкокортикоид воситалар), χ^2 (Бронхолистик воситалар) ва χ^3 (Антибиотиклар) тўплам остилари бўйича бронхиал астма касалини даволаш учун қўллаш бўйича берган хулосалари ўзаро таққосланган. Дориларнинг χ^1, χ^2 ва χ^3 тўплам остилари бўйича (10) формула билан ҳисобланган умумлашган баҳоларнинг (y^1, y^2, y^3 аломатлар) қийматлар тўплами (2) мезон бўйича таҳлил қилинди. Таҳлил натижаларини изоҳлаш учун лингвистик ўзгарувчининг 6 та қийматларидан фойдаланилди.

1-жадвал.

Аломатларнинг 3 та тўплам остиси бўйича ҳисоблаш эксперименти натижалари

| Аломатлар тўплам остилари | Умумлашган баҳоларнинг (2) бўйича қиймати | Лингвистик ўзгарувчи қиймати |
|---------------------------|---|------------------------------|
| χ^1 | 0.55 | Деярли мос келмайди |
| χ^2 | 0.51 | Деярли мос келмайди |
| χ^3 | 0.36 | Сезиларли мос келади |

Кўйилган масала моҳиятини ифодаловчи « K_1 ва K_2 гуруҳ врачларининг χ^1, χ^2 ва χ^3 тўплам остиларига бўлинган дори воситаларининг бронхиал астма касалини даволашда қўллаш бўйича фикрлари сезиларли даражада мос келади» гипотезасининг ростлигини текшириш натижалари 1-жадвалда келтирилган. Кўриниб турибдики, гипотезанинг ростлиги фақат χ^3 тўплам остисида тасдиқланади.

Таклиф қилинган иерархик агломератив қоидалари асосида E_0 объектлари тавсифи бўйича аломатларни ўзаро кесишмайдиган гуруҳларга бўлиш амалга оширилади. Алгоритм қоидалари бўйича талаб қилинади:

– E_0 объектлари турғунлиги қийматлари бўйича турли тоифадаги аломатларнинг ўзаро кесишмайдиган гуруҳларини шакллантириш;

– ўзаро кесишмайдиган гуруҳлар бўйича информатив аломатлар тўплам остисини аниқлаш.

Объектлар турғунлигини ҳисоблаш учун микдорий аломатлар қийматларини $[0..1]$ ораликқа нормалашда фойдаланилади.

Ўргатувчи E_0 танланманинг $S_a = (x_{a1}, \dots, x_{an})$ ва $S_b = (x_{b1}, \dots, x_{bn})$ объектлари ўртасидаги яқинлик ўлчови Журавлев метрикаси

$$\rho(S_a, S_b) = \sum_{i \in I} |x_{ai} - x_{bi}| + \sum_{i \in J} \begin{cases} 1, & x_{ai} \neq x_{bi}, \\ 0, & x_{ai} = x_{bi} \end{cases} \quad (11)$$

бўйича ҳисобланади.

Аломатларнинг $X(k) \subset X(n), k \leq n$ тўплам остиси учун (11) бўйича ҳосил қилинган $S_{d_0}, S_{d_1}, \dots, S_{d_{m-1}}, S_{d_0} = S_{d_1}$ камаймайдиган кетма-кетликдаги интервалнинг c_2 чегараси (2) ёрдамида аниқланади.

Курилган интервал учун

$$\lambda_1(p) = \left| \left\{ S_{d_i} \in K_p \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|,$$

$$\lambda_2(p) = \left| \left\{ S_{d_i} \in K_{3-p} \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|,$$

$$\theta_1(p) = \lambda_1(p) / |K_p| \text{ ва } \theta_2(p) = \lambda_2(p) / |K_{3-p}|$$

аниқланган бўлсин. У ҳолда аломатларнинг $X(k)$ тўплам остиси бўйича $S_d \in K_p, p = 1, 2$ объект турғунлиги

$$U(S_d, X(k)) = \theta_1(p)(1 - \theta_2(p)) \quad (12)$$

кўринишда ҳисобланади.

Қийматлари

$$b_{ij} = \begin{cases} \frac{\sum_{a=1}^m \sum_{b=1}^m \alpha(S_a, S_b) \beta(i, j, S_a, S_b)}{\sum_{p=1}^2 |K_p| (m - |K_p|)}, & i \neq j, \\ 0, & \text{акс холда} \end{cases}$$

формула билан ҳисобланадиган $B = \{b_{ij}\}_1^n$ яқинлик матричасини шакллантирайлик. Бу ерда

$$\beta(i, j, S_a, S_b) = \begin{cases} \rho(S_a, S_b), \{x_i, x_j\} \subset X(n), S_a, S_b \in E_0, \\ 0, & \text{акс холда,} \end{cases}$$

$$\alpha(S_a, S_b) = \begin{cases} 0, & S_a, S_b \in K_i, i = 1, 2, \\ 1, & S_a \in K_i, S_b \in K_{3-i}. \end{cases}$$

Аломатларнинг $X(k)$ тўплам остисидан $X(k+1), 1 \leq k \leq n$, $X(k) \subset X(k+1)$ тўплам остисига ўтишдаги E_0 объектларининг турғунлик баҳоларини ўзгариши

$$M(X(k), X(k+1)) = \frac{1}{m} \sum_{d=1}^m \begin{cases} 1, & U(S_d, X(k)) \leq U(S_d, X(k+1)), \\ 0, & \text{акс холда} \end{cases} \quad (13)$$

кўринишда ҳисобланади.

Агар (13) бўйича $G_p \cup \{x_d\}$ қиймати G_p гуруҳ қийматидан катта бўлса, $x_d \in X(n)$ аломат $G_p = X(k), p < n$ гуруҳ таркибига қўшилади.

Латент аломатлар тўплами бўйича қонуниятларни излаш усули таклиф қилинган. Танланма объектларини тавсифлаш учун латент аломатлар бошланғич аломатлардан арифметик амалларни бажариш орқали ёки

аломатлар тўплам ости бўйича умумлашган баҳолар қийматларини ҳисоблаш билан олиниши мумкин.

Эксперимент жараёнида аниқланган билимларни қатъиймас мантик формализми ва лингвистик ўзгарувчилар ёрдамида S объектнинг $K_l, l \in \{1, 2\}$ синфга тегишлиги ҳақидаги

Агар A у ҳолда B (<ишончлик коэффициентини>)

қатъиймас хулоса қоида кўринишида тавсифлаш мумкин. Бу ерда <ишончлик коэффициентини> – A шарт бажарилганда B хулосага ишончилилик коэффициенти.

Объектлар (пациентлар) сони 73 тадан иборат $E_0 = \{S_j\}_1^{73}$ танланма тадқиқ қилинди. Объектлар 31 та миқдорий аломатлар $X = (x_1, \dots, x_{31})$ билан тавсифланган. Эксперт-врачлар томонидан танланманинг 48 та объекти K_1 синфга – барқарор атеросклеротик бляшкали (АСБ) пациентларга, 25 таси K_2 синфга (бекарор АСБ) ажратилди.

2-жадвал.

Аломатлар интервалларининг (2) мезон бўйича чегаралари

| Аломатлар номи | Интерваллар ва улардаги синфлар вакиллари $[r_{i_1}, r_{i_p}] (r_{i_p}, r_{i_{73}}) \{u_{11}, u_{12}\} \{u_{21}, u_{22}\}$ | (2) мезон қиймати |
|--|---|-------------------|
| <i>Металлопротеиназ тўқима ингибитори 1 (ТИМП-1)</i> | [775.55, 1602.88](1602.88, 3677.22) {42,3} {3,25} | 0.717 |
| <i>Матрицали металлопротеиназ 9 (ММП-9)</i> | [28.1, 274.96](274.96, 1241.2) {42,4} {3,24} | 0.678 |
| <i>Чап қоринчанинг систолик чегаравий ўлчами (КСР)</i> | [2,3.1](3.1, 5.35) {4,24} {41,4} | 0.643 |
| <i>Каротид артериянинг стеноз даражаси (СССА)</i> | [0,55](55, 100) {37,1} {8,27} | 0.611 |
| <i>Ёш</i> | [46, 62](62, 88) {12,14} {33,14} | 0.284 |
| <i>Оғирлик</i> | [57, 72](72, 114) {14,11} {31,17} | 0.251 |
| <i>Сочдаги рух миқдори</i> | [29.74, 214.57] (214.57, 482.45) {21,17} {24,11} | 0.251 |
| <i>Диастолик артериал босим (ДАД)</i> | [60,80](80,120) {34,20} {11,8} | 0.238 |
| ... | | |

Тадқиқот учун (2) ва (3) бўйича бошланғич ва латент аломатларни ўзаро кесишмайдиган интервалларга бўлишнинг иккита усули ишлатилди. E_0 танланма объектлари тавсифидаги миқдорий аломатлар қийматларини (2) мезон экстремумида иккита интервалларга бўлиш натижалари 2-жадвалда келтирилган бўлиб, ундаги u_{ij} – K_j синф вакиллариининг i -интервалдаги сони, $i, j \in \{1, 2\}$.

Бир нечта аломатлар қийматлари бўйича (2-жадвалга қаранг) мумкин бўлган S объектнинг (пациентнинг) K_1 синфга (барқарор АСБ) тегишли эканлигининг қатъиймас хулоса қоидаларининг тўплам остиларидан бири қуйидаги кўринишга эга:

$$P_1: \text{Агар } ТИМП-1 \leq 1602.88 \text{ у ҳолда } S \in K_1 (0.717);$$

$$P_2: \text{Агар } ММП-9 \leq 274.96 \text{ у ҳолда } S \in K_1 (0.678);$$

$$P_3: \text{Агар } КСР \leq 31 \text{ у ҳолда } S \in K_1 (0.643).$$

Устунлик интерваллари қийматлари бўйича мумкин бўлган S объектнинг (пациентнинг) K_1 синфга (барқарор АСБ) тегишли эканлигининг қатъиймас хулоса қоидалари қуйидаги кўринишга эга:

$$P_1: \text{Агар } 775.55 \leq ТИМП-1 \leq 1602.88 \text{ у ҳолда } S \in K_1 (0.9);$$

$$P_2: \text{Агар } 65 \leq \text{Вазн} \leq 72 \text{ у ҳолда } S \in K_1 (0.35);$$

$$P_3: \text{Агар } 0 \leq СССА \leq 55 \text{ у ҳолда } S \in K_1 (0.96);$$

$$P_4: \text{Агар } 5.21 \leq КДР \leq 6.49 \text{ у ҳолда } S \in K_1 (1).$$

Агар хулоса қилиш жараёнида P_1 ва P_2 қоидаларнинг шарт қисмлари бажарилса, у ҳолда $S \in K_1$ ҳодисасига ишонч ўлчами қуйидаги кўринишни олади

$$\text{МД}[\langle S \in K_1 \rangle: 775.55 \leq ТИМП-1 \leq 1602.88, 65 \leq \text{Вазн} \leq 72] = \text{МД}[\langle S \in K_1 \rangle: 775.55 \leq ТИМП-1 \leq 1602.88] + \text{МД}[\langle S \in K_1 \rangle: 65 \leq \text{Вазн} \leq 72](1 - \text{МД}[\langle S \in K_1 \rangle: 775.55 \leq ТИМП-1 \leq 1602.88]).$$

Қоидаларга ишончлик коэффициентларини қўйган ҳолда ҳисоблаймиз

$$\text{МД}[\langle S \in K_1 \rangle: 775.55 \leq ТИМП-1 \leq 1602.88, 65 \leq \text{Вазн} \leq 72] = 0.9 + 0.35(1 - 0.9) = 0.935 .$$

Агар ҳодиса фойдасига гувоҳликлар иккитадан ортиқ бўлса, ҳодисага ишончлик ўлчами ҳисоби каскад шаклда амалга оширилади.

Аломатлар иккитадан ортиқ интервалларга бўлиниши қоидалар сонини ошишига олиб келади. Иккинчи томондан, бундай қоидаларни ишлатиш қабул қилинган қарорга ишонч коэффициентини кўтарилишига олиб келади. Қабул қилинган қарорнинг ишончлиги, ҳаттоки объект тавсифида айрим аломатларнинг қийматлари номаълум бўлган ҳолатларда ҳам юқори бўлиб қолади. Масалан, объектда (пациентда) *ТИМП-1*, *ММП-9* каби информатив аломатлар қийматлари йўқ бўлган ҳолда *ИМТ*, *Вазн*, *Сочдаги рух миқдори* ва *ДАД* бўйича мақбул ишонч коэффициентли хулосани қилиш мумкин:

$$\text{МД}[\langle S \in K_1 \rangle: \text{ИМТ} > 25.8, \text{Вазн} > 72, \text{Сочдаги рух миқдори} > 214.57, \text{ДАД} < 80] = 0.68 .$$

Қарор қабул қилиш дарахтларидан фарқли равишда қоидаларни шакллантиришда аломатларнинг боғлиқмас бўлиши талаб қилинмайди.

Диссертациянинг «**Информацион моделларни қуриш**» деб номланган бешинчи бобида берилганлар базасидан қонуниятларни ажратиб олиш

мақсадида берилганларни интеллектуал таҳлил қилиш алгоритмларини амалга оширувчи компьютер иловасининг (тизимнинг) тавсифи келтирилган. Тизим, ҳар бири алоҳида масала остини ечадиган ўзаро боғланган модуллар мажмуасидан иборат бўлиб, модулларнинг бир қисми визуал компоненталардан фойдаланган ҳолда амалга оширилган.

Ўргатувчи танланма селекциясини амалга ошириш учун <http://archive.ics.uci.edu/ml/machine-learning-databases> сайт репозиториясидан олинган берилганларда ҳисоблаш экспериментлар ўтказилди. Селекцияда синфлар конфигурациясини ажратувчи сиртлар билан адаптация қилиш ёки синфлар компактлик ўлчамини ошириш мақсадида шовқин объектларни излаб топиш ва уларни ўчириш амалга оширилди.

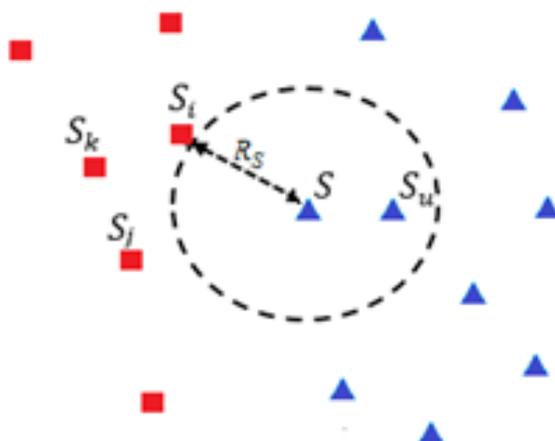
Умумлашган баҳоларни ҳисоблаш усуллари (стохастик ва детерминистик) ёрдамида латент аломатлардан ташкил топган R^2 фазони шакллантириш ва ушбу фазода танланманинг шовқин объектларини аниқлаш йўллари таклиф қилинган. R^2 фазода қўлланиладиган усулларнинг инвариантлиги танланма шовқин объектлар тўпламини бир қийматли равишда аниқлаш имконини беради. Шовқин объектлар тўплами берилган метрика бўйича чегаравий объектларнинг тўплам остиси сифатида қаралади. Чегаравий объектлар тўплами $B \in E_0$ қуйидагича аниқланади:

$$B = \left\{ S \in E_0 \mid \rho(S_i, S) = \min_{S_i \in K_t, S_j \in CK_t} \rho(S_i, S_j) \right\}.$$

Танланманинг $S \in B \cap K_t, t=1,2$ объекти K_t синфнинг N_t – шовқин объектлар тўплами тегишли бўлади, агар

$$\left| \left\{ S_i \in E_0 \mid \rho(S_i, S) = \min_{S_i \in CK_t, S_j \in K_t} \rho(S_i, S_j) \right\} \right| > \left| \left\{ S_i \in K_t \mid \rho(S_i, S) < \min_{S_i \in K_t, S_j \in CK_t} \rho(S_i, S_j) \right\} \right| \quad (14)$$

шарт бажарилса.

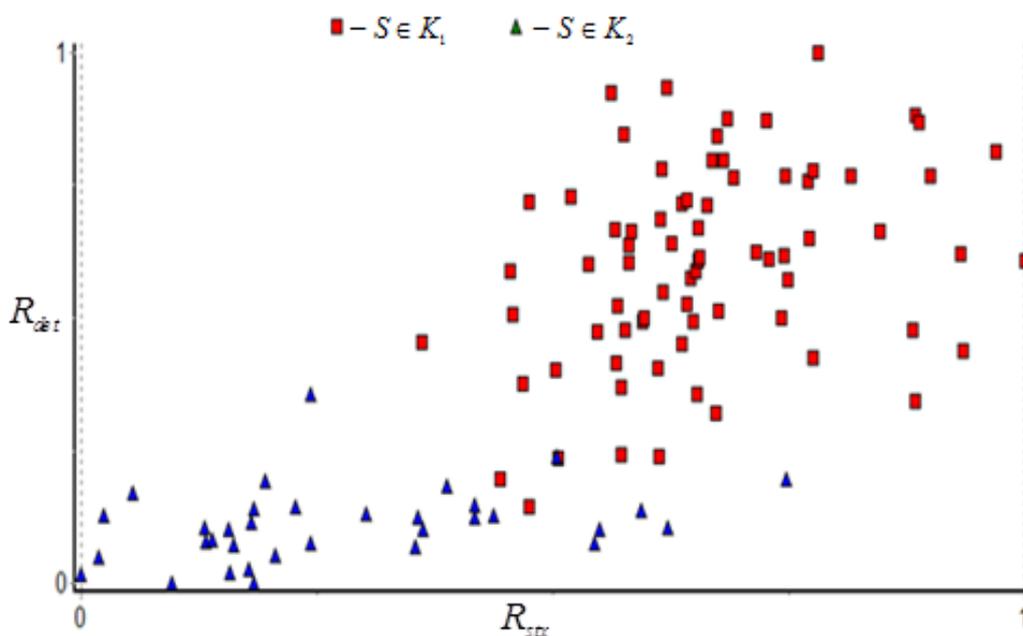


2-расм. Танланмадаги шовқин объектни ажратиб олишнинг геометрик кўриниши.

Танланмадаги шовқин объеклари жойлашувининг геометрик интерпретацияси 2-расмда кўрсатилган бўлиб, унда K_1 ва K_2 синфлар объеклари мос равишда ‘■’ ва ‘▲’ белгилари билан белгиланган. Масалан, $S \in K_2$ объекти учта $S_i, S_j, S_k \in K_1$ объектлар учун чегаравий бўлган ҳолда, S марказли ва R_S радиуси гипершарда эса K_2 синфнинг фақат битта S_u объекти борлиги сабабли у шовқин объект ҳисобланади.

Мазкур ишда қаралаётган умумлашган баҳоларни ҳисоблаш усуллариининг ўлчов масштабига нисбатан инвариантлиги сабабли, латентлар аломатлар фазосида яқинлик ўлчови сифатида эвклид метрикаси танланган.

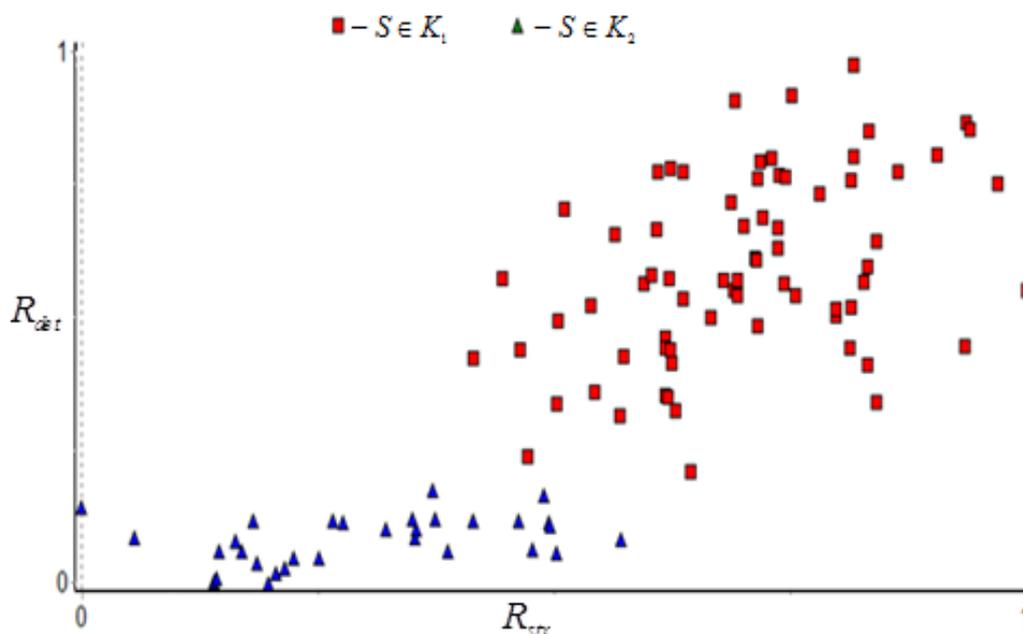
Эксперимент учун юрак хуружи бўлган пациентларни ҳолатини тавсифловчи «Echocardiogram» берилганлар танланмаси олинди. Танланма 108 объектдан иборат бўлиб, уларнинг 74 таси K_1 (1 йил ичида вафот этган пациент) синфга тегишли, 34 таси K_2 (тирик қолган ёки 1 йилдан кейин вафот этган пациент) синфига тегишли. Ҳар бир объект 8 микдорий ва 2 номинал аломатлар билан тавсифланган. E_0 объеклари стохастик ва детерминистик усуллар ёрдамида шакллантирилган $\{R_{stx}, R_{det}\}$ латент аломатлар қийматлари билан R^2 фазода тавсифланган.



3-расм. E_0 танланмалар объекларининг R^2 фазосида жойлашув тузилмаси.

Юқорида келтирилган (14) мезон бўйича шовқин объектлар тўплами $N = \{S_7, S_{38}, S_{71}, S_{78}, S_{101}\}$ ажратилди. Ушбу объектларни E_0 танланмадан ўчириш R^2 фазода синфлар объекларининг компактлиги (ажралганлиги) ошганлигини кўрсатди. Хусусан, R_{stx} аломат учун (2) мезон қиймати E_0 танланмадаги 0.75 дан $E_0 \setminus N$ танланмада 0.79 қийматига, R_{det} аломат учун мос равишда 0.86 дан 1 қийматига катталашди. 3 ва 4 – расмларда E_0 ва $E_0 \setminus N$

учун R^2 фазодаги танланма объектлар жойлашуви тузилмасининг график кўринишлари келтирилган.



4-расм. $E_0 \setminus N$ танланмалар объектларининг R^2 фазосида жойлашув тузилмаси.

4-расмда танланмадан шовқин объектлар ўчирилгандан кейин синфлар объектларини бўлакчи-чирикчи ажратиш имконияти юзага келганлиги кўрсатилган.

ХУЛОСА

«Умумлашган баҳоларни ҳисоблаш усуллари асосида яширин қонуниятларни аниқлаш тизимлари» мавзусидаги диссертация иши бўйича олиб борилган тадқиқотлар натижасида куйидаги хулосалар тақдим этилди:

1. Ўзаро кесишмайдиغان синфлар вакиллари устунлик қиладиган интерваллар бўйича аломатлар турғунлигини ҳисоблаш усули тақлиф қилинди. Усул миқдорий аломат бўйича кучли қонуниятларни аниқлаш имконини берди;

2. Аломатлар қийматларини ўзаро кесишмайдиغان интервалларга бўлишда синфлар вакиллариининг устун бўлиши шarti аниқланди. Бу шарт яширинган қонуниятларни излаш соҳаларини аниқлаш имконини берди;

3. Объектлар умумлашган баҳоларини стохастик ва детерминистик усулларда ҳисоблаш учун икки синфли образларни англаш масаласини ечишга амалий тавсиялар берилди. Тавсияларни қўлланилиши яширин қонуниятларни излаш учун аломатлар фазосини қискартиришга хизмат қилди;

4. Турли тоифадаги аломатлар кўрсатилган тўплам остиси бўйича умумлашган баҳоларни ҳисоблаш асосида қатъиймас тўпламга тегишлилик функциясини ҳисоблаш усули ишлаб чиқилди. Функция қиймати умумлашган баҳоларни интервалларга бўлиш мезонининг экстремуми бўйича аниқланди.

Усул «Синфлар объектларининг умумлашган баҳоларининг қийматлари ўзаро кесишади» мулоҳаза ростлиги даражасини ифодалаш имкониятини яратди;

5. Умумлашган баҳолар асосида латент аломатларни танлаш йўли билан қарорлар дарахти бўйича хулоса қоидаларини шакллантириш таклиф этилди. Усулдаги латент аломатларнинг боғлиқмаслик хусусияти инобатга олинган ҳолда хулоса қоидаларини танлашнинг лексографик тартиби классификация масалаларини ечишда йўналтирилган равишда қарор қабул қилиш имкониятини яратди;

6. Турли тоифадаги аломатларни гуруҳлашнинг иерархик агломератив усули ишлаб чиқилди. Аломатларни гуруҳга қўшиш учун Журавлев метрикаси бўйича ҳисобланган объектлар турғунлигидан фойдаланилди. Аломатларни гуруҳларга ажратиш умумлашган баҳоларни ҳисоблаш учун амалга оширилди. Умумлашган баҳолар объектлар тавсифидаги аломатлар фазосини қисқартириш учун латент аломатлар сифатида хизмат қилди;

7. Умумлашган баҳоларни ҳисоблашнинг стохастик ва детерминистик усуллари асосида шовқин объектларни ажратиш олиш мезонлари ишлаб чиқилди. Шовқин объектларни ажратиш олиш англаш алгоритмларининг умумлаштириш қобилиятини оширишга олиб келди;

8. Умумлашган баҳоларни ҳисоблаш усуллари асосида берилганлардан яширинган қонуниятларни ажратиш олиш ҳисоблаш экспериментларини ўтказиш учун компьютер иловаси яратилди. Компьютер иловаларидан заиф формаллашувчи предмет соҳалар учун инфорацион моделларни қуришда фойдаланилди;

9. Диссертация тадқиқоти доирасида ишлаб чиқилган усуллар, алгоритмлар ва дастурий маҳсулотлар «Ўзбекинвест» компанияси ва Ўзбекистон Республикаси монополияга қарши кураш қўмитаси фаолиятларига ҳамда Тошкент давлат стоматология институти илмий тадқиқотларига жорий қилинди. Берилганлар базаси бўйича аниқланган яширин қонуниятлар соҳа экспертларининг қарор қабул қилиш жараёнини 10-15 % га қисқариш имконини берди.

**РАЗОВЫЙ НАУЧНЫЙ СОВЕТ ПРИ НАУЧНОМ СОВЕТЕ
DSc.27.06.2017.Т.07.01 ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ
ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**

НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ УЗБЕКИСТАНА

МАДРАХИМОВ ШАВКАТ ФАЙЗУЛЛАЕВИЧ

**СИСТЕМЫ ОБНАРУЖЕНИЯ СКРЫТЫХ ЗАКОНОМЕРНОСТЕЙ НА
БАЗЕ МЕТОДОВ ВЫЧИСЛЕНИЯ ОБОБЩЕННЫХ ОЦЕНОК**

05.01.03 -Теоретические основы информатики

**АВТОРЕФЕРАТ ДОКТОРСКОЙ (DSc)
ДИССЕРТАЦИИ ПО ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2020

Тема докторской диссертации по техническим наукам (DSc) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за B2017.1.DSc/T27.

Докторская диссертация выполнена в Национальном университете Узбекистана.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице научного совета (www.tuit.uz) и Информационно-образовательном портале «ziyonet» (www.ziyonet.uz).

Научный консультант: **Игнатъев Николай Александрович**
доктор физико-математических наук, профессор

Официальные оппоненты: **Шарипбай Алтынбек Амирулы**
доктор технических наук, профессор

Гулямов Шухрат Маннопович
доктор технических наук, профессор

Маматов Нарзилла Салиджонович
доктор технических наук, старший научный сотрудник

Ведущая организация: **Казахский национальный университет**

Защита диссертации состоится «__» _____ 2020 г. в __ часов на заседании научного совета DSc.27.06.2017.T.07.01 при Ташкентском университете информационных технологий (Адрес: Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).

С докторской диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер №____). Адрес: Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-44).

Автореферат диссертации разослан «__» _____ 2020 года.
(протокол рассылки №__ от «__» _____ 2019 г.)

Р.Х.Хамдамов
Председатель научного совета по присуждению
учёных степеней, д.т.н., профессор

Ф. М.Нуралиев
Учёный секретарь научного совета по
присуждению учёных степеней, д.т.н., доцент

Ш.Х.Фазилов
Председатель разового научного семинара при
научном совете по присуждению учёных
степеней, д.т.н., профессор.

ВВЕДЕНИЕ (аннотация диссертации доктора наук (DSc))

Актуальность и востребованность темы диссертации. В современном мире в связи с интенсивным развитием компьютерных технологий и их применением в различных сферах человеческой деятельности наблюдаются потребность в качественной (интеллектуальной) обработке данных. Движущей силой инновационного развития человеческого общества является знания. Наряду с усвоением уже имеющихся знаний, актуальным является пополнение их объёмов за счёт открытия новых законов и закономерностей. В развитых странах мира, таких как США, Великобритания, Германия, Франция, Российской федерация, Япония, Китай, Южная Корея, в международных компаниях Amazon, Alibaba, Google, Yandex, Microsoft, SAN активно ведутся исследовательские работы по разработке и применению методов интеллектуального анализа данных в рамках систем искусственного интеллекта, ориентированных на решения задач прогнозирования, регрессии и классификации.

В мире проводится широкомасштабные научные исследования по разработке методов интеллектуального анализа данных для построения интеллектуальных систем, ориентированных на решение большого круга прикладных задач, не имея их аналитического описания и довольствуясь лишь наборами экспериментальных данных. Важнейшими задачами исследований являются разработка методов и алгоритмов восстановления пропусков в базах данных, удаление «шумовых» объектов выборки, предобработка и поиск скрытых закономерностей в базах данных, выбор нового признакового пространства в описании объектов, создание интеллектуального интерфейса с использованием аппарата нечёткой логики.

В нашей Республике идет повсеместное внедрение информационных технологий в социальной и производственной сферах. Для повышения эффективности системы государственного управления реализуется программа электронного правительства. В Стратегии действий по дальнейшему развитию Республики Узбекистан на 2017-2021 годы определены такие задачи, как «... внедрение информационно-коммуникационных технологий в экономику, социальную сферу, системы управления,... развитие и модернизацию дорожно-транспортной, инженерно-коммуникационной и социальной инфраструктуры»¹. При выполнении этих задач особая роль отводится решению проблем, связанных с выявлением скрытых закономерностей из баз данных, служащих источником новых знаний для интеллектуальных информационных систем.

Данное диссертационное исследование направлено на реализацию целей, сформулированных в Законе Республики Узбекистан «Об электронном правительстве» (2015), в Указах Президента Республики Узбекистан №УП-4947 от 7 февраля 2017 года «О стратегии действий по дальнейшему Развитию Республики Узбекистан», № УП-5264 от 29 ноября 2017 года

¹ Указ Президента Республики Узбекистан «О стратегии действий по дальнейшему Развитию Республики Узбекистан». УП-4947 от 7 февраля 2017 года

«Об образовании Министерства инновационного развития Республики Узбекистан» и № УП-5349 от 19 февраля 2018 года «О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций», постановлениями Президента № ПП-1730 от 21 марта 2012 года «О мерах по дальнейшему внедрению и развитию современных информационно-коммуникационных технологий», № ПП-2158 от 3 апреля 2014 года «О мерах по дальнейшему внедрению и развитию информационно-коммуникационных технологий в реальном секторе экономики» и других нормативно-правовых документах в сфере информационных технологий.

Соответствие исследования с приоритетными направлениями развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики IV. «Развитие информатизации информационно-коммуникационных технологий».

Обзор зарубежных научных исследований по теме диссертации². Во многих ведущих научных центрах и высших образовательных учреждениях мира проводятся научные исследования направленные на разработку новых методов и алгоритмов интеллектуального анализа данных, так и обоснованию и практическому применению уже существующих в рамках математических моделей из прикладных областей, в том числе Международные ассоциации искусственного интеллекта Европы, Америки и стран Азии, Российский научно-исследовательский институт искусственного интеллекта, Московский Государственный университет, Томский государственный университет, Санкт-Петербургский институт информатики и автоматизации (Российская Федерация), NYU Center for Data Science, National Center for Data Mining of the University of Illinois at Chicago, University of California, Department of Electrical and Computer Engineering University of Texas (США), Darmstadt University of Applied Sciences (Германия), University of Lions (США, Канада), Data mining and Big data institute (Израил), Center for Data Mining and Database Technology (Индия).

В качестве значимых научных достижений в мире по развитию интеллектуального анализа для выявления скрытых закономерностей из баз данных, машинному обучению и совершенствованию систем распознавания получены результаты: разработан метод «глубинного обучения» к решению задач диагностики диабетических болезней (Université Paris-Saclay, Франция), разработаны методы text mining для обработки текстов на естественном языке (Speech Research Group, Великобритания, University of Luxembourg, Люксембург, Bharathiar Universiteti, Индия, University of Oxford, Великобритания, Cornell Computational Linguistics Lab, Stony Brook University,

² Обзор научных исследований по теме диссертации составлен на основании www.braude.ac.il, sonatech.ac.in, www.msu.ru, cde.nyu.edu, www.ncdm.uic.edu, www.ece.utexas.edu, www.h-da.de/, www.ifmo.ru, www.universite-paris-saclay.fr, www.en.uni.lu, medicine.utah.edu, www.icyb.kiev.ua, www.titech.ac.jp, www.keele.ac.uk, www.cam.ac.uk, www.ox.ac.uk, www.conf.ling.cornell.edu, www.compling.stonybrook.edu, gtlnp.wordpress.com, ling.uni-konstanz.de, www.biominingbu.org, mi.eng.cam.ac.uk/mi/Main/Speech и других источников.

США, Konstanz universiteti, Германия, Georgia Tech Computational Linguistics Lab, Грузия, АБВУ, Российская Федерация), разработан метод экспериментальной валидации для систем машинного обучения (Center for Data Science and Public Policy at the University of Chicago, США), дано математическое обоснование методов и алгоритмов выбора признакового пространства (Институт проблем управления, Российская Федерация), предложены способы обнаружения и удаления «шумовых» объектов для селекции выборки данных (Томский государственный университет, Российская Федерация), разработаны методы интеллектуального анализа данных и машинного обучения для прогнозирования биомедицинских свойств наночастиц (Department of Biomedical Informatics, University of Utah, США), разработаны методы и средства построения интеллектуальных систем многоуровневого управления (Институт Кибернетики имени В.М.Глушкова, Украина), разработаны методы анализа ссылок и ассоциаций, визуализации и интеллектуальной аналитики для решения задачи быстрого поиска информации в тексте (Tokyo Institute of Technology, Япония), построены ансамбли алгоритмов на базе рандомизированного леса деревьев (Keele university, Великобритания), разработаны статистические алгоритмы для моделирования многопараметрических задач распознавания образов, а также для анализа данных (Cambridge university, Великобритания).

Дальнейшее совершенствование существующих и созданию новых методов интеллектуального анализа данных в мире осуществляется по следующим перспективным направлениям: разработка критериев и методов для оценки обобщающей способности алгоритмов распознавания; уменьшение комбинаторной сложности алгоритмов поиска логических закономерностей; машинного перевода с естественных языков с учётом контекста предметной области.

Степень изученности проблемы. Научные основы интеллектуального анализа данных для извлечения скрытых закономерностей из баз данных через отбор информативных наборов признаков и вычисление обобщающей способности алгоритмов распознавания были заложены в трудах Ю.И. Журавлева, Н.Г. Загоруйко, В.А. Дюка, В.П. Гладуна, А.С. Потапова, Ф.И. Андона, В.Н. Вапника, А.Ю. Зиновьева, К.В. Воронцова, Г.С. Лбова, В.Г. Царегродцева, П. Флаха, Г. Пятецкого-Шапиро, Х. Эдельштайна, А.А. Барсегяна, Т.А. Гаврилова, В.Ф. Хорошевского и других учёных. Ю.И. Журавлев разработал теорию алгебраического подхода к проблеме распознавания образов, в трудах В.А. Дюка предложен метод снижения комбинаторной сложности алгоритмов поиска логических закономерностей в данных, в работах Ю.А. Зиновьева и В.Г. Царегродцева приводится обоснование использования искусственных нейронных сетей для численного решения некорректных задач.

В трудах учёных нашей Республики М.М. Камилова, Т.Ф. Бекмуратова, Ф.Т. Адыловой, Ш.Х. Фазылова, Д.Т. Мухамадиевой, Р.Н. Усманова, А.Х. Нишанова предложены аналитические методы выбора информативных

наборов признаков, разработаны нейро-нечёткие логические модели для принятия решений и методы предобработки данных.

Поскольку для извлечения скрытых закономерностей из баз данных используются эвристические алгоритмы, построения информационных моделей для слабо структурированных предметных областей, как правило, не имеет строгого математического обоснования. По этой причине требуется совершенствование методик подбора критериев и алгоритмов для формирования моделей адекватно описывающих процессы и явления в предметных областях.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнена диссертация. Диссертационное исследование выполнено в соответствии с научно-исследовательскими планами Национального университета Узбекистана по проектам Ф-4-64 «Разработка и обоснование методов вычисления обобщённых оценок и индивидуальных метрик объектов в интеллектуальном анализе данных» (2012-2016), ОТ-Атех-2018-224 «Разработка информационной модели прогнозирования развития, течения и исхода язвенного колита и колоректального рака, оптимизации диагностики и лечения с учетом этиопатогенеза и факторов риска» (2018-2019).

Целью исследования является разработка и обоснование критериев и методов интеллектуального анализа данных для поиска скрытых закономерностей из баз экспериментальных данных.

Задачи исследования:

обоснование существования решения задачи разбиения значений количественных признаков на интервалы, в границах которых доминируют представители объектов одного из непересекающихся классов;

формирование нового признакового пространства в описании допустимых объектов на основе обобщённых оценок;

разработка и обоснование стохастического и детерминистического методов отображения описаний объектов выборки на числовую ось по указанным и определяемым наборам разнотипных признаков;

селекция обучающих выборок на основе обобщённых оценок объектов;

синтез деревьев решений с выбором в качестве атрибутов латентных признаков со значениями обобщённых оценок по определяемым наборам объектов;

разработка правил вывода для наполнения баз знаний на основе интервальных методов анализа данных.

Объектом исследования являются информационные модели на основе методов интеллектуального анализа данных для поиска скрытых закономерностей из баз данных слабо структурированных предметных областей.

Предметом исследования являются методы интеллектуального анализа данных, разработанные на основе вычисления обобщённых оценок объектов.

Методы исследования. Теоретические исследования в работе базировались на теории и алгоритмах интеллектуального анализа данных,

методах обработки информации, методах нечеткой логики, для вычислительных экспериментов использовалось компьютерное моделирование на основе извлечения скрытых закономерностей из баз данных.

Научная новизна диссертационного исследования заключается в следующем:

предложен способ (формула) оценивания устойчивости разбиения значений количественных признаков в описании допустимых объектов на интервалы доминирования представителей одного из двух классов;

разработан метод выбора в качестве атрибутов в узлы дерева решений латентных признаков на основе обобщённых оценок с учётом независимости исходных признаков по определяемому подмножеству объектов;

разработан алгоритм формирования нового пространства для описания объектов через латентные признаки;

определён критерий отбора шумовых объектов с использованием обобщённых оценок для селекции обучающих выборок;

на базе интервальных методов разработан способ объяснения процесса принятия решений экспертами в терминах нечёткой логики по указанным или определяемым наборам признаков.

Практические результаты исследования заключается в следующем:

разработан метод оценки качества (степени истинности гипотезы о компактности) выборки на основе вычисления значений устойчивости объектов классов;

разработан метод формирования нечётких правил вывода о принадлежности объектов к непересекающимся классам, основанный на разбиении значений количественных признаков в описании объектов выборки на интервалы;

разработан алгоритм построения дерева решений на основе обобщённых оценок;

предложена технология для решения задач классификации с использованием визуализации структуры отношений объектов выборок по их обобщённым оценкам (латентным признакам), вычисляемым стохастическим и детерминистическим методами;

предложено решение проблемы «проклятие размерности» путем формирования нового, сокращенного признакового пространства по правилам иерархической агломеративной группировки исходных признаков;

предложена методика выбора алгоритмов интеллектуального анализа данных при построения информационннх моделей для предметных областей.

Достоверность результатов исследования. Достоверность результатов исследования подтверждаются при проведении вычислительных экспериментов на выборках данных из предметных областей, на базах данных из репозитория, размещенных на специализированных сайтах глобальной сети Internet, использованием критериев устойчивости, вычисляющих степень истинности гипотезы о компактности при анализе структуры отношений

объектов классов, применением технологии формирования последовательности реализацией методов для выделения скрытых закономерностей.

Научная и практическая значимость результатов исследования. Теоретическая значимость результатов исследования заключается вкладом в развитие теории интеллектуального анализа данных, созданием новых методов и методологий выявления скрытых закономерностей из баз и хранилищ данных на основе обобщённых оценок. Разработанные методы и алгоритмы обосновываются их широким применением при разработке систем распознавания и классификации, используемых в технической диагностике, робототехнических системах, управлении технологическими процессами и производствами.

Практическая значимость результатов исследования определяется тем, что разработанные алгоритмы и программное обеспечение могут быть использованы при решении задач из таких слабоформализуемых предметных областей как медицина, фармакология, социология, учебный процесс. Применение методов вычисления обобщённых оценок позволяет сократить затраты ресурсов, требуемых для выявления скрытых закономерностей из баз данных по объектам, явлениям и процессам.

Внедрение результатов исследования. Разработанные программные обеспечения методов и алгоритмов выявления скрытых закономерностей из баз данных на основе методов вычисления обобщённых оценок позволило получить следующие результаты:

методы разбиения значений количественных признаков на интервалы и вычисления обобщённых оценок, критерий отбора шумовых объектов с использованием обобщённых оценок, а также методология объяснения процесса принятия решений экспертами в терминах нечёткой логики по указанному и определяемому набору показателей были внедрены в компании «Узбекинвест» (справка Национальная компания экспортно-импортного страхования «Узбекинвест» №132 от 8 июня 2018 г.). Результаты научного исследования выявили скрытые закономерности по оценкам рисков, принимаемых при страховании, дальнейшей оценки части рисков, подлежащих передаче на перестрахование и определение размеров страховых тарифов, а также прогнозирования вероятности наступления событий, которые перечисляются при составлении договора о страховании с физическими и юридическими лицами, что позволило на 15% сократить время обработки данных;

правила иерархической агломеративной группировки исходных признаков и методы поиска шумовых объектов на основе обобщённых оценок были внедрены в Ташкентский государственный институт стоматологии (справка Министерства Здравоохранения Республики Узбекистана № 8н-4/323 от 2 октября 2019г.). В результате выявленные скрытые закономерности в клинических базах данных позволила на 10-15% сократить процесс принятия решений экспертами за счет исключения из анализа неперспективных вариантов;

способ оценивания устойчивости разбиения значений количественных признаков в описании допустимых объектов на интервалы доминирования представителей одного из двух классов была внедрена в Комитет по борьбе с монополией Республики Узбекистан (справка Комитете по борьбе с монополией Республики Узбекистан №1443104-22 от 25 октября 2019 г.). В результате за счет выявления скрытых закономерностей по базам данных биржевых сделок позволил на 10-15% сократить процесс принятия решений экспертами.

Апробация результатов исследования. Теоретические и прикладные проблемы по теме диссертации обсуждались на 13 международных и 7 республиканских конференциях.

Опубликованность результатов исследования. По теме диссертации опубликованы 41 научных работ, из них 1 монография, 16 научных статей, в том числе 10 в республиканских, 6 в зарубежных журналах, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных результатов докторской диссертации. Получены 2 патента на изобретения, 2 свидетельства на программный продукт Агентства по интеллектуальной собственности Республики Узбекистан.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав, заключения, списка литературы, приложений и содержит 171 страниц текста.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность и востребованность темы диссертации, сформированы цели и задачи, выявлены объект и предмет исследования, определены соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан, изложена научная новизна и практические результаты исследования, обоснована достоверность полученных результатов, приведены перечень внедрений в практику результатов исследования, результаты апробации работы, сведения по опубликованным работам и структуре диссертации.

В первой главе диссертации **«Информационные модели на основе методов интеллектуального анализа данных»** отмечается, что в информационных моделях, основанных на знаниях, для решения задач из слабоструктурированных предметных областей используется информация из баз и хранилищ данных, экспертная информация в форме баз знаний, запись процессов, явлений в терминах математической символики. В рамках этих моделей производится поиск скрытых закономерностей из данных, являющихся источниками новых знаний. Описаны методы оценок для принятия экспертных решений. Обобщённые оценки представляют латентные явно неизмеримые признаки, которые могут использоваться при выборе нового признакового пространства для описания объектов. Доказана неопределённость такого понятия как допустимый объект, для описания которого используются латентные признаки. В качестве значений латентных

признаков предлагается использовать отображения описаний объектов по указанным или определяемым наборам признаков на числовую ось. В зависимости от выбора линейных или нелинейных методов отображения, структура отношений объектов классов в пространстве из латентных признаков меняется.

Произведен сравнительный анализ моделей алгоритмов вычисления оценок и обобщённых оценок, получаемых линейными и нелинейными методами.

Для вербализации знаний предлагается технология привязки значений функции принадлежности и обобщённых оценок к лингвистическим переменным для интерпретации результатов вычислений.

Во второй главе диссертации «**Интервальные методы анализа данных**» приводится описание форм представления логических закономерностей и способов их обнаружения. Предлагается для поиска закономерностей в базах данных использовать разбиение значений количественных признаков на интервалы.

Рассматривается задача распознавания в стандартной постановке. Считается, что задано множество $E_0 = \{S_1, \dots, S_m\}$ объектов непересекающихся классов $K_1, \dots, K_l, l > 1$. Описание объектов производится с помощью n разнотипных признаков $X(n) = (x_1, \dots, x_n)$, ξ из которых измеряются в интервальных шкалах, $n - \xi$ – в номинальной.

Считается, что существуют преобразования, позволяющие из наборов разнотипных признаков формировать латентные количественные признаки.

Заданы два условия разбиения значений количественных признаков на непересекающиеся интервалы:

1. Число интервалов равно числу классов объектов;
2. Число интервалов больше или равно числу классов.

Требуется вычислить значение границ интервалов по первому и второму условию.

Пусть $u_1^1, u_1^2 (u_2^1, u_2^2)$ – количество значений признака x_j класса $K_i, i = 1, 2$ соответственно в интервалах $[c_0, c_1]$ и $(c_1, c_2]$, p – порядковый номер элемента упорядоченной по неубыванию последовательности

$$r_{j_1}, \dots, r_{j_p}, \dots, r_{j_m} \quad (1)$$

По значениям $x_j \in X(n)$ у объектов из E_0 , определяющий границы интервалов как $c_0 = r_{j_1}, c_1 = r_{j_p}, c_2 = r_{j_m}$. Критерий

$$\left(\frac{\sum_{p=1}^2 \sum_{i=1}^2 u_i^p (u_i^p - 1)}{\sum_{t=1}^2 |K_t| (|K_t| - 1)} \right) \left(\frac{\sum_{p=1}^2 u_1^p (|K_2| - u_2^p) + u_2^p (|K_1| - u_1^p)}{2|K_1||K_2|} \right) \rightarrow \max_{c_0 < c_1 < c_2} \quad (2)$$

позволяет вычислять оптимальное значение границы между интервалами $[c_0, c_1], (c_1, c_2]$ и использовать её для выбора градаций количественного признака в номинальной шкале измерений. Выражение в левых скобках (2) представляет внутрикласовое сходство, в правых – межкласовое различие.

Для второго условия разбиение на непересекающиеся интервалы по (1), в границах которых доминируют значения класса $K_t, t = 1, 2$ производится по критерию

$$\left| \frac{d_t^i(u, v)}{|K_t|} - \frac{d_{3-t}^i(u, v)}{|K_{3-t}|} \right| \rightarrow \max, \quad (3)$$

где $d_t^i(u, v)$ – количество представителей класса K_t в интервале $[r_{c_u}, r_{c_v}]^i$.

Значение функции принадлежности c -го признака к K_t по интервалу $[r_{c_u}, r_{c_v}]^i$ определим как

$$f_{ci} = \frac{\eta_{1i}}{\eta_{1i} + \eta_{2i}}, \quad (4)$$

где $\eta_{ti} = \frac{d_t^i(u, v)}{|K_t|}, t \in \{1, 2\}$.

Устойчивость разбиения по всем интервалам лежит в $(0.5, 1]$ и вычисляется как

$$U(c) = \frac{1}{m} \sum_{\{[r_{c_u}, r_{c_v}]^i\}} \begin{cases} f_{ci}(v-u+1), & f_{ci} > 0.5, \\ (1-f_{ci})(v-u+1), & f_{ci} < 0.5, \end{cases} \quad (5)$$

Определены условия на множестве значений признаков, при которых разбиение на интервалы по (3) не существует.

Пусть для множества объектов обучающей выборки E_0 выполняется равенство $|K_1| = \alpha |K_2|, \alpha = \frac{c}{d}, c, d \in \{1, 2, \dots\}$. Существует количественный

признак, который в описании объектов E_0 представлен p ($2 \leq p \leq m$) разными значениями. Обозначим число одинаковых значений у класса K_1 как a_1, \dots, a_p

и b_1, \dots, b_p у класса K_2 . При выполнении условия $\frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_p}{b_p} = \alpha$ число

допустимых интервалов будет от 2 до p и значение функции принадлежности по классу K_1 или K_2 к любому интервалу равна 0.5, т. е. доминирования как такового не будет.

Интерпретация результатов разбиения значений признака по первому и второму условию на тестовом примере, показанном на рис. 1.

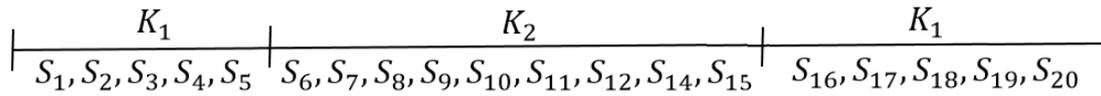


Рис. 1. Тестовый пример

Оптимальное значение критерия (1) равная 0.3611, при разбиении на интервалы $[1,5]$, $(5,20]$ указывает на относительно плохую разделимость (компактность) по признаку (см.рис.1). Устойчивость (4) по результатам разбиения на 3 интервала по критерию (3) для каждого интервала равна 1 и демонстрирует наличие хорошо обусловленных кластеров по значениям признака.

В третьей главе диссертации «**Методы вычисления обобщённых оценок**» рассматриваются методы вычисления обобщённых оценок для отображения отношений между объектами на числовую ось. Значение обобщённых оценок объектов вычисляются в рамках двухклассовой задачи распознавания. Объекты классов представлены в форме оппозиции вида – “*больной-практически здоровый*”, “*трудоголик-бездельник*”, “*богатый-бедный*”, “*отличник учёбы-задолжник*”, “*добрый-жадный*” и т.д. Описаны два метода вычисления обобщённых оценок: стохастический и детерминистический.

В стохастическом методе обобщённая оценка объекта $S_a \in E_0, S_a = (x_{a1}, \dots, x_{an})$ вычисляется как

$$R(S_a) = \sum_{i \in I} w_i t_i (x_i - c_1^i) / (c_2^i - c_0^i) + \sum_{i \in J} \mu_i(x_{ai}),$$

где $[c_0^i, c_1^i](c_1^i, c_2^i]$, w_i определяется по (2), $t_i \in \{-1, 1\}$, а значение элементов вектора $T = (t_1, \dots, t_n)$ вычисляется из условия

$$\min_{S_p \in K_1} R(S_p) - \max_{S_p \in K_2} R(S_p) \rightarrow \max.$$

Обозначим через p число градаций признака $r \in J, g_{dr}^t$ – количество значений t -й ($1 \leq t \leq p$) градации r -го признака в описании объектов класса K_d, l_{dr} – число градаций r -го признака в K_d . Различие по r -му признаку между классами K_1 и K_2 определяется как величина

$$\lambda_r = 1 - \frac{\sum_{t=1}^p g_{1r}^t g_{2r}^t}{|K_1||K_2|}. \quad (6)$$

Степень однородности (мера внутриклассового сходства) β_r значений градаций r -го признака по классам K_1, K_2 вычисляется по формулам:

$$D_{dr} = \begin{cases} (|K_d| - l_{dr} + 1)(|K_d| - l_{dr}), & p > 2, \\ |K_d|(|K_d| - 1), & p \leq 2; \end{cases}$$

$$\beta_r = \begin{cases} \frac{\sum_{t=1}^p g_{1r}^t (g_{1r}^t - 1) + g_{2r}^t (g_{2r}^t - 1)}{D_{1r} + D_{2r}}, & D_{1r} + D_{2r} > 0, \\ 0, & D_{1r} + D_{2r} = 0. \end{cases} \quad (7)$$

С помощью (6), (7) вес номинального признака $r \in J$ определяется как

$$v_r = \lambda_r \beta_r. \quad (8)$$

Очевидно, что множество чисел, идентифицирующих p градаций номинального признака, всегда можно взаимно однозначно отобразить в множество $\{1, \dots, p\}$. С учётом такого отображения для объекта $S = (x_1, \dots, x_n)$ вклад признака $x_i = j, i \in J, j \in \{1, \dots, p\}$ в обобщённую оценку определяется величиной

$$\mu_i(j) = v_i \left(\frac{\alpha_{ij}^1}{|K_1|} - \frac{\alpha_{ij}^2}{|K_2|} \right), \quad (9)$$

где $\alpha_{ij}^1, \alpha_{ij}^2$ – количество значений j -й градации i -го признака соответственно в классах K_1 и K_2 , v_i – вес i -го признака, вычисляемый по (8).

В детерминистическом методе обобщённая оценка объекта $S \in E_0 \cap K_d$, $S = (b_1, \dots, b_n)$ вычисляется по формуле

$$R(S) = \frac{1}{|K_{3-d}|} \sum_{S_j \in K_{3-d}} \left(\sum_{c \in I} \left\{ \begin{array}{l} f_{ci}, b_c \in [r_{cu}, r_{cv}]^i \text{ и } x_{jc} \notin [r_{cu}, r_{cv}]^i, \\ \frac{f_{ci} |b_c - x_{jc}|}{|r_{cu} - r_{cv}|}, r_{cu} \neq r_{cv} \\ 0, \quad r_{cu} = r_{cv} \end{array} \right\}, b_c, x_{jc} \in [r_{cu}, r_{cv}]^i + \sum_{c \in J} \left\{ \begin{array}{l} f_{ci}, b_c \neq x_{jc}, \\ 0, b_c = x_{jc}, \end{array} \right\} \right) \quad (10)$$

где $S_j = (x_{j1}, \dots, x_{jn})$ и значения τ_c градаций, $c \in J$, каждого номинального признака принадлежат множеству $\{1, 2, \dots, \tau_c\}$, $[r_{cu}, r_{cv}]^i$ – i -ый интервал доминирования $c \in I$ признака, f_{ci} – значение функции принадлежности c -го признака вычисляемая по (4). Если $c \in J$, то η_{1i}, η_{2i} в (4) рассматриваются как количество значений i -й градации соответственно в классах K_1, K_2 .

Предложен метод построения дерева решений на базе обобщённых оценок, где значение обобщённых оценок по паре информативных (латентных) признаков в узлах дерева используется для формирования классификатора по принятию решений, для которого формируется последовательность *if-then* правил лексикографического порядка. Метод

востребован в качестве средства интеллектуального анализа данных для извлечения новых, ранее неизвестных знаний об исследуемой предметной области.

В четвертой главе диссертации «Обнаружение скрытых закономерностей в базах данных» предлагаются методы извлечения знаний из баз данных с помощью логических закономерностей. Анализ логических закономерностей производится на основе свойства их устойчивости в разнотипном признаковом пространстве. По значениям устойчивости можно производить селекцию выборки для дальнейшего использования методов вычисления обобщённых оценок.

Процесс вычисления устойчивости объектов классов является инвариантным относительно масштабов измерений количественных признаков.

Предложена интерпретация обобщённых оценок в терминах нечёткой логики, значения которых вычисляются по указанным и определяемым наборам разнотипных признаков на примере экспертизы фармакологических препаратов.

Сравнивались выводы двух групп врачей K_1 и K_2 , представленных соответственно аллергологами ($|K_1|=23$), пульмонологами и терапевтами ($|K_2|=68$), по 3-м указанным наборам лекарственных препаратов: χ^1 (Глюкокортикоидные средства), χ^2 (Бронхолитические средства) и χ^3 (Антибиотики).

Множество значений, вычисленных по (10) и являющимися обобщёнными оценками (признаками y^1, y^2, y^3) по наборам χ^1, χ^2, χ^3 , анализировались по критерию (2). Для интерпретации результатов анализа использовались 6 значений лингвистической переменной.

Результаты проверки истинности гипотезы «Мнения врачей из группы K_1 и K_2 по использованию наборов лекарственных препаратов χ^1, χ^2 и χ^3 для лечения бронхиальной астмы существенно совпадают» приведены в табл. 1.

Таблица 1.

Результаты вычислительного эксперимента по 3 наборам признаков

| Набор признаков | Значение (2) по обобщённым оценкам | Значение лингвистической переменной |
|------------------------|---|--|
| χ^1 | 0,55 | <i>Мало совпадают</i> |
| χ^2 | 0,51 | <i>Мало совпадают</i> |
| χ^3 | 0,36 | <i>Существенно совпадают</i> |

Как видно из табл. 1 истинность гипотезы подтверждается только для набора χ^3 .

Предложен иерархический агломеративный алгоритм, на основе правил которого по описаниям объектов из E_0 реализуется разбиение признаков на непересекающиеся группы. Согласно правил алгоритма требовалось:

– сформировать непересекающиеся наборы разнотипных признаков по значениям их устойчивости на объектах из E_0 ;

– по непересекающимся наборам определить подмножество информативных латентных признаков.

Для вычисления устойчивости объектов используется нормирование значений количественных признаков в $[0,1]$.

Мера близости между двумя объектами и $S_b = (x_{b1}, \dots, x_{bn})$ выборки E_0 вычисляется по метрике Журавлева

$$\rho(S_a, S_b) = \sum_{i \in I} |x_{ai} - x_{bi}| + \sum_{i \in J} \begin{cases} 1, & x_{ai} \neq x_{bi}, \\ 0, & x_{ai} = x_{bi} \end{cases} \quad (11)$$

Определение границы интервала c_2 с помощью (11) в неубывающей последовательности $S_{d_0}, S_{d_1}, \dots, S_{d_{m-1}}, S_{d_0} = S_d$ по подмножеству признаков $X(k) \subset X(n), k \leq n$, производится по (2).

Пусть для интервалов построенных по критерию (2) определены

$$\lambda_1(p) = \left| \left\{ S_{d_i} \in K_p \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|,$$

$$\lambda_2(p) = \left| \left\{ S_{d_i} \in K_{3-p} \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|,$$

$$\theta_1(p) = \lambda_1(p) / |K_p| \text{ и } \theta_2(p) = \lambda_2(p) / |K_{3-p}|.$$

Тогда устойчивость объекта $S_d \in K_p, p = 1, 2$ по подмножеству признаков $X(k)$ вычисляется следующим образом

$$U(S_d, X(k)) = \theta_1(p)(1 - \theta_2(p)) \quad (12)$$

Сформируем матрицу близости $B = \{b_{ij}\}_1^n$, значение элементов которой вычисляется в следующем порядке

$$b_{ij} = \begin{cases} \frac{\sum_{a=1}^m \sum_{b=1}^m \alpha(S_a, S_b) \beta(i, j, S_a, S_b)}{\sum_{p=1}^2 |K_p| (m - |K_p|)}, & i \neq j, \\ 0, & \text{в противном случае,} \end{cases}$$

где

$$\beta(i, j, S_a, S_b) = \begin{cases} \rho(S_a, S_b), \{x_i, x_j\} \subset X(n), S_a, S_b \in E_0, \\ 0, & \text{в противном случае,} \end{cases}$$

$$\alpha(S_a, S_b) = \begin{cases} 0, & S_a, S_b \in K_i, i = 1, 2, \\ 1, & S_a \in K_i, S_b \in K_{3-i}. \end{cases}$$

Изменение оценки устойчивости объектов выборки E_0 при переходе из подмножеств признаков $X(k)$ в $X(k+1), 1 \leq k \leq n, X(k) \subset X(k+1)$ вычисляется следующим образом

$$M(X(k), X(k+1)) = \frac{1}{m} \sum_{d=1}^m \begin{cases} 1, & U(S_d, X(k)) \leq U(S_d, X(k+1)), \\ 0, & \text{в противном случае.} \end{cases} \quad (13)$$

Включение признака $x_d \in X(n)$ в группу (набор) $G_p = X(k), p < n$ выполняется, если оценка (13) на $G_p \cup \{x_d\}$ больше чем на G_p .

Предложен способ поиска закономерностей по наборам латентных признаков. Латентные признаки для описания объектов выборки могут быть получены с помощью арифметических операций над исходными количественными признаками, либо путем вычисления значений обобщенных оценок по подмножествам признаков.

Закономерности, выявленные в процессе эксперимента можно описывать с помощью формализмов нечеткого вывода и лингвистических переменных в виде нечеткого правила вывода о принадлежности объекта S к классу $K_l, l \in \{1, 2\}$ вида

Если А то В (<коэффициент уверенности>),

где <коэффициент уверенности> – коэффициент уверенности к заключению В при выполнении условия А.

Исследовалась выборка из 73 объектов (пациентов) $E_0 = \{S_j\}_1^{73}$, описанных 31 количественными признаками $X = (x_1, \dots, x_{31})$. Экспертом-врачом было выделено 48 объектов выборки с стабильной атеросклеротической бляшкой (АСБ), которые были отнесены к классу K_1 и 25 объектов к классу K_2 (нестабильные АСБ).

Для анализа использовались два метода разбиения исходных и латентных (явно неизмеримых) на непересекающиеся интервалы по (2) и (3).

В табл. 2 приведены результаты разбиения значений количественных признаков в описании объектов выборки E_0 на два интервала, при экстремуме критерия (2), где u_{ij} – количество представителей класса K_j в i -ом интервале, $i, j \in \{1, 2\}$.

Одно из подмножеств нечётких правил вывода о принадлежности допустимого объекта S (пациента) к классу K_1 (стабильному АСБ) по значениям нескольких признаков (см. табл. 2) имеют такой вид:

P_1 : Если $ТИМП-1 \leq 1602.88$ то $S \in K_1$ (0.717);

P_2 : Если $ММП-9 \leq 274.96$ то $S \in K_1$ (0.678);

P_3 : Если $КСР \leq 31$ то $S \in K_1$ (0.643).

Таблица 2.

Границы интервалов признаков по критерию (2)

| Признаки | Интервалы и количество представителей классов в интервалах | Значение критерия (2) |
|---|--|-----------------------|
| | $[r_{i_1}, r_{i_p}] (r_{i_p}, r_{i_{73}}] \{u_{11}, u_{12}\} \{u_{21}, u_{22}\}$ | |
| Тканевой ингибитор металлопротеиназы 1 (ТИМП-1) | [775.55,1602.88](1602.88,3677.22) {42,3} {3,25} | 0.717 |
| Матриксная металлопротеиназа 9 (ММП-9) | [28.1,274.96](274.96,1241.2){42,4} {3,24} | 0.678 |
| Конечный систолический размер ЛЖ(КСР) | [2,3.1](3.1,5.35) {4,24} {41,4} | 0.643 |
| Степень стеноза сонной артерии (СССА) | [0,55](55,100) {37,1} {8,27} | 0.611 |
| Возраст | [46, 62](62, 88) {12,14} {33,14} | 0.284 |
| Вес | [57, 72](72, 114) {14,11} {31,17} | 0.251 |
| Цинк в волосах | [29.74, 214.57] (214.57, 482.45) {21,17} {24,11} | 0.251 |
| Диастолическое АД (ДАД) | [60,80](80,120) {34,20} {11,8} | 0.238 |
| ... | | |

Правила вывода о принадлежности допустимого объекта S к классу K_1 по значениям признаков в интервале доминирования имеет вид:

P_1 : Если $775.55 \leq ТИМП-1 \leq 1602.88$ то $S \in K_1$ (0.9);

P_2 : Если $65 \leq Вес \leq 72$ то $S \in K_1$ (0.35);

P_3 : Если $0 \leq СССА \leq 55$ то $S \in K_1$ (0.96);

P_4 : Если $5.21 \leq КДР \leq 6.49$ то $S \in K_1$ (1).

Если в процессе вывода выполняются посылки правил P_1 и P_2 , то мера доверия к событию $S \in K_1$ принимает вид

$$\begin{aligned} & \text{МД}[\langle S \in K_1 \rangle: 775.55 \leq \text{ТИМП-1} \leq 1602.88, 65 \leq \text{Вес} \leq 72] = \\ & \text{МД}[\langle S \in K_1 \rangle: 775.55 \leq \text{ТИМП-1} \leq 1602.88] + \\ & \text{МД}[\langle S \in K_1 \rangle: 65 \leq \text{Вес} \leq 72] (1 - \text{МД}[\langle S \in K_1 \rangle: 775.55 \leq \text{ТИМП-1} \leq 1602.88]). \end{aligned}$$

Подставляя коэффициенты уверенности из правил, получим

$$\text{МД}[\langle S \in K_1 \rangle: 775.55 \leq \text{ТИМП-1} \leq 1602.88, 65 \leq \text{Вес} \leq 72] =$$

$$0.9+0.35(1-0.9)=0.935 .$$

Если число свидетельств в пользу события больше двух, то мера доверия вычисляется каскадным образом.

Наличие более двух интервалов разбиения признаков увеличивает число правил вывода. С другой стороны использование этих правил повышает коэффициент доверия к результатам принятия решений. Достоверность принятия решения остается высокой и при наличии у допустимого объекта пропусков в данных. Например, если у объекта (пациента) отсутствует значение по таким информативным признакам как *ТИМП-1*, *ММП-9*, то можно получить вывод на основе значений малоинформативных признаков (см. табл. 2) – *ИМТ*, *Вес*, *Цинк в волосах*, и *ДАД* с приемлемым коэффициентом доверия:

$$\text{МД}[\langle S \in K_1 \rangle: \text{ИМТ} > 25.8, \text{Вес} > 72, \text{Цинк в волосах} > 214.57, \text{ДАД} < 80] = 0.68 .$$

В отличие от деревьев решений при формировании правил вывода нет требований к независимости признаков.

В пятой главе диссертации «Разработка информационных моделей» даётся описание компьютерного приложения (системы) реализующей алгоритмы интеллектуального анализа данных с целью выявления скрытых закономерностей в базах данных. Система представляет из себя комплекс взаимосвязанных модулей, каждый из которых решает отдельную подзадачу. Часть модулей реализована с использованием визуальной формы.

Проведены вычислительные эксперименты на данных из репозитория сайта <http://archive.ics.uci.edu/ml/machine-learning-databases> для селекции обучающих выборок. При селекции производится поиск и удаления шумовых объектов с целью адаптации конфигурации классов поверхностями или увеличения меры компактности классов.

Предложены способы формирования пространства R^2 из латентных признаков с помощью методов (стохастического и детерминистического) вычисления обобщённых оценок и обнаружения шумовых объектов выборки в пространстве из латентных признаков.

Свойство инвариантности методов в R^2 позволяет единообразно выявить множество шумовых объектов выборки. Множество шумовых объектов рассматривается как подмножество граничных объектов классов по заданной метрике. Множество граничных объектов $B \in E_0$ определяется как

$$B = \left\{ S \in E_0 \mid \rho(S_i, S) = \min_{S_i \in K_t, S_j \in CK_t} \rho(S_i, S_j) \right\}.$$

Объект $S \in B \cap K_t, t = 1, 2$ принадлежит множеству шумовых объектов N_t класса K_t , если

$$\left| \left\{ S_i \in E_0 \mid \rho(S_i, S) = \min_{S_i \in CK_t, S_j \in K_t} \rho(S_i, S_j) \right\} \right| > \left| \left\{ S_i \in K_t \mid \rho(S_i, S) < \min_{S_i \in K_t, S_j \in CK_t} \rho(S_i, S_j) \right\} \right| \quad (14)$$

На рис.2 показано геометрическая интерпретация выделения шумового объектов выборки, где значкам ‘■’ обозначены объекты класса K_1 , а ‘▲’ – класса K_2 . Объект $S \in K_2$ является шумовым, так как он граничный для $S_i, S_j, S_k \in K_1$, а в гипершаре с центром S и с радиусом R_S содержится только один объект S_u класса K_2 .

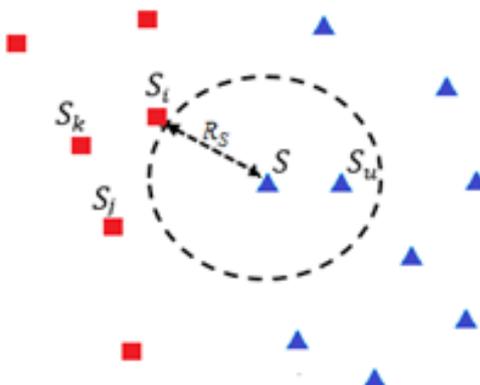


Рис. 2. Геометрическое представление выделения шумового объекта выборки.

Поскольку рассматриваемые в работе методы вычисления обобщённых оценок инвариантны к масштабам измерений в качестве меры расстояния в пространстве латентных признаков выбрана евклидова метрика.

Для эксперимента была взята выборка данных «Echocardiogram» из <http://archive.ics.uci.edu/ml/machine-learning-databases>, описывающая состояние пациентов, у которых был сердечный приступ.

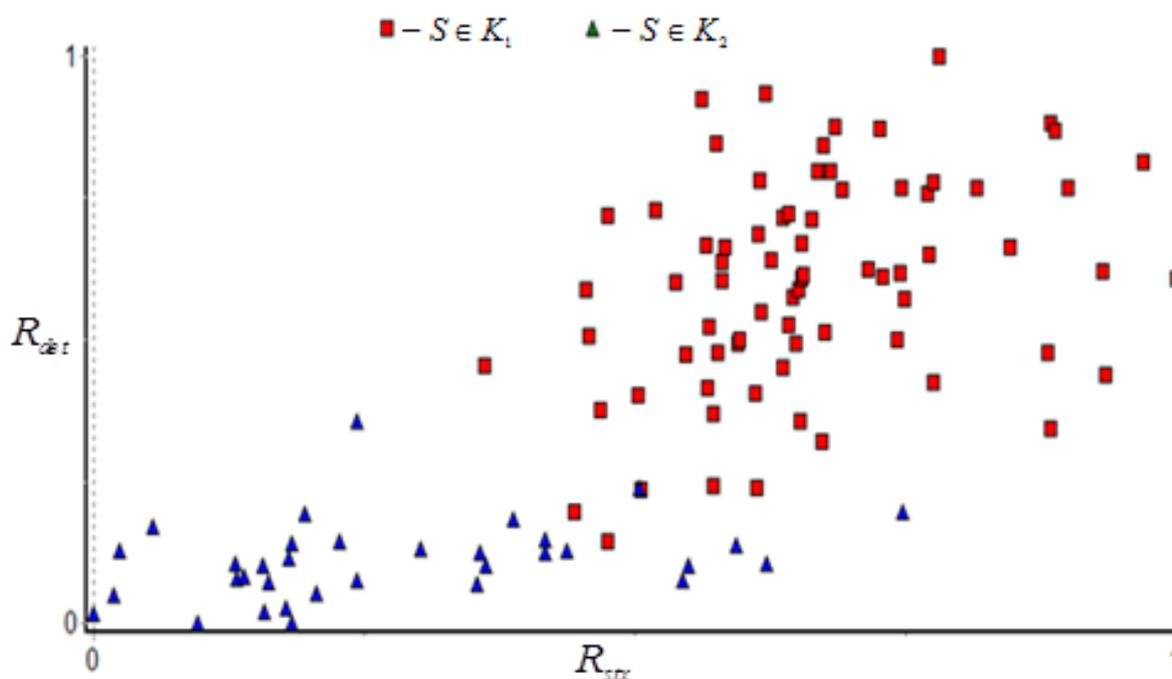


Рис. 4. Структура расположения объектов выборки E_0 в пространстве R^2 .

Выборка состоит из 108 объектов, из них 74 принадлежат к классу K_1 (пациент умер в течении 1 года), 34 к K_2 (пациент остался жив или умер после 1 года). Каждый объект описан набором из 8 количественных и 2 номинальных признаков.

Описание объектов E_0 в R^2 по значениям латентных признаков $\{R_{stx}, R_{det}\}$ сформировано с помощью стохастического и детерминистического методов.

По (14) выделено множество шумовых объектов $N = \{S_7, S_{38}, S_{71}, S_{78}, S_{101}\}$. Удаление этих объектов из E_0 показало улучшение компактности (разделимости) объектов классов выборки в пространстве R^2 . Значение критерия (2) для признака R_{stx} увеличилось от 0.75 в E_0 до 0.79 в $E_0 \setminus N$, а для признака R_{det} соответственно от 0.86 до 1. На рис. 3 и рис. 4 приведены графическое представление структуры расположения объектов выборки в пространстве R^2 по E_0 и по $E_0 \setminus N$.

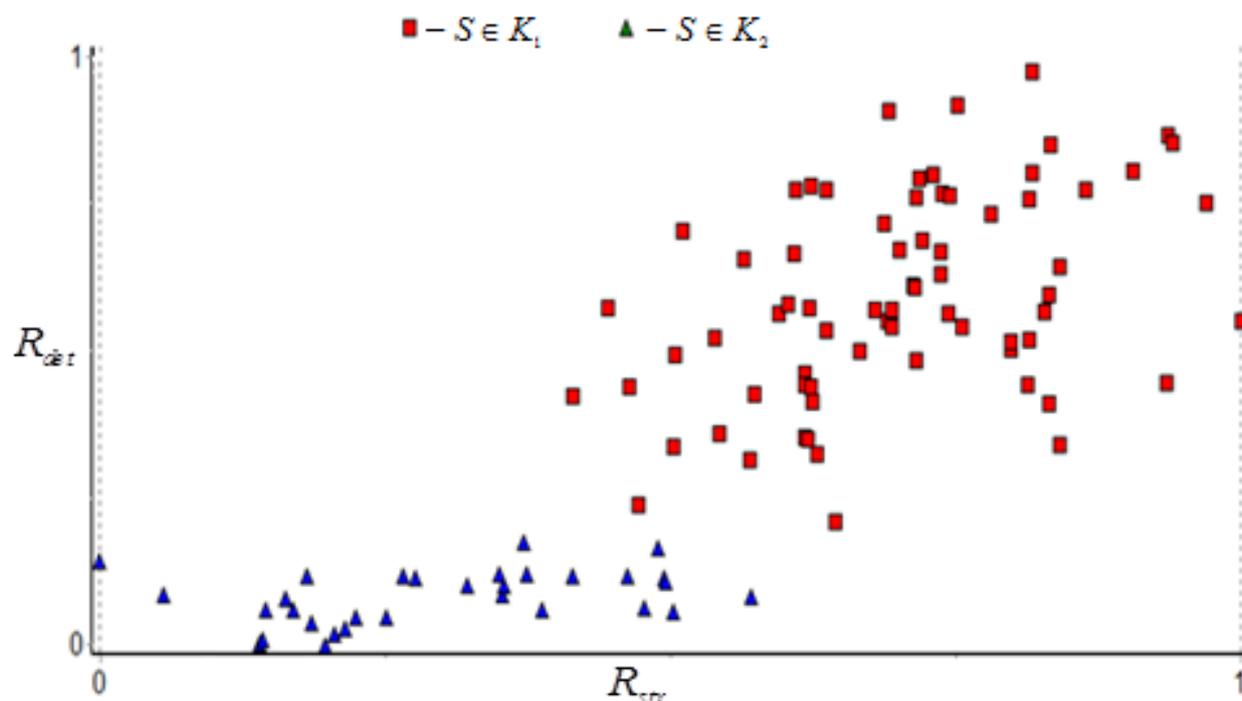


Рис. 4. Структура расположения объектов выборки $E_0 \setminus N$ в пространстве R^2 .

На рис. 4 демонстрируется возможность кусочно-линейного разделения объектов классов выборки после удаления шумовых объектов.

ЗАКЛЮЧЕНИЕ

Основные результаты, полученные в диссертационной работе на тему «Системы обнаружения скрытых закономерностей на базе методов вычисления обобщенных оценок», сводятся к следующему:

1. Предложен способ вычисления оценки устойчивости признаков в границах интервалов с доминированием значений из описаний объектов непересекающихся классов. Способ позволил выявлять наличие сильных закономерностей по наборам количественных признаков;

2. Определено условие отсутствия доминирования значений признаков объектов классов в границах непересекающихся интервалов. Это условие определяет область поиска скрытых закономерностей;

3. Разработаны практические рекомендации для вычисления обобщенных оценок объектов по стохастическому и детерминистическому методам в двухклассовой задаче распознавания. Применение рекомендации позволяет снизить размерности признакового пространства для поиска скрытых закономерностей;

4. Разработан метод вычисления значений функции принадлежности к нечётким множествам на основе обобщенных оценок по указанным наборам разнотипных признаков. При вычислении использовался экстремум критерия разбиения обобщенных оценок на интервалы. Данный метод позволяет определить степень истинности гипотезы «*Значение обобщенных оценок объектов классов пересекаются между собой*»;

5. Предложен метод формирования правил вывода по деревьям решений с использованием латентных признаков на основе обобщенных оценок. Лексикографический порядок сформирования правил вывода с учётом независимости латентных признаков позволил произвести направленный выбор принятия решений при классификации;

6. Разработан иерархический агломеративный метод группировки разнотипных признаков. В правилах группировки использовалась устойчивость объектов по метрике Журавлёва. Разбиение признаков на группы применялось для вычисления обобщенных оценок. Обобщенные оценки использовались в качестве латентных признаков для сокращения размерности признакового пространства в описании объектов.

7. Определены критерии отбора шумовых объектов на основе стохастического и детерминистического методов вычисления обобщенных оценок. Отбор шумовых объектов позволяет повысить обобщающую способность алгоритмов распознавания;

8. Разработаны компьютерные приложения для проведения вычислительных экспериментов с целью поиска скрытых закономерностей в данных на основе методов вычисления обобщенных оценок. Компьютерные приложения использовались для построения информационных моделей в слабоформулизованных предметных областях;

9. Методы, алгоритмы и компьютерные приложения, разработанные в рамках темы диссертационного исследования, были внедрены для

использования в компании «Узбекинвест» и антимонопольном Комитете Республики Узбекистан, а также при научных исследованиях в Ташкентском государственном институте стоматологии. Выявленные скрытые закономерности в базах данных позволили на 10-15% сократить процесс принятия решений экспертами.

**AD HOC SCIENTIFIC COUNCIL AT THE SCIENTIFIC COUNCIL
AWARDING SCIENTIFIC DEGREES DSc.27.06.2017.T.07.01 AT
TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES**

NATIONAL UNIVERSITY OF UZBEKISTAN

MADRAKHIMOV SHAVKAT FAYZULLAEVICH

**DETECTION SYSTEMS OF HIDDEN REGULARITY ON THE BASIS OF
THE METHODS OF CALCULATION OF GENERALIZED ESTIMATES**

05.01.03 – Theoretical foundations of informatics

**ABSTRACT OF THE DOCTORAL (DSc)
DISSERTATION OF TECHNICAL SCIENCES**

Tashkent-2020

The theme of doctoral (DSc) dissertation was registered with the number of B2017.1.DSc/T27 the Supreme Attestation Commission of the Cabinet of Ministers of the Republic of Uzbekistan.

The dissertation has been prepared at National University of Uzbekistan.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) on the website www.tuit.uz and on the website of «Ziyonet» Information and educational portal www.ziyonet.uz.

Scientific adviser: **Ignatyev Nikolay Aleksandrovich**
Doctor of Physical and Mathematical Sciences, Professor

Official opponents: **Sharipbay Altynbek Amiruly**
Doctor of Technical Sciences, Professor

Gulyamov Shukhrat Mannopovich
Doctor of Technical Sciences, Professor

Mamatov Narzulla Salizhonovich
Doctor of Technical Sciences, Senior Researcher

Leading organization: **Kazakh National University**

The defense will take place “_____” _____ 2020 y. at _____ on the meeting of Scientific council No. DSc.27.06.2017.T.07.01 at Tashkent University of Information Technologies (Address: 100202, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-43, fax: (+99871) 238-65-52, e-mail: tuit@tuit.uz).

The dissertation is available at the Information Resource Centre of the Tashkent University of Information Technologies (is registered under No. _____). (Address: 100202, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-43, fax: (+99871) 238-65-52).

Abstract of dissertation sent out on “_____” _____ 2020 y.
(mailing report No. ___ on “_____” _____ 2019 y.).

R.Kh.Khamdamov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor

F.M.Nuraliev
Scientific secretary of scientific
council awarding scientific degrees,
Doctor of Technical Sciences, Docent

Sh.H.Fazilov
Chairman of the one-time academic seminar under
the scientific council awarding scientific degrees,
Doctor of Technical Sciences, Professor

INTRODUCTION (abstract of the dissertation of doctor of science (DSc))

The aim of research work is to develop and justification of the criteria and methods of data mining for searching the hidden regularities on the bases of experimental data.

The object of the research work as is information models based on the methods of data mining to find hidden regularities in databases from different subject areas.

Scientific novelty of research work is as follows:

the method (formula) to evaluate the stability of the partition of values of quantitative features in the description of admissible objects on the intervals of dominance of representatives of one of the two classes is proposed;

a method has been developed for choosing latent features as attributes to the nodes of the decision tree based on generalized estimates, taking into account the independence of the initial features for a defined subset of objects;

an algorithm has been developed for the formation of a new features space from latent features;

the criterion for the selection of noise objects using generalized estimates for the selection of training samples is determined;

a method has been developed for explaining the decision-making process by experts in terms of fuzzy logic on the indicated or defined sets of features.

Implementation of the research results. The developed software methods and algorithms for identifying hidden regularities from databases based on methods for calculating generalized estimates allowed to obtain the following results:

a method of dividing values of quantitative features into the intervals and calculating generalized estimates, criterion of selecting noise objects based on generalized estimates and a methodology for explaining the decision-making process by experts in terms of fuzzy logic for the specified set of indicators were applied into Uzbekinvest company (Information №132 on 8.08.2018 of “Uzbekinvest” export – import national insurance company). The results of the scientific research were used to identify hidden patterns in assessing risks taken in insurance, further assessing part of the risks to be transferred for reinsurance and determining the size of insurance tariffs, as well as predicting the likelihood of events that are listed when drawing up an insurance contract with individuals and legal entities, which allowed to reduce data processing time by 15%;

the hierarchical agglomerative grouping of initial features and methods of isolation of noise objects based on generalized estimations were applied in Tashkent State Dental Institute (Reference of the Ministry of Health of the Republic of Uzbekistan № 8H-4.323 on October 2, 2019). The decision-making process has been reduced by 10-15% as a result of cutting off expert solutions that do not lead to expert analysis based on the internal laws found in clinical database processing;

the method of estimating the stability of intervals of quantitative features characteristic of possible objects, dominated by representatives of one of the two classes were applied in Antimonopoly Committee of the Republic of Uzbekistan was linked to the increase in real sales prices (Information of the Antimonopoly

Committee of the Republic of Uzbekistan № 1443/04-22 on October 25, 2019). Internal regulations revealed on the basis of exchange deals have led to a 10-15% reduction in decision-making by experts.

The structure and volume of the dissertation. The dissertation thesis consists of an introduction, five chapters, a conclusion, a list of used literature and appendices. The volume of the thesis is 171 pages of typewritten text.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (Часть I, Part I)

1. Игнатъев Н.А., Мадрахимов Ш.Ф. О некоторых способах повышения прозрачности нейронных сетей // Вычислительные технологии. - Российская Федерация, Новосибирск, 2003.-Т.8, №6.- С.31-37. (05.00.00; № 22)

2. Адылова З.Т., Игнатъев Н.А., Мадрахимов Ш.Ф. Методы защиты информации в компьютерных сетях, основывающиеся на многоагентных технологиях // Узб. журнал «Проблемы информатики и энергетики».- Ташкент:Фан, 2004.-№3- с.3-9. (05.00.00;№ 5)

3. Мадрахимов Ш.Ф. Построение функции принадлежности нечетких множеств для номинальных признаков на основе технологий искусственных нейронных сетей // ЎзМУ хабарлари, М.Улуғбек номидаги Ўзбекистон Миллий университети илмий журнали, Тошкент, «Университет», 2006, 103-105 б. (01.00.00; №8).

4. Игнатъев Н. А., Мадрахимов Ш.Ф. О мерах сложности и неопределённости решения задач классификации в искусственных нейронных сетях // Доклады АН РУз, 2007 г, №1, С.32-35.(05.00.00; № 9)

5. Игнатъев Н.А., Мадрахимов Ш.Ф. Устойчивость и обобщённые оценки классифицированных объектов в разнотипном признаковом пространстве // Вычислительные технологии. - Российская Федерация, Новосибирск, 2011.Т.16, № 2. С. 70 - 77. (05.00.00; № 22)

6. Игнатъев Н.А., Мадрахимов Ш.Ф. Обобщённые оценки и прогноз по деревьям решений //ЎзМУ хабарлари, М.Улуғбек номидаги Ўзбекистон Миллий университети илмий журнали, Тошкент, «Университет», 2015, №2/1, 115-121 б. (01.00.00; №8).

7. Мадрахимов Ш.Ф., Суюнов Н.Д., Хуррамов А.Х., Икромов Г.М. Интеллектуал таҳлил усуллари билан дори воситаларининг VEN гуруҳга тегишлигининг умумлашган баҳосини ҳисоблаш // ЎзМУ хабарлари, М.Улуғбек номидаги Ўзбекистон Миллий университети илмий журнали, Тошкент, «Университет», 2013, №2, 95-98 б. (01.00.00; №8)

8. Madrahimov Sh.F., Saidov D.Y. Sinf obyektlarining turg'unligi va alomatlarni guruhlash // Hisoblash va amaliy matematika muammolari ilmiy jurnali, Toshkent, №3(5) 2016, 50-54 б. (05.00.00;№ 23).

9. Мадрахимов Ш.Ф. Поиск закономерностей по наборам номинальных признаков в задачах распознавания с учителем // Узб. журнал «Проблемы информатики и энергетики».- Ташкент, «Фан ва технологиялар», 2016.-№2-с.45-49. (05.00.00;№ 5)

10. N.A. Ignatyev, Sh.F.Madrakhimov, D.Y.Saidov. Stability of object classes and selection of the latent features // International journal of engineering technology and sciences, 2017, Malaysia, Vol. 7, pp. 1-10, (№5, Global Impact Factor, IF=0,562).

11. Мадрахимов Ш.Ф. Отбор шумовых объектов на базе обобщённых оценок // Журнал Проблемы вычислительной и прикладной математики, Ташкент, №2(14) 2018, С.122-131. (05.00.00; № 23)

12. Мадрахимов Ш.Ф., Розыходжаева Г.А. Построение нечётких правил вывода для диагностики нестабильности атеросклеротической бляшки// Журнал Врач и Информационные технологии, Российская Федерация, Москва, № 3, 2018. С. 81-88. (№1, Web of science, IF=0,202)

13. Madrakhimov Sh. Calculation of the Generalized Estimations in Sets of Features and their Interpretation // International Journal of Software Engineering and Its Applications (South Korea). Vol.12, No.3 (2018), pp.29-38 (05.00.00; №19)

14. Розыходжаева Г.А., Игнатъев Н.А. Мадрахимов Ш.Ф. и др. Патент на изобретения Республики Узбекистан «Способ диагностики степени тяжести атеросклероза» // Агентство по интеллектуальной собственности РУз. № IAP 05362. Бюл. – 2014. № 8(160).

15. Розыходжаева Г.А., Игнатъев Н.А. Мадрахимов Ш.Ф. и др. Патент на изобретения Республики Узбекистан «Способ прогнозирования нестабильности атеросклеротической бляшки» //Агентство по интеллектуальной собственности РУз. № IAP 05363. Бюл. –2015. № 4(168).

II бўлим (Часть II, Part II)

16. Адылова З.Т., Мадрахимов Ш.Ф. О подходе к построению и программной реализации дерева решений для систем управления сетью связи// Вопросы кибернетики. Ташкент, 1995. Вып. С. 51-58.

17. Адылова З.Т., Мадрахимов Ш.Ф., Бассам Х.А. Применение семантической сети в задачах юриспруденции // Вопросы кибернетики. Ташкент, 1996, Вып. 154. С. 29-37.

18. Мадрахимов Ш.Ф. Билимларни тасвирлашнинг продукция ва фрейм моделларидан биргаликда фойдаланиш //Вопросы кибернетики.Ташкент, 1998, Вып. 156, С. 25-33.

19. Суюнов Н. Д., Игнатъев Н.А., Мадрахимов Ш.Ф., Икрамова Г. М. Фармакоэкономические исследования лекарственного обеспечения больных бронхиальной астмой в Узбекистане//Фармация, Российская Федерация, Москва, 2011, №3, С.33-36. (15.00.00, № 8).

20. Мадрахимов Ш.Ф., Хуррамов А.Х. Объектларнинг умумлашган баҳоладини ҳисоблаш учун турли тоифадаги аломатларнинг минимал тўпламини танлаш // ЎЗМУ хабарлари, М.Улуғбек номидаги Ўзбекистон Миллий университети илмий журнали, Тошкент, «Университет», 2013, №3, 112-115 б.

21. Мадрахимов Ш.Ф., Хуррамов А.Х. Умумлашган кўрсаткичлар ва улардан билимларни шакллантиришда фойдаланиш. Монография, Қарши: Қарши давлат университети, 2014. 92 бет.

22. Suyunov N.D., Madrakhimov Sh.F., Ikramova G.M. Analysis of drug provision of patients with bronchial asthma // News of Science and Education, NR 3 (27) 2015, Sheffield, Science And Education Ltd, p. 27-35.

23. Адылова З.Т., Мадрахимов Ш.Ф., Бассам Х.А. Экспертная система с многоуровневой базой знаний //Респ.конф. «Математическое моделирование и вычислительный эксперимент». Ташкент, 1996, С.80.

24. Адылова З.Т., Мадрахимов Ш.Ф., Хасанов Б. Об одном подходе к комбинированному использованию способов представления знаний// Международ. конф. «Проблемы информатики и управления, перспективы их решения». Тезиси докладов Ташкент, 1996, С.18.

25. Игнатъев Н.А., Мадрахимов Ш.Ф. Линейные оболочки как способ выбора минимальной конфигурации нейронных сетей // Межд. научная конф. по мат.логике. Тезисы докладов. Новосибирск, 1999, 10-11 август, С.56.

26. Игнатъев Н.А., Мадрахимов Ш.Ф. Экспертная система над множеством алгоритмов синтеза нейронных сетей // Труды Республиканской конференции «Проблемы алгоритмического программирования», Ташкент, 2000, С.4.

27. Мадрахимов Ш.Ф., Усмонов С. А. О подходе к построению и программной реализации вопросо-ответных систем// Международ. конф. по искусственному интеллекту. Тезиси докладов. Ташкент, 1994, С.72.

28. Мадрахимов Ш.Ф., Хуррамов А.Х. Объектнинг умумлашган баҳосини тавсифлашда лингвистик ўзгарувчилардан фойдаланиш// Амалий математика ва информатсион технологияларнинг долзарб муаммолари – Ал-Хоразмий 2012” Халқаро илмий анжуман материаллари, Т.2, Тошкент,2012. 111-113 б.

29. Madraximov Sh.F., Xurramov A.H. Umumlashgan baho asosida soha mutaxassislarining mulohazalarini qiyosiy tahlil qilish // Амалий математика ва информатсион технологияларнинг долзарб муаммолари – Ал-Хоразмий 2016” Халқаро илмий анжуман маърузалар тўплами, Бухоро, 2016, Т.1, 204-207 б.

30. Мадрахимов Ш.Ф., Саидов Д.Ю. Выбор латентных признаков по результатам иерархической агломеративной группировки данных // Труды конференции Актуальные проблемы прикладной математики и информационных технологий –Аль Хорезми 2016, Бухара, 9-10 ноябрь 2016, Том 1, с. 88-91.

31. Мадрахимов Ш.Ф., Саидов Д.Ю. Группировка признаков по критерию устойчивости объектов классов // Актуальные проблемы прикладной математики, информатики и механики. Сборник трудов Международной научно-технической конференции, Воронеж, 12–15 сентября 2016 г. С. 93-95

32. Мадрахимов Ш. Ф., Саидов Д. Ю. Анализ структуры выборки при описании объектов в пространстве из латентных признаков// Актуальные проблемы прикладной математики, информатики и механики. Сборник трудов Международной научно-технической конференции, Воронеж, 18–20 декабря 2017 г. С. 1519-1522.

33. Adilova Z.T., Ignatev N.A. Madrakhimov Sh.F. Intrusion detection system with artificial neural network and expert system // Proceedings: Second world conference on intelligent Systems for industrial Automation, Tashkent, CIS,Uzbekistan June 4 - 5, 2002 - P. 226 - 228.

34. Adilova Z.T., Ignatev N.A. Madrakhimov Sh.F. Synthesis of neural network with optimal configuration // Inter. Conference of Interprise Information System, Setubai, Portugal, 2001.

35. Madrakhimov Sh.F. Extraction of knowledge from experimental data // II International Conference «Science and Technology in XXI Century», Abstracts. Tashkent: 18-22 november, 2003, P.100-101.

36. Adilova F.T., Ignat'ev N.A., Madrakhimov Sh.F. The Approach to Individualized Teleconsultations of Patients with Arterial Hypertension // Global Telemedicine and eHealth Updates: Knowledge Resources, Vol. 3, 2010. –P.372-375.

37. Madraximov Sh. Censoring the sample by the stability of objects in classes // The 4th International Conference on BIG DATA APPLICATIONS AND SERVICES. 15-18 august 2017– Tashkent, Vol.4, № 1, p. 207-209.

38. Madrakhimov Sh.F., Khurramov A.Kh. Visualization of the structure of location of sample objects based on generalized estimates // Abstracts of the VI international scientific conference «Modern problems of the applied mathematics and information technology –Al-Khorezmiy 2018», NUU, Tashkent, September 13-15, 2018, p.28-29.

39. Madraximov Sh. Stochastic and deterministic methods for calculating generalized estimates of objects // Abstracts of the Joint International Conference «STEMM – Science – Technology – Education – Mathematics – Medicine», Bukhara – Samarkand – Tashkent, May 13–17, 2019, p.200.

40. Мадрахимов Ш.Ф., Суюнов Н.Д. Свидетельство об официальной регистрации программ для ЭВМ «Вычисление обобщённой оценки принадлежности лекарственных средств в VEN группам» // № DGU 02526. Агентство по интеллектуальной собственности РУз, -Ташкент. 18.06.2012 г.

41. Мадрахимов Ш.Ф. Свидетельство об официальной регистрации программ для ЭВМ «Системы обнаружения скрытых закономерностей на базе методов вычисления обобщенных оценок» // № DGU 04460. Агентство по интеллектуальной собственности РУз, -Ташкент. 09.06.2017 г.

Автореферат «Информатика ва энергетика муаммолари» илмий журнали тахририятида тахрирдан ўтказилди ва ўзбек, рус, инглиз тилларидаги матнларини мослиги текширилди.

Бичими 60x84¹/₁₆. Рақамли босма усули. Times гарнитураси.
Шартли босма табағи 3,5. Адади 100 нусха. Буюртма № 20.

Гувоҳнома реестр №10-3719.
«Тошкент кимё-технология институти» босмаҳонасида чоп этилди.
100011, Тошкент, Навоий кўчаси, 32-уй.