

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.27.06.2017.FM.01.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ

САИДОВ ДОНИЁР ЮСУПОВИЧ

АНГЛАШ МАСАЛАЛАРИДА АЛОМАТЛАР ФАЗОСИНИНГ
НОЧИЗИҚЛИ АЛМАШТИРИШЛАРИГА АСОСЛАНГАН АХБОРОТ
МОДЕЛЛАРИ

05.01.03 – Информатиканинг назарий асослари

ФИЗИКА-МАТЕМАТИКА ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент шаҳри–2017 йил

**Физика-математика фанлари бўйича фалсафа доктори (PhD)
диссертацияси автореферати мундарижаси**

**Оглавление автореферата диссертации
доктора философии (PhD) по физико-математическим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD)
on physical-mathematical sciences**

Саидов Дониёр Юсупович

Англаш масалаларида аломатлар фазосининг ночизиқли
алмаштиришларига асосланган ахборот моделлари 3

Саидов Дониёр Юсупович

Информационные модели на основе нелинейных преобразований
признакового пространства в задачах распознавания..... 17

Saidov Doniyor Yusupovich

Information models based on nonlinear transformations of feature space in
pattern recognition..... 31

Эълон қилинган ишлар рўйхати

Список опубликованных работ
List of published works 35

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.27.06.2017.FM.01.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ

САИДОВ ДОНИЁР ЮСУПОВИЧ

АНГЛАШ МАСАЛАЛАРИДА АЛОМАТЛАР ФАЗОСИНИНГ
НОЧИЗИҚЛИ АЛМАШТИРИШЛАРИГА АСОСЛАНГАН АХБОРОТ
МОДЕЛЛАРИ

05.01.03 – Информатиканинг назарий асослари

ФИЗИКА-МАТЕМАТИКА ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент шаҳри–2017 йил

Физика-математика фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2017.3.PhD/FM103 рақам билан рўйхатга олинган.

Диссертация Мирзо Улуғбек номидаги Ўзбекистон Миллий Университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгаш веб-саҳифаси (<http://ik-fizmat.nuu.uz/>) ва «ZiyoNet» Ахборот таълим порталида (www.ziynet.uz) жойлаштирилган.

Илмий раҳбар:	Игнатъев Николай Александрович физика-математика фанлари доктори, профессор
Расмий оппонентлар:	Усманов Ришат Ниязбекович техника фанлари доктори, профессор Ибрагимов Фарходжон Нурмухаммаджонович физика-математика фанлари номзоди
Етакчи ташкилот:	Ислом Каримов номидаги Тошкент давлат техника университети

Диссертация ҳимояси Ўзбекистон Миллий университети ҳузуридаги DSc.27.06.2017.FM.01.02 рақамли Илмий кенгашнинг «___»_____ 2017 йил соат___ даги мажлисида бўлиб ўтади. (Манзил: 100174, Тошкент ш., Олмазор тумани, Университет кўчаси, 4-уй. Тел.: (+99871) 227-12-24, факс: (+99871) 246-53-21, 246-02-24, e-mail: nauka@nuu.uz).

Диссертация билан Ўзбекистон Миллий университетининг Ахборот-ресурс марказида танишиш мумкин (___ рақами билан рўйхатга олинган). (Манзил: 100174, Тошкент ш., Олмазор тумани, Университет кўчаси, 4-уй. Тел.: (+99871) 246-02-24).

Диссертация автореферати 2017 йил «___» _____ куни тарқатилди.
(2017 йил «___» _____ даги _____ рақамли реестр баённомаси).

А. Р. Марахимов

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

З. Р. Рахмонов

Илмий даражалар берувчи Илмий кенгаш илмий котиби, ф.-м.ф.д.

Р.Ж. Алоев

Илмий даражалар берувчи Илмий кенгаш қошидаги илмий семинар раиси, ф.-м.ф.д., профессор

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳонда, рақамли технологиялардан фойдаланиш товарлар ишлаб чиқариш ва хизматлар самарадорлигини оширишнинг муҳим омили бўлиб ҳисобланмоқда. Ушбу технологияларнинг ривожланишида сунъий интеллект (СИ) усуллари муҳим роль касб этади. СИ усулларининг салмоқли қисми берилганларни интеллектуал таҳлили (БИТ) учун қўлланилмоқда. Берилганлардан БИТ ёрдамида ажратиб олинган яширин қонуниятлар ахборот моделлари учун янги билимлар ҳисобланиб, улар асосида заиф тузилмали предмет соҳалардаги қарор қабул қилиш жараёнлари асослаб берилмоқда.

Жаҳон амалиёти шуни кўрсатадики, заиф тузилмага эга предмет соҳалар учун ахборот моделларини қуришга уриниш аломатлар фазосининг катта ўлчамлиги ва мантиқий қонуниятларни қидириш учун мўлжалланган алгоритмларнинг юқори комбинаторик мураккаблиги сабабли самарасиз бўлиб қолмоқда. Шунингдек, англаш масаласини ечиш нуқтаи-назаридан, дастлабки аломатлар ва объектлар тўпламига нисбатан яхши сифат кўрсаткичларига эга информатив аломатлар тўпамини ва ўргатувчи объектларнинг қисм тўпамини қидириш муаммолари бўйича тақдиротлар давом этмоқда. Тақдиротлар шуни кўрсатадики, англаш тизимларининг сифатини ошириш йўлларида бири – объектларни тавсифлаш учун аниқланадиган аломатлар тўпамини сон ўқиға нозикли акслантириш йўли билан латент аломатларни синтез қилиш усули ҳисобланади. Берилганлар базасидан (сақлагичдан) яширин қонуниятларни ажратиб олувчи классификация алгоритмларининг аниқлигини ошириш мазкур тақдиротнинг долзарблигини аниқлайди.

Мустақиллик йилларида мамлакатимизда фундаментал ва амалий татбиққа эга бўлган долзарб илмий йўналишларга эътибор кучайтирилди. Хусусан, билимларни эгаллаш масалаларида билимларни тақдим этиш априор берилган бўлиб, фақат ушбу тақдимот доирасида тизимни қуриш кераклиги кўзда тутилса, метаўргатиш масалаларида эса тақдимотни ўзини автоматик тақдим этиш масаласи қўйиладики, унинг деталлари предмет соҳага боғлиқ равишда кескин ўзгариши мумкин. Метаўрганиш масаласини ечишнинг зарурати, машинали тизимлар учун тор предмет соҳада амал қилиш қобилиятлари билан боғлиқ чекловларни олиб ташлаш билан изоҳланади. Шунингдек, Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси асосида илмий-тақдирот ва инновация ютуқларини амалиётга жорий этиш маҳанизмларидан иқтисодиёт тармоқларининг самарадорлигини оширишда фойдаланиш муҳим аҳамиятга эга ҳисобланади.

Ўзбекистон Республикаси Президентининг 2008 йил 15 июлдаги ПҚ-916-сон «Инновацион лойиҳалар ва технологияларни ишлаб чиқаришга татбиқ этишни рағбатлантириш борасидаги қўшимча чора-тадбирлар тўғрисида»ги, 2017 йил 17 февралдаги ПҚ-2789-сон «Фанлар академияси

фаолияти, илмий-тадқиқот ишларини ташкил этиш, бошқариш ва молиялаштиришни янада такомиллаштириш чора-тадбирлари тўғрисида»ги Қарори ва 2017 йил 8 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича ҳаракатлар стратегияси тўғрисида»ги фармони ҳамда мазкур фаолиятга тегишли бошқа норматив-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Диссертация республика фан ва технологиялар ривожланишининг IV. «Математика, механика ва информатика» устувор йўналиши доирасида бажарилган.

Муаммонинг ўрганилганлик даражаси. Умумлашган қарор қилиш функциялари ёрдамида ночизикли классификаторлардан фойдаланишга бир неча бор ҳаракатлар бўлган. Амалда бу муаммо бошланғичга нисбатан катта бўлган янги аломатлар фазосига ўтиш орқали бажарилган. В.Н. Вапник ва Червоненкислар томонидан аломатлар фазоси ўлчамининг ўсишига мос равишда объектлар танланмасини синфларга тўғри ажратиш эҳтимоллигининг ошиши исбот қилинган.

Англаш усулларида фойдаланиш назарияси ва амалиётини ривожланишида чет эл ва юртимиз олимлари улкан ҳиссаларини қўшишган. Чет эл олимлари орасида Ю.И.Журавлев, Н.Г.Загоруйко, В.А.Дюк, К.В.Воронцова, В.Н.Вапникларни, юртимиз олимлари ичида М.М.Камилов, Т.Ф.Бекмуратов, Ф.Т.Адилова, Ш.Х.Фазылов, Д.Т.Мухаммадиевалар тадқиқотларини алоҳида қайд этиш мақсадга мувофиқдир.

Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқоти Ўзбекистон Миллий университетининг илмий-тадқиқот ишлари режасига мувофиқ Ф-4-64 «Берилганларни интеллектуал таҳлилида умумлашган баҳоларни ҳисоблаш ва объектларнинг индивидуал метрикасига асосланган усулларни ишлаб чиқиш ва асослаш» (2011-2016) илмий тадқиқот лойиҳаси доирасида бажарилган.

Тадқиқотнинг мақсади англаш масалаларида аломатлар фазосининг ночизикли алмаштиришларига асосланган сунъий интеллект мезонлари ва усулларини такомиллаштиришдан иборат.

Тадқиқотнинг вазифалари қуйидагилардан ташкил топган:

аломатларни гуруҳлаш, ҳамда уларнинг объектлар тавсифидаги қийматларини сон ўқига чизикли ва ночизикли акслантириш учун агрегатив иерархик усулларни ишлаб чиқиш ва асослаш;

чизикли ва ночизикли усуллардан фойдаланилган ҳолда берилганларни визуаллаштириш. Визуаллаштириш сифатини баҳолаш мезонини ишлаб чиқиш;

объектлар латент аломатлар билан тавсифланган ҳолатларда англаш алгоритмларининг умумлашириш қобилиятини баҳолаш;

дастлабки ишлов бериш инобатга олинган ва олинмаган ҳолатлар учун синфларга ажратилган объектларни тавсифловчи аломатлар қийматларини

ўзаро кесишмайдиган интервалларга ажратиш алгоритмининг мураккаблигини баҳолаш;

иерархик агломератив гуруҳлаш, берилганларни визуаллаштириш ва англаш алгоритмларининг умумлаштириш қобилиятини ҳисоблаш учун дастурий таъминот яратиш ва уни заиф тузилмага эга предмет соҳалар учун ахборот моделлар яратишда қўллаш.

Тадқиқотнинг объекти образларни англаш алгоритмларининг моделларини ишлаб чиқиш ва асослашдан иборат.

Тадқиқотнинг предмети – англаш масалаларидаги мумкин бўлган объектлар учун аломатлар фазосини танлаш методларидан иборат.

Тадқиқотнинг усуллари. Тадқиқот ишида дискрет математика, сунъий интеллект, математик таҳлил, алгоритмик тилларда дастурлаш асосида англашнинг оптимал алгоритмларини излаш усулларида фойдаланилган.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

берилганлар танланмасига асосланган ночизикли қарор қилиш қоидаларини шакллантириш усули такомиллаштирилган;

компактлик мезони асосида визуаллаштириш натижаларини қиёсий таққослаш усули тавсифланган бўлиб, компактлик қиймати ҳар бир синф объектларини минимал сондаги ўзаро кесишмайдиган гуруҳларга ажратиш натижалари бўйича аниқланган;

берилганларда тушуриб қолдирилган ва такрорланувчи қийматлар инобатга олинган ҳолда ҳамда берилганларни дастлабки ишлов беришни инобатга олган ва олмаган ҳолда миқдорий аломатларни ўзаро кесишмайдиган интервалларга бўлиш алгоритмининг мураккаблик баҳоси олинган;

объектлар тавсифидаги бошланғич аломатлар фазосидан латент аломатлар фазосига ўтилганда яқин қўшни алгоритмининг умумлаштириш қобилиятининг ошганлиги исботланган;

аломатларни агломератив иерархик гуруҳлаш ва уларни сон ўқиға ночизикли акслантиришда информативлиги бўйича тартибланган латент аломатлар тўпламларининг шаклланиши кўрсатилган.

Тадқиқотнинг амалий натижаси қуйидагилардан иборат:

объектлар тавсифини сон ўқиға чизикли ва ночизикли акслантиришдаги визуаллаштириш натижаларини баҳолаш усули ишлаб чиқилган;

аломатларни агломератив иерархик гуруҳлаш орқали англаш алгоритмининг умумлаштириш қобилиятини пасайтирмасдан ўлчанадиган бошланғич аломатлар сони камайтирилган;

аломатларни сон ўқиға акслантириш ва улар учун бўлиш мезонининг оптимал қийматидан фойдаланган ҳолда ҳисобланадиган чизикли дискрименант функциянинг бўсаға қийматини қўллаш натижасида англаш алгоритмларининг аниқлиги ва умумлаштириш қобилияти оширилган;

объектлар тавсифидаги аломатлар тўпламини сон ўқиға чизикли ва ночизикли акслантириш асосида латент аломатларни танлаш усуллари билан англаш масалаларини ечиш бўйича дастурий мажмуа ишлаб чиқилган.

Тадқиқот натижаларининг ишончлилиги. Олинган натижаларнинг ишончлилиги, модел ва тест берилганларда ўтказилган ҳисоблаш тажрибаларининг натижалари аломатлар фазосини ночизикли алмаштириш асосидаги алгоритмларнинг юқори умумлаштириш қобилияти билан тасдиқланади, латент аломатлар ёрдамида объектларни визуаллаштириш натижалари бўйича ўргатувчи танланманинг компактлик кўрсаткичи ошганлиги билан исботланган.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти англаш алгоритмларининг умумлаштириш қобилиятини оширувчи, заиф формаллашган предмет соҳалардаги берилганлар базаси ва сақлагичларидан яширинган билимларни ажратиб олиш имконини берувчи, мумкин бўлган объектларни янги аломатлар фазосини шакллантиришнинг чизикли ва ночизикли усулларини ривожлантириш, ҳамда ночизикли қоидаларни аналитик кўринишини келтириб чиқариш асосида қарор қабул қилишни асослаш, аломатлар фазосини визуаллаштириш ва компактлик мезони бўйича унинг баҳосини бериш, информатив аломатларни танлаб олишдан иборат.

Тадқиқот натижаларининг амалий аҳамияти заиф формаллашган предмет соҳалардан янги билимларни олишга имкон берувчи, таклиф қилинган алгоритмлар сифатини баҳолаш мезонлари асосида ахборот моделларини куриш учун дастурий мажмуадан фойланишдан иборат.

Тадқиқот натижаларининг жорий қилиниши. Яратилган тадқиқот усуллари ва алгоритмлари асосида:

аломатларни иерархик агломератив гуруҳлаш ва уларни сон ўқиға ночизикли акслантириш усули Ўзбекистон Республикаси Ер ресурслари, геодезия, картография ва давлат кадастри давлат қўмитасида тупроқ шўрланишининг ахборот моделини куришда фойдаланилган (Ўзбекистон Республикаси Ер ресурслари, геодезия, картография ва давлат кадастри давлат қўмитасининг 2017 йил 20 октябрдаги 03-05-8912–сон маълумотномаси). Илмий натижаларнинг қўлланилиши тупроқ шўрланишининг учта тури бўйича аломатларнинг ночизикли боғланишларини аниқлаш орқали яширин қонуниятларни топиш имконини берган;

ўргатувчи танланмадан шовқин объектларни қидириш ва уларни ўчириш орқали берилганларни селекция қилиш, компактлик ўлчовининг қийматларини ҳисоблаш ва берилганларни визуаллаштириш усуллари Ўзбекистон Республикаси Ер ресурслари, геодезия, картография ва давлат кадастри давлат қўмитасида тупроқ шўрланиши мониторинг қилишда қўлланилган (Ўзбекистон Республикаси Ер ресурслари, геодезия, картография ва давлат кадастри давлат қўмитасининг 2017 йил 20 октябрдаги 03-05-8912–сон маълумотномаси). Ушбу усулларнинг қўлланилиши тупроқ шўрланиши жараёнида миқдорий ва номинал аломатлар вазнларини аниқлаш ва объектлар ўртасидаги ўзаро муносабатни визуаллаштириш ҳамда аномал объектларни аниқлаш имконини берган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижалари 8 илмий-амалий анжуманларда, жумладан 5 та халқаро ва 3 та республика илмий-амалий анжуманларида муҳокамадан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Тадқиқот мавзуси бўйича жами 20 та илмий иш чоп этилган, шулардан, Ўзбекистон Республикаси Олий Аттестация комиссиясининг фалсафа доктори диссертациялари асосий илмий натижаларини чоп этиш тавсия этилган илмий нашрларда 8 та мақола, жумладан, 3 таси хорижий ва 5 таси республика журналларида нашр этилган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, учта боб, хулоса, фойдаланилган адабиётлар ва 3 та иловалардан ташкил топган. Диссертациянинг ҳажми 93 бетдан иборат.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида диссертация мавзусининг долзарблиги ва зарурати, тадқиқотнинг республика фан ва технологияларни ривожлантиришнинг устувор йўналишларига мос келиши асосланган, диссертация мавзусига бўйича чет элдаги илмий тадқиқотларнинг қисқача маълумоти ва муаммонинг ўрганилганлик даражаси келтирилган, тадқиқотнинг мақсад, вазибалари шакллантирилган, унинг объекти ва предмети кўрсатилган, тадқиқотнинг амалий натижалари ва илмий янгиликлари баён қилинган, олинган натижаларнинг назарий ва амалий аҳамияти очиқ берилган, тадқиқот натижаларининг қўлланилиши, диссертация тузилиши ва нашр қилинган илмий ишлар тўғрисида маълумотлар келтирилган.

Биринчи бобда англаш алгоритмларининг параметрларини ҳисоблаш учун ва объектларни тавсифлашда фазони танлашда аломатлар қийматларини ўзаро кесишмайдиган интервалларга ажратишдан фойдаланиш тавсифланган.

Синф объектларининг чизиқли ажралиш хусусияти яримтекислик шаклидаги мантиқий қонуният нуқтаи назаридан қаралган. Мантиқий қонуниятлар асосида масалани ечиш алгоритмлари ғоят катта комбинаторик мураккабликка эгаллиги билан характерланади. Мураккабликни камайтириш учун дастлабки ишлов бериш ва айрим эвристикалардан фойдаланиш таклиф этилади.

Ўзаро кесишмайдиган l та K_1, \dots, K_l синф вакилларида ташкил топган $E_0 = \{S_1, \dots, S_m\}$ объектлар тўплами берилган бўлсин. Объектлар n та турли тоифадаги аломатлар тўплами $X_n = (x_1, \dots, x_n)$ билан тасвирланган бўлиб, улардан δ ($\delta < n$) таси номинал шкалада, $n - \delta$ таси миқдорий шкалада ўлчанади. Берилганларда тушириб қолдирилган ва такрорланувчи қийматлар бўлиши мумкин. Миқдорий аломатлар (дастлабки ва латент) қийматларини ўзаро кесишмайдиган интервалларга ажратиш учун $F(*)$ мезон берилган деб ҳисобланади. Латент аломатлар миқдорий ва номинал аломатларнинг

комбинацияси орқали тасвирланиши мумкин. $F(*) = extr$ шартида l та ораликнинг чегаравий қийматларини аниқлаш талаб этилади.

Айтайлик, $u_i^p - (c_{2p-1}, c_{2p}]$ ораликдаги K_i синфнинг $x_j, j \in I, |I| = n - \delta$ миқдорий аломатнинг ўлчанган қийматлари тўплами, $A = (a_0, \dots, a_l), a_0 = 0, a_l = m_j, m_j - x_j$ бўйича қиймати тушириб қолдирилмаган объектлар сони ($2l \leq m_j \leq m$), $\alpha_{ij} - x_j$ аломатнинг K_i синфдаги такрорланувчи қийматлари сони, $a_p - E_0$ даги x_j аломат қийматларинини r_1, \dots, r_{m_j} ўсувчи кетма-кетлигидаги интервалнинг ўнг чегараси $c_{2p} = r_{a_p}$. аниқловчи элементнинг тартиб рақами бўлсин.

Қуйидаги мезон

$$\left(\frac{\sum_{p=1}^l \sum_{i=1}^l u_i^p (u_i^p - 1)}{\sum_{i=1}^l |\alpha_{ij}| (|\alpha_{ij}| - 1)} \right) \left(\frac{\sum_{p=1}^l \sum_{i=1}^l u_i^p (m_j - |\alpha_{ij}| - \sum_{t=1}^l u_i^t + u_i^p)}{\sum_{i=1}^l |\alpha_{ij}| (m_j - |\alpha_{ij}|)} \right) \rightarrow \max_{\{A\}} \quad (1)$$

$\{(c_{2p-1}, c_{2p}]\}, p = \overline{1, l}$ интерваллар чегараларининг оптимал қийматларини ҳисоблашга ва улардан миқдорий аломатларнинг номинал ўлчов шкаласида градацияларини аниқлаш учун фойдаланишга имкон беради.

Мезонни ҳисоблашдаги мураккабликни камайтиришга олиб келувчи берилганларни дастлабки ишлов беришдан фойдаланиш келтирилган. Тушириб қолдирилган ва такрорланувчи қийматлар бўлган ҳолда берилганларни дастлабки ишлов беришни инобатга олган ҳолдаги алгоритм мураккаблигини баҳолаш амалга оширилди. 1-расмда такрорланувчи қийматлар мавжуд ҳолат учун (1) мезонни ҳисоблаш алгоритмининг мураккаблигини ўзгариши кўрсатилган.



Расм 1. Аломатнинг такрорланувчи қийматларидаги мураккабликнинг ўзгариши.

Аломатларнинг ночизикли алмаштирилиши, одатда объектларни дастлабки фазога нисбатан каттароқ ўлчамга эга бўлган фазода (умумлашган фазода) тавсифлашга олиб келади. Умумлашган фазода информатив аломатларни танлаш олиш ва алгоритмларнинг умумлаштириш қобилиятларини ошириш учун бир қатор эвристикалар таклиф қилинган.

Айтайлик m_1, m_2 - аломатларнинг K_1, K_2 синфлардаги математик кутилмалар вектори бўлсин. Ҳар бир $x_i = X(u)$ аломат учун m_1, m_2 бўйича синф ичидаги ўхшашлик θ_i ва синфлараро фарқланиш γ_i қийматлари ҳисобланади.

Ушбу

$$\theta_j / \gamma_j \quad (2)$$

муносабат аломатларни синфларнинг математик кутилмалари атрофида уларнинг тарқалиш зичлиги бўйича тартибланиш ва баҳолаш имконини беради.

(2) даги каби ҳар бир $x_r \in X(n)$ аломат бўйича синф ичидаги ўхшашликни ҳисоблаш ушбу мезонида амалга оширилади:

$$\frac{|m_1 - m_2|^2}{\tilde{s}_1 + \tilde{s}_2}, \quad (3)$$

бу ерда $\tilde{s}_1 + \tilde{s}_2 = \theta_r$ синф ичидаги объектларнинг бир-биридан қанчалик узок жойлашганликлари йиғиндисини, $m_1 - m_2$ эса сонлар ўқидаги K_1 ва K_2 синфлар математик кутилмалари орасидаги фарқни ифодалайди.

Фишер мезони $\Phi(w) = \frac{|m_1 - m_2|^2}{\tilde{s}_1 + \tilde{s}_2} \rightarrow \max$, $Y(p) = (y_1, \dots, y_p)$, $p > n$ да чизикли

қарор қилувчи функциянинг (ЧҚҚФ) аломатлар жуфтликларини танлаш учун фойдаланилади. Бу ерда $(y_i, y_j) \subset Y(p)$, $X(n)$ даги аломатлар комбинацияси.

Муҳит табиати тўғрисидаги ҳар қандай фаразларни инобатга олмасдан ЧҚҚФ бўсаға қийматини танлаш усули қўлланилди. Бўсаға қиймати (1) даги $[c_1, c_2], (c_2, c_3]$ интервалнинг c_2 чегара қиймати бўйича қуйидагича ҳисобланди:

$$w_0 = (c_2 + u(S)) / 2 \quad (4)$$

бу ерда $u(S) = w_1 s_i + w_2 s_j$, ва $S = (s_1, \dots, s_p)$, $u(S) \in (c_2, c_3]$ - сон ўқидаги c_2 га яқин бўлган E_0 даги объектдир. Бўсаға қиймати (4) бўйича танланганда ЧҚҚФнинг умумлаштириш қобилияти, танланганинг нормал тарқалиши тўғрисидаги фараз бўйича танланган бўсаға қийматга нисбатан кўп ҳолларда юқори бўлиши ҳисоблаш эксперименти орқали исботланган.

Иккинчи бобда аломатлар жуфтликларини сон ўқида ночизикли акслантириш асосланган иерархик агломератив гуруҳлаш таклиф этилган. Гуруҳлаш натижалари бўйича аломатлар фазосини танлаш ва дастлабки аломатларнинг информатив тўплам остиларини ажратиб олиш муаммолари ечилади. Объектларни гуруҳлашдан мақсад – аломатлар фазосининг турли хилма-хилликлардаги кўринишларида компактлик ҳақидаги гипотезанинг

ростлигини текшириш ҳисобланади. §3.1 да аломатларни сон ўқида чизиқли акслантириш масаласи қаралади. Латент аломатларни чизиқли ва ночизиқли ҳисоблаш усуллари орасидаги фарқланишлар 1-жадвалда келтирилган.

1-жадвал.

Чизиқли ва ночизиқли усуллар орасидаги фарқланишлар

№	Латент аломатларни ҳисоблаш усули	
	Чизиқли	Ночизиқли
1.	Аломатлар гуруҳини шакллантириш иерархик агломератив гуруҳлаш қоидалари бўйича амалга оширилади. Ҳар бир гуруҳ аломатлари битта латент аломатга чизиқли акслантирилади.	Дастлабки аломатлардан латент аломатларни кетма-кет шакллантириш иерархик агломератив гуруҳлаш қоидалари бўйича амалга оширилади.
2.	Информатив латент аломатларни саралаш объектлар ўртасидаги яқинлик ўлчовини танлашга боғлиқ.	Информатив латент аломатларни танлаб олиш усули мавжуд. Усул аломатлар ўлчов масштабига ва объектлар орасидаги яқинлик ўлчовига нисбатан инвариант.

Ночизиқли алмаштириш миқдорий аломатлар учун қандай бўлса турли тоифадаги аломатлар учун ҳам худди шандай амалга оширилади. Миқдорий аломатлар учун масаланинг қўйилиши қуйидаги кўринишга эга.

Стандарт тарздаги икки синфли англаш масаласи қаралади. $E_0 = \{S_1, \dots, S_m\}$ танланма объектлари K_1 ёки K_2 ($E_0 = K_1 \cup K_2$) синфлардан бирига тегишли бўлиб, n та миқдорий аломатлар $X(n) = (x_1, \dots, x_n)$ билан тавсифланади. E_0 да $X(n)$ тўпламни ўзаро кесишмайдиган $X_1(k_1), \dots, X_\tau(k_\tau)$, $\tau \geq 1$, $k_1 + \dots + k_\tau \leq n$ қисм тўпламларга кетма-кет равишда ажратиш қоидаси берилган бўлсин. Қуйидагилар талаб этилади:

– ҳар бир $X_i(k_i)$ учун $S_j \in E_0$, $j = \overline{1, m}$ объект тавсифидаги $X_i(k_i)$ аломатлар қийматларини сон ўқидаги қийматга (умумлаширувчи баҳога) кетма-кет акслантирувчи A_i алгоритм (Ю.И. Журавлёвнинг образларни англашга алгебраик ёндошув терминларидаги англовчи оператор) аниқлансин;

– A_i англаш операторини ҳисоблаш учун аналитик кўринишни (формулани) келтириб чиқариш.

Айтайлик $u_i^1, u_i^2 = x_j^p$, $j \in I$, $I = \{1, 2, \dots, n\}$ аломатнинг K_i , $i = 1, 2$ синфга тегишли ва мос равишда $[c_1^{jp}, c_2^{jp}]$, $(c_2^{jp}, c_3^{jp}]$, $|K_i| > 1$ интервалларидаги қийматлари сони, $\mathcal{G} - E_0$ даги x_j^p нинг қийматларининг ўсувчи $r_{j1}, \dots, r_{jg}, \dots, r_{jm}$ кетма-кетлигидаги $c_1^{jp} = r_{j1}, c_2^{jp} = r_{jg}, c_3^{jp} = r_{jm}$ интерваллар чегараларини аниқловчи элементининг тартиб номери бўлсин.

Қуйидаги мезон

$$\left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (u_i^d - 1)}{\sum_{i=1}^2 |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i}| - u_{3-i}^d)}{2|K_1||K_2|} \right) \rightarrow \max_{c_1^{jp} < c_2^{jp} < c_3^{jp}} \quad (5)$$

$[c_1^{jp}, c_2^{jp}], (c_2^{jp}, c_3^{jp}]$ интерваллар орасидаги чегаралар қийматини баҳолаш имконини беради. Мезоннинг экстремуми x_j^p аломатнинг ω_j^p ($0 \leq \omega_j^p \leq 1$) вазни сифатида ишлатилади.

$S_r = (a_{r1}^p, \dots, a_{r,n-p}^p)$, $S_r \in E_0$ объектнинг (x_i^p, x_j^p) , $0 \leq p < n$, $i, j \in \{1, \dots, n-p\}$, $i \neq j$ аломатлар жуфтликлари бўйича b_{rij}^p умумлашган баҳосининг қиймати қуйидагича ҳисобланади:

$$b_{rij}^p = \mu_{ij} \left(t_i \omega_i^p \left(\frac{a_{ri}^p - c_2^{ip}}{c_3^{ip} - c_1^{ip}} \right) + t_j \omega_j^p \left(\frac{a_{rj}^p - c_2^{jp}}{c_3^{jp} - c_1^{jp}} \right) \right) + (1 - \mu_{ij}) t_{ij} \omega_{ij}^p \left(\frac{a_{ri}^p a_{rj}^p - c_2^{ijp}}{c_3^{ijp} - c_1^{ijp}} \right), i, j \in I,$$

$$t_i, t_j, t_{ij} \in \{-1, 1\}, \mu_{ij} \in [0, 1],$$

бу ерда $\omega_i^p, \omega_j^p, \omega_{ij}^p$ – мос равишда x_i^p, x_j^p аломатлар қийматлари тўпламидан ва уларнинг $x_i^p x_j^p$ кўпайтмаларидан бўйича (5) асосида аниқланган вазнлар, $t_i, t_j, t_{ij} \in \{-1, 1\}$, $\mu_{ij} \in [0, 1]$ ларнинг қийматлари қуйидаги функционал экстремумидан танланади:

$$\varphi(p, i, j) = \frac{\min_{S_r \in K_1} b_{rij}^p - \max_{S_r \in K_2} b_{rij}^p}{\max_{S_r \in E_0} b_{rij}^p - \min_{S_r \in E_0} b_{rij}^p}. \quad (6)$$

Ҳисоблаш эксперименти гипертония касаллиги кўрсаткичлари асосидаги тиббиёт берилганларида ўтказилган бўлиб, K_1 (деярли соғломлар), K_2 (касаллар) синфларидан ташкил топган. 2-жадвалда дастлабки аломатларни биринчи гуруҳга кетма-кет бирлаштириш натижалари, (5) мезон бўйича аломатлар комбинациясидан умумлашган баҳолар қийматларни (латент аломатларни) синтез қилиш тартиби ва ичма-ич кавсларни ҳисобга олган ҳолдаги (6) бўйича синф объектлари орасидаги чекинишлар келтирилган.

Аломатларни гуруҳлаш натижалари

Гуруҳ номери	Аломатларни бирлаштириш кетма-кетлиги	(5) мезон қиймати	(6) бўйича чекиниш
1	(x_4, x_{20})	0.9672	-0.0659
	$((x_4, x_{20}), x_9)$	0.9672	-0.0325
	$((((x_4, x_{20}), x_9), x_{18}))$	1.0000	0.0055
	$(((((x_4, x_{20}), x_9), x_{18}), x_8))$	1.0000	0.0076
	$((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2))$	1.0000	0.0079
	$((((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2), x_{10}))$	1.0000	0.0101
	$(((((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2), x_{10}), x_{12}))$	1.0000	0.0105

Биринчи гуруҳ бўйича латент аломат қийматларини ҳисоблаш кетма – кетлигининг аналитик кўриниши:

$$y_1^1 = -0.006947([АДС] - 140) + 0.394923([ДИАСТОЛА] - 0.42) - 0.0041346([АДС] * [ДИАСТОЛА] - 68.2);$$

$$y_2^1 = 1.602699(y_1^1 - 0.0047) + 3.128708([QRS] - 0.08);$$

$$y_3^1 = 0.581198y_2^1 + 0.551857([СИСПОК] - 0.485) + 0.892018y_2^1 * [СИСПОК];$$

$$y_4^1 = 0,21174(y_3^1 + 0.0388) + 0.2756([QT] - 0.36) + 2.13634(y_3^1 * [QT] + 0.0124);$$

$$y_5^1 = 0.005789(y_4^1 * [Рост] + 1.9623); \quad y_6^1 = 0.310793y_5^1 * [ИЛП];$$

$$y_7^1 = 0.1804337y_6^1 * [КДР].$$

Визуаллаштириш натижаларини таҳлил қилиш учун синф компактлиги баҳоси каби тузилмавий тавсифдан фойдаланиш таклиф этилган.

$$\theta_i = \frac{\sum_{j=1}^{\mu} m_{ij}^2}{m_i^2}, \quad (7)$$

бу ерда μ - $K_i, i=1,2$ даги гуруҳлар сони, $m_i = |K_i|$, m_{ij} - K_i синфнинг j - гуруҳидаги объектлари сони. Ўргатувчи танланма бўйича компактликнинг ўртача баҳоси қуйидаги формула ёрдамида ҳисобланади:

$$F(E_0, \rho) = \frac{\sum_{i=1}^2 m_i \theta_i}{m}, \quad (8)$$

бу ерда m - танланма объектлари сони.

Учинчи бобда компактлик кўрсаткичи аломатлар фазосининг турли хил алмаштиришларни таҳлил қилиш учун ишлатилиши кўрсатилган.

Объектларнинг турғунлик қийматлари танланманинг тузилмавий хилма-хиллигининг ўлчови сифатида хизмат қилади. Ундан шовқин объектларни излаш ва ўчириш орқали алгоритмнинг умулаштириш қобилятини ошириш мақсадида фойдаланиш мумкин. Синф объектлари турғунлик

қийматларининг турли-туманлиги метрикани танлашга бевосита боғлиқ. Ушбу ишда қуйидагилар амалга оширилган:

- дастлабки аломатлар тўпламини иерархик агломератив гуруҳлаш қоидалари бўйича ўзаро кесишмайдиган гуруҳларга ажратиш ва унинг асосида латент аломатлар тўпламини шакллантириш;
- информатив латент аломатлар тўпламини саралаш;
- англаш алгоритмларининг умумлаштириш қобилиятини ошириш мақсадида шовқин объектларини аниқлаш ва ўчириш.

(5) мезон бўйича қурилган ораликлар учун $\alpha_1(p) = \left| \left\{ S_{d_i} \in K_p \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|$, $\alpha_2(p) = \left| \left\{ S_{d_i} \in K_{3-p} \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|$, $\theta_1(p) = \alpha_1(p) / |K_p|$ ва $\theta_2(p) = \alpha_2(p) / |K_{3-p}|$ ни аниқлаймиз. $X(k)$ аломатлар тўплам остиси бўйича $S_d \in K_p, p=1,2$ объект турғунлиги қуйидаги тарзда ҳисобланади:

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

Қандайдир $x_a \in X(n)$ аломатни $X(k), X(k) \subset X(n)$ тўпламга (гуруҳга) қўшиш учун қуйидаги функционалдан фойдаланилади:

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

Шовқин объектларни ўчириш ҳамда (7) ва (8) бўйича компактликнинг ошишининг боғланиши кўрсатилган. (8) кўрсаткич бўйича ночизиқли усулнинг чизиқли усулдан афзаллиги эксперимент орқали тасдиқланди.

ХУЛОСА

Билимларга асосланган ахборот моделлар учун интеллектуал таҳлил усуллари яратиш ва асослаш илмий тадқиқотларнинг долзарб йўналишларидан ҳисобланади. Аломатларнинг тузилмавий хилма-хиллигида объектлар ўртасидаги муносабатлар кластер таҳлил усулларида фойдаланган ҳолда аломатлар фазосини ночизиқли алмаштиришлар орқали тадқиқ қилинган. Ночизиқли алмаштиришлардан мақсад ўзаро кесишмайдиган синф объектларининг идеал компактлигига интилиш ҳисобланади. Компактликни қатъиймас мантиқ атамаларида изоҳлаш берилганлар базасидан яширин қонуниятларни излашда фойдаланилади.

Ушбу ишда олинган асосий натижалар:

1. Турли тоифадаги аломатларни сон ўқида чизиқли ва ночизиқли акслантириш учун аломатларни иерархик агломератив гуруҳлаш қоидалари ишлаб чиқилди. Чизиқли ва ночизиқли акслантириш ўртасида жиддий фарқ мавжудлиги кўрсатилди. Таклиф этилган қоидалар ёрдамида дастлабки аломатлар фазоси ўлчамини сезиларли даражада қисқартиришга мувофиқ бўлинди;

2. Ораликларнинг оптимал чегарасини ҳисоблаш учун берилганларни дастлабки ишлов бериш усуллари таклиф этилган. Чегара қийматларини

аниқлашда «Ҳар бир оралик чегараларида фақат битта синфнинг объектларни тавсифлашдаги аломатлар қийматлари ётади» гипотезасининг ростлигини текширишдан фойдаланилади. Берилганларда такрорланишлар ва тушириб қолдирилган қийматлар бўлган ҳолатларда қидириш учун алгоритмлар мураккаблигининг баҳоси аниқланди. Интерваллар чегараларининг қийматларидан аломатлар гуруҳи бўйича объектлар турғунликларини ҳисоблаш ва аломатни ночизикли комбинацияга қўшиш учун иерархик агломератив гуруҳлаш қоидаларида фойдаланилган;

3. Объектлар тавсифидаги аломатларининг $(0,1]$ ораликқа тегишли мумкин бўлган қийматлари тўплами билан берилган ўргатувчи танланманинг компактлигини баҳолаш усули таклиф қилинди. Аломатларни ночизикли алмаштиришдаги компактлик қиймати чизикли алмаштиришга ва дастлабки «хом» аломатлар фазосидагига нисбатан юқори бўлиши ҳисоблаш экспериментлари ёрдамида исботланди. Компактлик баҳоси ҳамда алгоритмнинг умумлаштириш қобилияти ўртасида боғлиқлик мавжудлиги кўрсатилди;

4. Аломатларни ночизикли алмаштиришдан қарор қабул қилиш қоидаларининг аналитик кўринишини келтириб чиқаришда фойдаланилган. Иерархик агломератив алгоритмни амалга ошириш жараёнида ушбу алмаштириш асосида олинган латент аломатлар информативлик муносабати бўйича тартибланди;

5. Сувғориладиган ерлар шўрланишининг мотивининг ва саратон касаллигини ташхис қилиш учун иккита ахборот моделлар қурилди. Модел доирасида ўргатувчи танланманинг шовқин объектлари аниқланди. Шовқин объектларни ўчиришдан олинган самара синфнинг ва бутун танланма компактлик баҳоси орқали кўрсатилди.

**НАУЧНЫЙ СОВЕТ DSc.27.06.2017.FM.01.02
ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ
НАЦИОНАЛЬНОМ УНИВЕРСИТЕТЕ УЗБЕКИСТАНА**

НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ УЗБЕКИСТАНА

САИДОВ ДОНИЁР ЮСУПОВИЧ

**ИНФОРМАЦИОННЫЕ МОДЕЛИ НА ОСНОВЕ НЕЛИНЕЙНЫХ
ПРЕОБРАЗОВАНИЙ ПРИЗНАКОВОГО ПРОСТРАНСТВА В ЗАДАЧАХ
РАСПОЗНАВАНИЯ**

05.01.03 – Теоретические основы информатики

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ ДОКТОРА ФИЛОСОФИИ (PHD)
ПО ФИЗИКО-МАТЕМАТИЧЕСКИМ НАУКАМ**

г.Ташкент – 2017 год

Тема диссертации доктора философии (PhD) по физико-математическим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за № B2017.3.PhD/FM103.

Диссертация выполнена в Национальном Университете Узбекистана имени Мирза Улугбека. Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице Научного совета (www.ik-fizmat@nuu.uz) и на Информационно-образовательном портале «Ziyonet» (www.ziyonet.uz).

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

Официальные оппоненты: **Усманов Ришад Ниязбекович**
доктор технических наук, профессор

Ибрагимов Фарходжон Нурмухамеджонович
кандидат физико-математических наук

Ведущая организация: **Ташкентский государственный технический университет имени Ислама Каримова**

Защита диссертации состоится «___» _____ 2017 года в ___ часов на заседании Научного совета DSc.27.06.2017.FM.01.02 при Национальном университете Узбекистана. (Адрес: 100174, г. Ташкент, Алмазарский район, ул. Университетская, 4. Тел.: (+99871)227-12-24, факс: (+99871) 246-53-21, e-mail: nauka@nuu.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Национального университета Узбекистана (зарегистрирована за №___). (Адрес: 100174, г. Ташкент, Алмазарский район, ул. Университетская, 4. Тел.: (+99871) 246-02-24).

Автореферат диссертации разослан «___» _____ 2017 года.
(протокол рассылки №___ от «___» _____ 2017 года).

А. Р. Марахимов

Председатель Научного совета по присуждению ученых степеней, д.ф.-м.н., профессор

З. Р. Рахмонов

Ученый секретарь Научного совета по присуждению ученых степеней, д.ф.-м.н.

Р.Ж. Алоев

Председатель научного семинара при Научном совете по присуждению ученых степеней,
д.ф.-м.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии(PhD))

Актуальность и востребованность темы диссертации. Использование цифровых технологий является локомотивом повышения эффективности производства товаров и услуг в мире. Большая роль в развитии этих технологий отводится методам искусственного интеллекта (ИИ). Значительная часть методов ИИ применяется для интеллектуального анализа данных (ИАД). Скрытые закономерности, извлекаемые из данных с помощью ИАД, являются новыми знаниями в информационных моделях, в рамках которых обосновывается процесс принятия решений для слабо структурированных предметных областей.

Мировая практика свидетельствует, что большая часть усилий построения информационных моделей в слабо структурированных предметных областях становятся неэффективными из-за большой размерности признакового пространства и высокой комбинаторной сложности алгоритмов для поиска логических закономерностей. Продолжается исследование проблемы поиска информативных наборов признаков и подмножества объектов обучения, которые обладают лучшим качеством в смысле решения задач распознавания чем исходные множества признаков и объектов. Одним из путей повышения качества систем распознавания является синтез латентных признаков для описания объектов путём нелинейного отображения определяемых наборов признаков на числовую ось. Повышение точности алгоритмов классификации, извлечение скрытых закономерностей из баз (хранилищ) данных определяет актуальность данного исследования.

В нашей стране в годы независимости большое внимание уделяется научным направлениям, имеющим фундаментальное и прикладное значение. В частности, если в задачах приобретения знаний подразумевается, что представление знаний является заданными априори и нужно лишь построить систему в рамках этих представлений, то в задачах метаобучения ставится вопрос об автоматическом представлении самих представлений, детали которых могут сильно меняться в зависимости от предметной области. Решение задач метаобучения необходимо для снятия следующего ограничения для машинных систем – их способности функционировать только в узкой предметной области. На основе Стратегии Действий по развитию Республики Узбекистан особенно большое значение приобретают эффективные механизмы внедрения научных и инновационных достижений в целях повышения эффективности в сфере экономики страны.

Эта диссертация, в определенной степени, служит осуществлению задач, обозначенных в Постановлениях Президента Республики Узбекистан №-ПП-916 «О дополнительных мерах по стимулированию внедрения инновационных проектов и технологий в производство» от 15 июля 2008 года, №-ПП-2789 «О мерах по дальнейшему совершенствованию деятельности Академии наук, организации, управления и финансирования научно-исследовательской деятельности» от 17 февраля 2017 года и

№-УП- 4947 «О стратегии действий по дальнейшему развитию Республики Узбекистан» от 8 февраля 2017 года а также в других нормативно-правовых актах по данной деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан IV. «Математика, механика и информатика».

Степень изученности проблемы. Попытки использования нелинейных классификаторов предпринимались неоднократно с помощью обобщенных решающих функций. На практике это был уход в новое признаковое пространство, размерность которого была выше чем у исходного. В.Н. Вапником и Червоненкисом было доказано, что с ростом размерности признакового пространства увеличивается вероятность корректного разделения классов выборки объектов.

В развитие теории и практики использования методов распознавания большой вклад внесли известные зарубежные и отечественные учёные. Среди зарубежных учёных можно выделить Журавлева Ю.И., Загоруйко Н.Г., Дюк В.А., Воронцова К.В., Вапника В.Н., среди отечественных особо следует отметить труды Камилова М.М., Бекмуратова Т.Ф., Адыловой Ф.Т., Фазылова Ш.Х., Мухамадиевой Д.Т.

Связь темы диссертации с научно-исследовательскими работами учреждением высшего образования, где выполнялась диссертация. Диссертационное исследование выполнено в рамках научного гранта согласно плану научно-исследовательских работ Национального университета Узбекистана Ф-4-64 «Разработка и обоснование методов вычисления обобщённых оценок и индивидуальных метрик объектов в интеллектуальном анализе данных» (2011-2016 гг.).

Целью исследования. Совершенствование критериев и методов искусственного интеллекта на основе нелинейных преобразований признакового пространства в задачах распознавания образов.

Задачи исследования состоит в следующем:

разработка и обоснование агломеративных иерархических методов группировки признаков для линейного и нелинейного отображения их значений в описании объектов на числовую ось;

визуализация данных с использованием линейного и нелинейного методов. Разработка критерия для оценки качества результатов визуализации;

оценка обобщающей способности алгоритмов распознавания при описании объектов с помощью латентных признаков;

оценка сложности алгоритма разделения значений признаков в описании классифицированных объектов на непересекающиеся интервалы с учётом и без учёта предобработки;

разработка комплекса программ для иерархической агломеративной группировки, визуализации данных и вычисления обобщающей способности

алгоритмов распознавания. Использование комплекса для построения информационных моделей в слабо структурированных предметных областях.

Объект исследования. Разработка и обоснование моделей алгоритмов распознавания образов.

Предмет исследования. Методы выбора признакового пространства для описания допустимых объектов в задачах распознавания.

Методы исследования. Поиск оптимальных алгоритмов распознавания на основе методов дискретной математики, теории искусственного интеллекта, математического анализа, программирования на алгоритмических языках.

Научная новизна исследования заключается следующем:

Совершенствован метод формирования нелинейных решающих правил на основе обучающих выборок данных;

описана методика сравнения результатов визуализации на основе критерия компактности. Значения компактности определяются по результатам разбиения объектов каждого класса на минимальное число непересекающихся групп;

получена оценка сложности алгоритма разбиения значений количественных признаков на непересекающиеся интервалы: с учетом пропусков и повторяющихся значений; с учётом и без учёта предобработки данных;

доказано повышение обобщающей способности алгоритма ближайший сосед при переходе от исходного описания объектов в пространство из латентных признаков;

показано, что при иерархической агломеративной группировке признаков и нелинейном отображении их на числовую ось формируются наборы латентных признаков, упорядоченные по отношению их информативности.

Практические результаты исследования заключаются в следующем:

разработан способ оценки результатов визуализации при линейном и нелинейном отображении описаний объектов на числовые оси;

значительно уменьшился число измеряемых исходных признаков без снижения обобщающей способности алгоритмов распознавания при использовании иерархической агломеративной группировки признаков;

применение порогов линейных дискриминантных функций, вычисленных с использованием оптимального значения критерия разбиения значений результатов отображения признаков на числовую ось повысил точность и обобщающую способность алгоритмов распознавания;

разработан программный комплекс по решению задач распознавания методами отбора латентных признаков на основе линейного и нелинейного отображений наборов признаков в описании объектов на числовую ось.

Достоверность результатов исследования. Достоверность полученных результатов обосновывается тем, что итоги вычислительных экспериментов на модельных и тестовых данных подтверждают высокую обобщающую способность алгоритмов распознавания на основе нелинейных

преобразований признакового пространства, доказывається повышение показателей компактности обучающих выборок по результатам визуализации объектов с помощью латентных признаков.

Научная и практическая значимость результатов исследования. Научная значимость результатов исследования заключается в развитии методов линейного и нелинейного формирования нового признакового пространства с описанием допустимых объектов, позволяющих повысить обобщающую способность алгоритмов распознавания, извлекать скрытые закономерности из баз и хранилищ данных в виде новых знаний из слабо формализованных предметных областей, обоснование принятия решений на основе вывода аналитического представления нелинейных правил, визуализации признакового пространства и его оценки по критерию компактности, отбора информативных наборов признаков.

Практическая значимость результатов исследования заключается в использовании программного комплекса для извлечения скрытых закономерностей из баз и хранилищ данных, позволяющих получить новые знания из слабо формализованных предметных областей, строить информационные модели принятия решений на основе предложенных критериев оценки качества алгоритмов.

Внедрение результатов исследования. На основе разработанных методов исследования:

метод иерархической агломеративной группировки признаков и нелинейное отображения их на числовую ось внедрен в Государственного комитета Республики Узбекистан по земельным ресурсам, геодезии, картографии и государственному кадастру для построения информационной модели засоленности почв (справка от 20 октября 2017 года №03-05-8912 Государственного комитета Республики Узбекистан по земельным ресурсам, геодезии, картографии и государственному кадастру). Предлагаемый в результате исследований метод отображения позволил выявить скрытые закономерности по трём типам засоленности почв путем отбора связанных нелинейной зависимостью признаков;

методы селекции данных посредством обнаружения и удаления шумовых объектов выборки, вычисления значений меры компактности и визуализации данных были применены в Государственного комитета Республики Узбекистан по земельным ресурсам, геодезии, картографии и государственному кадастру для мониторинга засоленности почв (справка от 20 октября 2017 года №03-05-8912 Государственного комитета Республики Узбекистан по земельным ресурсам, геодезии, картографии и государственному кадастру). В результате внедрение этих методов позволил определять веса номинальных и количественных признаков в процессе засоления почвы и визуализировать отношений между объектами, а также выявить аномальные объекты выборки.

Апробация результатов исследования. Результаты данного исследования были обсуждены, в том числе, на 5 международных и 3 республиканских научно-практических конференциях.

Публикация результатов исследования. По теме диссертации опубликовано 20 научных работ, из них 8 входят в перечень научных изданий, предложенных Высшей аттестационной комиссией Республики Узбекистан для защиты диссертаций доктора философии, в том числе из них 3 опубликованы в зарубежных журналах и 5 в республиканских научных изданиях.

Структура и объем диссертации. Структура диссертации состоит из введения, трех глав, заключения, списка использованной литературы и 3 приложений. Объем диссертации составляет 93 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность и востребованность темы диссертации, определено соответствие исследования приоритетным направлениям развития науки и технологий республики, приведены обзор зарубежных научных исследований по теме диссертации и степень изученности проблемы, сформулированы цели и задачи, выявлены объект и предмет исследования, изложены научная новизна и практические результаты исследования, раскрыта теоретическая и практическая значимость полученных результатов, даны сведения о внедрении результатов исследования, об опубликованных работах и о структуре диссертации.

В первой главе описывается использование разбиение значений признаков на непересекающиеся интервалы для вычисления параметров распознающих алгоритмов и выбора пространства в описании объектов.

Свойство линейной делимости объектов классов рассматривается с точки зрения логических закономерностей в форме полуплоскостей. Алгоритмы решения задач на базе логических закономерностей характеризуются огромной комбинаторной сложностью. Для уменьшения сложности предлагается использовать предобработку данных и некоторые эвристики.

Пусть задано множество объектов $E_0 = \{S_1, \dots, S_m\}$, содержащее представителей l непересекающихся классов K_1, \dots, K_l . Описание объектов производится с помощью набора из n разнотипных признаков $X_n = (x_1, \dots, x_n)$, $\delta (\delta < n)$ из которых измеряются в номинальной, $n - \delta$ в интервальной шкалах. Допускается наличие пропусков и повторяющихся значений в данных. Считается, что задан критерий $F(*)$ для разбиения значений количественного признака (как исходного так и латентного) на непересекающиеся интервалы. Требуется определить значения границ l интервалов при $F(*) = extr$.

Пусть u_i^p - множество измеренных значений количественного признака $x_j, j \in I, |I| = n - \delta$ класса K_i в интервале $(c_{2p-1}, c_{2p}]$, $A = (a_0, \dots, a_l)$, $a_0 = 0, a_l = m_j$, m_j - число объектов без пропусков ($2l \leq m_j \leq m$) по x_j , α_{ij} - число повторяющихся значений признака x_j в

классе K_i , a_p - порядковый номер элемента упорядоченной по возрастанию последовательности r_1, \dots, r_{m_j} значений x_j из E_0 , определяющий правую границу интервала $c_{2p} = r_{a_p}$.

Критерий

$$\left(\frac{\sum_{p=1}^l \sum_{i=1}^l u_i^p (u_i^p - 1)}{\sum_{i=1}^l |\alpha_{ij}| (|\alpha_{ij}| - 1)} \right) \left(\frac{\sum_{p=1}^l \sum_{i=1}^l u_i^p (m_j - |\alpha_{ij}| - \sum_{t=1}^l u_i^t + u_i^p)}{\sum_{i=1}^l |\alpha_{ij}| (m_j - |\alpha_{ij}|)} \right) \rightarrow \max_{\{A\}} \quad (1)$$

позволяет вычислять оптимальные значения границ интервалов $\{(c_{2p-1}, c_{2p})\}$, $p = \overline{1, l}$ и использовать их для определения градаций количественного признака в номинальной шкале измерений.

Описана предобработка данных для уменьшения сложности вычисления (1). Сделана оценка сложности алгоритма с учётом предобработки данных, наличия пропусков и повторяющихся значений. На рис. 1 показано изменение сложности алгоритма вычисления (1) при наличии повторяющихся значений.



Рис.1. Изменение сложности при наличии повторяющихся значений признака.

Нелинейные преобразования признаков, как правило, приводили к описанию объектов в пространстве (обобщённом пространстве) более высокой размерности, чем исходное. Предлагается ряд эвристик для отбора информативных признаков в обобщенном пространстве и повышения обобщающей способности алгоритмов.

Пусть m_1, m_2 - векторы математических ожиданий признаков из класса K_1, K_2 . Для каждого признака $x_i \in X(n)$ по m_1, m_2 вычисляются значения внутриклассового сходства θ_i и межклассового различия γ_i . Соотношение

$$\theta_j / \gamma_j \quad (2)$$

даёт возможность оценивать и упорядочивать признаки по плотности их распределения вокруг математических ожиданий классов.

Также как и в (2) вычисление внутриклассового сходства по отдельному признаку $x_r \in X(n)$ используется в критерии

$$\frac{|m_1 - m_2|^2}{\tilde{s}_1 + \tilde{s}_2}, \quad (3)$$

в котором сумма внутриклассового разброса $\tilde{s}_1 + \tilde{s}_2 = \theta_r$, а $m_1 - m_2$ есть разность математических ожиданий классов K_1 и K_2 на числовой оси.

Критерий Фишера $\Phi(w) = \frac{|m_1 - m_2|^2}{\tilde{s}_1 + \tilde{s}_2} \rightarrow \max$ использовался для выбора пар признаков линейной решающей функции (ЛРФ) из $Y(p) = (y_1, \dots, y_p)$, $p > n$, где $y_i \in Y(p)$ - комбинация признаков из $X(n)$.

Применялся способ выбора порога ЛРФ без всяких предположений о природе среды. Значение порога вычислялось по границе c_2 интервалов $[c_1, c_2], (c_2, c_3]$ по (1) как

$$w_0 = (c_2 + u(S)) / 2 \quad (4)$$

где $u(S) = w_1 s_i + w_2 s_j$, и $S = (s_1, \dots, s_p)$, $u(S) \in (c_2, c_3]$ - ближайший к c_2 объект E_0 на числовой оси. Доказано, что обобщающая способность ЛРФ в большинстве случаев выше при выборе порога по (4) чем при выборе порога в предположении о нормальном распределении выборки.

Во второй главе предлагается иерархическая агломеративная группировка на основе нелинейного отображения пар признаков на числовую ось. По результатам группировки решается проблема селекции признакового пространства и отбора информативных наборов исходных признаков. Целью группировки объектов является проверка истинности гипотезы о компактности при различных многообразиях представления признакового пространства. В §3.1 рассматривается задача о линейном отображении признаков на числовую ось. Различия между линейным и нелинейным методами вычисления латентных признаков приводятся в табл. 1.

Таблица 1.

Различия между линейным и нелинейным методами

№	Метод вычисления латентных признаков	
	Линейный	Нелинейный
1.	Формирование групп признаков производится по правилам иерархической агломеративной группировки. Признаки каждой группы линейно отображаются в один латентный признак.	Последовательное формирование латентных признаков из исходных производится по правилам иерархической агломеративной группировки.
2.	Отбор информативных латентных признаков зависит от выбранной меры близости между объектами.	Существует способ отбора информативных латентных признаков. Метод инвариантен масштабам измерения признаков и мерам близости между объектами.

Нелинейные преобразования используются как для количественных так и разнотипных признаков. Для количественных признаков постановка задачи выглядит так.

Рассматривается двухклассовая задача распознавания в стандартной постановке. Объекты выборки $E_0 = \{S_1, \dots, S_m\}$ принадлежит одному из классов K_1 или K_2 , ($E_0 = K_1 \cup K_2$) и описываются с помощью n количественных признаков $X(n) = (x_1, \dots, x_n)$. На E_0 задано правило последовательного разбиения набора $X(n)$ на непересекающиеся подмножества $X_1(k_1), \dots, X_\tau(k_\tau)$, $\tau \geq 1$, $k_1 + \dots + k_\tau \leq n$. Требуется:

- для каждого $X_i(k_i)$ определить алгоритм A_i (распознающий оператор в терминологии алгебраического подхода к распознаванию образов Ю.И. Журавлёва) для последовательного отображения значений признаков из $X_i(k_i)$ в описании объекта $S_j \in E_0$, $j = \overline{1, m}$ в значение (обобщённую оценку) на числовой оси;

- получить аналитический вид (формулы) для вычисления распознающего оператора A_i .

Пусть u_i^1, u_i^2 – количество значений признака x_j^p , $j \in I$, $I = \{1, 2, \dots, n\}$ класса K_i , $i = 1, 2$ соответственно в интервалах $[c_1^{jp}, c_2^{jp}]$, $(c_2^{jp}, c_3^{jp}]$, $|K_i| > 1$, g – порядковый номер элемента упорядоченной по возрастанию последовательности $r_{j1}, \dots, r_{jg}, \dots, r_{jm}$ значений x_j^p из E_0 определяющий границы интервалов как $c_1^{jp} = r_{j1}$, $c_2^{jp} = r_{jg}$, $c_3^{jp} = r_{jm}$. Критерий

$$\left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (u_i^d - 1)}{\sum_{i=1}^2 |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i}| - u_{3-i}^d)}{2|K_1||K_2|} \right) \rightarrow \max_{c_1^{jp} < c_2^{jp} < c_3^{jp}} \quad (5)$$

позволяет вычислять оптимальное значение границы между интервалами $[c_1^{jp}, c_2^{jp}]$, $(c_2^{jp}, c_3^{jp}]$. Экстремум критерия (5) используется в качестве веса ω_j^p ($0 \leq \omega_j^p \leq 1$) признака x_j^p .

Значение обобщённой оценки b_{rij}^p объекта $S_r = (a_{r1}^p, \dots, a_{r,n-p}^p)$, $S_r \in E_0$ по паре (x_i^p, x_j^p) , $0 \leq p < n$, $i, j \in \{1, \dots, n-p\}$, $i \neq j$, вычисляется как

$$b_{rij}^p = \mu_{ij} \left(t_i \omega_i^p \left(\frac{a_{ri}^p - c_2^{ip}}{c_3^{ip} - c_1^{ip}} \right) + t_j \omega_j^p \left(\frac{a_{rj}^p - c_2^{jp}}{c_3^{jp} - c_1^{jp}} \right) \right) + \\ + (1 - \mu_{ij}) t_{ij} \omega_{ij}^p \left(\frac{a_{ri}^p a_{rj}^p - c_2^{ijp}}{c_3^{ijp} - c_1^{ijp}} \right), i, j \in I,$$

$$t_i, t_j, t_{ij} \in \{-1, 1\}, \mu_{ij} \in [0, 1],$$

где $\omega_i^p, \omega_j^p, \omega_{ij}^p$ – веса признаков, определяемые по (5) соответственно по множеству значений признаков x_i^p, x_j^p и их произведения $x_i^p x_j^p$, значения $t_i, t_j, t_{ij} \in \{-1, 1\}$, $\mu_{ij} \in [0, 1]$ выбирается по экстремуму функционала

$$\varphi(p, i, j) = \frac{\min_{S_r \in K_1} b_{rij}^p - \max_{S_r \in K_2} b_{rij}^p}{\max_{S_r \in E_0} b_{rij}^p - \min_{S_r \in E_0} b_{rij}^p}. \quad (6)$$

Рассматривается эксперимент на медицинских данных с показателями гипертонической болезни, где класс: K_1 (практически здоровые); K_2 (больные). В табл. 2 приведены результаты последовательного объединения исходных признаков в первую группу и порядок синтеза значений обобщенных оценок (латентных признаков) из комбинаций признаков по критерию (5) и отступами между объектами классов по (6) с учётом вложенных скобок.

Таблица 2.

Результаты группировки признаков.

Номер группы	Последовательность объединения признаков	Значение критерия (5)	Отступ по (6)
1	(x_4, x_{20})	0.9672	-0.0659
	$((x_4, x_{20}), x_9)$	0.9672	-0.0325
	$((((x_4, x_{20}), x_9), x_{18}))$	1.0000	0.0055
	$(((((x_4, x_{20}), x_9), x_{18}), x_8))$	1.0000	0.0076
	$((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2))$	1.0000	0.0079
	$((((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2), x_{10}))$	1.0000	0.0101
	$(((((((((x_4, x_{20}), x_9), x_{18}), x_8), x_2), x_{10}), x_{12}))$	1.0000	0.0105

Аналитический вид последовательности вычисления значений латентного признака по первой группе:

$$y_1^1 = -0.006947([АДС] - 140) + 0.394923([ДИАСТОЛА] - 0.42) -$$

$$0.0041346([АДС] * [ДИАСТОЛА] - 68.2);$$

$$y_2^1 = 1.602699(y_1^1 - 0.0047) + 3.128708([QRS] - 0.08);$$

$$y_3^1 = 0.581198y_2^1 + 0.551857([СИСПОК] - 0.485) + 0.892018y_2^1 * [СИСПОК];$$

$$y_4^1 = 0.21174(y_3^1 + 0.0388) + 0.275601([QT] - 0.36) + 2.136341(y_3^1 * [QT] + 0.0124);$$

$$y_5^1 = 0.005789(y_4^1 * [Рост] + 1.9623); \quad y_6^1 = 0.310793y_5^1 * [ПЛП];$$

$$y_7^1 = 0.1804337y_6^1 * [КДР].$$

Для анализа результатов визуализации предлагается использовать такую структурную характеристику как оценка компактности класса

$$\theta_i = \frac{\sum_{j=1}^{\mu} m_{ij}^2}{m_i^2}, \quad (7)$$

где μ - число непересекающихся групп в $K_i, i=1, l$, $m_i = |K_i|$, m_{ij} - число объектов в j - ой группе класса K_i . Усреднённая оценка компактности по обучающей выборке вычисляется с помощью формулы

$$F(E_0, \rho) = \frac{\sum_{i=1}^l m_i \theta_i}{m}. \quad (8)$$

где m - количество объектов в выборке.

В третьей главе показатели компактности используются для анализа различных преобразований признакового пространства.

Значения устойчивости объектов служат мерой структурного разнообразия выборки, которую можно использовать для поиска и удаления шумовых объектов с целью повышения обобщающей способности алгоритмов. Многообразие значений устойчивости объектов классов напрямую зависит от выбора метрики. В работе реализовано:

- разбиение исходного набора признаков на непересекающиеся группы по правилам иерархической агломеративной группировки и формирование на их основе набора латентных признаков;
- отбор набора информативных латентных признаков;
- обнаружение и удаление шумовых объектов с целью повышения обобщающей способности алгоритмов распознавания.

Для интервалов, построенных по критерию (5), определим $\alpha_1(p) = \left| \left\{ S_{d_i} \in K_p \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|$, $\alpha_2(p) = \left| \left\{ S_{d_i} \in K_{3-p} \mid \rho(S_d, S_{d_i}) \in [c_1, c_2] \right\} \right|$, $\theta_1(p) = \alpha_1(p) / |K_p|$ и $\theta_2(p) = \alpha_2(p) / |K_{3-p}|$. Устойчивость объекта $S_d \in K_p, p=1,2$ по подмножеству признаков $X(k)$ вычисляется следующим образом:

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

Для включения признака $x_a \in X(n)$ в набор (группу) $X(k), X(k) \subset X(n)$ используется функционал

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

Показана связь удаления шумовых объектов и увеличение компактности по (7) и (8). Экспериментально подтверждено преимущество нелинейного метода перед линейным по показателю (8).

ЗАКЛЮЧЕНИЕ

Разработка и обоснование методов интеллектуального анализа для информационных моделей, основанных на знаниях, является актуальным направлением научных исследований. Отношения между объектами на многообразиях структур признаков исследуются с помощью нелинейных преобразований пространства с использованием методов кластерного анализа. Цель нелинейных преобразований выражается в стремлении к идеальной компактности объектов из непересекающихся классов. Интерпретация компактности в терминах нечётной логики используется для поиска скрытых закономерностей в базах данных.

Основные результаты, полученные в данной работе:

1. Разработаны правила для иерархической агломеративной группировки признаков, используемые для линейного и нелинейного отображения разнотипных признаков на числовую ось. Описаны существенные различия между линейным и нелинейным отображением. С помощью предложенных правил удалась существенно сократить размерность исходного признакового пространства;

2. Предложены способы предобработки данных для вычисления оптимальных границ интервалов. При поиске границ используется проверка истинности гипотезы «В границах каждого интервала лежат значения признаков в описании объектов одного класса». Определена оценка

сложности алгоритмов для поиска при наличии повторяющихся значений и пропусков в данных. Значения границ интервалов использовались в правилах иерархической агломеративной группировки для вычисления устойчивости объектов по группе признаков и включения признака в нелинейную комбинацию;

3. Предложен способ оценки компактности обучающей выборки с множеством допустимых значений в $(0, 1]$. С помощью вычислительного эксперимента доказано, что при нелинейном преобразовании признаков компактность выше чем при линейном и в исходном пространстве из «сырых» признаков. Показано, что оценка компактности имеет связь с показателями обобщающей способности алгоритмов;

4. Нелинейное преобразование признаков было использовано для вывода аналитического представления решающих правил. Латентные признаки на основе этого преобразования были упорядочены по отношению информативности в процессе реализации иерархического агломеративного алгоритма;

5. Построены две информационные модели для диагностики рака и мониторинга засоленности орошаемых земель. В рамках моделей определены шумовые объекты обучающих выборок. Эффект от удаления шумовых объектов показан с помощью оценок компактности классов и выборки в целом

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.27.06.2017.FM.01.02 NATIONAL UNIVERSITY OF UZBEKISTAN**

NATIONAL UNIVERSITY OF UZBEKISTAN

SAIDOV DONIYOR YUSUPOVICH

**INFORMATION MODELS BASED ON NONLINEAR
TRANSFORMATIONS OF FEATURE SPACE IN PATTERN
RECOGNITION**

05.01.03 – Theoretical foundations of informatics

**ABSTRACT OF DISSERTATION OF THE DOCTOR OF
PHILOSOPHY (PhD) ON PHYSICAL AND MATHEMATICAL SCIENCES**

Tashkent–2017

The theme of dissertation of doctor of philosophy (PhD) on physical and mathematical sciences was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2017.3.PhD/FM103.

Dissertation has been prepared at the National University of Uzbekistan named after Mirzo Ulugbek.

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

Scientific supervisor: **Ignatyev Nikolay Aleksandrovich**
doctor of physical and mathematical sciences, professor

Official opponents: **Usmonov Rishad Niyazbekovich**
doctor of technical sciences, professor

Ibragimov Farkhadjan Nurmuhamadjonovich
candidate of physical and mathematical sciences

Leading organization: **Tashkent state technical university named after Islam Karimov**

Defense will take place « ____ » _____ 2017 at ____ at the meeting of Scientific Council number DSc.27.06.2017.FM.01.02 at National University of Uzbekistan, Institute of Mathematics. (Address: University str. 4, Almazar area, Tashkent, 100174, Uzbekistan, Ph.: (+99871) 227-12-24, fax: (+99871) 246-53-21, e-mail: nauka@nuu.uz).

Dissertation is possible to review in Information-resource centre at National University of Uzbekistan (is registered № ____) (Address: University str. 4, Almazar area, Tashkent, 100174, Uzbekistan, Ph.: (+99871) 246-02-24).

Abstract of dissertation sent out on « ____ » _____ 2017 year
(Mailing report № _____ on « ____ » _____ 2017 year)

A. R. Maraximov
Chairman of scientific council on award of scientific degrees, D.T.S., Professor

Z. R. Raxmonov
Scientific secretary of scientific council on award of scientific degrees, D.F.-M.S.

R. J. Alov
Chairman of scientific Seminar under Scientific Council on award of scientific degrees, D.F.-M.S., professor

INTRODUCTION (abstract of PhD thesis)

The aim of research work is development of criteria and methods of artificial intelligence based on nonlinear transformation of feature space in pattern recognition problems.

The object of the research work is development and validation of the model of pattern recognition algorithms.

Scientific novelty of the research work is as follows:

- the method of forming nonlinear decision rules based on training data samples is developed;

- a methodology for comparing visualization results based on the compactness criterion is described. The compactness values are determined by the results of partitioning the objects of each class into the minimal number of disjoint groups;

- an estimation of the complexity of the algorithm for dividing the values of quantitative features into disjoint intervals is obtained: taking into account the missing values and repeated values; with and without taking into account the preprocessing of data.

- the increase of the generalizing ability of the algorithm of the nearest neighbor during the transition from the initial description of objects to the space of latent features is proved;

- it is shown that with a hierarchical agglomeration grouping of features and non-linear mapping of their values onto the numerical axis, formed the set of latent features and ordered in relation of their informativeness.

Implementation of the research results. On the basis of the developed methods of research:

agglomerative hierarchical method of grouping features and nonlinear mapping them on the numerical axis introduced in the State Committee of the Republic of Uzbekistan on Land Resources, Geodesy, Cartography and State Cadaster for building information model of soil salinity (Certificate of October 20, 2017 №03-05-8912 State Committee of the Republic of Uzbekistan Land Resources, Geodesy, Cartography and State Cadaster). The mapping method proposed as a result of the research made it possible to obtain the hidden regularities in three types of salinity of the soil by selecting the features associated with nonlinear dependence;

methods of data selection by detecting and removing noisy objects of the sample, calculating compactness values, and visualizing data were applied to the State Committee of the Republic of Uzbekistan for Land Resources, Geodesy, Cartography and the State Cadaster for Monitoring Soil Salinity (Certificate of October 20, 2017 No. 03-05 -8912 of the State Committee of the Republic of Uzbekistan for Land Resources, Geodesy, Cartography and State Cadaster). As a result of the introduction of these methods allowed to determine the weight of the nominal and quantitative features in the process of soil salinization and visualize the relationships between the objects, and to identify abnormal sample objects.

The structure and volume of the thesis. The thesis consists of an introduction, three chapters, conclusion and bibliography. The volume of the thesis is 93 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙЎХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Игнатъев Н.А., Саидов Д.Ю. Вычисление сложности эффективных алгоритмов выбора оптимальных границ интервалов // Проблемы информатики и энергетики, 2014, – Ташкент, N 6, С. 35-40. (05.00.00; № 5).

2. Саидов Д.Ю. Аналитическое представление распознающих операторов для вычисления обобщённых оценок // Вестник НУУз, 2016, – Ташкент, N 2/2, С. 102-107. (01.00.00; № 8).

3. Madraximov Sh.F., Saidov D.Y. Sinf obyektларining turg'unligi va alomatlarıni guruhlash // Проблемы вычислительной и прикладной математики, 2016, – Ташкент, N 3(5), С. 50-55. (01.00.00; № 9).

4. Саидов Д. Ю., Нуржанов Ш. Обобщающая способность алгоритмов распознавания с учётом нелинейности // Проблемы информатики и энергетики, 2016, – Ташкент, N 1, С. 33-39. (05.00.00; № 5).

5. Saidov D.Y. Analytical representation of recognition operators to calculate the generalized estimation //International Journal of Innovative Science, Engineering & Technology, 2016, India, Vol. 3, Issue 10, pp. 197-200, (Scientific Journal Impact Factor=4.332, №23).

6. Saidov D.Y. Data visualization and its proof by compactness criterion of objects of classes //International Journal of Intelligent Systems and Applications, 2017, Hang Kong, Vol 9, No. 8, pp. 51-58, (Global Impact Factor=0,743, №5).

7. 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, (Global Impact Factor=0,562, №5).

8. Саидов Д.Ю. Нелинейное отображение наборов признаков на числовую ось и дискриминантный анализ данных // Проблемы вычислительной и прикладной математики, 2017, – Ташкент, N 1(7), С. 18-23. (01.00.00; №9).

II бўлим (II часть; II part)

9. Саидов Д.Ю. Алгоритмы выбора оптимальных границ интервалов по критерию компактности // Материалы научно-технической конференции «Прикладная математика и информационная безопасность», 2014, Ташкент, с. 242.

10. Саидов Д.Ю. Нелинейные преобразования в признаковом пространстве // Труды научной конференции «Неклассические уравнения математической физики и их приложения», 2014, Ташкент, с. 332.

11. Саидов Д.Ю. Нелинейные преобразования признакового пространства и их аналитические представления // Международный молодежный научный форум «Ломоносов-2015», 2015, - Россия

12. Мадрахимов Ш.Ф., Саидов Д. Ю. Группировка признаков по критерию устойчивости объектов классов // Материалы тезисов Международной научно-технической конференции «Актуальные проблемы прикладной математики, информатики и механики», 2016 – Воронеж С. 93-96.

13. Саидов Д.Ю. Нелинейные преобразования в задачах распознавания с учителем и их Визуаллаштириш // Материалы Республиканской научно-технической конференции «Проблемы информационных и телекоммуникационных технологий», 2016, Ташкент, с. 62-65.

14. Саидов Д.Ю. Выбор порогов в дискриминантный анализ данных // Материалы тезисов Международной научно-технической конференции «Актуальные проблемы прикладной математики и информационных технологий – Аль-Хорезми 2016», 2016, Бухара, с. 30.

15. Мадрахимов Ш.Ф., Саидов Д.Ю. Выбор латентных признаков по результатам иерархической агломеративной группировки данных // Материалы тезисов Международной научно-технической конференции «Актуальные проблемы прикладной математики и информационных технологий – Аль-Хорезми 2016», 2016, Бухара, с. 88-91.

16. Игнатъев Н.А., Саидов Д.Ю. Анализ данных и принятие решений с помощью логических закономерностей в форме полуплоскостей // Известия СамНЦ, 2017, Том 19, № 4(2), С. 294-300. (РИНЦ Импакт-фактор 0,347)

17. Saidov D.Y. Calculation and analysis the compactness values of classes and the training sample as a whole // The 4th International Conference on BIG DATA APPLICATIONS AND SERVICES, Vol.4, No.1, 2017, Tashkent, pp. 324-327

18. Игнатъев Н.А., Саидов Д.Ю. Линейные дискриминантные функции и выбор спрямляющего пространства для их реализации // Известия Академии Вооруженных сил Республика Узбекистана, Ташкент, 2017. № 1 (24). С. 128-135.

19. Игнатъев Н.А., Саидов Д.Ю. Guruhlanuvchi alomatlar to`plamini son o`qiga nochiziqli akslantirish, № DGU 03183, Ўзбекистон Республикаси Интеллектуал Мулк Агентлиги, Тошкент , 2015

20. Игнатъев Н.А., Саидов Д.Ю. Разбиение признаков на непересекающиеся интервалы по максимуму критерия внутриклассового сходства и межклассового различия их значений, № DGU 04461, Ўзбекистон Республикаси Интеллектуал Мулк Агентлиги, Тошкент , 2017

Авторефератнинг ўзбек, рус ва инглиз тилларидаги нусхалари
«ЎЗМУ хабарлари» таҳририятида таҳрирдан ўтказилди.

Босишга рухсат этилди: 07.12.2017 йил
Бичими 60x44 ¹/₁₆, «Times New Roman»
гарнитурда рақамли босма усулида босилди.
Шартли босма табағи 2,4. Адади: 100. Буюртма: № _____.

Ўзбекистон Республикаси ИИВ Академияси,
100197, Тошкент, Интизор кўчаси, 68

«АКАДЕМИЯ НОШИРЛИК МАРКАЗИ»
Давлат унитар корхонасида чоп этилди.