

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSC.13/05.05.2023.T.07.03 RAQAMLI ILMIY KENGASH**

**“TIQXMMI” MILLIY TADQIQOT UNIVERSITETI HUZURIDAGI
FUNDAMENTAL VA AMALIY TADQIQOTLAR INSTITUTI**

TURSUNOV ALISHER TULKUNOVICH

**TURLI XILDAGI MA'LUMOTLARNI TAHLIL QILISH
MASALALARIDA NOMIQDORIY BELGILARGA RAQAMLI ISHLOV
BERISH USULLARI VA ALGORITMLARINI ISHLAB CHIQISH**

05.01.03- Informatikaning nazariy asoslari

**TEXNIKA FANLAR BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent-2025

**Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati
mundarijasi**

**Оглавление автореферата диссертации доктора философии(PhD) по
техническим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD) on technical
sciences**

Tursunov Alisher Tulkunovich

Turli xildagi ma'lumotlarni tahlil qilish masalalarida nomiqdoriy belgilarga raqamli
ishlov berish usullari va algoritmlarini ishlab chiqish.....3

Турсунов Алишер Тулкунович

Разработка методов и алгоритмов оцифровки неколичественных признаков в
задачах анализа разнотипных данных.....21

Tursunov Alisher Tulkunovich

Development of methods and algorithms for digitizing non-quantitative features in
problems of analyzing different types of data.....40

E'lon qilingan ishlar ro'yxati

Список опубликованных работ

List of published works.....44

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSC.13/05.05.2023.T.07.03 RAQAMLI ILMIY KENGASH**

**“TIQXMMI” MILLIY TADQIQOT UNIVERSITETI HUZURIDAGI
FUNDAMENTAL VA AMALIY TADQIQOTLAR INSTITUTI**

TURSUNOV ALISHER TULKUNOVICH

**TURLI XILDAGI MA'LUMOTLARNI TAHLIL QILISH
MASALALARIDA NOMIQDORIY BELGILARGA RAQAMLI ISHLOV
BERISH USULLARI VA ALGORITMLARINI ISHLAB CHIQISH**

05.01.03- Informatikaning nazariy asoslari

**TEXNIKA FANLAR BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent-2025

Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasida B2023.4.PhD/T4167 raqam bilan ro'yxatga olingan.

Dissertatsiya "TIQXMMI" milliy tadqiqot universiteti huzuridagi Fundamental va amaliy tadqiqotlar institutida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, rus, ingliz (rezyume)) Ilmiy kengash veb-sahifasida (www.tuit.uz) va "Ziyonet" axborot ta'lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar:

Nishanov Axram Hasanovich
texnika fanlari doktori, professor

Rasmiy opponentlar:

Primova Xolida Anorboyevna
texnika fanlari doktori, dotsent

Dadaxanov Musoxon Xoshimxonovich
texnika fanlari falsafa doktori (PhD), dotsent

Yetakchi tashkilot:

Islom Karimov nomidagi Toshkent davlat texnika universiteti

Dissertatsiya himoyasi Toshkent axborot texnologiyalari universiteti huzuridagi DSc.13/05.05.2023.T.07.03 raqamli Ilmiy kengashning 2025-yil "22" yanvar soat 14⁰⁰ dagi majlisida bo'lib o'tadi. (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel.: (99871) 238-64-43, e-mail: ilmiy_kengash@tuit.uz).

Dissertatsiya bilan Toshkent axborot texnologiyalari universitetining Axborot-resurs markazida tanishish mumkin (316 raqam bilan ro'yxatga olingan). (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel: (99871) 238-64-70).

Dissertatsiya avtoreferati 2025-yil "7" yanvar da tarqatildi.
(2025-yil "6" yanvar dagi 6 raqamli reyestr bayonnomasi).



M.M.Kamilov

Ilmiy darajalar beruvchi Ilmiy kengash raisi,
texnika fanlari doktori, professor, O'ZR FA akademigi

N.A.Egamberdiyev

Ilmiy darajalar beruvchi Ilmiy kengash ilmiy kotibi,
texnika fanlari bo'yicha falsafa doktori(PhD)

N.O.Raximov

Ilmiy darajalar beruvchi Ilmiy kengash qoshidagi
ilmiy seminar raisi, texnika fanlari doktori, dotsent

KIRISH (falsafa doktori (PhD) dissertatsiyasining annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahonda zamonaviy axborot-kommunikatsiya texnologiyalarining jadal rivojlanishi va katta hajmdagi ma'lumotlarga ishlov berish, ma'lumotlarni tahlil qilish, kerakli ma'lumotlarni obyektardan ajratib olish, timsollarni tanib olish kabi masalalarni hal qilish zarurati tobora dolzarb bo'lmoqda. AQSh, Xitoy, Yaponiya, Rossiya, Germaniya kabi rivojlangan davlatlarda ushbu yo'nalishlarda ilmiy asoslarni rivojlantirish katta ahamiyat kasb etmoqda. Tadqiqotlarning ushbu yo'nalishida turli xildagi ma'lumotlarni tahlil qilishda nomiqdoriy belgilarga raqamli ishlov berish usul va algoritmlarini ishlab chiqishga alohida e'tibor qaratilmoqda.

Jahonda turli xil ko'rinishda berilgan obyektlarning o'xshashlik darajalarini aniqlashga asoslangan klasterlash algoritmlari ishlab chiqish va ular asosida ma'lumotlarni tahlil qilish tizimini yaratishni taqozo etmoqda. Shu sababdan soha mutaxassislari tomonidan qaror qabul qilish jarayoni juda murakkab jarayonga aylanib bormoqda. Jumladan, katta hajmdagi ma'lumotlarga dastlabki ishlov berish, raqamli bo'lmagan ma'lumotlarga ishlov berishda katta o'lchamdan kichik muhimroq o'lchamga o'tish, tasniflash usullari, belgilarni turlarini o'zgartirish, timsollarni tanib olishda belgilar maydonining o'lchamini kamaytirish usullari, ma'lumotlarni tahlil qilishda informativ belgilar majmuasini tanlash va tanlangan informativ belgilar fazosida sinflashtirish masalasini yechish va zamonaviy tanib olish tizimlarini ishlab chiqish dolzarb masalalardan hisoblanmoqda.

Respublikamizda ham ushbu yo'nalishda ma'lumotlarga dastlabki ishlov berish, tahlil qilish va timsollarni tanib olishga mo'ljallangan apparat-dasturiy majmualarni ishlab chiqish va joriy etishga alohida e'tibor qaratilmoqda. 2017- 2021 -yillarda O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasida, jumladan "... iqtisodiyot, ijtimoiy soha va boshqaruv tizimiga axborot-kommunikatsiya texnologiyalarini joriy etish, ... axborot xavfsizligini ta'minlash va axborotni himoya qilish tizimlarini takomillashtirish, axborot sohasidagi tahdidlarga o'z vaqtida va munosib qarshi harakatlarni tashkil etish"¹ vazifalari belgilangan.

O'zbekiston Respublikasi Prezidentining 2018-yil 19-fevralda "Axborot texnologiyalari va kommunikatsiya sohasini yanada takomillashtirish chora-tadbirlari to'g'risida"gi PF-5349-sonli, 2020-yil 5-oktabrda "Raqamli O'zbekiston-2030 strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to'g'risida"gi PF- 6079-sonli, 2020-yil 28-aprelda "Raqamli iqtisodiyot va elektron hukumatni keng joriy etish chora-tadbirlari to'g'risida"gi PQ- 4699- sonli, 2022-yil 28-yanvarda "2022 - 2026 yillarga mo'ljallangan yangi O'zbekistonning taraqqiyot strategiyasi to'g'risida"gi PF-60-sonli, 2023-yil 11- sentabrda "O'zbekiston-2030 strategiyasi to'g'risida"gi PF-158-sonli Farmonlari hamda O'zbekiston Respublikasi Prezidentining 2023-yil 28-dekabrda "Sog'liqni saqlash tizimini raqamlashtirishni jadallashtirish hamda ilg'or raqamli texnologiyalarni joriy etish bo'yicha qo'shimcha chora-tadbirlar to'g'risida"

¹ O'zbekiston Respublikasi Prezidentining 2017-yil 7-fevraldagi "O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasi to'g'risida"gi PF-4947-son Farmoni

PQ– 415–sonli Qarori va mazkur faoliyatga tegishli boshqa meyoriy–huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi. Mazkur tadqiqot respublika fan va texnologiyalar rivojlanishining IV. “Axborotlashtirish va axborot-kommunikatsiya texnologiyalarini rivojlantirish” ustuvor yo‘nalishi doirasida bajarilgan.

Muammoning o‘rganilganlik darajasi. So‘nggi yillarda ma’lumotlarni tahlil qilish masalalarida belgilarga raqamli ishlov berish, informativ belgilarni tanlab olish, tasvirlarga dastlabki ishlov berish va timsollarni tanib olish usullarini ishlab chiqish va takomillashtirish hamda ularni amaliyotga joriy etish masalalari Y.I. Juravlev, M.A.Ayzerman, V.V.Aleksandrov, E.M.Braverman, M.B. Aydarxanov, N.D.Gorskiy, V.I.Vasilev, R.Gonsales, R.Vuds, A.G. Ivaxnenko, E.Patrik, U.Prett, L.Roberts, V.A.Soyfer, G.S.Lbov, Sh.Ulman, P.Xart, M.I.Shlezinger va boshqa xorijiy olimlar tomonidan tadqiq etilgan.

Respublikamizda timsollarni tanib olish va tasvirlarga dastlabki ishlov berish, ma’lumotlarni tahlil qilish masalalarida belgilarga raqamli ishlov berish hamda informativ belgilar fazosini tanlab olishning nazariy asoslarini rivojlantirishga M.M. Kamilov, Sh.X.Fozilov, F.T.Adilova, N.M.Mirzayev, A.X.Nishanov, B.B.Akbaraliyev, N.S.Mamatov, S.S.Radjabov va boshqalar katta hissa qo‘shib kelmoqdalar.

Hozirgi kunda ma’lumotlarni tahlil qilish tizimlari nazariyasidagi yutuqlarga qaramay, ma’lumotlarga dastlabki ishlov berish, timsollar va belgilarni tanib olish hamda informativ belgilarni tanlab olish usul va algoritmlarini ishlab chiqish va ular asosida ma’lumotlarga ishlov berish dasturiy ta’minotini yaratish yetarli darajada o‘rganilmagan.

Tadqiqotining dissertatsiya bajarilgan oliy ta’lim yoki ilmiy–tadqiqot muassasasining ilmiy–tadqiqot ishlari rejalari bilan bog‘liqligi. Dissertatsiya tadqiqoti Muhammad Al–Xorazmiy nomidagi Toshkent axborot texnologiyalari universitetining ilmiy–tadqiqot ishlari №BV– V– F4– 011– “Noaniqlik sharoitlarida ma’lumotlarni intellektual tahlilining noravshan–nokorrekt masalalarini yechish usul va algoritmlari” mavzularidagi ilmiy tadqiqot loyihalari doirasida bajarilgan.

Tadqiqotning maqsadi turli xildagi ma’lumotlarni tahlil qilishda nomiqdoriy belgilarga raqamli ishlov berishning obyektlar o‘xshashliklarini aniqlashga asoslangan usullari va algoritmlarini ishlab chiqish va takomillashtirishdan iborat.

Tadqiqotning vazifalari:

turli xildagi ma’lumotlarni tahlil qilishda sifatiy belgilardan miqdoriy belgilarga o‘tish, ma’lumotlarga dastlabki ishlov berish va belgilarning turlarini o‘zgartirish usullari hamda tadqiqot masalasini shakllantirish;

belgilarning juftlashgan o‘zaro bog‘liqliklarini tadqiq qilishda empirativ tartib shkalasini aniqlashda foydalaniladigan g almashtirishlar sonining butun son gradatsiya darajalarini optimal almashtirishni topishning hisoblash algoritmlarini ishlab chiqish;

ma'lumotlarni tahlil qilish jarayonida informativ belgilar majmuasini tanlash va tanlangan informativ belgilar fazosida sinflashtirish masalasini yechimini beruvchi usullar va algoritmlarni ishlab chiqish;

turli xil belgilar fazosida berilgan obyektlarning o'xshashlik darajalarini aniqlashga asoslangan klasterlash algoritmlarini ishlab chiqish;

nomiqdoriy belgilarga raqamli ishlov berishda obyektlar o'xshashliklariga asoslangan klasterlash algoritmlarni ishlab chiqish va mavjudlarini takomillashtirish.

Tadqiqotning obyeksi sifatida ma'lumotlarni tahlil qilishda foydalaniladigan, obyektlarning muhim xususiyatlarini ifodalovchi turli xildagi belgi va atributlar, obyektlar o'xshashliklarini aniqlash va ularni baholashga asoslangan yondashuvlar olingan.

Tadqiqotning predmetini nomiqdoriy belgilarga raqamli ishlov berish, belgilar turlarini almashtirish, timsollarni tanib olish hamda informativ belgilar majmuasini tanlash usul va algoritmlardan iborat.

Tadqiqotning usullari. Tadqiqotlar jarayonida ma'lumotlarni intellektual tahlil qilish, belgilarga raqamli ishlov berish, ehtimollar va statistika nazariyasi, diskret matematika va timsollarni tanib olish usullaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

turli xildagi ma'lumotlarni tahlil qilishda sifatiy belgilardan miqdoriy belgilarga o'tish, ma'lumotlarga dastlabki ishlov berish va belgilarning turlarini o'zgartirish usul va algoritmlari obyektlarni empirik munosabatlardan tashkil topgan mantiqiy funksiyalari qiymatlarining o'xshashligi asnosida ishlab chiqilgan;

belgilarning juftlashgan o'zaro bog'liqliklari, belgilar aro korrelyasiya va regressiyani tadqiq etishda empirativ tartib shkalasini aniqlashning g almashtirishlar soni butun son gradatsiya darajalarini optimal almashtirishni hisoblash algoritmi obyekt belgilarning o'zaro bog'liqliklarini tadqiqi asosida ishlab chiqilgan;

ma'lumotlarni tahlil qilishda informativ belgilar majmuasini tanlash va tanlab olingan informativ belgilar fazosida sinflashtirish masalasini yechish usul va algoritmlari Evklid fazoda obyektlararo λ kesishuv nazariyasiga asoslangan holda ishlab chiqilgan;

turli xil belgilar fazosida berilgan obyektlarning o'xshashlik darajalarini aniqlashga asoslangan klasterlash algoritmi sonli belgilar fazosida obyektlarning o'xshashligini mazmunan ifodalovchi mezonga asosan ishlab chiqilgan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

tadqiqot natijasida ishlab chiqilgan usul va algoritmlar kesimida ma'lum fan soha mutaxassisi, dasturchi bo'lmagan foydalanuvchilar uchun mo'ljallangan dasturiy majmua ishlab chiqilgan;

"Timsollarni tanib olishda har xil tipda berilgan ma'lumotlarni klasterlash" nomli dasturiy majmua ishlab chiqilgan;

"Har xil tipda berilgan ma'lumotlarni sinflashtirishda sun'iy neyron modeliga asoslangan hal qiluvchi qoida" hamda "Har xil tipli ma'lumotlar fazosida informativ belgilar majmuasini tanlash" nomli dasturiy majmualar ishlab chiqilgan.

Tadqiqot natijalarining ishonchliligi. Olingan natijalarning ishonchliligi va asosliligi nomiqdoriy belgilarga raqamli ishlov berish usullari va algoritmlarini

to'g'ri qo'llash, xulosalar va ma'lumotlarning tajribaviy tadqiqotlar natijalariga muvofiqligi, shuningdek amaliy masalalarni yechilishi bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqotning natijalarining ilmiy ahamiyati ishlab chiqilgan algoritmlar ma'lumotlarni tahlil qilish asosida nomiqdoriy belgilarga raqamli ishlov berish texnologiyalarining nazariy asoslarini yanada rivojlantirishga xizmat qilishi bilan izohlanadi.

Tadqiqot natijalarining amaliy ahamiyati ishlab chiqilgan algoritmlar va usullar hamda ular asosida yaratilgan dasturiy majmua raqamli bo'lmagan ma'lumotlarni tahlil qilish masalalarini yechishda soha mutaxassislari yakuniy qaror qabul qilishlarida ko'mak beruvchi zamonaviy tanib olish tizimi ishlab chiqilganligi bilan izohlanadi.

Tadqiqot natijalarning joriy qilinishi. Ma'lumotlarni tahlil qilishda informativ belgilarni tanlab olish, nomiqdoriy belgilarga raqamli ishlovlar berish, ko'p o'zgaruvchilik chiziqli regressiya tahlili masalalarini yechishda belgilar turlarini o'zgartirish algoritmlari va dasturiy majmualari bo'yicha olingan natijalar asosida:

turli xildagi farmasevtik ma'lumotlarni tahlil qilish, timsollarni tanib olish masalalari uchun berilgan belgilar fazosi o'lchamini qisqartirish usullari va algoritmlar asosida ishlab chiqilgan "Timsollarni tanib olishda har xil tipda berilgan ma'lumotlarni klasterlash" nomli dasturiy majmua "Enrich" XK farmasevtik korxonasida amaliyotda joriy etilgan (O'zbekiston Respublikasi Sog'liqni saqlash vazirligi muvofiqlashtiruvchi Ekspertlar Kengashining 2024-yil 26-apreldagi 01/01- 836- sonli ma'lumotnomasi). Ilmiy tadqiqot natijasida nomiqdoriy belgilarga raqamli ishlov berish hamda belgilar turlarini o'zgartirish usullari orqali miqdoriy belgilarga almashtirish asosida tajriba ma'lumotlarni tahlil qilish tizimining yagona integrallashgan tizimi yaratildi. Natijada ish samaradorligini o'rtacha 10-12% ga oshirishga imkon bergan;

ma'lumotlarni (dori vositalari) tahlil qilishda informativ belgilarni tanlab olish, sinflashtirish hamda belgilar maydoni o'lchamini qisqartirish usullari va algoritmlari asosida ishlab chiqilgan "Har xil tipda berilgan ma'lumotlarni sinflashtirishda sun'iy neyron modeliga asoslangan hal qiluvchi qoida" hamda "Har xil tipli ma'lumotlar fazosida informativ belgilar majmuasini tanlash" nomli dasturiy majmualar "Innovatsion farmasevtik birikmalar ilmiy laboratoriyasi"da amaliyotga joriy etilgan (O'zbekiston Respublikasi Sog'liqni saqlash vazirligi muvofiqlashtiruvchi Ekspertlar Kengashining 2024-yil 26-apreldagi 01/01- 836- sonli ma'lumotnomasi). Ilmiy tadqiqot natijasida timsollarni tanib olishda belgilar maydonining o'lchamini kamaytirish, informativ belgilarni tanlab olish hamda belgilar turlarini almashtirish usullari va algoritmlari, shuningdek ma'lumotlarni tahlil qilishning klassik usullari asosida tajriba ma'lumotlarni tahlil qilish tizimi yaratildi. Natijada ish unumdorligini o'rtacha 10-15% ga oshirishga imkon bergan;

turli xilda berilgan ma'lumotlarni tahlil qilish masalalarini yechishda informativ belgilar majmualarini tanlab olish orqali tajriba ma'lumotlarini tahlil qilish usullari va algoritmlari asosida ishlab chiqilgan "Har xil tipli ma'lumotlar fazosida informativ belgilar majmuasini tanlash" nomli dasturiy majmua Toshkent farmasevtika instituti "Farmasevtik ishlab chiqarishni tashkil qilish va boshqarish"

kafedrasida laboratoriya mashg'ulotlarida amaliyotga joriy etilgan (O'zbekiston Respublikasi Sog'liqni saqlash vazirligi muvofiqlashtiruvchi Ekspertlar Kengashining 2024-yil 26-apreldagi 01/01-836-sonli ma'lumotnomasi). Natijada tajriba ma'lumotlarini tahlil qilishda ish unumdorligini taxminan 1,2 barobarga oshirgan.

Tadqiqot natijalarining aprobatsiyasi. Dissertatsiyaning asosiy nazariy hamda amaliy natijalari 5 ta xalqaro va 8 ta respublika ilmiy-texnik hamda ilmiy-amaliy anjumanlarda muhokamadan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Dissertatsiya mavzusi bo'yicha jami 23 ta ilmiy ishlar chop etilgan, ulardan 1 ta monografiya, 6 ta O'zbekiston Respublikasi Oliy attestatsiya komissiyasining doktorlik dissertatsiyalarining asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda, jumladan 1 tasi xorijiy va 5 tasi respublika jurnallarida nashr qilingan hamda 3 ta EHM uchun yaratilgan dasturiy vositalarini qaydlash guvohnomalari O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligidan olingan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya kirish, to'rt bob, xulosa va foydalanilgan adabiyotlar ro'yxati hamda ilovalardan iborat. Dissertatsiyaning umumiy hajmi 116 betni tashkil etadi.

DISSERTATSIYANING ASOSIY MAZMUNI

Dissertatsiyaning kirish qismida mavzuning dolzarbligi va zarurati, tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo'nalishlariga mosligi, muammoning o'rganilganlik darajasi, dissertatsiya bajarilgan oliy ta'lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi, tadqiqotning maqsadi, vazifalari, obyekti, predmeti, usullari, ilmiy yangiligi, amaliy natijalari, tadqiqot natijalarining ishonchliligi, ilmiy va amaliy ahamiyati, amaliyotga joriy qilinishi, aprobatsiyasi, natijalarning e'lon qilinganligi, dissertatsiyaning tuzilishi va hajmi haqidagi ma'lumotlar keltirilgan.

Dissertatsiyaning **“Turli xildagi ma'lumotlarni tahlil qilish usullarining zamonaviy holati”** deb nomlangan birinchi bobida ma'lumotlarga dastlabki ishlovlar berishda o'quv tanlanmalarni shakllantirish hamda turli xildagi tajriba ma'lumotlariga ishlov berishda ma'lumotlarni tahlil qilish usullari va algoritmlari ko'rib chiqilgan. Mazkur bobning 1.1-paragrafida jahonda keng tarqalgan ma'lumotlarni tahlil qilish usullarining tarixi va zamonaviy holati bayon qilingan. 1.2-paragrafida ma'lumotlarga dastlabki ishlovlar berishda o'quv tanlanmalar shakllantirilgan va timsollarni tanib olish masalalarida eng ko'p tarqalgan usullar ko'rib chiqilgan. Amaliy statistikaning dasturiy ta'minoti turlari va modellari keltirilgan. 1.3-paragrafida belgilarning turlarini o'zgartirishning mavjud usullari bayon etilgan bo'lib, ushbu usullarning afzalliklari va kamchiliklari ko'rsatilgan. “Belgilarning turlarini o'zgartirish” tushunchasi ostida turli xildagi belgilardan bir turga ya'ni miqdoriy, sifat yoki tasnifli ko'rinish shakllariga o'tkazilishi ko'rib chiqilgan.

Mazkur bobning so'nggi paragrafida dissertatsiya ishining maqsadi, vazifalari keltirilib o'tilgan bo'lib, tadqiqot masalasi shakllantirilgan.

Tadqiqot masalasini shakllantirishda p - o'lchovli belgilar fazosidagi $X_i \in A$, $i = \overline{1, m}$; umumiy tanlanmadan foydalangan holda o'quv tanlanmalarni aniqlashtirish talab etiladi. Bunda $X_{k1}, X_{k2}, \dots, X_{kN_k} \in A_k, k = \overline{1, m}$; sinflarni shakllantirish talab etiladi. Bu yerda A_k sinf N_k ta $X_{k1}, X_{k2}, \dots, X_{kN_k}$ obyektlardan tashkil topgan $A = \bigcup_{k=1}^m A_k$ bo'ladi.

Bu klasterlash masalasini yechish uchun obyektlar aro yaqinlik funksiyasi kiritiladi. Agar obyektlarning belgilari nominal bo'lsa, u holda ularning o'xshashligini ko'rsatuvchi kattalik $\rho^j(x_{ki}, x_{kq})$ orqali belgilanib va quyidagi formula asosida hisoblanadi:

$$\rho_{ki}^j(x_{ki}, x_{kq}) = \begin{cases} 1, & \text{agar } (x_{ki}^j - x_{kq}^j) = 0 \\ 0, & \text{aks holda} \end{cases} \quad (1)$$

Agar obyektlar sonli belgilar orqali ifodalangan bo'lsa, u holda ularning o'xshashligini o'lchovchi kattalik $\rho^j(x_{pi}, x_{pq})$ orqali belgilanib, (2) formula asosida hisoblanadi:

$$\rho_{ki}^j(x_{ki}, x_{kq}) = \begin{cases} 1, & \text{agar } |x_{ki}^j - x_{kq}^j| \leq \varepsilon^j \\ 0, & \text{aks holda} \end{cases} \quad (2)$$

bu yerda $k = \overline{1, m}; i \neq q = \overline{1, m_k}; j = \overline{1, p}; \rho_{ki}^j(x_{ki}, x_{kq}) = (\rho_{ki}^1(x_{ki}, x_{kq}), \rho_{ki}^2(x_{ki}, x_{kq}), \dots, \rho_{ki}^p(x_{ki}, x_{kq}))$.

Yuqorida keltirilgan (2) formuladagi j -belgiga mos ε^j —bosag'aviy qiymat har bir sinf obyektlarining belgilari uchun quyidagicha hisoblanadi:

$$\varepsilon^j = \frac{1}{N-1} \sum_{i=1}^{N-1} |x_{ki}^j - x_{ki+1}^j|,$$

bu yerda $j = \overline{1, p}; k = \overline{1, m}; i = \overline{1, N-1}$; ga teng.

Dissertatsiyaning **“Turli xildagi ma'lumotlarni tahlil qilish usullari va algoritmlari”** deb nomlangan ikkinchi bobida turli xildagi ma'lumotlarni tahlil qilishda nomiqdoriy belgilarga raqamli ishlov berish usullari va algoritmlari ko'rib chiqilgan. Mazkur bobning 2.1-paragrafida belgilarning o'zaro bog'liqliklarini tadqiq etish masalalari o'rganilgan.

$m * p$ o'lchamdagi X matritsani quyidagi ko'rinishida ifodalab olamiz:

$$X = [X_1, X_2, X_3] \quad (3)$$

bu yerda $X_1 - \ell$ ta miqdoriy belgilar $x^{(1)}, x^{(2)}, \dots, x^{(\ell)}$ ga ega bo'lgan qism-matritsa, $X_2 -$ faqat $i = \overline{\ell+1, m}$; indeksli sifatli belgilarga ega va $X_3 -$ esa X dan olingan, faqat $i = \overline{m+1, p}$ indeksli tasnifli belgilarga ega bo'lgan qism-matritsa. Natural sonlar qatori kesmasi $K_g = \{1, 2, \dots, g\}$ tanlab olinadi. Bunda ξ ko'rinishi har bir i - gradatsiyaga K_g dan olingan d_i butun sonli metka qiymatini beradi, u gradatsiya darajasi deb ataladi. U holda ξ ko'rinishi $D \in P^g$ darajalarning o'rinlarini almashtirish bilan beriladi. Bu yerda $D = (d_1, d_2, \dots, d_g)$, $P_g - g$ natural sonlarning o'rinlarini almashtirishlar to'plamidir.

Belgilar turlarini o'zgartirish uchun empirativ shkalalarni tanlash mezoni sifatida $1 - r_{12}^2$ qiymatni minimallashtirish orqali foydalanish taklif etiladi. Bu yerda $r_{12}^2 - x_1$ miqdoriy belgi orasidagi korrelatsiya koeffitsienti va x_2^ξ raqamli ishlov

berilgan miqdoriy bo'lmagan belgi. R_{12} uchun, quyidagi munosabatni olishimiz mumkin:

$$r_{12}^2 = \frac{\bar{a}^T C \bar{a}}{\bar{a}^T T \bar{a}}$$

$$C = \bar{\mu} \lambda_{11}^{-1} \mu^{-T}, \mu^T = [\mu_1^1, \mu_1^2, \dots, \mu_1^g] \text{ ushbu munosabatlarni olamiz:}$$

$$1 - r_{12}^2 = \frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} \quad (4)$$

bu yerda $H = T - C$.

Shunday qilib, o'zaro juftlashgan bog'lanishlarni o'rganishda raqamli belgilarni qidirish mezonini quyidagicha ko'rinishiga ega:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} \rightarrow \min \quad (5)$$

bu yerda $\bar{a} \in B^g$.

Optimal butun sonli raqamli belgilarni topish uchun (5) shart bo'yicha masalani yechishga o'tamiz. Bu holda (5) masala quyidagicha ko'rinishga keladi:

$$\frac{\bar{D}^T H \bar{D}}{\bar{D}^T T \bar{D}} \rightarrow \min \quad (6)$$

bu yerda $D \in P^g$.

Faraz qilaylik, $x_{\ell+1}$ tasniflash belgilarining g gradatsiyalari soni oltitadan katta bo'lmaydi, $g \leq 6$. Bunday holda, har bir darajali empirativ shkalaning g butun son darajalarining o'rniga $\bar{D}^T = (d_1, \dots, d_g)$, $d_i \in K_g$ mos keladi. Barcha almashtirishlar soni $g!$ ga teng bo'lib, ko'rib chiqilayotgan jarayon uchun nisbatan kichik bo'lsa, u holda (6) ifodaning optimal qiymati barcha almashtirishlarni to'liq sanab o'tish orqali izlash usuli taklif etiladi. Bunda g ga teng bo'lgan, $x_{\ell+1}$ tasniflash belgisining berilgan gradatsiya sonidan barcha almashtirishlarni olish usulini aniqlaymiz.

Dastlab $k_1 = 1$ deb olamiz. Birinchi navbatda bu qiymatga $k_2 = 2$ ni qo'shib, so'ngra ikkinchisini, natidjida $2!$ hosil qilamiz. Ya'ni $\{2,1\}\{1,2\}$ o'rin almashtiramiz. Uchunchi qadamda $3!$ ni almashtiramiz, g qadamda $g!$ almashtirishlarni hosil qilamiz.

Empirativ tartib shkalasini aniqlashda foydalaniladigan, $g \leq 6$ ning butun son gradatsiya darajalarini optimal almashtirishini topishning hisoblash algoritmlari quyidagi qadamlardan iborat bo'ladi.

Birinchi qadam. Barcha $g!$ aniqlashda yuqorida taklif yetilgan sxema ko'rinishida almashtirish kerak.

Ikkinchi qadam. $\bar{D}_0 = (1, 2, \dots, g)$ gradatsiyalarini dastlabki almashtirishlarini tanlash va $F(\bar{D}_0) = \frac{\bar{D}_0^T H \bar{D}_0}{\bar{D}_0^T T \bar{D}_0}$; $F_{min} = F(\bar{D}_0)$ ni qiymatni hisoblash.

Uchunchi qadam. $\bar{D}_k$, - keyingi almashtirishlarga o'tish. $F(\bar{D}_k)$ qiymatini hisoblash. Agar $F(\bar{D}_k) < F_{min}$, u holda $F_{min} = F(\bar{D}_k)$, $k = \overline{1, g}$;

To'rtinchi qadam. Agar barcha $g!$ almashtirishlar amalga oshirilsa, u holda beshinchi qadamga o'tiladi, aks holda uchunchi qadamga o'tiladi.

Beshinchi qadam. F_{min} ning minimal qiymatiga mos keladigan $\bar{D}_k$ vektori va izlangan empirativ tartibli shkalasi hisoblanadi.

Endi $g > 6$ holatni ko'rib chiqamiz. Bunday holda g ning butun son darajalarini mumkin bo'lgan almashtirishlarning umumiy soni ($g!$) dan juda katta bo'ladi va (6) ifodaga mos keladigan optimal empirativ shkalani olish uchun to'liq qidirish usulidan foydalanish samarasiz bo'ladi.

Shuning uchun darajalarni optimal almashtirishni ketma-ket yaqinlashish usuli yordamida izlash taklif etiladi. Bunday holda dastlabki almashtirishlar ixtiyoriy tarzda amalga oshiriladi va yakuniy natija $\bar{D}_0$ ga bog'liq bo'ladi.

Shunday qilib, turli xil boshlang'ich qiymatlar uchun turli xil (6) ifodaning mahalliy minimal qiymatlari olinadi. Birinchi navbatda $\bar{D}_0$ ni hisoblash usulini aniqlaymiz, buning uchun yordamchi masalani yechamiz.

Faraz qilaylik, (6) ifodada minimal haqiqiy sonlar to'plamida topilishi kerak, ya'ni B^g uchun mumkin bo'lgan yechimlar to'plamini kengaytiramiz va optimal vektorni topish masalasi (6) ko'rinishiga ega bo'ladi.

Umumlashgan (5) ifodasi Releya munosabati deyiladi. Ma'lumki odatdagi Releya munosabati quyidagicha ko'rinishga ega:

$$\frac{(Ax, x)}{(x, x)} \quad (7)$$

haqiqiy tengsizlik

$$v_1 \leq \frac{(Ax, x)}{(x, x)} \leq v_n \text{ nisbati esa}$$

$$v_1 = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)} \quad (8)$$

bu yerda x ixtiyoriy nolga teng bo'lmagan vektor, $v_1, \dots, v_n$ - A matritsasining xos qiymati, v_1 -raqamlarning minimal qiymati.

$\bar{a}$ vektorni markazlashtirish shartini qo'llaymiz, ya'ni:

$$\frac{1}{N} \sum_{i=1}^g a_i n_i = 0$$

Bu holda T matritsa belgilarning o'zaro bog'liqliklarini tahlil qilish masalasi uchun $\frac{n_i}{N}$ ta elementga ega bo'lgan holda diagonal ko'rinishni oladi va shuning uchun degeneratsiya bo'lmagan hisoblanadi. U holda (5) munosabatni (7) ko'rinishga keltirish mumkin, chunki:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} = \frac{(H^T \bar{a}, \bar{a})}{(T^T \bar{a}, \bar{a})} = \frac{(T^{T-1/2} H^T T^{T-1/2} \bar{b}, \bar{b})}{(\bar{b}, \bar{b})}$$

bu yerda $\bar{b} = T^{T-1/2} \bar{a}$.

$B = T^{T-1/2} H^T T^{T-1/2}$ kabi belgilaymiz. U holda quyidagiga ega bo'lamiz:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} = \frac{(B \bar{b}, \bar{b})}{(\bar{b}, \bar{b})}$$

Ushbu tenglik uchun (8) munosabatdan foydalangan holda B matritsaning minimal xos soniga mos keluvchi $\bar{b}_0$ xos vektorni topish mumkin. Keyin izlanayotgan vektor quyidagicha tarzda hisoblaniladi:

$$\bar{a} = B^{-1/2} \bar{b}$$

Biroq $B^{-1/2}$ matritsani topish jarayoni anchagina katta hisoblashlarni talab qiladi, shuning uchun yana bitta qo'shimcha cheklanishni kiritamiz:

$$\bar{a}^T T \bar{a} = 1 \quad (9)$$

U holda (5) minimallashtirish (9) shart bilan masalasi quyidagi kvadratik shaklning shartli minimumini qidirishga keltiriladi:

$$\bar{a}^T H \bar{a} \rightarrow \min \quad (10)$$

(10) masalani yechish uchun Lagranj ko'paytiruvchilari usulidan foydalanamiz, unga ko'ra optimal vektor $\bar{a}$ quyidagi Lagranj funksiyasi birinchi hosilasini nolga tenglashtiradi:

$$F(\bar{a}) = \bar{a}^T H \bar{a} - \nu \bar{a}^T T \bar{a}$$

bu yerda ν - kiritilayotgan Lagranj ko'paytiruvchisi.

Quyidagiga ega bo'lamiz:

$$\frac{\partial F(\bar{a})}{\partial \bar{a}} = 2H\bar{a} - 2\nu T\bar{a} = 0$$

yoki quyidagi chiziqli bir jinsli tenglamalar sistemasini olamiz:

$$H\bar{a} - \nu T\bar{a} = 0 \quad (11)$$

Ushbu tenglamani chapdan $\bar{a}$ ga ko'paytirib va (9) ni hisobga olgan holda quyidagini olamiz:

$$\bar{a}^T H \bar{a} = \nu \quad (12)$$

Ya'ni (5) ning qiymati (9) shart bilan kiritilgan Lagranj ko'paytiruvchisi bilan aniqlanadi. Shuning uchun ν sifatida minimal xos sonni tanlab olib va (11) tenglamalar sistemasini yechgan holda (5) ifodani minimallashtiruvchi $\bar{a}$ vektorni olamiz.

(11) ni chapdan T^{-1} ko'paytirgan holda quyidagini olamiz:

$$T^{-1}H\bar{a} = \nu \bar{a}$$

Ya'ni bu holda izlanayotgan optimal yechim (5) $T^{-1}H$ matritsaning minimal xos son ν ga mos keluvchi xos vektori bilan aniqlanadi. $\bar{a}_0$ orqali (5) ni minimallashtiruvchi topilgan optimal vektorni belgilaymiz.

$\bar{a}_0$ vektori komponentlarini o'sish tartibida joylashtiramiz, ya'ni:

$$a_{i1} \leq a_{i2} \leq \dots a_{ig}$$

bu yerda i_j -indeks, $\bar{a}_0$ vektorning a_{ij} komponentining joylashishini ko'rsatib, $\{1, 2, \dots, g\}$ to'plamidan qiymatlarni oladi.

$\bar{D}_0$ vektorining komponentlarini quyidagi sxema bo'yicha butun son qiymatlarini belgilaymiz:

$$d_{i1} = a_{i1} = 1; d_{i2} = a_{i2} = 2; \dots d_{ig} = a_{ig} = g$$

Natijada biz quyidagicha vektorni hosil qilamiz:

$$\bar{D}_0 = (d_1, \dots, d_g)$$

Buni darajalarning dastlabki almashinuvi sifatida ko'rib chiqamiz.

Ketma-ket yaqinlashish usuli bilan empirativ tartibli shkalani aniqlaymiz, ularning mohiyati quyidagicha tavsiflanadi:

(6) sur'at va mahrajini mos ravishda M_1 va M_2 bilan belgilaymiz. k va ℓ -chi gradatsiyalarini ko'chirishda sur'at va maxrajning o'sishi quyidagi formulalar yordamida hisoblanadi:

$$\Delta M_1 = \bar{c} \bar{d}, \quad \bar{c} = (c_1, \dots, c_g)$$

bu yerda

$$c_i = \begin{cases} 2(d_k - d_\ell)(h_{i\ell} - h_{ik}), & i \neq k, \ell; \\ (d_k - d_\ell)(h_{\ell\ell} - h_{kk}), & i = k, \ell; \end{cases}$$

$$\Delta M_2 = \bar{f}\bar{d}, \bar{f} = (f_1, \dots, f_g)$$

bu yerda

$$f_i = \begin{cases} 2(d_k - d_\ell)(t_{i\ell} - t_{ik}), & i \neq k, \ell; \\ (d_k - d_\ell)(t_{\ell\ell} - t_{kk}), & i = k, \ell; \end{cases} \quad (13)$$

$$\frac{1 + \Delta M_1/M_1}{1 + \Delta M_2/M_2} < 1$$

shartga ko'ra k va ℓ -gradatsiyalar darajalarini transpozitsiyasi (6) qiymatini kamaytiradi. Endi $g > 6$ holati uchun (6) ifodani minimallashtiradigan butun sonli gradatsiya darajalarining optimal almashtirishni qidirish uchun hisoblash algoritmlarini tavsiflaymiz:

Birinchi qadam. Yuqorida tavsiflangan usul bo'yicha $\bar{D}_0$ gradatsiya darajalarining dastlabki almashtirishlarni aniqlaymiz.

Ikkinchi qadam. Avvalgi k va ℓ gradatsiyasidagi interperatsiya juftliklari afzalini tanlash.

Uchinchi qadam. d_k va d_ℓ darajalarining transpozitsiyasida ΔM_1 va ΔM_2 larni aniqlash (13) ifoda qiymatini hisoblash.

To'rtinchi qadam. (13) ifodasida tengsizlik bajarilsa, $\bar{D}_0$ da yangi almashirishlarni hosil qilamiz, aks holda beshinchi qadamga o'tamiz.

Beshinchi qadam. Agar hosil bo'lgan (6) ifodadagi tartibning har qanday juft darajalarining transpozitsiyasi pasayishga olib kelmasa, u holda chiqish amalga oshiriladi. Aks holda ikkinchi qadamga o'tiladi.

Shunday qilib, yuqorida tavsiflangan algoritm $x^{\ell+1}$ belgisi uchun (6) ifodasini minimallashtirgan holda, $\bar{D}$ darajalarining optimal almashtirishni topishga imkon beradi. Agar bir nechta miqdoriy bo'lmagan belgilar mavjud bo'lsa, u holda $x^{\ell+2}, \dots, x^p$ belgilar bilan amallar bajariladi.

2.2- paragrafda timsollarni tanib olish masalalarida belgilar turlariga ishlovlar berish usul va algoritmlari, ularning turlari va ta'riflari keltirilgan. Timsollarni tanib olish masalasi uchun empirativ shkalalarni dispersiya nisbati minimallashtiriladigan sonli belgilar bilan berish ko'zda tutiladi.

2.3- paragrafda belgilar turlarini almashtirish algoritmlari asosida nomiqdoriy belgilar uchun ko'p o'zgaruvchili chiziqli regressiya tahlili masalalarni yechish usullari ko'rib chiqilgan. Regressiya tahlili masalasi uchun empirativ shkalani qidirish mezoni quyidagi ko'rinishni oladi:

$$\frac{\bar{a}^T s \bar{a}}{\bar{a}^T s' \bar{a}} \rightarrow \min \quad (14)$$

Faraz qilaylik, bog'liq bo'lgan belgi sifatida miqdoriy bo'lmagan $x_{\ell+1}^\xi$ belgisi qaralyapti. U holda $x_{\ell+1}$ uchun empirativ shkalani beruvchi sonli belgilar $1 - r_{\ell+1.1\dots\ell}^2$ kattalikni minimallashtirishi kerak. Ko'rsatish mumkinki, bu holda nomiqdoriy $x_{\ell+1}$ belgi uchun empirativ o'lchovni qidirish mezoni quyidagi shaklga ega bo'ladi:

$$\frac{\bar{a}^T s \bar{a}}{\bar{a}^T T \bar{a}} \quad (15)$$

bu yerda $T = [t_{ij}], i, j = \overline{1, g}$;

$$t_{ij} = \begin{cases} -\frac{n_i n_j}{N}, i \neq j \\ \frac{n_i}{N^2}(N - n_i), i = j \end{cases}$$

bu yerda $\bar{a} \in B^g$

Butun sonli belgilarni qidirishda (14) va (15) mezonlar quyidagi shaklda bo‘ladi:

$$\frac{\bar{D}^T S \bar{D}}{\bar{D}^T S' \bar{D}} \rightarrow \min \text{ va } \frac{\bar{D}^T S \bar{D}}{\bar{D}^T T \bar{D}} \rightarrow \min \quad (16)$$

bu yerda $\bar{D} \in P^g$.

(16) ifodadagi mezonlarni optimallashtirish uchun ko‘rsatilgan yondashuv ham qo‘llaniladi.

Dissertatsiyaning “**Ma’lumotlarni tahlil qilish masalalarida informativ belgilarni tanlab olish**” deb nomlangan uchinchi bobida timsollarni tanib olish masalalarida informativ belgilar fazosini tanlash usullari va algoritmlarini ishlab chiqish taklif qilinadi.

Ta’rif 1. $R^p = \{x = (x^{(1)}, x^{(2)}, \dots, x^{(p)})\}$ to‘plamni qisqartiramiz.

λ - orqali maydonni belgilab olamiz:

$$R^p|_{\lambda} = \{x|_{\lambda} = (\lambda^1 x^{(1)}, \dots, \lambda^p x^{(p)})\}, \lambda^k \in \{0,1\}$$

Ikki obyektlar $x, y \in R^p$ orasidagi qisqartirilgan oraliqni, $R^p|_{\lambda}$ to‘plamdagi y ’ni $x|_{\lambda}$ va $y|_{\lambda}$ orasidagi Yevklid oralig‘ini tushunamiz:

$$\|x - y\|_{\lambda} = \sqrt{\sum_{k=1}^p (x^{(k)} - y^{(k)})^2 \lambda^k} \quad (17)$$

belgilab olamiz:

$$S_q|_{\lambda} = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} (\|\bar{x}_q - x_{qi}\|_{\lambda})^2} \quad (18)$$

bu yerda $\bar{x}_q = \frac{1}{m_q} \sum_{i=1}^{m_q} x_{qi}$.

$S_q|_{\lambda}$ - λ bilan kesishadigan X_q sinfidagi obyektlarning tarqalishi o‘rtacha kvadrati.

λ dagi kesishishga nisbatan X_q, X_t sinflari orasidagi yaqinlik o‘lchovlarini aniqlaymiz:

$$R_{qt}|_{\lambda} = \sqrt{\frac{1}{m_q m_t} \sum_{i=1}^{m_q} \|\bar{x}_t - x_{qi}\|_{\lambda}^2 \sum_{i=1}^{m_q} \|\bar{x}_q - x_{ti}\|_{\lambda}^2} \quad (19)$$

λ dagi kesishishga nisbatan X_q, X_t sinflari orasidagi ajratish o‘lchovi sifatida quyidagi qiymatni tushunamiz:

$$L_{qt}|_{\lambda} = \frac{R_{qt}|_{\lambda}}{S_q|_{\lambda} S_t|_{\lambda}} \quad (20)$$

λ - informativlik quyidagi formula yordamida aniqlanadi:

$$I(\lambda) = \min_{q, t, q \neq t} L_{qt}|_{\lambda} \quad (21)$$

Usulning maqsadi λ vektorining ℓ informativligini topishdan iborat bo‘lib, bunda $I(\lambda)$ ning informativligi maksimal darajaga yetadi.

Barcha qiymatlarni qidirish usuli orqali $I(\lambda)$ fuksiyasini maksimumini topamiz:

$\lambda = (\lambda^1, \lambda^2, \dots, \lambda^k), \lambda^k \in \{0,1\}$ buning uchun,

$$\sum_{k=1}^p \lambda^k = \ell$$

belgilab olamiz:

$$\bar{x}_q^{(k)} = \frac{1}{m_q} \sum_{i=1}^{m_q} x_{qi}^{(k)}, q = \overline{1, m}; \quad (22)$$

$$S_q^k = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} (\bar{x}_q^{(k)} - x_{qi}^{(k)})^2} \quad (23)$$

$$R_{qt}^k = \sqrt{\frac{1}{m_q m_t} \sum_{i=1}^{m_q} (\bar{x}_t^{(k)} - x_{qi}^{(k)})^2 \sum_{i=1}^{m_q} (\bar{x}_q^{(k)} - x_{ti}^{(k)})^2} \quad (24)$$

(17) – (24) lardan shunday xulosa kelib chiqadi:

$$S_q \ell_\lambda = \sum_{k=1}^p \lambda^k S_q^k \quad (25)$$

$$R_{qt} \ell_\lambda = \sum_{k=1}^p \lambda^k R_{qt}^k \quad (26)$$

$I(\lambda)$ funksiyani maksimallashtirish C_p^ℓ varianti orqali qidirishni talab etadi.

Tasdiq 1. Har qanday $M, 0 \leq M \leq 2^p$ butun son uchun $\lambda \in \Lambda$ shunday ifoda mavjud, bunda $(\lambda) = M$.

$\Lambda^\ell - \Lambda$ ning qism-to‘plami bo‘lgani uchun, bu Λ^ℓ ham kiritilgan tartibga nisbatan qat’iy tartiblangan hisoblanadi.

Tasdiq 2. $\lambda_1(1, \dots, 1, 0, \dots, 0)$, berilgan bo‘lsin, u holda Q qoidasini aniqlaymiz.

$$\lambda = (*, *, \dots, *, 1, 0, 0, \dots, 0, 1, \dots, 1),$$

$$Q(\lambda) = (*, *, \dots, *, 0, 1, 1, \dots, 1, 0, \dots, 0)$$

$$\lambda_2 = Q(\lambda_1), \lambda_3 = Q(\lambda_2) \text{ bo‘lsin, holda } \{\lambda_i\} = \Lambda^\ell$$

Λ^ℓ to‘plamini maksimal va minimal elementlarini $\lambda_{\max}$ i $\lambda_{\min}$ orqali, ya’ni kiritilgan tartib ma’nosida belgilab olamiz:

$$1. \lambda_{\max} \in \Lambda^\ell, \lambda_{\min} \in \Lambda^\ell.$$

$$2. \text{Ixtiyoriy } \lambda \in \Lambda^\ell / \{\lambda_{\max}\}, \lambda_{\max} > \lambda$$

$$3. \text{Ixtiyoriy } \lambda \in \Lambda^\ell / \{\lambda_{\min}\}, \lambda_{\min} < \lambda$$

Bunda elementlar mavjud, chunki Λ^ℓ to‘plam cheklangan va qat’iy tartiblangan hisoblanadi.

Endi belgilarning ℓ -informativ qism-tizimini aniqlash algoritmini qadamlar bo‘yicha bayon qilamiz:

Birinchi qadam. Quyidagi qiymatlarni hisoblash: $x_q^{(k)}, k = 1, p, q = 1, m;$

$$S_q^k, k = 1, p; q = 1, m; R_{qt}^k, k = 1, p; q, t = 1, m (q \neq t);$$

Ikkinchi qadam. $\lambda = (0, \dots, 0, 1, \dots, 1), I_* = 0, \lambda_* = \lambda.$

Uchinchi qadam. Quyidagini hisoblash: $s_q|_\lambda, q = 1, m; R_{qt}|_\lambda, q, t = 1, m;$

$(q \neq t); L_{qt}|_\lambda, q, t = 1, m (q \neq t); I(\lambda)$ formulalar bo‘yicha (25), (26), (20),

(21).

To‘rtinchi qadam. Agar $I(\lambda) > I_*$, bo‘lsa, u holda $I(\lambda), \lambda_* = \lambda$ bo‘ladi, aks holda keyingi bosqichga o‘tamiz.

Beshinchi qadam. Agar $\lambda \neq (1, \dots, 1, 0, \dots, 0)$ bo‘lsa, u holda keyingi λ tanlanadi va 3- bosqichga o‘tish amalga oshiriladi. Aks holda jarayon tugaydi va ℓ -informativ vektor λ_* hisoblash natijasi amalga oshiriladi.

Yuqorida keltirilgan usul bizga ℓ -belgilardan iborat bo'lgan informativ qism tuzilmasini aniqlashni imkon beradi. O'quv tanlanma obyektlarni tanib olish maqsadida berilgan mezonidagi yuqori darajada informativ belgilarga ega bo'lgan belgilarning eng kichik qism tizimini dastlabki belgilar tizimida aniqlash usulini ko'rib chiqamiz.

Masalaning qo'yilishi. $X = \{x\}$ o'quv obyektlar to'plami berilgan bo'lsin.

$x|_{\lambda}$ va $y|_{\lambda}$ hamda $S_q|_{\lambda}$ orasidagi masofa (15) va (16) formulalari orqali aniqlanadi.

$x = (x^{(1)}, \dots, x^{(p)})$ -timsollarni tanib olish obyekti bo'lsin.

$$d_q(x)|_{\lambda} = \frac{\|x - \bar{x}_q\|}{S_q|_{\lambda}} \quad (27)$$

$d_q(x)|_{\lambda}$ da X_q sinfidagi x obyekti λ ga qisqartirishga nisbatan masofasining o'lchovi sifatida tushuniladi.

Qisqartirish uchun tanib olish tamoyillarini λ orqali aniqlaymiz. x obyektni X_t sinfiga tegishli deb hisoblaymiz, agar:

$$d_t(x)|_{\lambda} = \min_{q=1, \dots, m} d_q(x)|_{\lambda} \quad (28)$$

Tanib olishning ishonchligi o'lchovi sifatida biz quyidagi qiymatlardan foydalanamiz:

$$Q(\lambda) = M(\lambda) / \sum_{q=1}^m m_q \cdot 100\% \quad (29)$$

Yuqorida berilgan $\ell \leq p$ holat uchun ℓ belgilarning informativ kombinatsiyalarini aniqlash usuli taklif qilingan.

Bunday holda, har bir $\ell \in \{1, 2, \dots, p\}$ uchun ℓ -informativ vektor aniqlanadi.

$$\lambda(\ell) = (\lambda^1, \dots, \lambda^p) \\ \sum_{k=1}^p \lambda^k = \ell, \lambda^k \in \{0; 1\}$$

ℓ -ning eng kichigi keyingi mezon bo'yicha tanlanishi taklif etiladi:

$$Q(\lambda(\ell)) = \max_{1 \leq n \leq p} Q(\lambda(n)) \quad (30)$$

Qo'yilgan masalani yechish uchun algoritm quyidagi ko'rinishda ishlab chiqildi:

Birinchi qadam. $\bar{x}_q^{(k)}$, $k = \overline{1, p}$; $q = \overline{1, m}$; S_q^k , $k = \overline{1, p}$; $q = \overline{1, m}$; $\ell = 0$ (22), (23) formulalar orqali qiymatlarni hisoblaymiz.

Ikkinchi qadam. (25) va (26) formulalaridan foydalangan holda $\ell = \ell + 1$ uchun λ vektorni topamiz.

Uchinchi qadam. (29) formula orqali $Q(\lambda)$ aniqlaymiz.

To'rtinchi qadam. Agar $\ell = p$ bo'lsa, u holda 5-qadamga o'tiladi. Aks holda 2-qadamga qaytiladi.

Beshinchi qadam. (30) formuladan ℓ ni eng kichigini aniqlaymiz.

Ushbu algoritm natijasida (30) shartini qanoatlantiradigan eng kichik ℓ -informativ vektor olinadi.

Dissertatsiya ishining yakuniy bobi “**Nomiqdoriy ma'lumotlarga raqamli ishlov berish algoritmlari va amaliy masalalarni yechish**” deb nomlangan bo'lib, mazkur bobda dissertatsiya ishida ishlab chiqilgan algoritmlar asosida insonlarning

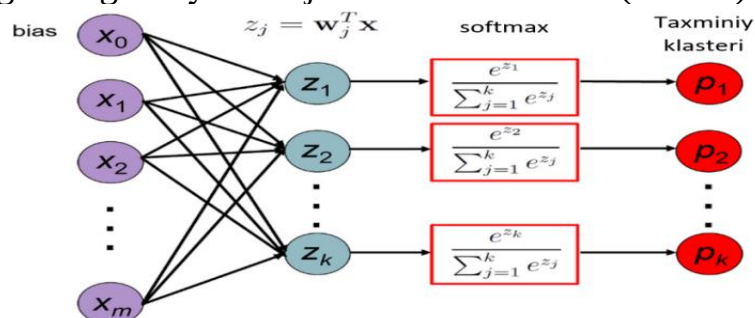
qon bosimlariga ta'sir etuvchi turli xil belgilar fazosida ifodalangan dori vositalarini guruhlarga ajratish algoritmi va dasturiy majmuasini ishlab chiqish tadqiq etilgan. "Timsollarni tanib olishda har xil tipda berilgan ma'lumotlarni klasterlash", "Har xil tipda berilgan ma'lumotlarni sinflashtirishda sun'iy neyron modeliga asoslangan hal qiluvchi qoida", "Har xil tipli ma'lumotlar fazosida informativ belgilar majmuasini tanlash" nomli dasturiy majmualar yaratilgan bo'lib, natijalar inson qon bosimiga ta'sir etuvchi 1100 tadan ortiq dori vositalari kesimida tajriba sinovdan o'tkazilgan. Taklif etilgan algoritmdan foydalanib klasterizatsiya masalasi yechilgan va o'quv tanlanma shakllantirilgan. O'quv tanlanma o'nta sinflardan iborat. Tanlanma belgilari turli xil belgilar fazosida berilgan bo'lib, ular nominal va qiymatli belgilardan iborat bo'lgan. Ishlab chiqilgan algoritm ikkala turdagi belgilarni inobatga olgan holda ishlaydi. O'quv tanlanma obyektlarining o'z sinflarini to'g'ri topishi neyron to'rlariga asoslangan algoritmi va dasturiy majmuadan foydalangan holda tajriba-sinovdan o'tkazilgan(1-jadvalga qarang).

1-jadval

Tadqiq etilgan uchta algoritmlarning qiyosiy tahlili

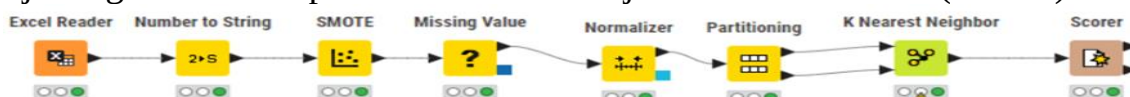
Sinflar	Neyron tarmoq orqali tanib olish aniqligi(%)	KNIME da yaratilgan model orqali tanib olish aniqligi(%)	Har xil turli belgilar fazosida berilgan obyektlarni klassifikatsiyalash algoritmi aniqligi(%)
1-Sinf	82.62	94.23	95.52
2-Sinf	81.40	93.48	90.18
3-Sinf	84.80	88.03	97.36
4-Sinf	86.23	90.26	88.07
5-Sinf	82.02	94.11	95.19
6-Sinf	83.79	91.48	90.27
7-Sinf	81.83	93.49	94.13
8-Sinf	85.61	88.43	93.47
9-Sinf	87.32	94.13	94.06
10-Sinf	82.55	90.88	96.47

Qon bosimiga ta'sir etuvchi turli xil turdagi dori-vositalarni klasterlash terapevtik ta'siridagi o'xshashliklari asosida tasniflashni o'z ichiga oladi. Klasterlash algoritmi orqali olingan natija ustida neyron to'ri yordamida obyektning qaysi sinfga tegishliligi bo'yicha tajriba-sinov o'tkazildi(1-rasm).



1-rasm. Neyron tarmoqning namunaviy tuzilishi

Bundan tashqari neyron tarmoqlarni **scratch-blok** shaklida qurishga mo'ljallangan **KNIME** platformasida ham tajriba-sinov o'tkazildi(2-rasm).



2-rasm. k-eng yaqin qo'shnilar KNIME modeli

Tahlilda klasterlash masalasi tadqiq etilgan bo'lib, bunda 1116 ta qon bosimiga ta'sir etuvchi har xil turdagi dori-vositalari jadval ko'rinishda olingan. Bunda bitta nominal belgi va to'qqizta miqdoriy belgi tadqiq etilgan.

Klasterlash masalasini yechimini topish uchun quyidagi algoritm taklif etilgan:

Birinchi qadam. Dori vositalar ma'lumotlari $x_i = (x_i^1, x_i^2, \dots, x_i^N) \in X$, $i = \overline{1, M}$; ga dastlabki ishlov beriladi. Bunda tushib qolgan ma'lumotlar to'ldiriladi, anomal ma'lumotlar o'rta qiymatga almashtiriladi va miqdoriy belgilar $\varepsilon^j = \frac{1}{M-1} \sum_{i=1}^{M-1} |x_{pi}^j - x_{pi+1}^j|$, asosida normirovka qilinadi.

Bu yerda $j = \overline{1, N}$; $p = \overline{1, r}$; $i = \overline{1, M-1}$; ga teng.

Ikkinchi qadam. O'quv tanlanma obyektlari ma'lumotlar bazasiga kiritiladi. Boshlang'ich ma'lumotlar bazasi barcha X_p , $p = \overline{1, r}$; sinf obyektlari kesimida shakllantiriladi;

Uchinchi qadam. X_p sinf obyektlarining o'z sinfini shakllanishiga qo'shgan hissasini baholashni aniqlashda ishlatiladigan nominal belgilar fazosida obyektlarning o'xshashligini ko'rsatuvchi kattalik (31) sonli belgilar fazosida obyektlarning o'xshashligini ko'rsatuvchi kattalik (32) formulalarga asosan hisoblanadi:

To'rtinchi qadam. Tanlanma belgilarida ixtiyoriy p – sinfdagi i – obyektни qolgan, shu sinfdagi $m_p - 1$ ta obyektlar majmuasidagi tutgan o'rni barcha $p = \overline{1, r}$; $i \neq q = \overline{1, m_p}$; $j = \overline{1, N}$; lar uchun $\rho_{pi}(x_{pi}, x_{pq})$ vektorning barcha parametrlarini hisoblanadi.

Ya'ni, $\rho_{pi}(x_{pi}, x_{pq}) = (\rho_{pi}^1(x_{pi}, x_{pq}), \rho_{pi}^2(x_{pi}, x_{pq}), \dots, \rho_{pi}^N(x_{pi}, x_{pq}))$ vektor belgilari $p = \overline{1, r}$; $i \neq q = \overline{1, m_p}$; $j = \overline{1, N}$;

Nominal belgilar uchun:

$$\rho_{pi}^j(x_{pi}, x_{pq}) = \begin{cases} 1, & \text{agar } (x_{pi}^j - x_{pq}^j) = 0 \\ 0, & \text{aks holda} \end{cases} \quad (31)$$

Sonli belgilar uchun:

$$\rho_{pi}^j(x_{pi}, x_{pq}) = \begin{cases} 1, & \text{agar } |x_{pi}^j - x_{pq}^j| \leq \varepsilon^j \\ 0, & \text{aks holda} \end{cases} \quad (32)$$

formulalar asosida hisoblanadi.

Beshinchi qadam. Ixtiyoriy p – sinfdagi i – obyektни qolgan, shu sinfdagi $m_p - 1$ ta obyektlar majmuasidagi tutgan o'rni foizda quyidagicha hisoblanadi:

$$v(X_p) = \frac{1}{m_p} \sum_{i=1}^{m_p} v(x_{pi}, X_p) = \frac{\frac{1}{m_p} \sum_{i=1}^{m_p} \kappa(x_{pi}, X_p) * 100\%}{N}$$

Oltinchi qadam. Tanlanma belgilarida ixtiyoriy p – sinfdagi i – obyekt ushbu sinfga keltiradigan foydasi hisoblanadi. Bunda yangi obyekt sinfga kelib qo'shilgandan keyingi sinfdagi obyektlarnin+

o'xshashlik darajasidan

$$v(X_q \oplus \bar{x}) = \frac{1}{m_q+1} \sum_{i=1}^{m_q+1} v(x_{qi}, X_q) = \frac{\frac{1}{m_q+1} \sum_{i=1}^{m_q+1} \kappa(x_{qi}, X_q) * 100\%}{N}$$

yangi obyekt kelmasidan oldingi qiymat ayiriladi,

$$v(X_q) = \frac{1}{m_q} \sum_{i=1}^{m_q} v(x_{qi}, X_q) = \frac{\frac{1}{m_q} \sum_{i=1}^{m_q} \kappa(x_{qi}, X_q) * 100\%}{N}$$

$v(X_q \oplus \bar{x}) - v(X_q)$ shu tarzda barcha sinflarning yangi obyektga nisbatan foydalarining eng katta qiymatga ega sinfga obyekt o'tqaziladi. Agarda foydalari teng bo'lsa, o'z sinfida qoldiriladi. Agar p – sinfda emas boshqa ikkita sinfda yuqori qiymatlar teng bo'lsa, u holda ushbu obyekt q ko'rsatkichi kichigiga o'tkaziladi.

XULOSA

“Turli xildagi ma'lumotlarni tahlil qilish masalalarida nomiqdoriy belgilarga raqamli ishlov berish usullari va algoritmlarini ishlab chiqish” mavzusidagi dissertatsiya ishi bo'yicha olib borilgan tadqiqotlar natijasida quyidagi xulosalar taqdim etildi:

1. Turli xil turdagi ma'lumotlarni tahlil qilishda sifatiy belgilardan miqdoriy belgilarga o'tish algoritmlari ma'lumotlarga dastlabki ishlov berish va belgilarning turlarini o'zgartirish usullari va algoritmlari asosida tadqiq yetildi;

2. Belgilarning juftlashgan o'zaro bog'liqliklari, belgilar aro korrelyatsiya va regressiyani tadqiq etishda empirativ tartib shkalasini aniqlashning g almashtirishlar soni butun son gradatsiya darajalarini optimal almashtirishini hisoblash algoritmlari ishlab chiqildi;

3. Ma'lumotlarni tahlil qilishda informativ belgilar majmuasini tanlash va tanlangan informativ belgilar fazosida sinflashtirish masalasini yechish usul va algoritmlari ishlab chiqildi;

4. Turli xil ko'rinishda berilgan obyektlarning o'xshashlik darajalarini aniqlashga asoslangan klasterlash algoritmlari ishlab chiqildi. Ular mavjud mashhur algoritmlar bilan dori vositalarini klasterlash amaliy masalasini yechishdagi qiyosiy tahlili keltirildi. O'quv tanlama sinf obyektlarni xarakterlovchi belgilar fazosi orqali klasterlash masalasini yechishning yangicha yondashuvi va algoritmlari taklif etildi;

5. Tanlama sinf obyektlarini xarakterlovchi belgilar fazosi orqali klasterlash masalasini yechishning yangicha yondashuvi va o'quv tanlanma obyektlarining o'z sinflarini to'g'ri topishi neyron to'rlariga asoslangan algoritm taklif etilgan;

6. Nomiqdoriy belgilarga raqamli ishlov berishda obyektlar o'xshashliklariga asoslangan klasterlash algoritmlari ishlab chiqildi;

7. Sinflashtirish, klasterlash, timsollarni tanib olish va ma'lumotlarni intellektual tahlilida ma'lumotlar belgilari metka ko'rinishda berilgan holatlar uchun dastlabki ishlov berishda belgilar majmuasini kamaytirish va ular asosida taklif qilingan masalalarni yechish usullari keltirildi;

8. O'zaro bog'langan muloqot oynasi rejimida ishlab chiqilgan ma'lumotlarni tahlil qilish tizimidan foydalangan holda turli xil turda berilgan tajriba ma'lumotlarni qayta ishlash usuli joriy etildi;

**НАУЧНЫЙ СОВЕТ DSC.13/05.05.2023.T.07.03 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

**ИНСТИТУТ ФУНДАМЕНТАЛЬНЫХ И ПРИКЛАДНЫХ
ИССЛЕДОВАНИЙ НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО
УНИВЕРСИТЕТА “ТИИИМСХ”**

ТУРСУНОВ АЛИШЕР ТУЛКУНОВИЧ

**РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ОЦИФРОВКИ
НЕКОЛИЧЕСТВЕННЫХ ПРИЗНАКОВ В ЗАДАЧАХ АНАЛИЗА
РАЗНОТИПНЫХ ДАННЫХ**

05.01.03- Теоретические основы информатики

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ
ДОКТОРА ФИЛОСОФИИ (PhD) ПО ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2025

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Министерстве высшего образования, науки и инноваций Республики Узбекистан за № В2023.4.PhD/T4167.

Диссертация выполнена в институте фундаментальных и прикладных исследований Национального исследовательского университета «ТИИИМСХ»

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-сайт Научного совета (www.tuit.uz) и на Информационно-образовательном портале «Ziynet» (www.ziynet.uz).

Научный руководитель:

Нишанов Ахрам Хасанович
доктор технических наук, профессор

Официальные оппоненты:

Примова Холида Анорбоевна
доктор технических наук, доцент

Дадаханов Мусохон Хошимхонович
доктор философии технических наук (PhD), доцент

Ведущая организация:

Ташкентский государственный технический университет имени Ислама Каримова

Защита диссертации состоится «22» января 2025 года в 14⁰⁰ часов на заседании Научного совета № DSc.13/05.05.2023.T.07.03 при Ташкентском университете информационных технологий. (Адрес: 100084, г.Ташкент, ул. Амира Темура, 108. Тел.: (99871)238-64-43, e-mail: ilmiy_kengash@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (зарегистрирована № 316). (Адрес: 100084, г. Тошкент, ул. Амира Темура. Тел.: (99871) 238-64-70).

Автореферат диссертации разослан «7» января 2025 года.
(протокол реестра № 6 от «6» января 2025 г.).



М.М.Камилов
Председатель Научного совета
по присуждению ученых степеней,
доктор технических наук, профессор, академик АН РУз

Н.А.Эгамбердиев
Учёный секретарь Научного совета
по присуждению ученых степеней,
доктор философии технических наук (PhD)

Н.О.Рахимов
Председатель научного семинара при Научном
совете по присуждению учёных степеней,
доктор технических наук, доцент

Введение (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В мире все более актуальными становятся бурное развитие современных информационно-коммуникационных технологий и необходимость решения таких задач, как обработка больших объемов данных, извлечение необходимой информации из объектов, распознавание образов. В развитых странах, таких как США, Китай, Япония, Россия и Германия, развитие научных основ в этих направлениях приобретает большое значение. В этом направлении исследований особое внимание уделяется разработке методов и алгоритмов цифровой обработки нечисловых признаков при анализе различных типов данных.

В мире востребована необходимость разработки алгоритмов кластеризации, основанных на определении степени сходства объектов, заданных в разных формах, и создание на их основе систем анализа данных. По этой причине процесс принятия решений экспертами отрасли становится очень сложным процессом. Актуальными задачами считаются предварительная обработка больших объемов данных, переход от большей размерности к меньшей при обработке нечисловых данных, методы классификации, изменение типов признаков, методы сокращения размерности признакового пространства в распознавании образов, выбор информативных признаков, решение задачи классификации в пространстве выбранных информативных символов и разработка современных систем распознавания.

В нашей республике также в этом направлении особое внимание уделяется разработке и внедрению аппаратно-программных комплексов, предназначенных для предварительной обработки и анализа данных и распознавания образов. В Стратегии действий по дальнейшему развитию Республики Узбекистан на 2017-2021 годы в том числе определено “.. внедрение информационно-коммуникационных технологий в систему экономики, социальной сферы и управления,... совершенствование систем обеспечения информационной безопасности и защиты информации, организация своевременных и адекватных действий против угроз в информационной сфере”¹

Данное диссертационное исследование служит реализации задач, определенных Указом Президента Республики Узбекистан от 19 февраля 2018 года УП №5349 “О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций”, УП №6079 “Об утверждении Стратегии “Цифровой Узбекистан-2030” и мерах по ее эффективной реализации” от 5 октября 2020 года, Указом № ПП №4699 “О мерах по широкому внедрению цифровой экономики и электронного правительства” от 28 апреля 2020 года, УП №60 “О новой Стратегии развития Узбекистана на 2022 - 2026 годы” от 28 января 2022 года, УП №158 “О стратегии Узбекистан-2030” от 11 сентября 2023 года и Постановлением Президента Республики

¹ Указ Президента Республики Узбекистан, от 07.02.2017 г. № УП-4947 “О Стратегии действий по дальнейшему развитию Республики Узбекистан”

Узбекистан от 28 декабря 2023 года ПП № 415 “О дополнительных мерах по ускорению цифровизации системы здравоохранения и внедрению передовых цифровых технологий” и другими нормативно-правовыми актами, касающимися данной деятельности.

Соответствие исследований приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий IV. “Развитие информатизации и информационно-коммуникационных технологий”.

Степень изученности проблемы. В последние годы вопросы анализа данных в части оцифровки признаков, отбора информативных признаков, разработки и совершенствования методов предварительной обработки изображений и распознавания образов, а также их внедрения в практику исследованы Ю.И.Журавлёвым, М.А.Айзерманом, В.В.Александровым, М.Браверманом, М.Б.Айдархановым, Н.Д.Горским, В.И. Васильевым, Р.Гонсалесом, Р.Вудсом, А.Г.Ивахненко, Э.Патриком, У. Преттом, Л.Робертсом, В.А.Сойфером, Г.С.Лбовым, Ш.Ульманом, П. Хартом, М.И. Шлезингером и другими зарубежными учеными.

В нашей республике большой вклад в разработку теоретических основ распознавания образов и предварительной обработки изображений, методов оцифровки при анализе данных, выбора информативных наборов признаков внесли М.М.Камилов, Ш.Х.Фазылов, Ф.Т.Адилова, Н.Т.Мирзаев, А.Х.Нишанов, Б.Б. Акбаралиев, Н.С.Маматов, С.С.Раджабов и другие.

В настоящее время, несмотря на прогресс в теории систем анализа данных, не достаточно изучены вопросы разработки методов и алгоритмов предварительной обработки данных, распознавания образов, выбора информативных наборов признаков и создания на их основе программного обеспечения для обработки данных.

Связь диссертационного исследования с планами научно-исследовательских работ научно-исследовательского учреждения, где выполнена диссертация. Диссертационное исследование выполнено в рамках научно-исследовательских проектов Ташкентского университета информационных технологий имени Мухаммада Ал-Хоразмий № BV- V- F4- 011–“Методы и алгоритмы решения некорректных задач интеллектуального анализа данных в условиях неопределенности”.

Цель исследования: разработка и совершенствовании методов и алгоритмов, основанных на выявлении сходства объектов, при анализе разнотипных данных.

Задачи исследования:

переход от качественных типов признаков к количественным при анализе разнотипных данных, методы предварительной обработки данных и преобразования типов признаков, формирование направления исследования;

разработка алгоритмов расчета оптимального выбора императивной шкалы порядка для целочисленных цифровых g меток для задачи исследования парных взаимосвязей признаков;

разработка методов и алгоритмов выбора информативных наборов признаков в процессе анализа данных и решение задачи классификации в пространстве информативных признаков;

разработка алгоритмов кластеризации, основанных на определении степени сходства объектов, заданных различными типами признаков;

разработка и совершенствование существующих алгоритмов кластеризации, основанных на сходстве объектов, при преобразовании типов признаков.

Объектом исследования были взяты различные признаки и атрибуты, отражающие важные особенности объектов, используемые при анализе информации, и подходы, основанные на определении и оценке сходства объектов.

Предметом исследования являются методы и алгоритмы цифровой обработки неколичественных признаков, преобразования типов признаков, распознавания образов и выбора набора информативного признаков.

Методы исследования. В ходе исследований были использованы методы интеллектуального анализа данных, цифровой обработки признаков, теории вероятностей и статистики, дискретной математики и распознавания образов.

Научная новизна исследования заключается в следующем:

при анализе разнотипных данных переход от качественных признаков к количественным признакам, первичная обработка данных и методы изменения типов признаков разработаны на основе логических функций, состоящих из эмпирических отношений объектов, с учетом схожести их значений;

изучение взаимных связей признаков, корреляции и регрессии признаков, а также определение эмпирической порядковой шкалы при вычислении перестановок g целочисленных градаций и алгоритм оптимальной замены целочисленных степеней градации основаны на исследовании взаимных связей признаков объектов;

разработаны методы и алгоритмы выбора информативных признаков и решения задачи классификации в пространстве выбранных информативных признаков на основе теории пересечений объектов λ в эвклидовом пространстве;

разработан алгоритм кластеризации, основанный на определении степени сходства объектов при наличии различных типов признаков на основе критерия, который количественно выражает сходства объектов в пространстве числовых признаков.

Практические результаты исследования заключаются в следующем:

в результате исследования в разрезе разработанных методов и алгоритмов разработан программный комплекс, предназначенный для пользователей, не являющихся специалистами в данной области, программистами;

разработано программное обеспечение под названием “Кластеризация разнотипных данных при распознавании образов”;

разработаны программные комплексы под названием “Решающее правило, основанное на модели искусственного нейрона при классификации

разнотипных данных” и “Выбор информативных признаков в пространстве разнотипных данных”.

Достоверность результатов исследования объясняется правильным применением методов и алгоритмов оцифровки неколичественных признаков, соответствием выводов и информации результатам экспериментальных исследований, а также решением практических вопросов.

Научная и практическая значимость результатов исследования.

Научная значимость результатов исследования объясняется тем, что разработанные алгоритмы на основе анализа информации способствуют дальнейшему развитию теоретических основ технологий цифровой обработки неколичественных признаков.

Практическая значимость результатов исследования объясняется тем, что разработанные алгоритмы и методы, а также созданный на их основе программный комплекс, позволили создать современную систему распознавания, способствующую принятию специалистами отрасли окончательного решения при анализе неколичественных данных.

Внедрение результатов исследования. На основе результатов, полученных по алгоритмам и программным комплексам выбора информативных признаков при анализе данных, оцифровки неколичественных признаков, преобразования типов признаков при решении задач множественного линейного регрессионного анализа:

на фармацевтическом предприятии ЧП «Enrich» внедрен в практику программный комплекс “Кластеризация разнотипных данных при распознавании образов”, разработанный на основе методов и алгоритмов снижения пространства описания, предоставляемых для анализа различных видов фармацевтической информации (Справка № 01/01-836 от 26 апреля 2024 года, выданная координирующим Экспертным советом министерства здравоохранения республики Узбекистан). В результате научного исследования была создана единая интегрированная система анализа экспериментальной информации, основанная на оцифровке неколичественных признаков и преобразования различных видов признаков в количественные. В результате удалось повысить эффективность работы в среднем на 10-12%;

программные комплексы “Решающее правило, основанное на модели искусственного нейрона при классификации разнотипных данных” и “Выбор информативных признаков в пространстве разнотипных данных”, разработанные на основе методов и алгоритмов отбора, классификации информативных признаков и сокращения размерности пространства описания при анализе информации (лекарственные средства), внедрены в практику в “Научной лаборатории инновационных фармацевтических соединений” (Справка № 01/01-836 от 26 апреля 2024 года, выданная координирующим Экспертным советом министерства здравоохранения республики Узбекистан). В результате научного исследования создана единая интегрированная система анализа экспериментальной информации на основе методов и алгоритмов снижения размерности пространства признаков, выбора информативных

признаков и преобразования типов признаков, а также классических методов анализа информации при распознавании образов. Это позволило повысить производительность труда в среднем на 10-15%.

программный комплекс “Выбор информативных признаков в пространстве разнотипных данных”, разработанный на основе методов и алгоритмов анализа экспериментальной информации путем выбора комплексов информативных признаков при решении задач анализа различных видов информации внедрена в практику проведения лабораторных работ кафедры “Организация и управление фармацевтическим производством” Ташкентского фармацевтического института (Справка № 01/01-836 от 26 апреля 2024 года, выданная координирующим Экспертным советом министерства здравоохранения республики Узбекистан). В результате анализа экспериментальных данных производительность работы была увеличена примерно в 1,2 раза.

Апробация результатов исследования. Основные теоретические и практические результаты диссертации были обсуждены на 5 международных и 8 республиканских научно-технических и научно-практических конференциях.

Опубликованность результатов исследования. Всего по теме исследования диссертации опубликованы 23 научные работы, из них 1 монография, 6 статей опубликованы в научных изданиях, рекомендованных к публикации основных научных результатов докторских диссертаций ВАК Республики Узбекистан, в том числе 1 зарубежном и в 5 республиканских журналах, 3 свидетельства на программные средства для ЭВМ получено от агентства интеллектуальной собственности при Министерстве юстиции Республики Узбекистан.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложений. Общий объем диссертации составляет 116 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

В введении диссертации рассматриваются актуальность и необходимость темы, соответствие исследования приоритетным направлениям развития науки и технологий республики, степень изученности проблемы, связь диссертации с планами научно-исследовательских работ высшего образовательного заведения, цель, задачи, объект, предмет, методы, научная новизна, практическая значимость результатов, достоверность, научная и практическая значимости, внедрение, апробация результатов исследования, публикация результатов, структура и объем работы.

В первой главе диссертации под названием “Современное состояние методов анализа разнотипных данных” рассматриваются методы и алгоритмы анализа данных при формировании обучающей выборки и компьютерной обработке различных видов экспериментальной информации. В параграфе 1.1 данной главы описаны история и современное состояние

широко распространенных в мире методов анализа данных. В параграфе 1.2 сформированы обучающие выборки для первичной обработки информации и рассмотрены наиболее распространенные методы распознавания образов. Приведены виды и модели программного обеспечения прикладной статистики. В параграфе 1.3 описаны существующие способы изменения типов признаков, а также указаны преимущества и недостатки этих методов. Под понятием “преобразование типов признаков” рассматривается преобразование разнотипных признаков в один вид: количественный, качественный или классификационный.

В последнем параграфе этой главы изложены цели и задачи диссертационной работы, сформулирована задача исследования:

При формулировании исследовательской задачи необходимо из общей выборки $X_i \in A$, $i = \overline{1, m}$ в p -мерном пространстве определить обучающие выборки. Здесь необходимо сформировать классы $X_{k1}, X_{k2}, \dots, X_{kN_k} \in A_k$, $k = \overline{1, m}$, где класс A_k состоит из N_k объектов, $A = \bigcup_{k=1}^m A_k$.

Для решения этой задачи кластеризации будет введена функция близости объектов. Если признаки объектов номинальны, то величина, показывающая их сходство, определяется через $\rho^j(x_{ki}, x_{kq})$ и измеряется по формуле:

$$\rho_{ki}^j(x_{ki}, x_{kq}) = \begin{cases} 1, & \text{если } (x_{ki}^j - x_{kq}^j) = 0 \\ 0, & \text{иначе} \end{cases} \quad (1)$$

Если объекты представлены числовыми знаками, то их сходство определяется величиной $\rho^j(x_{pi}, x_{pq})$ и рассчитывается по формуле (2):

$$\rho_{ki}^j(x_{ki}, x_{kq}) = \begin{cases} 1, & \text{если } |x_{ki}^j - x_{kq}^j| \leq \varepsilon^j \\ 0, & \text{иначе} \end{cases} \quad (2)$$

Здесь $k = \overline{1, m}$; $i \neq q = \overline{1, m_k}$; $j = \overline{1, p}$; $\rho_{ki}^j(x_{ki}, x_{kq}) = (\rho_{ki}^1(x_{ki}, x_{kq}), \rho_{ki}^2(x_{ki}, x_{kq}), \dots, \rho_{ki}^p(x_{ki}, x_{kq}))$.

В соответствии с индексом j в приведенной выше формуле (2) ε^j — для обозначений объектов каждого класса рассчитывается следующим образом:

$$\varepsilon^j = \frac{1}{N-1} \sum_{i=1}^{N-1} |x_{ki}^j - x_{ki+1}^j|,$$

здесь $j = \overline{1, p}$; $k = \overline{1, m}$; $i = \overline{1, N-1}$.

Во второй главе диссертации “**Методы и алгоритмы анализа разнотипных данных**” рассмотрены методы и алгоритмы оцифровки неколичественных признаков при анализе различных видов информации. В параграфе 2.1 этой главы обсуждаются вопросы, связанные с изучением взаимосвязи признаков.

Выражаем матрицу X размером $m * p$ в виде:

$$X = [X_1, X_2, X_3], \quad (3)$$

где $x_1 - \ell$ - количественные признаки $x^{(1)}, x^{(2)}, \dots, x^{(\ell)}$, X_2 - только качественные признаки с индексами $i = \overline{\ell + 1, m}$ и X_3 - полученными от X имеет только классификационные признаки с индексами $i = \overline{m + 1, p}$; часть, которая имеет. Будет выбран ряд натуральных чисел $K_g = \{1, 2, \dots, g\}$. При этом императивная шкала ξ присваивает каждой i -й градации целое число d_i

из множества K_g , которая называется меткой градации. В этом случае ξ выражается перестановкой градаций $D \in P^g$, где $D = (d_1, d_2, \dots, d_g)$, P_g – набор перестановок g натуральных чисел.

В качестве критерия выбора императивных шкал для преобразования типов признаков предлагается использовать минимизацию значения $1 - r_{12}^2$. Здесь r_{12}^2 – коэффициент корреляции между количественным признаком x_1 и оцифрованным нечисловым признаком x_2^ξ . Для R_{12} мы можем получить следующее отношение:

$$r_{12}^2 = \frac{\bar{a}^T C \bar{a}}{\bar{a}^T T \bar{a}},$$

где $C = \bar{\mu} \lambda_{11}^{-1} \mu^{-T}$, $\mu^T = [\mu_1^1, \mu_1^2, \dots, \mu_1^g]$. Получим

$$1 - r_{12}^2 = \frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} \quad (4)$$

Здесь $H = T - C$.

Таким образом, критерий для поиска действительных меток при изучении взаимосвязей признаков выглядит следующим образом:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} \rightarrow \min, \quad (5)$$

где $\bar{a} \in B^g$.

Перейдём к решению вопроса поиска целочисленных меток в соответствии с критерием (5), который в этом случае выглядит так:

$$\frac{\bar{D}^T H \bar{D}}{\bar{D}^T T \bar{D}} \rightarrow \min, \quad (6)$$

где $D \in P^g$.

Предположим, что число градаций классификационного признака $x_{\ell+1}$ не превышает шести, $g \leq 6$. В этом случае каждый уровень императивной шкалы g вместо целых числовых уровней соответствует $\bar{D}^T = (d_1, \dots, d_g)$, $d_i \in K_g$. Количество всех перестановок $g!$. Если максимальное значение g относительно невелико для рассматриваемого процесса, то предлагается полностью пересчитать (6) для всех значения перестановок. При этом определяем способ получения всех перестановок из заданного числа градаций классификационного признака $x_{\ell+1}$.

Зададимся начальным числом $k_1 = 1$. Добавляя к этому значению второе число $k_2 = 2$, а затем добавим вторую перестановку, получим $2!$. Иначе, поменяем местами $\{2,1\}$ $\{1,2\}$. На третьем шаге получим $3!$. Перестановок, на шаге с номером g получим искомые $g!$ перестановок.

Вычислительный алгоритм поиска оптимальной перестановки целочисленных рангов градаций при $g \leq 6$, используемый при определении императивной шкалы порядка, состоит из следующих шагов:

Первый шаг. Определение всех $g!$ перестановок по предложенной выше схеме.

Второй шаг. Определение начальной перестановки градаций $\bar{D}_0 = (1, 2, \dots, g)$ и вычисление значения $F(\bar{D}_0) = \frac{\bar{D}_0^T H \bar{D}_0}{\bar{D}_0^T T \bar{D}_0}$; $F_{\min} = F(\bar{D}_0)$.

Третий шаг. Переход к следующей перестановке $\bar{D}_k$, Вычисление значения $F(\bar{D}_k)$. Если $F(\bar{D}_k) < F_{min}$, тогда $F_{min} = F(\bar{D}_k), k = \overline{1, g}$.

Четвертый шаг. Если рассмотрены все $g!$ перестановок, то переход к пятому шагу, в противном случае, к третьему шагу.

Пятый шаг. Определение вектора $\bar{D}_k$, соответствующего минимальному значению F_{min} . Этот вектор принимается за искомую императивную шкалу.

Рассмотрим теперь случай, когда $g > 6$. При этом условии общее число возможных перестановок из g целочисленных рангов оказывается довольно большим ($g!$) и применение метода полного перебора для получения оптимальной императивной шкалы, удовлетворяющей критерию (6), становится неэффективным вследствие огромных затрат машинного времени. Поэтому оптимальную перестановку рангов предлагается искать с использованием метода последовательных приближений. При этом начальная перестановка $\bar{D}_0$ выбирается произвольным способом и конечный результат зависит именно от $\bar{D}_0$. Таким образом для различных начальных значений получаются различные локальные минимумы (6). Определим прежде всего способ вычисления начального приближения $\bar{D}_0$, для чего решим вспомогательную задачу. Предположим, что минимум (6) необходимо найти на множестве действительных чисел, т.е. расширим множество допустимых решений до B^g и задача поиска оптимального вектора будет иметь вид (6).

Отношение (5) называется обобщённым отношением Релея. Известно, что для обычного отношения Релея вида:

$$\frac{(Ax, x)}{(x, x)} \quad (7)$$

справедливы неравенства

$$v_1 \leq \frac{(Ax, x)}{(x, x)} \leq v_n, \text{ а также соотношение} \quad (8)$$

$$v_1 = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)}$$

где x - любой ненулевой вектор, $v_1, \dots, v_n$ - собственные числа матрицы A , а v_1 - минимальное из этих чисел.

Зададимся условием центрированности вектора $\bar{a}$, т.е.: $\frac{1}{N} \sum_{i=1}^g a_i n_i = 0$.

В этом случае матрица T для задач анализа взаимозависимости признаков примет диагональный вид с элементами $\frac{n_i}{N}$ и поэтому будет невырожденной. Тогда отношение (5) можно свести к виду (7), так как:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} = \frac{(H^T \bar{a}, \bar{a})}{(T^T \bar{a}, \bar{a})} = \frac{(T^{T-1/2} H^T T^{T-1/2} \bar{b}, \bar{b})}{(\bar{b}, \bar{b})},$$

где $\bar{b} = T^{T-1/2} \bar{a}$.

Обозначим $B = T^{T-1/2} H^T T^{T-1/2}$. Тогда имеем:

$$\frac{\bar{a}^T H \bar{a}}{\bar{a}^T T \bar{a}} = \frac{(B \bar{b}, \bar{b})}{(\bar{b}, \bar{b})}.$$

Используя для данного равенства соотношение (8), можно найти собственный вектор $\bar{b}_0$, соответствующий минимальному собственному числу матрицы B . Далее, искомый вектор вычисляется следующим образом:

$$\bar{a} = B^{-1/2} \bar{b}$$

Однако, процесс нахождения матрицы $B^{-1/2}$ является довольно громоздким, поэтому введём ещё одно дополнительное ограничение:

$$\bar{a}^T T \bar{a} = 1 \quad (9)$$

Тогда задача минимизации (5) при условии (9) сведётся к поиску условного минимума квадратичной формы:

$$\bar{a}^T H \bar{a} \rightarrow \min \quad (10)$$

Для решения задачи (10) используем метод множителей Лагранжа, по которому оптимальный вектор $\bar{a}$ обращает в нуль первую производную функции Лагранжа

$$F(\bar{a}) = \bar{a}^T H \bar{a} - v \bar{a}^T T \bar{a},$$

где v - вводимый множитель Лагранжа.

Имеем:

$$\frac{\partial F(\bar{a})}{\partial \bar{a}} = 2H\bar{a} - 2vT\bar{a} = 0,$$

или получим систему линейных однородных уравнений.

$$H\bar{a} - vT\bar{a} = 0 \quad (11)$$

Умножим это уравнение слева на $\bar{a}$ и, учитывая (9), получим:

$$H\bar{a} = v \quad (12)$$

т.е. значение (5) определяется введённым множителем Лагранжа при условии (9). Следовательно, выбрав в качестве v минимальное собственное число и решив систему уравнений (11), получим вектор $\bar{a}$, минимизирующий (5).

Умножив (11) слева на T^{-1} , получим: $T^{-1}H\bar{a} = v\bar{a}$ т.е. в этом случае искомое оптимальное решение (5) определяется собственным вектором матрицы $T^{-1}H$, соответствующим минимальному собственному числу v .

Обозначим через $\bar{a}_0$ найденный оптимальный вектор, минимизирующий (5). Упорядочим компоненты вектора по возрастанию, т.е.

$$a_{i1} \leq a_{i2} \leq \dots a_{ig},$$

где i_j - индекс, указывающий на место компонента a_{ij} вектора $\bar{a}_0$, принимает значения из множества $\{1, 2, \dots, g\}$.

Присвоим компонентам вектора $\bar{D}_0$ целочисленные значения по следующей схеме:

$$d_{i1} = a_{i1} = 1; d_{i2} = a_{i2} = 2; \dots d_{ig} = a_{ig} = g.$$

Получим в итоге некоторый вектор $\bar{D}_0 = (d_1, \dots, d_g)$, который будем рассматривать в качестве начальной перестановки рангов, используемой при определении императивной шкалы порядка методом последовательных приближений, суть которого излагается ниже.

Обозначим числитель и знаменатель отношения (6) через M_1 и M_2 соответственно. При транспозиции k -й и ℓ -й градаций приращение числителя и знаменателя вычисляется по формулам:

$$\Delta M_1 = \bar{c} \bar{d}, \quad \bar{c} = (c_1, \dots, c_g),$$

здесь

$$c_i = \begin{cases} 2(d_k - d_\ell)(h_{i\ell} - h_{ik}), & i \neq k, \ell; \\ (d_k - d_\ell)(h_{\ell\ell} - h_{kk}), & i = k, \ell; \end{cases}$$

$$\Delta M_2 = \bar{f}\bar{d}, \bar{f} = (f_1, \dots, f_g),$$

здесь

$$f_i = \begin{cases} 2(d_k - d_\ell)(t_{i\ell} - t_{ik}), & i \neq k, \ell; \\ (d_k - d_\ell)(t_{\ell\ell} - t_{kk}), & i = k, \ell; \end{cases} \quad (13)$$

$$\frac{1 + \Delta M_1/M_1}{1 + \Delta M_2/M_2} < 1$$

$$\frac{1 + \Delta M_1/M_1}{1 + \Delta M_2/M_2} < 1$$

Транспозиция рангов k -й и ℓ -й градаций уменьшит значение (6). Опишем теперь вычислительный алгоритм поиска оптимальной перестановки целочисленных рангов градаций, минимизирующей отношение (6) для случая $g > 6$:

Первый шаг. Определение начальной перестановки рангов градаций $\bar{D}_0$ по способу, описанному выше.

Второй шаг. Выбор некоторой отличной от предыдущей итерации пары градаций k и ℓ

Третий шаг. Определение ΔM_1 и ΔM_2 при транспозиции рангов d_k и d_ℓ . Вычисление значения отношения (13).

Четвертый шаг. При выполнении неравенства (13) получаем новую перестановку $\bar{D}_0$ в противном иначе переход к шагу 5.

Пятый шаг. Проверка условия останова. Выход осуществляется в том случае, если транспозиция любой пары рангов полученного упорядочения не приводит к уменьшению (6), иначе переход к шагу 2.

Таким образом, описанный выше алгоритм позволяет найти оптимальную перестановку рангов $\bar{D}$, минимизирующую отношение (6) для признака $x^{\ell+1}$. В случае наличия нескольких неколичественных признаков те же операции выполняются с признаками $x^{\ell+2}, \dots, x^p$.

В параграфе 2.2 приведены методы и алгоритмы обработки преобразования импов признаков в задачах распознавания образов, их виды и описания. Для распознавания символов предусмотрена выдача императивных шкал с числовыми знаками, минимизирующими дисперсионное соотношение.

В параграфе 2.3 предусмотрены способы решения вопросов, решаемых методами многомерного регрессионного анализа для неколичественных признаков на основе алгоритмов оцифровки признаков. Критерий поиска императивной шкалы для задачи регрессионного анализа выглядит следующим образом:

$$\frac{\bar{a}^T S \bar{a}}{\bar{a}^T S \bar{a}} \rightarrow \min \quad (14)$$

Предположим, что в качестве связанного знака рассматривается неколичественный признак $x_{\ell+1}^\xi$. Тогда число символов, дающих императивную шкалу для $x_{\ell+1}$, должно минимизировать величину $1 - r_{\ell+1.1 \dots \ell}^2$. Можно показать, что критерий поиска императивной шкалы для неколичественного признака $x_{\ell+1}$ имеет следующую вид:

$$\frac{\bar{a}^T S \bar{a}}{\bar{a}^T T \bar{a}} \quad (15)$$

здесь $T = [t_{ij}], i, j = \overline{1, g}$

$$t_{ij} = \begin{cases} -\frac{n_i n_j}{N}, i \neq j \\ \frac{n_i}{N^2} (N - n_i), i = j \end{cases}$$

где $\bar{a} \in B^g$

Критерии поиска целочисленных меток (14) и (15) имеют вид:

$$\frac{\bar{D}^T S \bar{D}}{\bar{D}^T S \bar{D}} \rightarrow \min \text{ в } \frac{\bar{D}^T S \bar{D}}{\bar{D}^T T \bar{D}} \rightarrow \min, \quad (16)$$

где $\bar{D} \in P^g$

Для оптимизации критериев (16) используется описанный выше подход.

В третьей главе диссертации «**Выбор информативных признаков в задачах анализа данных**» предлагается разработать методы и алгоритмы выбора пространства информативных признаков в вопросах распознавания образов.

Определение 1. Сократим множество $R^p = \{x = (x^{(1)}, x^{(2)}, \dots, x^{(p)})\}$.

Определим площадь через λ :

$$R^p|_{\lambda} = \{x|_{\lambda} = (\lambda^1 x^{(1)}, \dots, \lambda^p x^{(p)})\}, \lambda^k \in \{0, 1\}.$$

Под укороченным расстоянием между двумя объектами $x|_{\lambda}$ и $y|_{\lambda}$ мы понимаем Евклидово расстояние между $x, y \in R^p$ в множестве $R^p|_{\lambda}$:

$$\|x - y\|_{\lambda} = \sqrt{\sum_{k=1}^p (x^{(k)} - y^{(k)})^2 \lambda^k} \quad (17)$$

определим:

$$S_q|_{\lambda} = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} (\|\bar{x}_q - x_{qi}\|_{\lambda})^2}, \quad (18)$$

где $\bar{x}_q = \frac{1}{m_q} \sum_{i=1}^{m_q} x_{qi}$.

$S_q|_{\lambda}$ —средний квадрат распределения объектов класса X_q , усечённых с помощью λ .

Определим меры близости между классами X_q, X_t относительно усечения по λ :

$$R_{qt}|_{\lambda} = \sqrt{\frac{1}{m_q m_t} \sum_{i=1}^{m_q} \|\bar{x}_t - x_{qi}\|_{\lambda}^2 \sum_{i=1}^{m_t} \|\bar{x}_q - x_{ti}\|_{\lambda}^2} \quad (19)$$

Следующее значение мы понимаем как меру разделения классов X_q, X_t относительно усечения по λ :

$$L_{qt}|_{\lambda} = \frac{R_{qt}|_{\lambda}}{S_q|_{\lambda} S_t|_{\lambda}} \quad (20)$$

λ —информативность определяется по следующей формуле:

$$I(\lambda) = \min_{q, t \ q \neq t} L_{qt}|_{\lambda} \quad (21)$$

Цель метода — найти ℓ информативность λ вектора, при которой функция $I(\lambda)$ максимальна. Находим максимум функции $I(\lambda)$ методом перебора всех значений: $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^k)$, $\lambda^k \in \{0, 1\}$, для этого,

$$\sum_{k=1}^P \lambda^k = \ell$$

определим:

$$\bar{x}_q^{(k)} = \frac{1}{m_q} \sum_{i=1}^{m_q} x_{qi}^{(k)}, q = \overline{1, m} \quad (22)$$

$$S_q^k = \sqrt{\frac{1}{m_q} \sum_{i=1}^{m_q} (\bar{x}_q^{(k)} - x_{qi}^{(k)})^2} \quad (23)$$

$$R_{qt}^k = \sqrt{\frac{1}{m_q m_t} \sum_{i=1}^{m_q} (\bar{x}_t^{(k)} - x_{qi}^{(k)})^2 \sum_{i=1}^{m_q} (\bar{x}_q^{(k)} - x_{ti}^{(k)})^2} \quad (24)$$

из (17)-(24) вытекает следующее:

$$S_q \ell_\lambda = \sum_{k=1}^P \lambda^k S_q^k \quad (25)$$

$$R_{qt} \ell_\lambda = \sum_{k=1}^P \lambda^k R_{qt}^k \quad (26)$$

Максимизация функции $I(\lambda)$ требует поиска по вариантам C_p^e .

Утверждение 1. Для любого целого числа $M, 0 \leq M \leq 2^p$, существует выражение $\lambda \in \Lambda$ такое, что $(\lambda) = M$.

Поскольку Λ^ℓ - является подмножеством Λ^ℓ , это Λ^ℓ также строго упорядочен относительно порядка ввода.

Утверждение 2. Если дано $\lambda_1(1, \dots, 1, 0, \dots, 0)$, тогда определим правило Q :

$$\lambda = (*, *, \dots, *, 1, 0, 0, \dots, 0, 1, \dots, 1),$$

$$Q(\lambda) = (*, *, \dots, *, 0, 1, 1, \dots, 1, 0, \dots, 0)$$

$$\lambda_2 = Q(\lambda_1), \lambda_3 = Q(\lambda_2), \text{ тогда } \{\lambda_i\} = \Lambda^\ell.$$

Определим максимальный и минимальный элементы множества Λ^ℓ через $\lambda_{\max}$ и $\lambda_{\min}$, то есть в смысле введенного порядка:

1. $\lambda_{\max} \in \Lambda^\ell, \lambda_{\min} \in \Lambda^\ell$.
2. Для любого $\lambda \in \Lambda^\ell / \{\lambda_{\max}\}, \lambda_{\max} > \lambda$.
3. Для любого $\lambda \in \Lambda^\ell / \{\lambda_{\min}\}, \lambda_{\min} < \lambda$.

Он содержит элементы, поскольку множество Λ^ℓ конечно и строго упорядочено.

Теперь опишем пошагово алгоритм определения ℓ -информативной подсистемы признаков.

Первый шаг. Вычисление следующих значений: $x_q^{(k)}, k = 1, p, q = 1, m$;

$$S_q^k, k = 1, p; q = 1, m; R_{qt}^k, k = 1, p; q, t = 1, m (q \neq t).$$

Второй шаг. $\lambda = (0, \dots, 0, 1, \dots, 1), I_* = 0, \lambda_* = \lambda$.

Третий шаг. Вычисление значений: $s_q|_\lambda, q = 1, m; R_{qt}|_\lambda, q, t = 1, m (q \neq t); L_{qt}|_\lambda, q, t = 1, m (q \neq t); I(\lambda)$ по формулам (25), (26), (20), (21).

Четвертый шаг. Если $I(\lambda) > I_*$, тогда $I(\lambda), \lambda_* = \lambda$, иначе переход к следующему шагу.

Пятый шаг. Если $\lambda \neq (1, \dots, 1, 0, \dots, 0)$, тогда выбирается следующий λ и переход к шагу. Иначе процесс заканчивается и находится ℓ -информативный вектор λ_* .

Таким образом, описанный метод позволяет определить структуру информативной части, состоящей из ℓ признаков.

Рассмотрим метод определения наименьшей подсистемы признаков с высокоинформативными признаками по данному критерию в исходной системе с целью распознавания выбранных обучающих объектов.

Постановка задачи. Пусть дано множество $X = \{x\}$ обучающих объектов.

Расстояние между $x|_\lambda$ и $y|_\lambda$, а также $S_q|_\lambda$ определяются по формулам (15) и (16).

Пусть $x = (x^{(1)}, \dots, x^{(p)})$ – объекты распознавания.

$$d_q(x)|_\lambda = \frac{\|x - \bar{x}_q\|}{S_q|_\lambda} \quad (27)$$

$d_q(x)|_\lambda$ понимается как мера расстояния объекта x класса X_q по отношению к сокращению λ .

Для краткости определим принципы распознавания через λ . Мы считаем, что объект x принадлежит классу X_t , если:

$$d_t(x)|_\lambda = \min_{q=\overline{1,m}} d_q(x)|_\lambda \quad (28)$$

В качестве меры достоверности распознавания мы используем следующие значения:

$$Q(\lambda) = M(\lambda) / \sum_{q=1}^m m_q \cdot 100\% \quad (29)$$

Предлагается метод определения информативных комбинаций из ℓ признаков для случая $\ell \leq p$, приведенного выше.

В этом случае для каждого $\ell \in \{1, 2, \dots, p\}$ определяется ℓ – информативный вектор.

$$\begin{aligned} \lambda(\ell) &= (\lambda^1, \dots, \lambda^p) \\ \sum_{k=1}^p \lambda^k &= \ell, \lambda^k \in \{0; 1\} \end{aligned}$$

Предлагается выбирать наименьшее из ℓ по следующему критерию:

$$Q(\lambda(\ell)) = \max_{1 \leq n \leq p} Q(\lambda(n)) \quad (30)$$

Разработан алгоритм решения данной задачи в следующем виде:

Первый шаг. Вычисляем значения $\bar{x}_q^{(k)}$, $k = \overline{1, p}$; $q = \overline{1, m}$; S_q^k , $k = \overline{1, p}$; $q = \overline{1, m}$; $\ell = 0$ по формулам (22), (23).

Второй шаг. Используя формулы (25) и (26), находим вектор λ при $\ell = \ell + 1$.

Третий шаг. В соответствии с формулой (29) определим $Q(\lambda)$.

Четвертый шаг. Если $\ell = p$, тогда переход к шагу 5, иначе переход к шагу 2.

Пятый шаг. По формуле (30) находим наименьшее ℓ .

В результате работы данного алгоритма получается наименьший ℓ - информативный вектор, удовлетворяющий условию (30).

Четвертая глава диссертации «Алгоритмы оцифровки не количественных данных и решение практических задач» является заключительной главой диссертационной работы и на основе предложенных в диссертационной работе методов и алгоритмов классифицируются лекарственные средства, выраженные в фазе различных признаков, влияющих

на артериальное давление людей на основе разработанного алгоритма и программного комплекса. Созданы программные комплексы “Программа кластеризации разнотипных данных при распознавании образов”, “Программа решающего правила, основанная на модели искусственного нейрона при классификации разнотипных данных”, “Программа выбора комплекса информативных признаков в пространстве разнотипных данных” в результате проведены опытные испытания в разрезе более 1100 действующих лекарственных средств. С помощью предложенного алгоритма решен вопрос кластеризации и сформирован обучающий набор. Этот набор состоит из десят классов. Признаки представлены различными типами, состоящими из номинальных и количественных признаков. Разработанный алгоритм работает с учетом обоих типов признаков. Было проведено опробирование образовательных объектов с использованием алгоритма и программного комплекса, основанного на нейронах(см. таблицу 1).

1-таблица

Сравнительный анализ трех изученных алгоритмов

Класс	Точность распознавания по нейронной сети(%)	Точность распознавания по модели, созданной в KNIME(%)	Точность алгоритма классификации объектов, заданных в признаками разнотипных данных(%)
1-Класс	82.62	94.23	95.52
2- Класс	81.4	93.48	90.18
3- Класс	84.8	88.03	97.36
4- Класс	86.23	90.26	88.07
5- Класс	82.02	94.11	95.19
6- Класс	83.79	91.48	90.27
7- Класс	81.83	93.49	94.13
8- Класс	85.61	88.43	93.47
9- Класс	87.32	94.13	94.06
10- Класс	82.55	90.88	96.47

Кластеризация различных типов антигипертензивных препаратов предполагает их классификацию на основе сходства их терапевтических эффектов. По результату, полученному алгоритмом кластеризации, была проведена экспериментальная проверка принадлежности объекта к классу с помощью нейронной сети(см. рис. 1).

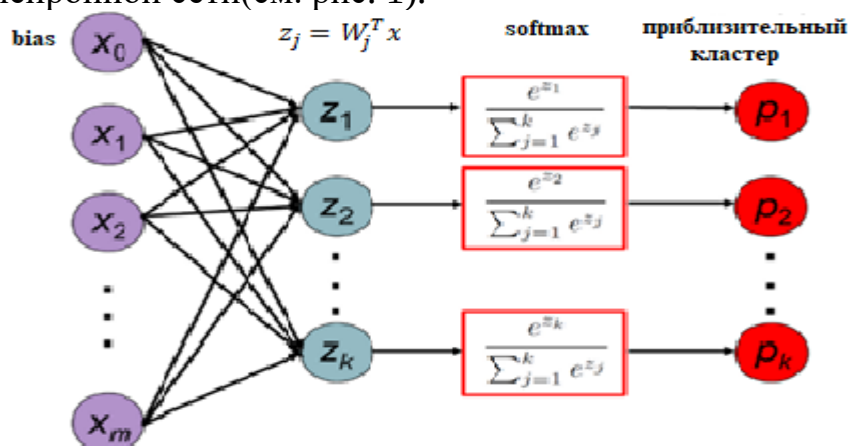


Рис. 1. Типовая структура нейронной сети.

Кроме того, был проведён экспериментальный тест на платформе **KNIME**, предназначенной для построения нейронных сетей в виде **scratch-блок**(см. рис. 2).

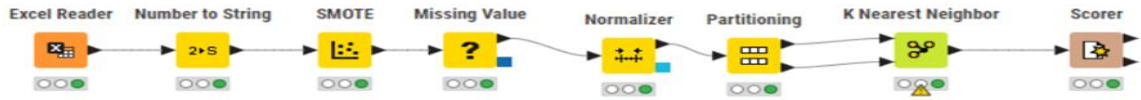


Рис. 2. Модель KNIME k-ближайших соседей.

В анализе изучен вопрос кластеризации, при котором в таблице взяты 1116 различных видов лекарственных средств, влияющих на артериальное давление. Здесь исследуется один номинальный и девять количественных признаков.

Для решения вопроса кластеризации предложен следующий алгоритм:

Первый шаг. Информация о лекарственных средствах предварительно обрабатывается при $x_i = (x_i^1, x_i^2, \dots, x_i^N) \in X$, $i = \overline{1, M}$. При этом пропущенная информация заполняется, аномальная информация заменяется на среднее значение и нормируется на основе количественных признаков $\varepsilon^j = \frac{1}{M-1} \sum_{i=1}^{M-1} |x_{pi}^j - x_{pi+1}^j|$. Здесь $j = \overline{1, N}$; $p = \overline{1, r}$; $i = \overline{1, M-1}$.

Второй шаг. Объекты обучающей выборки включаются в базу данных. База исходной внутренней информации формируется в разрезе всех X_p , $p = \overline{1, r}$.

Третий шаг. Величина, показывающая сходство объектов в пространстве номинальных признаков, используемых для определения вклада объектов класса X_p в формирование своего класса, и величина, показывающая сходство объектов в пространстве количественных признаков, рассчитывается по формуле сходства.

Четвертый шаг. В опознавательных знаках объект добровольного p – класса, оставшийся в комплексе $m_p - 1$ объектов того же класса, занимает для всех $p = \overline{1, r}$; $i \neq q = \overline{1, m_p}$; $j = \overline{1, N}$; для всех параметров вектора $\rho_{pi}(x_{pi}, x_{pq})$.

Т.е. $\rho_{pi}(x_{pi}, x_{pq}) = (\rho_{pi}^1(x_{pi}, x_{pq}), \rho_{pi}^2(x_{pi}, x_{pq}), \dots, \rho_{pi}^N(x_{pi}, x_{pq}))$ признаки вектора $p = \overline{1, r}$; $i \neq q = \overline{1, m_p}$; $j = \overline{1, N}$.

Для номинальных признаков:

$$\rho_{pi}^j(x_{pi}, x_{pq}) = \begin{cases} 1, & \text{если } (x_{pi}^j - x_{pq}^j) = 0 \\ 0, & \text{иначе} \end{cases}$$

Для количественных признаков:

$$\rho_{pi}^j(x_{pi}, x_{pq}) = \begin{cases} 1, & \text{если } |x_{pi}^j - x_{pq}^j| \leq \varepsilon^j \\ 0, & \text{иначе} \end{cases}$$

Пятый шаг. Место объекта в добровольном p -классе в комплексе из $m_p - 1$ объектов того же класса оценивается в процентах следующим образом и предварительно рассчитывается по формуле:

$$v(X_p) = \frac{1}{m_p} \sum_{i=1}^{m_p} v(x_{pi}, X_p) = \frac{\frac{1}{m_p} \sum_{i=1}^{m_p} \kappa(x_{pi}, X_p) * 100\%}{N}$$

Шестой шаг. В признаках выборки объектом добровольного i -го класса является польза, приносимая этому классу. При этом из-за степени сходства объектов в классе после добавления нового объекта:

$$v(X_q \oplus \bar{x}) = \frac{1}{m_{q+1}} \sum_{i=1}^{m_{q+1}} v(x_{qi}, X_q) = \frac{\frac{1}{m_{q+1}} \sum_{i=1}^{m_{q+1}} \kappa(x_{qi}, X_q) * 100\%}{N}$$

вычитается значение, предшествующее приходу нового объекта,

$$v(X_q) = \frac{1}{m_q} \sum_{i=1}^{m_q} v(x_{qi}, X_q) = \frac{\frac{1}{m_q} \sum_{i=1}^{m_q} \kappa(x_{qi}, X_q) * 100\%}{N}$$

$v(X_q \oplus \bar{x}) - v(X_q)$ - таким образом, к классу, имеющему наибольшую ценность в прибылях и убытках всех классов по отношению к новому объекту. Если они равны, то оставляются в своем классе. Если в двух других классах, а не в p -классе, более высокие значения равны, то q показателю этого объекта переводится в меньшую величину.

ЗАКЛЮЧЕНИЕ

На основе результатов исследований в рамках темы “Разработка методов и алгоритмов оцифровки нечисловых признаков при анализе разнотипных данных” можно сделать следующие выводы:

1. Разработаны алгоритмы перехода от качественных признаков к количественным для анализа различных типов данных, включая методы предварительной обработки данных и преобразования типов признаков.

2. Разработаны алгоритмы для определения императивной шкалы порядка, вычисления количества возможных перестановок и оптимизации целочисленных меток при исследовании взаимосвязей признаков, корреляции и регрессии.

3. Разработаны методы и алгоритмы для выбора информативных признаков и решения задачи классификации в пространстве выбранных информативных признаков.

4. Разработаны алгоритмы кластеризации, основанные на определении степени сходства объектов, представленных в различных формах. Проведен сравнительный анализ с существующими популярными алгоритмами кластеризации в задаче кластеризации лекарственных средств. Предложен новый подход и алгоритмы для кластеризации объектов обучающего набора на основе характеристик их признаков.

5. Предложен новый подход к решению задачи кластеризации объектов обучающего набора с использованием нейронных сетей, обеспечивающих правильное распределение объектов по своим классам в пространстве признаков.

6. Разработаны алгоритмы кластеризации, основанные на сходстве объектов, для оцифровки номинальных признаков.

7. Предложены методы уменьшения количества признаков при предварительной обработке данных в задачах классификации, кластеризации, распознавания образов и интеллектуального анализа данных, где признаки представлены в виде меток, а также предложены методы решения этих задач.

8. Внедрен метод обработки экспериментальных данных, представленных в различных форматах, с использованием системы анализа данных, разработанной в режиме взаимосвязанного окна.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSC.13/05.05.2023.T.07.03 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

**INSTITUTE OF FUNDAMENTAL AND APPLIED RESEARCH AT THE
NATIONAL RESEARCH UNIVERSITY “TIAME”**

TURSUNOV ALISHER TULKUNOVICH

**DEVELOPMENT OF METHODS AND ALGORITHMS FOR DIGITIZING
NON-QUANTITATIVE FEATURES IN PROBLEMS OF ANALYZING
DIFFERENT TYPES OF DATA**

05.01.03 – Theoretical basis of computer science

**DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY (PhD)
ON TECHNICAL SCIENCES**

Tashkent – 2025

The theme of doctor of philosophy (PhD) on technical sciences was registered at the Supreme attestation commission at the Ministry of higher education, science and innovation of the Republic of Uzbekistan under number №B2023.4.PhD/T4167.

The dissertation has been prepared at the institute of Fundamental and applied research under the national research university "TIAME".

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-educational portal (www.ziyonet.uz).

Scientific adviser:

Nishanov Ahram Hasanovich
Doctor of Technical Sciences, Professor

Official opponents:

Primova Xolida Anorboyevna
Doctor of Technical Sciences, Docent

Dadaxanov Musoxon Xoshimxonovich
Doctor of Philosophy in technical sciences (PhD), Docent

Leading organization:

Tashkent state technical university named after Islam Karimov

The defense of dissertation will take place "22" january 2025 at 14⁰⁰ at the meeting of Scientific council No. DSc.13/05.05.2023.T.07.03 at Tashkent University of Information Technologies (Address: 100084, Tashkent city, Amir Temur street, 108. Ph: (+99871) 238-64-43, e-mail: ilmiy_kengash@tuit.uz).

The dissertation can be reviewed at the Information Resource Centre of Tashkent University of Information Technologies (is registered under No 316). (Address: 100084, Tashkent city, Amir Temur street, 108. Ph.: (+99871) 238-64-70, e-mail: ilmiy_kengash@tuit.uz).

Abstract of dissertation sent out on "7" january 2025 y.
(mailing report No. 6 on "6" january 2025 y.).



M.M.Kamilov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor,
Academician of the Academy of Sciences
of the Republic of Uzbekistan

N.A.Egamberdiev
Scientific secretary of the scientific council
awarding scientific degrees,
Doctor of Philosophy in Technical Sciences(PhD)

N.O.Rakhimov
Chairman of the scientific seminar under the
scientific council awarding scientific degrees,
Doctor of Technical Sciences, Docent

INTRODUCTION (abstract of PhD dissertation)

The aim of the research the task involves the development and improvement of methods and algorithms based on the identification of object similarities through digital processing in the analysis of various types of data.

The object of the research as supported in data analysis, various symbols and attributes representing the important means of objects, identifying the similarities of objects and elements that can be used from them are obtained.

The scientific innovation of the research is as follows:

in the analysis of various types of data, the methods and algorithms for transitioning from qualitative attributes to quantitative attributes, initial data processing, and changing the types of attributes have been developed based on the similarity of the values of the logical functions composed of empirical relationships of objects;

the algorithm for calculating the number of g substitutions and the optimal substitution of integer gradation levels in determining the empirical order scale during the study of paired correlations, correlations, and regression of attributes has been developed based on the study of the interrelationships of object attributes;

the methods and algorithms for selecting a set of informative attributes and solving the classification problem in the space of selected informative attributes have been developed based on the theory of " λ " intersection between objects in Euclidean space;

the clustering algorithm, based on determining the similarity degree of objects given in various attribute spaces, has been developed based on a criterion that meaningfully expresses the similarity of objects in a numerical attribute space.

Implementation of the research results. Based on the results obtained from selecting informative features in data analysis, applying digital processing to quantitative features, and solving multivariate linear regression analysis problems, methods and algorithms for modifying feature types were developed:

for analyzing various types of pharmaceutical data and solving recognition problems, methods and algorithms for reducing the dimensionality of the given feature space have been developed, resulting in the software suite 'Clustering program for recognizing various types of data in symbol recognition.' This program has been implemented in practice at 'Enrich' Pharmaceutical company (The reference from the Ministry of Health of the Republic of Uzbekistan, Coordination Expert Council, dated April 26, 2024, with the number 01/01-836). As a result of the scientific research, a unified integrated system for experimental data analysis was created, based on the digital processing of quantitative features and modifying feature types to convert qualitative features. This led to an average increase in work efficiency by 10-12%.

based on the analysis of data (pharmaceutical products), methods and algorithms for selecting informative features, classification, and dimensionality reduction of the feature space have been developed. The software packages "Decision rule program based on artificial neural model for classifying various types of data" and "Feature selection program in various types of data spaces" have been

implemented in practice at the “Innovation pharmaceutical compounds scientific laboratory” (The reference from the Ministry of Health of the Republic of Uzbekistan, Coordination Expert Council, dated April 26, 2024, with the number 01/01-836). As a result of the scientific research, methods and algorithms for reducing the dimensionality of the feature space, selecting informative features, and modifying feature types in symbol recognition, as well as classical methods of data analysis, have been developed. A system for experimental data analysis was created based on these approaches. This led to an average increase in work efficiency by 10-15%.

based on methods and algorithms for analyzing experimental data through the selection of informative feature sets in solving various types of data analysis problems,² the software package “Feature selection program in various types of data spaces” has been developed. This program has been implemented in laboratory exercises at the “Pharmaceutical production organization and management” department of Tashkent Pharmaceutical institute (The reference from the Ministry of Health of the Republic of Uzbekistan, Coordination Expert Council, dated April 26, 2024, with the number 01/01-836). As a result, the work efficiency was increased by approximately 1.2 times when analyzing the experimental data.

The structure and scope of the dissertation. The dissertation consists of an introduction, four chapters, a conclusion, a list of references, and appendices. The total volume of the dissertation is 116 pages.

E'LON QILINGAN ISHLAR RO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I bo'lim (Часть I; Part I)

1. Narzullaev D.Z., Tursunov A.T. Обработка разнотипной информации в задачах анализа данных // Монография Т.: «Ilmiy-texnika axboroti press nashriyoti». Toshkent-2022. -152 с.

2. Nishanov A.X., Akbarova M.Kh., Tursunov A.T., Ollamberganov F.F., Rashidova D.E. Algorithm for clustering different types of drugs affecting blood pressur. Journal of Mathematics, Mechanics and Computer Science. ISSN1563– 0277. JMMCS. №3 (123).2024. -P.108-120 (Web of Science, №1).

3. Nishanov A.X., Tursunov A.T., Ollamberganov F.F. Алгоритм кластеризации, основанный на сходстве объектов. Kimyoviy texnologiya.nazorat va boshqaruv. Xalqaro ilmiy-texnikaviy jurnal. 2024 yil № 4. – Б.10-21 (05.00.00; № 12).

4. Tursunov A.T., Nishanov A.X. Применение методов оцифровки в задачах распознавания образов. “ТАТУ хабарлари” №3/2023 (67) 2023. ISSN 2010-9857. -Б.79-93 (05.00.00; № 31).

5. Nishanov A.X., Narzullayev D.Z., Tursunov A.T. Разработка методов и алгоритмов снижения размерности пространства описания для задачи распознавания образов. “ТАТУ хабарлари” №1/2024 (67) 2024 ISSN 2010-9857 -Б.87-99 (05.00.00; № 31).

6. Narzullayev D.Z., Abduraxmonov B.A., Tursunov A.T., Urinov E.M. Оцифровка признаков для задачи регрессионного анализа. “Informatika va energetika va muammolari” ilmiy-jurnal 2022 yil № 6. (05.00.00; № 5)

7. Nishanov A.X., Tursunov A.T., Ollamberganov F.F. Qon bosimiga ta'sir etuvchi har xil tipda berilgan dori vositalarini klasterlash algoritmi. O'zbekiston Milliy axborot agentligi Ilm-Fan bo'limi elektron jurnal. 2024 yil № 9 (59). _- Б.115- 123 (263/7.1 va 263/7.4-son OAK Rayosatining qarori).

II bo'lim (Часть II; Part II)

8. Narzullaev D.Z., Baydullaev A.S., Abdurakhmanov B.A., Tursunov A.T., Ilhamov Kh.Sh. Development of Methods and Algorithms for Dimension Reduction of Space Description for Pattern Recognition Problem. Information Technologies and Intelligent Decision Making Systems. Third International Scientific and Practical Conference, ITIDMS 2023 Moscow, Russia, December, 12–14, 2023 Revised Selected Papers.92-102 pages (Scopus conference).

9. Tursunov A.T. Принципы организации вычислений в системе анализа данных. Toshkent farmatsevtika institutining 85-yilligiga bag'ishlangan «Farmatsevtika sohasining bugungi holati: muammolar va istiqbollar» mavzusida xalqaro ilmiy-amaliy anjuman konferensiyasi materiallari to'plami. Toshkent, 2022 -yil 25-26 noyabr. –Б.29-30.

10. Tursunov A.T., Narzullayev D.Z. Разработка методов и алгоритмов анализа данных разнотипной природы. Материалы IV международной научно-практической конференции, “Современное состояние фармацевтической отрасли: проблемы и перспективы” Toshkent. 2023-yil 23-oktyabr. -Б. 46-47.

11. Tursunov A.T., Narzullayev D.Z. Методы и алгоритмы анализа данных при обработки экспериментальные данные различной природы. Toshkent farmatsevtika institutining 85-yilligiga bag‘ishlangan «Farmatsevtika sohasining bugungi holati: muammolar va istiqbollar» mavzusida xalqaro ilmiy-amaliy anjuman konferensiyasi materiallari to‘plami. Toshkent, 2022-yil 25-26 noyabr. – Б.26-27.

12. Tursunov A.T. Разработка алгоритмов преобразования типов признаков для задачи распознавания образов. Материалы IV международной научно-практической конференции, “Современное состояние фармацевтической отрасли: проблемы и перспективы” Toshkent. 2023-yil 23- oktabr. -Б.45-46.

13. Tursunov A.T. Farmasevtik korxonalarda ma’lumotlarni tahlil qilish asosida ishlab chiqarish samaradorligini oshirish. Abu Ali Ibn Sino va zamonaviy farmasevtikada innovatsiyalar V xalqaro ilmiy amaliy anjuman maqolalar to‘plami-Toshkent-2022. -Б.80-82.

14. Narzullaev D.Z., Baydullaev A.S., Mamatkulov Z.U., Abdurahmonov B.A. Tursunov A.T. Information system “Farms of the Republic of Uzbekistan” International Conference on Remote Sensing of the Earth: Geoinformatics, Cartography, Ecology, and Agriculture (RSE 2022) 151-15512296 (Scopus conference).

15. Narzullayev D.Z., Ilkhamov H.Sh., Tursunov A.T., Baydullayev A.S., Akhemod Yu.A. Automation of the Agricultural Sector of the Republic of Uzbekistan. E3S Web of Conferences 392,01037(2023). II International Conference on Agriculture, Earth Remote Sensing and Environment (RSE-II-2023). <https://doi.org/10.1051/e3sconf/202339201037> (Scopus conference).

16. Narzullayev D.Z., Shadmanov K.K., Baydullayev A.S., Radjabov E.E., Tursunov A.T. Automated farm management system in Uzbekistan. IOP Conf. Series: Earth and Environmental Science 723 (2021)032036. doi:10.1088/1755- 1315/723/3/032036 (Scopus conference).

17. Narzullaev D.Z., Samigova N.A., Tursunov A.T., Shadmanov K.K. Methods for assessing risk factors to the health of highly qualified athletes. E3S Web of conferences 284.01010 2021 tracee 2021(Scopus conference).

18. Shadmanov K.K., Narzullaev D.Z., Tursunov A.T., Karshiboev Sh.U. Системный подход к проблемам математического моделирования. Материалы Республиканской научно-практической конференции «Наука и инновации в современных условиях Узбекистана», часть I, Нукус, 20 мая 2020. -С.43-47.

19. Nishanov.A.X., Shadmanov K.K., Tursunov A.T. Qon bosimiga ta’sir etuvchi dori vositalarini klasterlash algoritmi. “Zamonaviy farmatsevtika sohasini rivojlanishining dolzarb masalalari va tendensiyalari” mavzusidagi II xalqaro ilmiy-amaliy anjuman materiallari. Toshkent.2024-yil 17-18-oktabr. -Б.377-379.

20. Narzullayev D.Z., Shadmanov K.K., Tursunov A.T. Turli xildagi ma'lumotlarni tahlil qilishda timsollarni tanib olish usullari. "Zamonaviy farmatsevtika sohasini rivojlanishining dolzarb masalalari va tendensiyalari" mavzusidagi II xalqaro ilmiy-amaliy anjuman materiallari. Toshkent. 2024-yil 17- 18 oktabr. -B.448-450.

21. Nishanov A.X., Tursunov A.T., Ollamberganov F.F., Rahmanov A.T. Timsollarni tanib olishda har xil tipda berilgan ma'lumotlarni klasterlash dasturi. O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligining EHM uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnomasi №DGU 202402525. Toshkent. 07.03.2024.

22. Nishanov A.X., Tursunov A.T., Ollamberganov F.F., Rahimova D.E. Har xil tipda berilgan ma'lumotlarni sinflashtirishda sun'iy neyron modeliga asoslangan hal qiluvchi qoida dasturi. O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligining EHM uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnomasi №DGU 202403123. Toshkent. 20.03.2024.

23. 8Nishanov A.X., Tursunov A.T., Ollamberganov F.F., Rahimova D.E. Har xil tipli ma'lumotlar fazosida informativ belgilar majmuasini tanlash dasturi. O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligining EHM uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnomasi № DGU 202403124. Toshkent. 20.03.2024.

«Muhammad al-Xorazmiy avlodlari» ilmiy jurnali
tahririyatida tahrirdan o'tkazildi va o'zbek, rus va ingliz
tillaridagi matnlarini mosligi tekshirildi