

УДК 004.9
МРНТИ 20.53.19

ЛОРЕНЦ МЕТРИКАСЫН ҮЛГІ ТАНУДА ҚОЛДАНУ

Е.Р. КЕРІМБЕКОВ, Е.С. СЕЙІТКАМАЛ

Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті

Аңдатпа: Техника мен технологиялардың дамуы ғылым әлемінде төңкеріс жасауда, атап айтсақ бір ғана есептеуіш техникасының пайда болуы кез келген ғылым саласына жаңа серпін беріп, түрлі жаңалықтардың ашылуына жол ашты. Қазіргі таңда ғылым салаларының бірлесуі ғылым әлемінде тың ойлардың туындауын және бұрынғы әдістердің қаншалықты дәрежеде оңтайлы екенін бағалауға мүмкіндік береді. Бұл мақалада ең әуелі зерттеліп отырған тақырып бойынша шетелдік жарияланымдарға әдебиеттерге шолу жасадық, Лоренц метрикасының Евклид кеңістігімен салыстырып айырмашылығын формулалар мен суреттерде мысал ретінде келтірдік. Лоренц метрикасын пайдалана отырып үлгі тануда жаңа алгоритм құрдық және бұл алгоритмнің тиімділігін тексеру үшін деректер базасында түрлі сынақтардан өткіздік. Жүргізілген тәжірибе нәтижесінде Лоренц метрикасы арқылы құрылған алгоритм классикалық алгоритмдермен салыстырылды, атап айтсақ Bayes, kNN және осы секілді алгоритмдер, содан соң нақты түрде сараптамалық нәтижелерді ұсындық.

Түйінді сөздер: Лоренц метрикасы, Евклид кеңістігі, ақылды алгоритмдер, үлгі тану, жіктеу

APPLICATION LORENTZ METRICS IN PATTERN RECOGNITION

Abstract: The development of technologies and technologies has revolutionized the world of science, in particular, the emergence of only one computing technology has given new impetus to science and led to the discovery of various innovations. Nowadays, the integration of science into the world of science allows us to judge the emergence of new ideas and the optimality of the past. In this article, we first reviewed the literature for foreign publications on the topic under study and compared the Lorentz metric with Euclidean space as an example in formulas and illustrations. Using the Lorentz metric, we created a new model recognition algorithm and checked the database to verify the effectiveness of this algorithm. As a result of the experiment, the algorithm created by the Lorentz metric was compared with classical algorithms, namely Bayes algorithms, kNN and similar ones, and then presented specific results.

Keywords: Lorentz metric, Euclidean space, intellectual algorithms, pattern recognition, classification

ПРИМЕНЕНИЕ ЛОРЕНЦОВОЙ МЕТРИКИ В РАСПОЗНАВАНИИ ОБРАЗОВ

Аннотация: Развитие техники и технологий произвело революцию в мире науки, в частности, появление только одной вычислительной техники дало новый импульс науке и привело к открытию различных инноваций. В наши дни интеграция науки в мир науки позволяет нам судить о появлении новых идей и оптимальности прошлого. В этой статье мы впервые рассмотрели литературу для зарубежных публикаций по исследуемой теме и сравнили метрику Лоренца с Евклидовым пространством в качестве примера в формулах и иллюстрациях. Используя метрику Лоренца, мы создали новый алгоритм распознавания моделей и проверили базу данных, чтобы проверить эффективность этого алгоритма. В результате эксперимента алгоритм, созданный метрикой Лоренца, сравнивался с классическими алгоритмами, а именно с алгоритмами Байеса, kNN и аналогичными, затем представлял конкретные результаты.

Ключевые слова: метрика Лоренца, Евклидово пространство, интеллектуальные алгоритмы, распознавание образов, классификация

Кіріспе

Бұл жұмыста біз Лоренц кеңістігінің қашықтықты өлшеуін жіктеу мәселелерінде қолдануды ұсынамыз. Осы тәсіл метрикалық оқытудың және көп қырлы оқытудың жалпы парадигмасына жатады [1,2]. Лоренц кеңістігі өзінің жеке қасиеттері бар классификация мәселесінде тиімді шешімдерді табуға жоғары әлеуетке ие [3]. Лоренц кеңістіктегі қашықтықты метрика оң болып табылмайды, сол себепті біз деректер нүктелерін жіктеу жылдамдығын арттыру үшін ерекше түрде ұйымдастыруға мүмкіндік аламыз. Сондықтан Лоренц қашықтығы жіктеу мәселелерінде пайдалы. Лоренц кеңістігі әдетте физикада жалпы және арнайы салыстырмалық теорияларының өрістерінде, сондай-ақ ғарыш теориясы мен дифференциалдық салалардағы математикада пайдаланылады [4]. Физикада, әсіресе салыстырмалы теория тақырыбында, Лоренц кеңістігі мен оның қасиеттері туралы теориялық материалдар жеткілікті. Лоренц кеңістігінде математикалық операцияларды қолдану зерттелді [5].

Лоренц кеңістігін пайдалану туралы кейбір ғылыми зерттеулер бар [6,7]. Жалпы алғанда, бұл зерттеулерде Лоренц кеңістігінің артықшылығын пайдаланып, жаңа функционалды экстракция және өлшемді азайту әдістері жасалды. Содан кейін зерттеушілер Лоренц метрикалық тензорды жартылай бақыланатын өлшемді азайту әдісін ұсынды. Авторлар жаңа үлгіде таңбаланбаған үлгілердің жергілікті құрылымын сақтауы мүмкін екендігін атап өтті [8]. Осылайша, ұсынылған алгоритм жоғары дисперсиялық кимылдарды зерделеу мүмкіндігіне ие.

1. Лоренц кеңістігіндегі қашықтық пен аймақ метрикасы

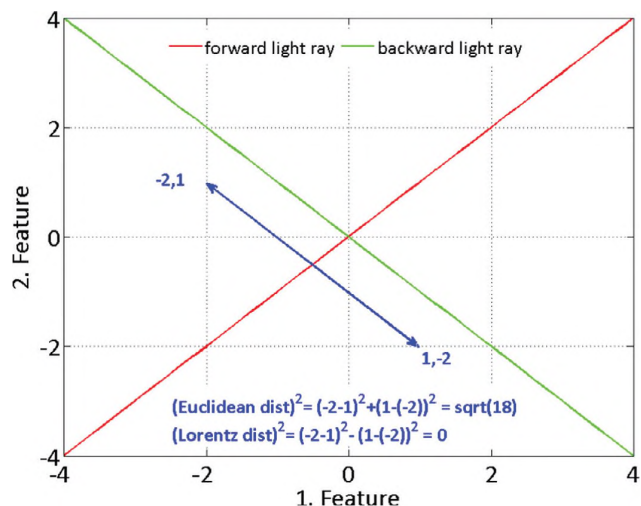
Лоренц кеңістігі математика, кеңістіктік теория және дифференциалдық геометрияда кеңінен қолданылатын кеңістік болып табылады. Дегенмен, инженерлік қосымшалар бойынша зерттеулер жеткіліксіз. Сондықтан, Лоренц кеңістігіндегі математикалық білімдерді әдебиетте тануды және жіктеуді тану сияқты жаңа салалар ретінде қолданылады.

Лоренц кеңістігі ерекше кеңістік және Евклидтік емес кеңістік деп аталады. Лоренц кеңістігіндегі ішкі айнымалылар Евклид кеңістігінен ерекшеленеді, өйткені Лоренц кеңістігінде он анықтылықтың жағдайы қолайсыз болып табылады. Евклид кеңістігіндегі тең қашықтықтың нүктелерінің орналасуы шеңберді құрайды. Дегенмен, Лоренц кеңістігіндегі бірдей қашықтықтардың нүктелерінің құрылымы басқаша. Лоренц кеңістігіндегі аймақ тұжырымдамасын зерттеу үшін ең алдымен осы кеңістіктегі екі нүктенің арасындағы қашықтықты түсіну қажет. Әр кеңістіктегі қашықтықты есептеу кезінде арнайы метрикалар анықталуы керек [9].

$$d(X,Y) = \sqrt{\left(\sum_{i=1}^{l-1} |x_i - y_i|^2\right) - |x_l - y_l|^2} \quad (1)$$

мұнда 1 – кеңістіктің өлшемі (элементтер саны) және қашықтықтың өлшеміне соңғы өлшемнің қосындысы теріс екенін ескеріңіз.

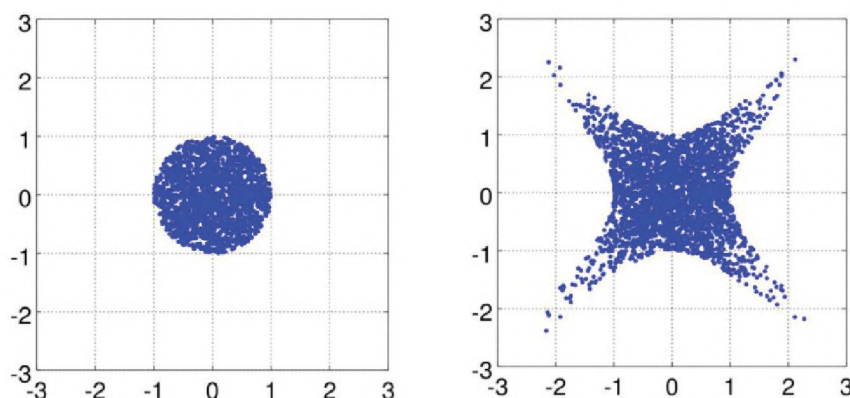
1-формуладан көріп отырғанымыздай, Лоренц метрикасы уақыттың өсіне сәйкес келетін екінші мерзімде кері таңбалы белгіге ие. Лоренц метрикасының негізгі айырмашылығы – екі нүктенің арасындағы қашықтық нөлге тең болуы мүмкін.



1-сурет. Лоренц пен Евклид қашықтықтарын салыстыру

Лоренц пен Евклид метрикаларын пайдалана отырып, екі нүкте арасындағы қашықтықты есептеуді салыстыру 1-суретте көрсетілген. Бұл суретте 2 нүкте берілген: $(-2, 1)$ және $(1, -2)$. Бірінші координат бірінші нысанға, екін-

шісі екінші нысанға тиесілі. Егер бұл нүктелер Евклид кеңістікте болса, онда қашықтық тең №18. Егер бұл нүктелер Лоренц кеңістігінде болған деп санасақ, онда (1) формула бойынша арақашықтық нөлге тең болады.



2-сурет. а) Евклид кеңістігінде Гаусс тарату арқылы блоктардың орналасуы және ә) Лоренц кеңістігінде

Лоренц кеңістігінде Лоренц аралығындағы екі нүктенің арасындағы қашықтық көлденең бағытта параллель 45° -та нөлге тең. Содан кейін аймақты бейнелеу үшін, біз Гаусс таратуына сәйкес кездейсоқ 5000 нүкте аламыз. Біріншіден, Евклид қашықтығы $d \leq 1$ бойынша координат жазығының басына ең жақын нүктелері таңдалады. Бұл 2.а суретте көрсетілген Евклид бірлігінің шеңбері. Екіншіден, Лоренц аралығындағы $d \leq 1$ қашықтығымен координат жазығының басына ең жақын нүктелер таңдалады. Бұл 2.ә суретте көрсетілген крестің гипербола-лық пішіні. Осылайша Евклид және Лоренц кеңістіктерінде аймақтар әртүрлі. Бұл құбылыс бізге Лоренц кеңістігінде қолдануға арналған нүктелерді өзгерту идеясын береді.

2. Қолданылған мәліметтер қоры

Бұл зерттеуде Лоренц кеңістігін жіктеуде әзірленген классификациялық алгоритмдердің дәлдігі мен сенімділігін тексеру үшін CLIMATE, GESTURE, PARKINSON, RELAX, VERTEBRAL және WINE деректер жиынтығы ұсынған мәліметтерімен зерттеу жұмысы жүргізілді. Бұл деректер жиындарындағы атрибуттардың саны 6 дан 500-ге дейін болады және екі сыныптың деректер жиыны әдетте қолайлы болады. Эксперименттерде деректер жинақтарындағы үлгілердің 30%-ы оқу үшін, қалғаны – тестілеу үшін жұмсалды. Қосымшада келтірілген мәліметтер жиынтығы туралы кейбір статистикалық деректер 1 кестеде көрсетілген.

1 кесте – Деректер жинағының статистикасы

	Атрибуттардың саны	Сыныптар саны	Үлгілердің саны	Оқыту	Тест
CLIMATE	8	2	540	120	420
GESTURE	18	2	448	150	298
PARKINSON	29	2	1040	200	60
RELAX	12	2	182	60	122
VERTEBRAL	6	3	310	110	200
WINE	13	2	130	44	86

3. Ұсынылған әдістер

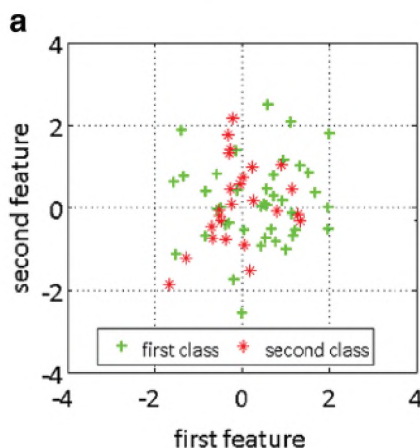
Мақалада Лоренц кеңістігінде жаңа жіктеу әдісі және оның екі нұсқасы ұсынылған. Аталған әдістің негізгі мәні Лоренц аралығын шектеу критерийі ретінде пайдаланылады. Бұл әдіс Лоренц қашықтыққа сәйкес ең жақын үлгілерді немесе көршілерді анықтау арқылы жіктеуді жүзеге асырады.

3.1. Алдын ала өңдеу

Жіктеудің тиімділігін арттыру үшін алдын ала өңдеу қадамы барлық үлгілерге қолданылуы керек. Алдын ала өңдеу сатысы математика мен алгебрадағы матрицаларды аудару және көбейту сияқты жақсы белгілі негізгі операциялардан тұрады (қысу). Осылайша, үлгілер Евклидтік бір шеңбердің пішініне сәйкес келетін гиперболалық крест пішінінің бір кідірісінде бөлу ретінде жарамды болады. Алдын ала өңдеу сатысында матрицалық көбейту Лоренц қашықтығы үшін үлгілерді мағыналы түрде түрлендіру әдісінде қолданылады. Матрицаны көбейту немесе қысуды елеулі пайдалану үшін ауысымдық жұмыс қосымша құрал ретінде таңдалады [10].

Үлгілердің (деректердің) ауысуы қарапайым, олардың сомасын вектордың нүктесін δ -мен ауыстыру жеткілікті және бұл төменде көрсетілгендей оператордың көмегімен жүзеге асуы мүмкін:

$$T_{\delta}f(\vartheta) = f(\vartheta + \delta), \quad (2)$$



Алдын ала өңдеу сатысында тек матрицалық көбейту (қысу) саналы түрде таңдалады, бірақ бұл матрицалық теорияда диагональды матрицаның қасиеттеріне байланысты. Диагональды матрица егер $i \neq j \forall i, j \in \{1, 2, \dots, n\}$ болса, онда $\lambda = 0$, бұл екі өлшемді деректер үшін келесідей:

$$\lambda = \begin{pmatrix} w & 0 \\ 0 & q \end{pmatrix} \text{ немесе } \lambda = \begin{pmatrix} 0 & w \\ q & 0 \end{pmatrix} \quad (3)$$

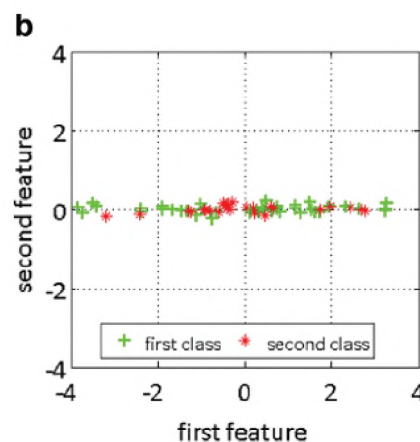
мұнда $w, q \in \mathbb{R}$. Бұл жұмыста біз матрицаның бірінші формасын пайдаланамыз, ал матрицаның λ параметрінің w, q параметрінің арақатынасы $w = 20 * q$ ретінде анықталады. Бастапқы іс ретінде қабылданады

$$\lambda = \begin{pmatrix} 2 & 0 \\ 0 & 0.1 \end{pmatrix} \quad (4)$$

Осы құндылықтар, ең алдымен, біздің байқаулар барысында болжамдар ретінде таңдалады. Бірақ осыдан кейін бөліктің соңында осы параметрлерді өзгерту әсерін зерттейміз. Үлгілер матрицаның геометриялық центроидтер класына сәйкес көбейтіледі:

$$X' = X\lambda \quad (5)$$

3-суретте екі сыныпта орындалатын қысу әсерін көре аламыз. Қысуды пайдалана отырып алдын ала өңдеу кезеңі проекцияның бір түрі болып табылады [11].



3-сурет. Деректер жиынтығынан екі сипаттамалық нүктені бөлу:
а) түпнұсқа және б) λ матрицасында қысылғаннан кейін

4. Зерттеулер мен нәтижелер

Осы мақалада ұсынылған әдістердің классификациялық проблемалардағы нәтижелерін тексеру үшін кейбір қоғамдық деректер жинақтары таңдалған. CLIMATE, GESTURE, PARKINSON, RELAX, VERTEBRAL және WINE эксперименттерінде пайдаланылады. Таңдалған деректер жиынтығындағы мүмкіндіктер саны 6-29 аралықта өзгереді. Бұл кестелер туралы 1-кестеде статистикалық ақпарат бар.

Эксперименттерде екі өлшемді Лоренц кеңістігі зерттелді, сондықтан барлық ерекшеліктер жұптарға бөлінеді. Мысалы, RELAX деректер жиынтығынан функционалды топ үшін, 73.50% ең жақсы жіктеу жылдамдығы басқа әдістермен салыстырғанда Лоренц метрикасы бойынша жіктеу (ЛМЖ) әдісімен алынады. Сол функционалдық топта kNN 62,39% қатынасын құрады. Евклид қашықтығын Лоренцке ауыстырғаннан кейін, kNN классификациясы нәтижесіне шамамен 10% табыс деңгейі қосылды. Әдебиеттерден белгілі болғандай, SVM екі кластағы деректер жиынтығымен тиімді жұмыс істейді, алайда біздің жағдайымызда жіктеу коэффициенті SVM арқылы 47,86% құрайды. Сондай-ақ, Bayes классификаторы жіктеу коэффициентін 72,65% құрайды, ал ЛМЖ арқылы алынған нәтижелерге өте жақын. Лоренц метрикасы бойынша жіктеуде айналу әдісін қолдану (ЛМЖА) бұл жағдайда Bayes сияқты нәтиже береді. Мұндай кезде ең жақсы жіктеу коэффициенті оңтайлы қысу матрицасы бойынша жұмыс істейтін Лоренц метрикасы бойынша жіктеуде қысу матрицасының оңтайлы параметрлері (ЛМЖП) алгоритмі бойынша алынған, ол 82,05% құрайды.

ЛМЖА ЛМЖ-ге қарағанда ЛМЖА жылдамдығы аз болса, CLIMATE және GESTURE деректер жиынтығынан көруге болады, бірақ ЛМЖА нәтижелері басқа классификаторларға қарағанда оңтайлы шешім. Бұл жағдайларда ЛМЖП әдісімен үздік нәтижелер 93,99% және 91,38% сәйкес келеді. Сонымен қатар, ЛМЖП GESTURE ішінде 20% шамасында айтарлықтай айырмашылық бар. Бұл жағдайда ЛМЖ шығысы үшін ең

жақын қаражат kNN-мен 69,83% -ды құрайды, сонымен қатар CLIMATE деректер жинағында ең жақын жылдамдық 90,14% -ды құрайды. Эксперименттерде, WINE деректер жиынтығынан алынған функционалды топ үшін, барлық үш ЛМЖ, ЛМЖА және ЛМЖП әдістерінің ең жақсы жіктеу жылдамдығы 96,20% деп жазады. Осындай жағдайға ең жақын нәтиже kNN бойынша 93,67%, Bayes және SVM нәтижесі сәйкесінше 84,81% және 88,61% құрады.

PARKINSON үшін ЛМЖ және ЛМЖА әдістеріне тиісінше 73,30% және 73,33% жіктеу жылдамдығы. Жоғары SVM шығарылымына қарамастан, ЛМЖ және ЛМЖА 2% мөлшерінде оң нәтиже береді. Дегенмен, ЛМЖП үздік жіктеу нәтижесін 80,00%-ға дейін, ал, бұл өндірісті айтарлықтай арттырады. VERTEBRAL деректер жиынтығы үшін нәтиже ЛМЖА бойынша 76,50% деңгейінде алынды, онда ЛМЖ және ЛМЖП тиісінше 70,00% және 72,00% өндіреді. Сол жағдайда ЛМЖ-ге ең жақын ставка kNN-мен 60,00% -ды алады, ал ең нашар шығыс SVM-да 38,50% -ды құрайды. Бұл ЛМЖП ЛМЖА-дан гөрі жіктеу жылдамдығын азайтқан жағдайда ғана, бірақ басқа жағдайларда артықшылық сақталады.

Мұнда ең нашар жағдай үшін оңтайлы параметрлер алынады, содан кейін оны барлық жұп жұптарда қолданамыз. Осылайша есептеу қиындықтары едәуір азайды. Сонымен қатар, қысу матрицасының оңтайлы параметрлері әрбір функция үшін жеке есептелуі мүмкін. Біз жіктеу ставкалары бұдан да арта түсетінін күтудеміз, бірақ есептік күрделілігі де жағымсыз түрде өседі. Классификация нәтижелеріне негізделе отырып, ЛМЖП әдісі көп жағдайларда kNN және ЛМЖ, ЛМЖА-мен салыстырғанда дұрыс нәтиже береді.

Осы кестелердегі әртүрлі k мәндеріндегі ауысулар сәйкес kNN нәтижелерімен сәйкес келмейді. CLIMATE деректер жинағының тәуірлеу CL мәнінің төмен мәнінде, ЛМЖА және ЛМЖП мәндері 77%-дан 87%-ға дейін өзгереді, бірақ k = 5 бастап, жіктеу нәтижелері 92% және 94% арасында өзгереді. Сол жағдайларда, kNN аз өзгерістері бар өзгеріс-

терді көрсетеді. ЛМЖ, ЛМЖА және ЛМЖП әдістерінің жіктелуі нашар k мәндері үшін нашар, содан кейін жоғары k мәндеріне тез көтеріледі. Осы жағдайға қарамастан, ЛМЖ, ЛМЖА және ЛМЖП әдістері kNN және басқа классикалық әдістерге қарағанда тиімді нәтиже береді.

5. Қорытынды

Мақалада Лоренц қашықтықты метрикасына негізделген жаңа жіктеу әдісі ұсынылды. Кейбір қоғамдық деректер жиынтығын қолдана отырып, ЛМЖ әдісінің жарамдылығы тексерілді. Эксперименттік нәтижелерге сүйенсек, ЛМЖ әдісі деректер жиынтығының бөлінуіне сәйкес әртүрлі сомалар бойынша жіктеу жылдамдығын арттырды. Басқа әдістермен жақсы жіктеу коэффициенттерін алу жағдайында, ЛМЖ әдісі жақсартылған шағын нәтижелермен үздік баға береді. Дегенмен, басқа әдістермен төменгі классификацияның көрсеткіштері алынған жағдайда, ЛМЖ әдісі жоғары жетілдірулермен тиімдірек шешім шығарады. Дифференциалды іргелендіру 20%-дан жоғары болуы мүмкін.

Қалған зерттеулерде ротациялық операцияны қосу және қысу матрицасының оңтайлы параметрлерін табу арқылы алдын ала өң-

деу қадамы күшейтілді және екі кеңейтілген ЛМЖА және ЛМЖП алгоритмдері ұсынылды. Жаңа ЛМЖ, ЛМЖА және ЛМЖП әдістерінің нәтижелері CLIMATE, GESTURE, PARKINSON, RELAX, VERTEBRAL, WINE деректер жинақтары арқылы жасалды. Алдын ала өңдеу қадамындағы үлгілердің айналу бұрышы тандалған және 45° деп бекітілген. Сығымдау матрицасының оңтайлы параметрлерін анықтағанда, нашар жіктеу жылдамдығын беретін ерекшелікті жұп сілтеме ретінде тандалды.

Жалпы ЛМЖ әдісі және ЛМЖА және ЛМЖП кеңейтілген нұсқалары классикалық классификаторларға қарағанда оңтайландырылған шешімдерді береді. Осылайша Лоренц қашықтығы метрикасын жіктеу мәселелерінде қолдану дәлелденді. Лоренц қашықтығы қазіргі кездегі екі қиындыққа берілген жауап. Әрине, біз көп өлшемді деректер жиынтығын көптеген екі өлшемді ішкі жиындарға азайта отырып, содан кейін осы ішкі жиындарды жіктеуге қолдана аламыз. Болашақта жұмыста, сыныптардың үлкен саны мен жоғары өлшемді сипаттамалары, Лоренц қашықтықты метрикасын қарапайым жіктеу мәселелері бойынша айрықша критерий ретінде қарастырып, егжей-тегжейлі зерттелетін болады.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР

1. Y. Deng , Y. Li , Y. Qian , X. Ji , Q. Dai , Visual words assignment via information-theoretic manifold embedding, IEEE Trans. Cybern. 44 (10) (2014) 1924–1937.
2. Tan X., et al. “Face recognition from a single image per person: A survey”, Pattern recognition 39(9):1725-1745, 2006.
3. Theodoridis S., Koutroumbas K., Pattern Recognition, 4th ed., Elsevier, 2009.
4. Kerimbekov Y., et al., “The use of Lorentz distance metric in classification problems.” Pattern Recognition Letters, 84: 170-176, 2016.
5. R. Liu , Z. Su , Z. Lin , X. Hou , Lorentzian discriminant projection and its applications, in: ACCV’09, 3, 2009, pp. 311–320.
6. H.S. Bilge , Y. Kerimbekov , Classification with Lorentzian distance metric, in: 23th Signal Processing and Communications Applications Conference (SIU2015), 2015, pp. 2106–2109.
7. H.S. Bilge , Y. Kerimbekov , H.H. Ugurlu , A new classification method by using Lorentzian distance metric, in: Innovations in Intelligent SysTems and Applications 2015 International Symposium on, 2015, pp. 1–6.
8. Y. Deng , Q. Dai , R. Liu , Z. Zhang , S. Hu , Low-rank structure learning via non-convex heuristic recovery, IEEE Trans. Neural Netw. Learn. Syst. 24 (3) (2013) 383–396.

9. Gündoğan H., Keçioğlu O., “Lorentz matrix multiplication and the motions on Lorentz plane”, Glasnik Matematicki, 41: 329-334, 2006.
10. Brualdi R., Introductory Combinatorics, 5th ed., Pearson Prentice Hall, 2010.
11. Abate A.F., et al. “2D and 3D face recognition: A survey”, Pattern recognition letters, 28(14):1885-1906, 2007.