

УДК 621.396.41

О.Қ. Хўжаев, Н.А. Эгамбердиев

ТИББИЁТ СОҲАСИ АХБОРОТ ТИЗИМЛАРИДА ЎХШАШ ТАШХИСЛАРНИ ТОПИШ УЧУН БАҲОЛАРНИ ХИСОБЛАШ АЛГОРИТМИНИ ҚЎЛЛАШ УСУЛИ

Мазкур моқолада тиббиёт соҳаси ахборот тизимлари маълумотлар базаси асосида маълумотларнинг интеллектуал таҳлили масалаларидан бири бўлган ранжирлаш масаласини ечиш қараб чиқилган. Ранижирлаш масаласини умумий ҳолда математик қўйилиши келтириб ўтилган ва ечиш учун мавжуд бўлган RankBoost, RankSVM, IR-SVM алгоритмлари таҳлил қилинган ҳамда қайси соҳаларда қўлланилиши айтиб ўтилган. Шу билан бир қаторда эвристик усуллардан ҳисобланган нейрон тўрининг тўғри тақсимланган нейрон тўри моделини қўллаб ранжирлаш масаласини ечишнинг устун ва камчилик жиҳатлари келтириб ўтилган. Ундан ташқари классификация масаласини ечиш учун ишлаб чиқилган олти босқичли баҳолашни ҳисоблаш алгоритмининг дастлабки учта босқичидан фойдаланган ҳолда ранжирлаш масаласини ечиш орқали ўхшаш ташхисларни топишда қўлланилган ва натижалар олинган. Олинган натижалар тиббиёт соҳаси ахборот тизими маълумотлар базаси маълумотлари асосида текширилган ва алгоритм ишончилиги солиштирилган. Баҳолашни ҳисоблаш алгоритмини ранжирлаш масаласини ечиш учун мослаштрилган вариантнинг блок схемаси келтириб ўтилган.

Калит сўзлар: Ранжирлаш, Классификация, RankBoost, RankSVM, IR-SVM, ListNet, ListMLE, кўп қатламли тўғри тақсимланган нейрон тўри, баҳолашни ҳисоблаш алгоритми.

Кириш

Тиббиёт соҳаси ахборот тизимлари маълумотлар базасида асосан шифохонага келган беморлар, уларнинг турли лаборатория текширувлари натижалари, уларга қўйилган шифокорлар ташхислари ва даволаш жараёнидаги муолажалар тўғрисидаги маълумотлар релацион маълумотлар базасида сақланади. Аксарият ҳолларда шифокорларга бошқа шифокорларнинг беморларга қўйган ташхис натижалари зарур бўлиб қолади. Маълумотлар базасида минглаб беморлар маълумотлари бўлган ҳолда унинг ичидан энг мақбулини танлаб олиш, уларни ичидан мақбул ташхисларни саралаб олиш кўп вақт талаб қиладиган мураккаб жараён ҳисобланади. Бундай вазиятда маълумотлар интеллектуал таҳлилининг ранжирлаш масаласини ечиш орқали мақбул ечимларни танлаб олиш мумкин.

Масаланинг қўйилиши

Ранжирлаш масаласининг математик қўйилиши куйидаги:

Бизга $I = \{i_1, i_2, \dots, i_j, \dots, i_n\}$ тиббиёт ахборот тизими маълумотлар базасидан олинган беморларнинг лаборатория текширувлари асосида шакллантирилган объектлар тўплами берилган бўлсин. Хар бир қаралаётган i_j объект битта беморнинг турли лаборатория текширувлари натижалари ва улар куйидаги белгилар билан характерлансин. $i_j = \{x_1, x_2, \dots, x_h, \dots, x_m\}$. Хар қандай $(i, j) \in \{1, \dots, n\}$ объектлар индекслари жуфтлиги учун $k < l$ тартиблаш берилган бўлсин. Бизга шундай $a: X \rightarrow R$ функция қуриш талаб қилинади, хар қандай $k < l$ учун $a(i_k) < a(i_l)$ шарт бажарилсин. a функцияни биз ранжирлаш функцияси деб атаймиз[13].

Бизнинг масаламизда эса янги келган i_{n+1} объект учун энг яқин m та объектни топиш масаласи қўйилади. Бунда биз янги келган i_{n+1} объектга нисбатан қолган n та объектлар билан яқинлик мезонини ўрнатамиз ва ражнрлаймиз. Натижада ҳосил бўлган рўйхатдаги дастлабки L та объект биз қидираётган объектлар бўлади.

Мавжуд усуллар таҳлили.

Ранжирлаш масаласини асосан хужжатларни маълум бир белгилар асосида тартиблаш, интернетдаги ҳавоаларни рейтинглаш каби масалаларда жуда кўп қўлланилади ва Google, Yandex, Yahoo, Bing катта қидириш тизимлари ўзларининг қидирув машиналарида шу масала ечимларини қўллашади[9, 10, 11].

Ранжирлаш масаласини ечиш учун шу вақтгача бир қанча усул ва алгоритмлар таклиф қилинган. Булар оқимли ёндашув асосидаги ординал регрессия ва классификация алгоритмлари, жуфтлик ёндашув асосидаги RankNet, FRank, RankBoost, RankSVM, IR-SVM алгоритмлари ва рўйхатли ёндашув асосидаги SoftRank, SVMmap, AdaRank, RankGP, ListNet, ListMLE алгоритмлари[12].

Ундан ташқари ранжирлаш масаласини ечишда нейрон тўрининг кўп қатламли тўғри тақсимланган нейрон тўри моделини қўллаш самарали натижа беради. Нейрон тўрининг мазкур моделини электрон хужжатларни рейтинглашда қўллаш [7,8] мақолаларда кўриб чиқилган. Биз ҳозирда ечишни мақсад қилган масаламизда нейрон тўрининг кўп қатламли тўғри тақсимланган нейрон тўри моделини қўллаш яхши самара беради, лекин камчилик жиҳати шундаки, нейрон тўрини хатоликларни қайта тақсимлаш алгоритми асосида ўқитишимизни инобатга олсак, бу модел билан ишлаганда ўқитиш жараёнида жуда катта хажмдаги ҳисоблашлар

бажарилади. Шу сабабли ҳам бу усулни жуда самарали деб бўлмайди [7].

Ранжирлаш масаласида асосий жиҳат бу ранжирлаш белгиларини танлаб олишдир. Бизнинг масаламизда белгилар кардиология шифохоналари ахборот тизими маълумотлар базасидаги турли лаборатория текширувлари натижалари бўлиб, улар сонли кўринишда ифодаланган.

Таклиф этилаётган усул.

Маълумотларнинг интеллектуал таҳлилида классификация масалаларини ечишда баҳоларни ҳисоблаш алгоритми ишлатилади. Баҳоларни ҳисоблаш алгоритми олти босқичдан иборат бўлиб, ҳар бир босқич маълум бир вазифа бажарувчи амаллар кетма - кетлигидан иборат [1,2]. Бу босқичлар куйидагилар:

1. *Таянч тўпламлар тизими*
2. *Яқинлик функцияси*
3. *Фиксирланган таянч тўпламнинг қаторлари*
4. *бўйича баҳоларни ҳисоблаш*
5. *Фиксирланган таянч тўплам бўйича бир синфнинг баҳоларини ҳисоблаш*
6. *Таянч тўпламлар тизимлари бўйича K_u синфнинг баҳоси*
7. *A алгоритм учун ҳал қилувчи қоида*

Классификация масаласини ечиш учун мазкур босқичларнинг ҳаммаси бажарилиши зарур. Лекин бизнинг масаламизда дастлабки учта яъни таянч тўпламлар тизими, яқинлик функцияси ва фиксирланган таянч тўпламнинг қаторлари бўйича баҳоларни ҳисоблаш босқичларини бажариш етарли ҳисобланади. Бу босқичларда бажариладиган ҳисоблаш жараёнларини таҳлил қилиб чиқамиз:

1. *Таянч тўпламлар тизими.* Бу босқичда биз беморларни турли лаборатория текширувлардан ўтгандаги натижаларни маълумотлар базасидан ўқиб оламиз ва матрица кўринишидаги таянч тўпламни ҳосил қиламиз. n та объектлар сони ва m та ҳар бир объект белгилари учун матрица куйидагича бўлади:

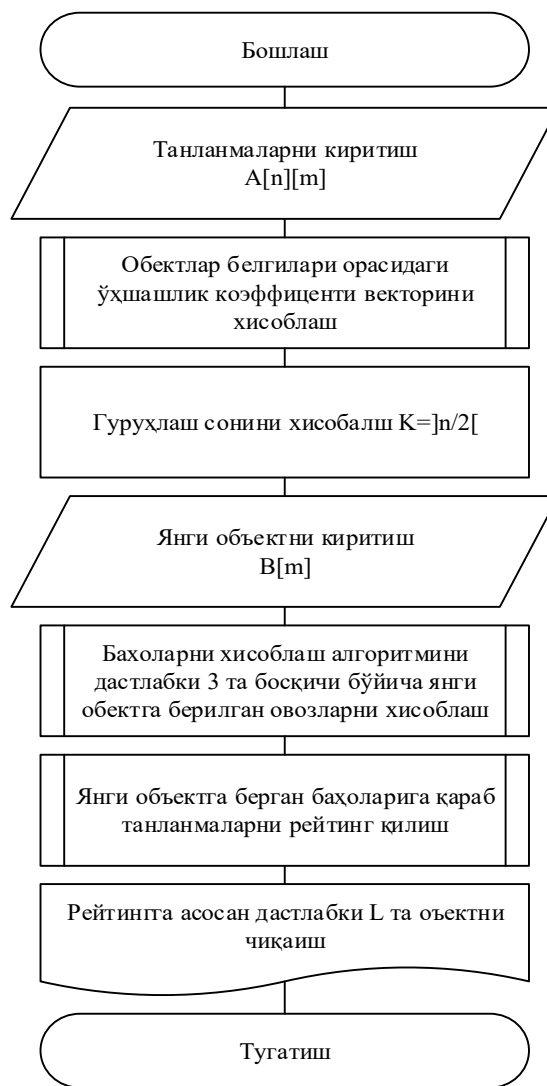
$$X_1 = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \quad (1)$$

2. *Яқинлик функцияси.* Бу босқичда таянч тўплам объектларини белгилари асосида ҳар бир белги учун яқинлилик параметри куйидагича ҳисобланади:

$$\varepsilon_j = \frac{1}{n-1} \sum_{i=1}^{n-1} |x_{ij} - x_{i,j+1}| \quad (2)$$

3. *Фиксирланган таянч тўпламнинг қаторлари бўйича баҳоларни ҳисоблаш.* Бу босқич

алгоритмизининг асосий босқичи бўлиб бунда янги келган объектга нисбатан ҳар бир объектнинг овозлари ҳисоблаб чиқилади. Баҳоларни ҳисоблаш алгоритмида объектларнинг овозлари нафақат объектни ҳар бир белгисини синов объектига берган баҳолари асосида балки белгиларни гуруҳлаш асосида ҳам олинди. Унда гуруҳлаш коэффиценти $k = \{1, 2, \dots, n\}$ бўлиши мумкин.



Лекин k нинг ҳамма қиймати учун овозларни ҳисоблаш ва уни ичидан энг мақбулини танлаш жуда кўп ҳисоблашларни талаб қилади. Бундай вазиятда алгоритмни ишлаш мураккаблиги нейрон тўрларидагига қараганда ҳам кўпроқ ҳисоблашларни келтириб чиқаради. Шу сабабли k нинг оптимал қийматини танлаш олиш масалани ечиш жараёнида муҳим аҳамият касб этади. Таҷриба натижалари шуни кўрсатадики, гуруҳлаш сонини белгиловчи k ни $k = \lfloor \frac{n}{2} \rfloor$ кўринишида олиш мақсадга мувофиқ.

k танлаганимиздан кейин, таянч тўпламдаги ҳар бир объектлар бўйича синов объектига бўлган баҳоларни ҳисоблаймиз ва $s[n]$ векторни ҳосил қиламиз. Кейин $s[n]$ векторни камайиш бўйича

тартиблаймиз ва ҳосил бўлган векторга асосан дастлабки L та объектими масаланинг ечими ҳисобланади.

Масалани ечиш алгоритмини қуйидаги блок-схема орқали ифодалаймиз:

Тажриба синовлари натижалари таҳлили. Тажриба маълумотлари сифатида Республика ихтисослашган кардиология маркази Хоразм вилояти бўлими “iCardio” ахборот тизими маълумотлар базасидаги 90 нафар беморнинг 6 та лаборатория текширувларидан 21 та белги бўйича берилган маълумотларидан фойдаланилди. Тажриба натижалари қуйидагича:

Дастлаб яқинлилик векторини ҳисоблаб оламиз, у қуйидагича бўлади:

1-жадвал.

Яқинлилик векторини ҳисоблаш натижалари

e[1]	7.2045	e[8]	4.7193	e[15]	41.3795
e[2]	0.4380	e[9]	16.7750	e[16]	12.6136

e[3]	133.2772	e[10]	55.0840	e[17]	0.4102
e[4]	0.0579	e[11]	21.35	e[18]	0.1795
e[5]	0.0250	e[12]	0.8352	e[19]	0.1806
e[6]	0.4318	e[13]	0.9727	e[20]	1.8352
e[7]	2,75	e[14]	55.4545	e[21]	1.0488

Ундан кейн синов объекти киритилади, ва унга объектлар томонидан берилган баҳолар $k=10$ ҳолат учун ҳисобланади.

Натижалардан қўришимиз мумкин жами овозларни ҳисоблаш учун жами комбинациялар 352716 та, натижада овоз олишлар бўйича 1-ўринни 48-объект тўплаган бўлиб, у тўплаган овозлар сони 187756 та, шу сабабли мазкур объект синов объектими энг яқин объект ҳисобланади. Шундай тарзда бизга зарур бўлган L та синов объектими энг яқин объектларни топишимиз мумкин.

Atabaev_Odilbek, 134.0, 4.25, 4.4, 1.4, 4.4, 35.8, 5.0, 1.9, 62.0, 286.0, 28.8, 5.8, 4.9, 164.0, 113.0, 31.0, 0.6, 0.8, 3.6, 4.8, 2.8

$k=10$

352716 ta kombinatsiya

0-testlanuvchi ob'yektning o'quv tanlanma ob'yektlari uchun ovozlari

1 eng yaqin ob'yekt: 48 184756.0 ovoz bilan

Madrimov_Baxtiyor_141.0 4.47 5.7 1.4 4.4 35.8 7.0 1.9 62.0 321.0
28.6 6.6 5.6 226.0 154.0 31.0 1.0 0.9 3.5 4.9 3.0

stenokardiya

2 eng yaqin ob'yekt: 60 92378.0 ovoz bilan

Yakubova_Xoljon2_ 109.0 3.91 4.1 1.4 4.4 35.8 6.0 1.7 68.0
302.0 23.0 5.7 4.7 161.0 105.0 35.0 1.0 1.2 3.6 4.2 2.1

stenokardiya

3 eng yaqin ob'yekt: 9 43758.0 ovoz bilan

Allamov__amat_ 104.0 3.46 6.3 1.4 4.4 35.8 5.0 2.3 59.0
288.0 34.4 6.4 5.2 212.0 130.0 38.0 0.5 0.5 3.6 4.3 2.4

ikks

4 eng yaqin ob'yekt: 46 43758.0 ovoz bilan

Ruzmetov_Kodir_ 140.0 5.1 8.4 1.4 4.4 35.8 7.0 2.6 52.0
280.0 36.7 6.1 4.7 187.0 105.0 42.0 1.2 1.0 3.5 4.7 3.5

stenokardiya

Bajarilish vaqti: 2.629 sekund

1- расм. Дастур натижалари

Хулоса

Хулоса ўрнида шуни айтиш мумкинки, ҳозирда нафақат кардиология соҳаси, балки деярли барча тиббиёт муассасаларида иш жараёни самарадорлигини ошириш ахборот тизимлари ишлаб чиқилиб улардан фойдаланилмоқда. Кўпгина тиббиёт ахборот тизимларига эса интеллектуал таҳлил модуллари қўшилиб тизим имкониятлари оширилиб борилмоқда. Ранжирлаш масаласи бу жараёнда муҳим ўрин тутувчи масалалардан

ҳисобланади. Бундай жараёнда баҳоларни ҳисоблаш алгоритмининг биз юқорида келтириб ўтган модификацияланган вариантини қўллаш тизим ишлаш ишончилигини оширишда яхши самара беради.

Адабиётлар:

1. Журавлев Ю.И. Избранные научные труды. – М.: Магистр, 1998. –420 с.

2. Журавлев Ю.И., Камилов М.М., Туляганов Ш.Е. Алгоритмы вычисления оценок и их применение. Ташкент: Фан, 1974. -119с.

3. Kamilov M.M., Hudaiberdiev M.X., Khamroev A.Sh. Methods of Computing Epsilon Thresholds in the Estimates' Calculation's Algorithms. International Conference "Problems of Cybernetics and Informatics" (PCI'2012), Volume III. September 12-14, 2012. – Baku, Azerbaijan. – Pp. 133-135.

4.Хамроев А.Ш. Алгоритм выбора оптимального метода вычисления значений е-порогов в алгоритмах вычисления оценок. Международный научно-технический журнал. Химическая технология. Контроль и управление. – Ташкент, 2012. – № 3. – С. 78-82.

5.Худайбердиев М.Х., Хамроев А.Х., Мамиева Д.З. Объектлар ҳақидаги ўқув ва назорат танланмаларини шакллантиришда баҳолашни ҳисоблаш алгоритми. Республика илмий-техник анжумани: “Ахборот ва телекоммуникация технологиялари муаммолари”. - Тошкент, 2016 йил 10-11 март. – 185-187 бб.

6. Худайбердиев М.Х., Хамроев А.Ш. О взаимосвязи параметров в моделей алгоритмов вычисления оценок. Интеллектуальные системы (INTELS'-2014): Десятый международный симпозиум. 30 июня - 4 июля – Москва, 2014. – С. 49-52.

7. Хўжаев О.Қ., Абдуллаев Ф.О., Раҳманова М.Р. “Методы определение веса документов в электронных ресурсах” // НАУКА И МИР Международный научный журнал, № 3 (19), 2015, Том 1 ISSN 2308-4804.

8. Хўжаев О.Қ., Абдуллаев Ф.О., Артиков М.Э., Бобожанов Б. “Применение нейронных сетей поисковых машинах типа электронного правительства”// НАУКА И МИР Международный научный журнал, № 3 (19), 2015, Том 1 ISSN 2308-4804.

9.<http://developer.yahoo.net/blogs/hadoop/2008/02/yahoo-worlds-largest-production-hadoop.html>

10. <https://blogs.bing.com/search/2009/06/01/user-needs-features-and-the-science-behind-bing>

11.https://academy.yandex.ru/events/data_analysis/g rant2009/

12. Tie-Yan Liu (2009), Learning to Rank for Information Retrieval, Foundations and Trends in Information Retrieval: Vol. 3: No 3, cc. 225-331, ISBN 978-1-60198-244-5, DOI 10.1561/15000000016.

13. <http://www.MachineLearning.ru/wiki>

Хўжаев Отабек Қадамбоевич

Тошкент ахборот технологиялари университети Урганч филиали “Ахборот таълим технологиялари” кафедраси мудири

Эл. почта: otabek.hujaev@gmail.com

Эгамбердиев Нодир Абдуназарович

Тошкент ахборот технологиялари университети “Ахборот технологияларининг дастурий таъминоти” кафедраси катта ўқитувчиси

Эл. почта: nodir0188@mail.ru

O.Q. Xo'jayev, N.A. Egamberdiyev

Method of applying algorithm of calculating ranks for searching similar diagnosis on medical information systems

This paper focuses on the problem of ranking, which is one of the issues of intelligent data analysis based on the medical information system database. Given commonly mathematical description of ranking task and analyzed RankBoost, RankSVM, IR-SVM algorithms for solving ranking task and also said its applications. In addition, the neuronal network, which is heuristic method, has been designed to solve the problem of ranking to the feed forward neural network model. And so on six stages calculating ranks algorithm for classification task is applied and is taken results for searching similar diagnosis in database of medical information system and redesigned in three stages. Taken results checked and compared algorithm accuracy on based database of medical information system. Given block-scheme of adopted variant of calculating ranks algorithm for solving ranking task.

Keywords: *Ranking, RankBoost, RankSVM, IR-SVM, ListNet, ListMLE, feed forward neural network, algorithm of calculating ranks.*

УДК 621.396.41

О.Ж. Бабомурадов, Н.С.Маматов, Л.Б. Бобоев, Б.И.Отахонова

МАТНЛАРГА ДАСТЛАБКИ ИШЛОВ БЕРИШ ВА ЭҲТИМОЛЛИ БАҲОЛАШ МОДЕЛЛАРИ АСОСИДА ТАСНИФЛАШ

Мақолада турли табиий тилларда берилган матнли маълумотларни таснифлаш масаласи доирасида олиб борилган назарий ва амалий тадқиқот натижалари таҳлили келтирилган. Амалга оширилган манбалар таҳлил асосида тадқиқот учун усткурма ёндашувлар белгилаб олинган. Матнли электрон ҳужжатларни эҳтимолли баҳолаш моделлари асосида таснифлаш учун матнга дастлабки ишлов бериш орқали информатив белгилар фозосини ҳосил қилиш ёндашуви келтирилган. Матнли ҳужжатлардан ташкил топган модел масалани ечиш орқали Бернулли модели самарадорлиги баҳоланган ва тажрибавий тадқиқот натижалари ёрдамида асосланган.