

TASAVVURLI, UMUMIY-TASAVVURLI VA RAQAMLI- TASAVVURLI QOIDALARNI TAHLILI

Sherzod Rajabov

Toshkent davlat iqtisodiyot universiteti “Raqamli iqtisodiyot va axborot texnologiyalari” kafedrası katta o`qituvchisi

mr.sherzod.uz@mail.ru

Annotatsiya

Ushbu maqolada Data Mining ning turli algoritm va metodlari qo`llanilishi va ular orasida neyron tarmoqlari, shajaraviy yechimlar, klasterizatsiya algoritmlari, shu jumladan masshtablanadigan, hodisalar orasidagi assotsiativ aloqalarni va xakozolarni aniqlaydigan algoritmlar haqida yozilgan.

Аннотация

В данной статье описывается использование различных алгоритмов и методов интеллектуального анализа данных, включая нейронные сети, древовидные решения, алгоритмы кластеризации, в том числе масштабируемые алгоритмы, определяющие ассоциативные связи между событиями и причинно-следственными связями.

Abstract

This article describes the use of various algorithms and methods of Data Mining, including neural networks, tree solutions, clustering algorithms, including scalable algorithms that identify associative relationships between events and causality.

Kalit soʻzlar

Data mining, SQL-soʻrovlari, algoritm, assotsiativ qoidalar, bozor savatchasi, klasterizatsiya algoritmlar.

Ключевые слова

Интеллектуальный анализ данных, SQL-запросы, алгоритм, ассоциативные правила, рыночная корзина, алгоритмы кластеризации.

Keywords

Data mining, SQL queries, algorithm, associative rules, market basket, clustering algorithms.

Kirish

Har bir bilim ayrim yo`nalishda qo`llanilish orqali foyda keltirishi kerak. Bu bilimlar matematik bo`lmagan foydalanuvchi uchun tushunarli va sodda ko`rishga ega bo`lishi kerak. Masalan inson tomonidan yengil qabul qilinadigan logik-konstruksiyalar orkali «agar... u holda...». Bundan tashqari ushbu qoidalar har xil MBBT uchun SQL-so`rovlari orqali qo`llanilishi mumkin. Mabodo olingan ma'lumotlar tushunarli bo`lmasa, foydalanuvchi uchun ma'lumotlarni tushunarli darajaga olib keladigan kayta ishlash metodlari mavjud bo`lishi kerak. Data Miningda qo`llaniladigan algoritmlar ko`p sonli xisoblashni talab kiladi. Ilgari bu fakt Data Mining uchun qiyin masala sifatida ko`rilar edi, biroq xozirgi paytda zamonaviy protsessorlarning rivojlanishi bu masalani ahamiyatini susaytirdi.¹

Masalalarni yechishda Data Mining ning turli algoritmlar va metodlari qo`llaniladi. Ular orasida keng qo`llaniladiganlari bu: neyron tarmoqlari, shajaraviy yechimlar, klasterizatsiya algoritmlari, shu jumladan masshtablanadigan, hodisalar orasidagi assotsiativ aloqalarni va xakozolarni aniklaydigan algoritmlar.

Adabiyotlar tahlili

Ma'lumotlar ombori kundan kunga kengayishi sababli assotsiativ qoidalarni topish uchun samarali masshtablanadigan algoritmlar talab qilinadi. Bu qoidalar masalalarni qisqa vaqt ichida yechish imkonini beradi.

Bilimlarni aniqlash metodlari orasida keng qo`llaniladigan bu assotsiativ qoidalarni topish algoritmi. AIS deb ataluvchi birinchi assotsiativ qoidalarni topish usuli 1993 yil, IBM Almaden tadqiqot markazi xodimlari tomonidan ishlab chiqilgan. Shundan sung bu yualishga katta e'tiborkaratilgan. 90-yillar o`rtasi shu yo`nalishdagi ochilishlarning yuqori davri hisoblanadi. Bugungi kunda assotsiativ koidalarni aniklash uchun asosan Apriori algoritmi qo`llaniladi. Uning muallifi Rakesh Agravaldir (Rakesh Agrawal).²

Assotsiativ qoidalar va bozor savatlarining tahlili bo'yicha chet ellik olimlarning, jumladan R. Agrawal, T. Imielinski,³ A. Swami⁴, R. Srikant⁵, A. Savasere,⁶ J. Hipp,⁷ U. Guntzer. ishlarida ko`plab natijalar keltirilgan.

Effektiv algoritmlardan biri sifatida Apriora va genetik algoritmlar ko`rsatilgan.

Tadqiqot metodologiyasi

Qo'yilgan masalani yechish uchun ma'lumotlarni tahlili qilishdagi assotsiativ qoidalar va algoritmlardan hamda Cell protsessorlar bazasidagi hisoblash tizimi uchun moslashtirilgan bozor savatlarini tahlil qilishning parallel algoritmidan foydalanilgan.

Tahlil va natijalar

¹ https://www.researchgate.net/publication/353743345_Educational_Data_Mining_A_Review

² <https://sigmodrecord.org/publications/sigmodRecord/0006/pdfs/Privacy-Preserving%20Data%20Mining.pdf>

³ https://en.wikipedia.org/wiki/Tomasz_Imieli%C5%84ski

⁴ <https://scholar.google.com/citations?user=3dfMSRMAAAAJ&hl=en>

⁵ <https://ece.illinois.edu/about/directory/faculty/rsrikant>

⁶ <https://www.researchgate.net/profile/Ashok-Savasere>

⁷ <https://int.soccerway.com/players/jenny-hipp/362729/>

Assotsiativ koidalar xodisalar orasidagi muntazamlilikni aniklashda qo'llaniladi. Bunga misol tariqasida kuyidagi tasdiqni keltirish mumkin: non sotib olmoqchi bulgan ist'emolchi sut ham sotib olishi mumkinligining ehtimoli 75%ni tashkil kiladi.

Bozor savatchalarini tahlil qilish (market basket analysis) -bu supermarketlarda eng tipik, shablonli xaridlarni qidirish (tasavvurli qoidalarni qidirish) dir. Bozor savatchalarini tahlil qilish bir-biriga bog'liq bo'lgan tovarlar kombinatsiyalarini aniqlash maqsadida ma'lumotlar bazasini tahlil qilish yo'li orqali amalga oshiriladi. Boshqacha qilib aytganda, bunda "juftli tovarlar" aniqlanadi.

Ushbu juft tovarlarning biri kalitli, u bilan xarid qilinadigan tovar esa – xamroh bo'ladigan tovar hisoblanadi. Mazkur tahlil juftli tovarlarni xarid qilish tezligini hamda hamroh bo'luvchi tovar kalitli tovarlar bilan xarid qilinishi ehtimolini aniqlashga yordam beradi.

Bozor savatchasini tahlil qilish yordamida olingan natijalar quyidagilarni beradi:

- Tovar va zaxiralar assortimentini optimallashtirish;
- Savdo zallarida tovarlarni joylashtirish;
- Promoaksiyalarni rejalashtirish (juft tovarlarga chegirmalarni taqdim etish);
- Hamroh bo'luvchi tovarlarning siljishi hisobiga sotish hajmini oshirish.

Masalan, tahlil natijasida shu aniqlandiki, makaron va ketchup (pivo va chipslar, shampun va balzam, printer va qog'oz)ni birgalikda xarid qilish shablonli xarid hisoblanadi. Xarid qilish tezligini va hamroh bo'luvchi tovar kalitli tovar bilan birga xarid qilinishi ehtimolini bilgan holda, tovarlarni yonma-yon qo'yish va ularning birini reklama qilish yo'li orqali xaridorlarni ushbu tovarlarni birgalikda xarid qilishga "majbur qilish" mumkin.

Mazkur tahlil asosida qoidalar tuziladi, ular quyidagilar uchun qo'llanilishi mumkin:

- Savdo zalida tegishli bezashlarni shakllantirish yo'li bilan sotish hajmini oshirish;
- Xizmat ko'rsatish sohasida yangilikka ijobiy munosabatda bo'lgan mijozlarning ulushini aniqlash;
- Xaridorlarning profilini aniqlash;
- Mijozlarga tegishli mahsulotlarni taklif etish yo'li bilan ularning sodiqligini oshirish va yo'qotishlar xavfini kamaytirish;
- Fribgarlikni qidirish.

Oxirgi paytlarda ma'lumotlar bazasida bilimlarni aniqlash usullariga qiziqish tobora oshib bormoqda. Juda katta hajmga ega bo'lgan zamonaviy ma'lumotlar bazasi yangi miqyosi bo'yicha kengayib borilayotgan ma'lumotlarni tahlil qilish algoritmlariga barqaror talabni qo'ygan. Bilimlarni aniqlash bo'yicha eng mashhur usullardan biri tasavvurli qoidalarni qidirish algoritmi hisoblanadi.

Tasavvurli qoidalar bir-biriga bog'langan voqealar o'rtasidagi qonuniyatni topishga imkon beradi. Ushbu qoidaning misoli sifatida "Non"ni xarid qilgan odam 72 % ehtimoli bilan "Sut"ni xarid qilishi mumkin degan ibora xizmat qilishi mumkin. AIS deb nomlangan birinchi tasavvurli qoidalarni qidirish algoritmi IBM Almaden tadqiqot markazi xodimlari tomonidan 1993 yilda ishlab chiqilgan. Ushbu dastlabki ishdan boshlab tasavvurli qoidalarga bo'lgan qiziqish oshib borgan; o'tgan asr 90 yillar

o'rtalarida ushbu sohadagi tadqiqot ishlari cho'qqiga yetgan bo'lib, o'shandan keyin har yili bir nechta algoritmlar paydo bo'lgan.

Ma'lumotlar tahlili quyidagi amaliy masalalar uchun qo'llaniladi:

- Bozorni segmentlash – kompaniyangizning bir xil mahsulotlarini xarid qiluvchi mijozlarning umumiy tavsiflari;
- mijozlarning ketishi – qaysi mijozlar kompaniyangizdan boshqa raqobatchi kompaniyaga o'tishini pognoz qilish;
- fribgarlikni aniqlash – sohta bo'lishga ehtimol ko'proq bo'lgan tranzaksiyalarni aniqlash;
- to'g'ridan-to'g'ri marketing – maksimal javob olish uchun pochta jo'natmalariga joylashish lozim bo'lgan materiallarni aniqlash;
- iste'mol savatchasini tahlil qilish – qaysi tovar va xizmatlar ko'pincha birga xarid qilinayotganligini baholash;
- tendensiyaning tahlil qilish – o'tgan va joriy oylarda oddiy xaridorlar o'rtasidagi farqlarni baholash.

Tasavvurli qoidalar (Association Rules). Dastlab ushbu masala supermarketlarda amalga oshirilayotgan xaridlarning oddiy shablonini topish uchun tasavvurli qoidalarni qidirish uchun taklif etilgan, shu bois ba'zida u bozor savatchasini tahlil qilish (market basket analysis) deb ham yuritiladi.

Faraz qilaylik, xarid tranzaksiyalardan tarkib topgan ma'lumotlar bazasi mavjud. Har bir tranzaksiya – xaridor tomonidan bir tashrif buyurganda xarid qilingan tovarlar to'plamidir. Ushbu tranzaksiya boshqacha qilib aytganda bozor savatchasi deb ham ataladi.

1- aniqlash usuli. Faraz qilaylik, $I = \{i_1, i_2, i_3, \dots, i_n\}$ – elementlar deb nomlanuvchi tovarlarning ko'pligi (to'plami)dir. Faraz qilaylik, D – tranzaksiyalarning ko'pligi, bu yerda har bir tranzaksiya $T \subseteq I$, $T \subseteq I$ dan tarkib topgan elementlar to'plamidir. Har bir tranzaksiya binar vektorini o'zida namoyon etadi, bu yerda $t[k] = 1$, agar ik elementi tranzaksiyada bo'lsa, boshqacha qilib aytganda $t[k] = 0$. Ta'kidlanganimizdek, T tranzaksiya I dan tarkib topgan elementlarning qandaydir to'plami X dan iborat, agar $X \subseteq T$. Tasavvurli qoida deb $X \Rightarrow Y$ implikasiyaga aytiladi, bu yerda $X \subseteq I$, $Y \subseteq I$ va $X \cap Y = \emptyset$. $X \Rightarrow Y$ qoidasi s (support) ko'magiga ega, agar D tranzaksiyadan tarkib topgan tranzaksiyalar $X \cup Y$ dan iborat, $\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y)$. Qoidaning haqqoniyligi X dan Y kelib chiqish ehtimollik darajasi qandayligini ko'rsatadi. $X \Rightarrow Y$ qoidasi (confidence) c haqqoniyligi bilan adolatli hisoblanadi, agar X dan tarkib topgan D dan iborat bo'lgan tranzaksiyalar Y dan tarkib topgan bo'lsa, $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$.

Buni aniq misolda ko'rib chiqamiz: non bilan bog'liq bo'lgan tranzaksiyalarning 75% sutni ham o'z ichiga oladi. Barcha tranzaksiyalarning 3% ikkita tovarni o'z ichiga oladi. 75% – bu qoidaning haqqoniyligi (confidence), 3% bu ko'mak (support), yoki “Non” “Sut” 75% ehtimollik bilan xarid qilinadi. Boshqacha qilib aytganda, tahlilning maqsadi quyidagi qonuniyatlarni o'rnatishdan iborat: agar tranzaksiyada qandaydir X elementlar to'plami to'qnashsa, buning asosida boshqa Y elementlar to'plami ushbu tranzaksiyada ham to'qnashmasligi lozim degan xulosa qilishimiz mumkin. Ushbu bog'liqliklarni o'rnatish bizga juda oddiy va intuitiv tarzda tushunarli bo'lgan qoidalarni topishga imkon beradi.

Tasavvurli qoidalarni qidirish algoritmlari barcha X Y qoidalarni topish uchun mo'ljallangan bo'lib, bunda ushbu qoidalarning ko'magi va haqqoniyligi ayrim oldindan aniqlangan chegaralar, ularga mos ravishda minimal ko'mak (minsupport) va minimal haqqoniylik (minconfidence) deb nomlangan ko'rsatkichdan yuqori bo'lishi lozim.

Tasavvurli qoidalarni topish masalasi ikkita kichik masalalarga bo'linadi:

1. Minsupport chegarasi qanoatlantiradigan barcha elementlar to'plamini topish. Ushbu elementlar to'plamlari ko'pincha to'qnashuvchi to'plam deb ataladi.

2. Minconfidence chegarasi qanoatlantiradigan haqqoniyligi bilan 1 bandiga ko'ra topilgan elementlar to'plamidan tarkib topgan qoidalarni to'plash.

Minimal ko'mak va minimal haqqoniylik parametrlari uchun qiymatlar shunday tarzda tanlab olinadiki, topilgan qoidalarning soni cheklangan bo'lishi kerak. Agar ko'mak katta qiymatga ega bo'ladigan bo'lsa, algoritmlar tahlilchilarga yaxshi ma'lum bo'lgan yoki mazkur tahlilni olib borishda hech qanday mantiq bo'lmaydigan darajada ayon bo'lgan qoidalarni topadi. Boshqa tomondan, ko'makning kichik qiymati ko'p sonli qoidalarni to'plashga olib keladi, bu albatta ko'p hisoblash resurslarini talab etadi. Shunga qaramay, ko'plab qiziqarli qoidalar aynan ko'mak chegarasining aynan kichik qiymatida joylashgan. Zero ko'makning kichik qiymatlari statistik jihatdan asoslanmagan qoidalarning to'planishiga olib keladi.

Tasavvurli qoidalarni qidirish bir ko'rishda be'manilik masala hisoblanmaydi. Muammolardan biri – tez-tez uchrayotgan elementlar to'plamini topishda algoritmik murakkablik hisoblanadi, chunki elementlarning soni o'sishi bilan $I(I)$ da salohiyatli elementlar to'plami soni eksponensial tarzda o'sadi.

Umumlashgan tasavvurli qoidalar (Generalized Association Rules)

Tasavvurli qoidalarni qidirishda biz tahlil qilinayotgan elementlarni bir xil deb faraz qilganmiz. Bozor savatchasini tahlil qilishga yana bir bor qaytadigan bo'lsak, bu mutlaq bir hil atributlarga ega bo'lgan tovarlardir, ularning nomlari bundan mustasno. Biroq tranzaksiyani tovarlar qaysi guruhga kirganligi to'g'risida axborot bilan to'ldirish va tovarlar ierarxiasini tuzishga qiyinchilik tug'dirmaydi. Ushbu guruhlashtirish (taksomaniya) misolini ierarxik modeli ko'rinishda keltiramiz.



1-rasm. Sotiladigan mahsulotlarning ketma ketligi

Faraz qilaylik, bizga D tranzaksiya bazasi berilgan va qaysi guruhlar (taksonlar)ga elementlar kirganligi ma'lum. Shunda guruhlarni guruhlar bilan, alohida elementlarni guruhlar bilan va x.k. bog'lovchi qoidalar ma'lumotlaridan hosil chiqarish mumkin.

Masalan, agar Xaridor "Alkogolsiz ichimliklar" guruhidan tovarni xarid qilgan bo'lsa, u "Sut mahsulotlari" yoki "Sharbat" "Sut mahsulotlari" guruhidagi tovarni ham xarid qiladi. Ushbu qoidalar umumiy tasavvurli qoidalar deb ataladi.

Aniqlash usuli 2. Umumiy tasavvurli qoida deb $X \Rightarrow Y$ ga aytiladi. bunda $X \subset I$, $Y \subset I$ va $X \cap Y = \emptyset$ va bu yerda Y to'plamga kirgan elementlardan biri X ga kirmagan bironta element avlodiga kirmaydi. Ko'mak va haqqoniylik tasavvurli qoidalar holati singari hisoblanadi (Aniqlash usuli 1 ga qarang.)

Ierarxiya ko'rinishdagi elementlarni guruhlashtirish to'g'risidagi qo'shimcha axborotni kiritish quyidagi afzalliklarni beradi:

Bu nafaqat alohida elementlar o'rtasida, balki ierarxiya (guruh)ning turli pog'onalari bilan tasavvurli qoidalarni o'rnatishga yordam beradi.

Alohida elementlar yetarli darajada ko'makka ega bo'lmasligi mumkin, ammo butun guruh minsupport chegarasiga qanoatlanishi mumkin.

Ushbu qoidalarni topish uchun yuqorida nomi keltirilgan har qanday algoritmdan foydalanish mumkin. Buning uchun har bir tranzaksiyani tranzaksiyaga kirgan har bir elementning barcha avlodlari bilan to'ldirish lozim. Biroq, ushbu algoritmlarni to'g'ridan-to'g'ri qo'llash quyidagi muammolarga olib keladi.

Xulosa

Yuqorida izohlangan tasavvurli qoidalardan tashqari bilvosita tasavvurli qoidalar, inkor etish bilan tasavvurli qoidalar, vaqt bilan bog'liq bo'lgan voqealar uchun vaqt tasavvurli qoidalar va x.k.

Yuqorida ta'kidlanganidek, tasavvurli qoidalarni qidirish masalasi dastlab bozor savatchasini tahlil qilish uchun tavsiya etilgan. Tasavvurli qoidalar xaridlar, mijozlarning xohishlarini tahlil qilish, supermarketlarda tovarlarni joylashtirishni rejalashtirish, kross-marketing, manzilli jo'natishda xaridorlarning hatti-harakatlari bo'yicha segmentlashtirishda samarali qo'llanilmoqda. Biroq, ushbu algoritmlarning qo'llanilish sohasi faqatgina savdo bilan cheklanmaydi. Ular boshqa sohalarda ham muvaffaqiyatli tarzda qo'llanilmoqda: tibbiyotda, veb-sahifalarga kirishni tahlil qilish (Web Mining), matnni tahlil qilish (Text Mining), aholini ro'yxatga olish bo'yicha ma'lumotlarni tahlil qilish uchun, telekommunikatsion uskunalarning ishdan chiqishini tahlil qilish va prognozlashtirishda va x.k.

Foydalanilgan adabiyotlar ro'yxati

1. Agrawal R., Imieliński T., Swami A. Mining association rules between sets of items in large databases //Proceedings of the 1993 ACM SIGMOD international conference on Management of data. – 2017. – C. 207-216.
2. Agraval, Rakesh va boshqalar. "Grafiklar yordamida o'xshash hujjatlarni aniqlash." AQSh patenti № 10,025,783. 2018 yil 17 iyul.
3. Sun X. va boshqalar. Muti-xatti-harakat xususiyatlariga asoslangan bilimlarni kuzatish qarorlar daraxti takomillashtirilgan DKVMN // ACM Turingni nishonlash konferentsiyasi materiallari-Xitoy. – 2019. – S. 1-6.

4. Rajabov, Sherzod Baxtiyorovich. "The role of backend and frontend information systems infrastructure." *Science and Education* 4.3 (2023): 212-216.
5. 2. Sadinov A., Rajabov S., Samieva M. Improving waste recycling in Uzbekistan in digital technologies //E3S Web of Conferences. – EDP Sciences, 2023. – T. 452. – C. 05023.
6. Safarovich, Kuchkarov Tohir, Kobilov Alisher Urinovich, and Rajabov Sherzod Bakhtiyorovich. "INNOVATION AND THE DIGITAL ECONOMY." *Gospodarka i Innowacje*. 23 (2022): 122-128.
7. Kobilov A. U. et al. Development of the digital economy in the Republic of Uzbekistan //E3S Web of Conferences. – EDP Sciences, 2023. – T. 402. – C. 08038.
8. Abdulaxatov M., Zoirov U. The implementation process of designing a digital organization. The problem to determine //Science and Education. – 2023. – T. 4. – №. 3. – C. 168-173.
9. Kobilov A. et al. ARTIFICIAL INTELLIGENCE AS A TECHNOLOGICAL INNOVATION FOR ECONOMIC DEVELOPMENT OF THE REPUBLIC OF UZBEKISTAN //Proceedings of the 6th International Conference on Future Networks & Distributed Systems. – 2022. – C. 292-297.
10. Tufail S, Riggs H, Tariq M, Sarwat AI. Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms. *Electronics*. 2023; 12(8):1789. <https://doi.org/10.3390/electronics12081789>