

Abduganieva M.
“Raqamli iqtisodiyot va axborot texnologiyalari” kafedrası
katta o`qituvchisi

ASSOTSIIATIV QOIDALAR VA BOZOR SAVATLARI

Annotatsiya: Ushbu maqolada xizmat ko'rsatish va tijorat yo'nalishidagi zamonaviy tashkilotlar plastik kartochkalar va nazorat qiluvchi kompyuter tizimlari orqali qilingan har bir buyurtma to'g'risida aniq ma'lumotlarni yig'ib, ma'lumotlarni yozish va saqlash texnologiyasi yordamida iste'molchilar tomonidan qilingan xarid, buyurtma va xizmatlar haqida katta xajmdagi ma'lumotlarni to'planish, menejment va marketing sohasidagi mutaxassislar tomonidan xaridorlarning xatti xarakatlarida qonuniyatlarni aniqlash, ularning iste'molchilik bilimlari, xatti xarakatlari tashkilotning marketing va mahsulotlar siyosatini boshqarish va tashkilotning daromadi va raqobatbardoshligini oshirishda, zamonaviy axborot texnologiyalari sohasida ma'lumotlarni intellektual tahlil qilish yordamida yig'ilgan statistik ma'lumotlarni tahlil qilish vositalari tahlili masalalari yoritib berilgan.

Kalit so'zlar: Assotsiativ qoidalar, Apriora, Bozor savatchalari, Genetik algoritmlar, Cell protsessorlar, Intel protsessorlari.

Abduganieva M.
"Digital economy and information technologies" department
senior teacher

ASSOCIATIVE RULES AND MARKET BASKETS

Annotation: In this article, modern organizations in the field of Service and Commerce collect accurate data on each order made through plastic cards and controlling computer systems, collect data on purchases, orders and services made by consumers with the help of data writing and storage technology, determine the laws in the behavior of buyers by specialists in the field of management and marketing, , the issues of analysis of statistical data collected using intellectual data analysis in the field of modern Information Technology have been highlighted in the management of marketing and product policies of the organization and increasing the income and competitiveness of the organization.

Keywords: associative rules, a Priora, market baskets, genetic algorithms, Cell processors, Intel processors.

KIRISH

Kompyuter texnologiyalarining rivojlanishi saqlash kerak bo'lgan ma'lumotlar hajmining ko'payishiga olib kelmoqda. Bu esa o'z o'rnida insonning ma'lumotlar bilan ishlashini murakkablashtiradi. Ma'lumotlar ustida ishlashda tahlilning ahamiyati shubxasiz juda katta, chunki bu «ishlov berilmagan ma'lumotlar» orasida bilimlar tuzilishiga olib keladi. Bu bilimlar qaror qabul qilishda qo'llanilishi mumkin. Shu sababli oxirgi payt «ma'lumotlar bazasidan bilim olish» (knowledge discovery in databases) yo'nalishi keskin suratlarda rivojlanib bormoqda. Hozirgi kunda ma'lumotlar omborining xajmi, yangi masshtablanuvchi algoritmlarning kelib chiqishi uchun asosiy sabab bo'lib xizmat qilmoqda.

Data Mining – bu «ishlov berilmagan» ma'lumotlar ichidan kerakli, ilgari taniqli bo'lmagan, amaliy jihatdan foydali va interpretatsiyaga loyiq bilimlarni topish jarayoni hisoblanadi. Bu bilimlar inson hayotining turli yunalishlarida qaror qabul qilish jarayonida muhim rol o'ynaydi.[1]

Data Mining metodlarini qo'llash orqali aniqlangan axborotlar ilgari notrivial va notanish bo'lishi kerak, masalan, o'rtacha sotish bunga misol bo'la olmaydi. Bilimlar xossalari orasida yangi bog'liqliklarni aniqlanishi, biri ikkinchisining hossalari oldindan aniqlashi va hakoza. Aniqlangan bilimlar yangi ma'lumotlarda ayrim ishonchlilik darajasi bilan qo'llanilishi kerak.

Har bir bilim ayrim yunalishda 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 qiladi. Ilgari bu fakt Data Mining uchun qiyin masala sifatida ko'rilar edi, biroq hozirgi paytda zamonaviy protsessorlarning rivojlanishi bu masalani ahamiyatini susaytirdi.[2]

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.

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

Bilimlarni aniqlash metodlari orasida keng qo'llaniladigani bu assotsiativ qoidalarni topish algoritmi. AIS deb ataluvchi birinchi assotsiativ koidalarni 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 aniqlash uchun asosan Apriori algoritmi qo'llaniladi. Uning muallifi Rakesh Agrawaldir (Rakesh Agrawal). [3]

Assotsiativ qoidalar xodisalar orasidagi muntazamlikni aniklashda qo'llaniladi. Bunga misol tariqasida quyidagi tasdiqni keltirish mumkin: non sotib olmoqchi bulgan ist'emolchi sut ham sotib olishi mumkinligining ehtimoli 75%ni tashkil qiladi.[4]

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.

Tadqiqot mavzusi bo'yicha adabiyotlar tahlili

Assotsiativ qoidalar va bozor savatlarinining tahlili bo'yicha chet ellik olimlarning, jumladan R. Agrawal, T. Imielinski, A. Swami, R. Srikant. A. Savasere, E. Omiecinski, and S. Navathe, J.S. Park, M.-S. Chen, and S.Y. Philip, J. Hipp, U. Guntzer, and G. Nakaeizadeh. 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.

Tadqiqot maqsadi va vazifalari. Tanlangan metod va algoritmlar asosida bozor savatchalarini tahlili qilishning optimallashtirish. Ushbu maqsadga erishish uchun quyidagi masalalar yechildi:

- Assotsiativ qoidalar o‘rganish va mavjud adabiyotlardan nazariy ma’lumotlarni to‘plash;
- Apriora - assotsiativ qoidasining algoritmini va unga oid nazariy ma’lumotlarni tahlil qilish;
- Bozor savatchalarining tahlil qilish masalalarini yechishning Genetik algoritmi tahlil qilish;
- Bozor savatchalarining taxlil qilish masalalarini yechishning parallel algoritmi tahlil qilish;
- O‘rganilgan algoritmlar asosida qilingan hisoblashlarni tahlil qilish;
Algoritm va metodlarni yaxshilash bo‘yicha asosli taklif va tavsiyalar ishlab chiqish.

Tadqiqot natijalari

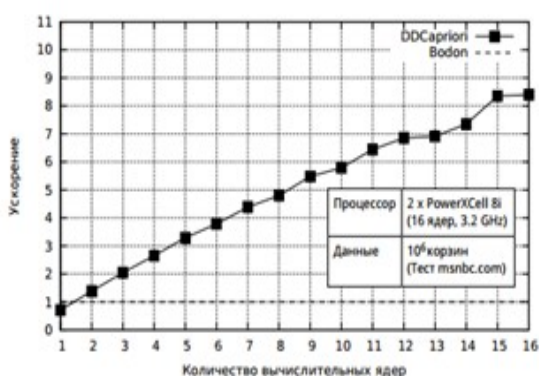
Tadqiqotning nazariy ahamiyati, bozor savatlarini tahlili qilish masalalarini yechish uchun assotsiativ qoidalarni izlash va algoritmlarni ishlab chiqishdan iborat.

Dissertatsiyani amaliy ahamiyati, assotsiativ qoidalar asosida ishlab chiqilgan parallel algoritm yaratish va ushbu algoritm ma’lumotlar tahlilini Cell protsessorlar bazasila hisoblashlarni amalga oshirish.

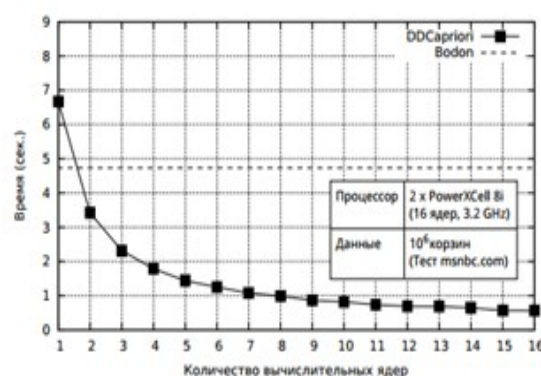
Ishlab chiqilgan algoritm samaradorligini baholash uchun biz uch seriyadagi hisoblash tajribalarini o‘tkazdik. Tajribalarning tashqi ma’lumotlari sifatida Data Mining algoritmi samaradorligini baholash uchun foydalanilgan msnbc.com veb sayti sahifalariga tashrif buyurishlar to‘g‘risidagi standart test to‘plam ma’lumotlari olindi . B to‘plam test vazifasi sayt sahifalariga tashrif buyurish bo‘yicha qaydlarni aks ettiradi. Har bir qayd qanday mazmuniy toifaga

tegishliligi to‘g‘risidagi belgidan iborat. Tajribalarda tez-tez tashrif buyuriladigan sahifalarning qiduruv to‘plamlari amalga oshirildi .

Tajribalarning birinchi seriyasida biz hisoblash yadrosiga bog‘liq holda ish vaqti va algoritm tezlanishini aniqladik. Izchil algoritmlarning qabul qilingan unumdorlik birligi uchun tezlanishni hisoblashda bugungi kunda bozor savatini tahlil vazifalarini yechishning eng yaxshi algoritmlaridan biri sanaladi. Tajribalar shuni ko‘rsatdiki, DDCapriori algoritmi chiziqliga yaqin tezlanishni namoyon qildi..



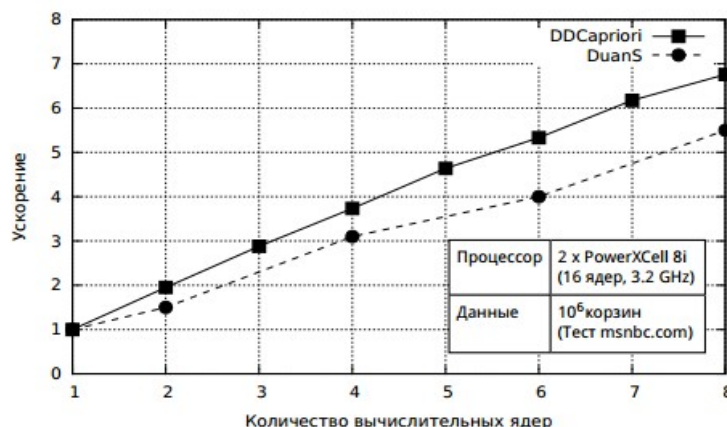
a) tezlanish



b) vaqt bo‘ychida tahlil

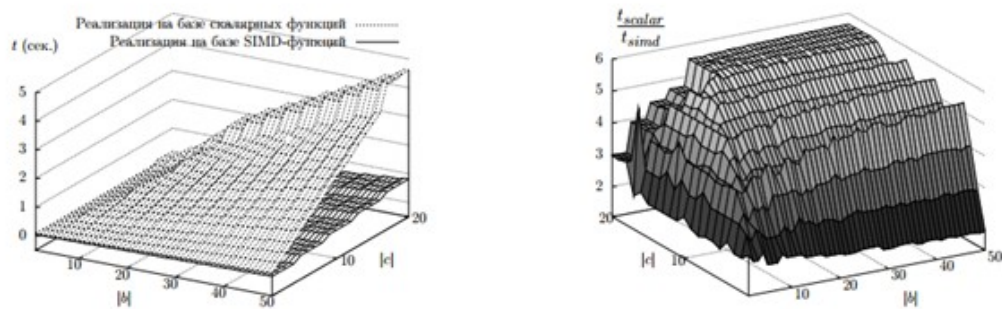
1-rasm. DDCapriori algoritmining ishlash kuchi

Bundan tashqari biz ishlab chiqilgan algoritm va Count Distribution dlya Cell uchun Count Distribution algoritmini o‘lchamlarini taqqosladik. Taqqoslash shuni ko‘rsatdiki ishlab chiqilgan algoritm bir qancha yaxshi o‘lchamlarga ega



2-rasm. DataDistribution va CountDistribution mastshablarni taqqoslash.

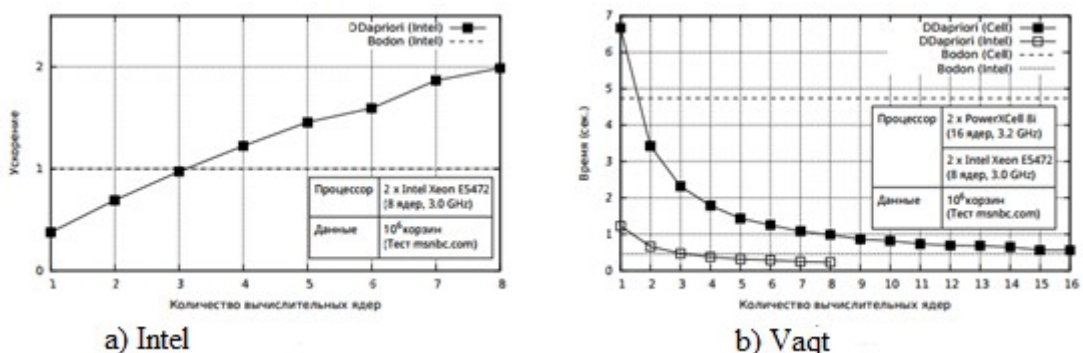
Tajribalarning ikkinchi seriyasida biz nomzod uzunligi va savatga bog‘liq holda savatga nomzodning kirishini tekshirishda skalyar bilan birga vektor operatsiyalarini qo‘llashdan yutuqni taqqosladik. Tajribalar natijalari shuni ko‘rsatdiki, vektor operatsiyalaridan keladigan yutuq nomzod va savat uzunligiga to‘g‘ri proporsional..



a) Vaqt bo‘yicha tekshirish b) vektorli orqali tekshirish

3-rasm. Kandidatning korzinaga kirishini vektorli funktsiya orqali tekshirish.

Tajribalarning uchinchi seriyasi Cell va Intel platformasidagi ishlab chiqilgan algoritmni unumdorligini taqqoslashga yo‘naltirilgan. Bu tajribalarni o‘tkazish uchun biz Intel protsessoriga mo‘ljallangan DDCapriori algoritmini ishlab chiqdik. SPE-tizimi bilan birga buni amalga oshirishda POSIX- tizimidan foydalaniladi va vektor funksiyalari ishlatilmaydi. Tajribalar natijalari 10-rasmda keltirilgan .



a) Intel

b) Vaqt

3-rasm. Cell va Intel protsessorlarini ishlash jarayonini taqqoslash

Tajribalar shuni ko'rsatdiki, Cell protsessorlaridagi algoritmlar Intel protsessorlaridagi algoritmlarga qaraganda bir qancha yaxshi tezlanishlarni namoyon qildi. Ammo Intel protsessorlari Cell protsessorlariga qaraganda ancha yuqori tezkor faoliyatni ta'minlaydi.

Xulosa

Ushbu tadqiqotda ishlov berilmagan ma'lumotlar orasida bilimlar tuzilishiga olib keladi. Bu bilimlar qaror qabul qilishda qo'llanilishi mumkin. Shu sababli so'nggi yillarda «ma'lumotlar bazasidan bilim olish» (knowledge discovery in databases) yo'nalishi keskin suratlarda rivojlanib borayotganini yaqol misol sifatida ko'rishimiz mumkin.

Data Mining – bu «ishlov berilmagan» ma'lumotlar ichidan kerakli, ilgari taniqli bo'lmagan, amaliy jihatdan foydali va interpretatsiyaga loyiq bilimlarni topish jarayoni hisoblanadi. Bu bilimlar inson xayotining turli yunalishlarida qaror kabul qilish jarayonida muhim rol uynaydi.

Xususan, bozor savatchasining asosiy tushunchalari izohlangan, katta hajmdagi ma'lumotlar qayta ishlash va zarur axborotni samarali topishga imkon beruvchi Data Mining dastaklaridan biri sifatida assotsiativ qoidalar ko'rib chiqilgan. Yuqorida ta'kidlaganidek, assotsiativ qoidalarni izlash masalasi dastlab bozor savatchasini tahlil qilish uchun taqdim etilgan.

Assotsiativ 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.

Tadqiqot natijalaridan kelib chiqqan ma'lumotlar yordamida parallelashtirishning umumiy usuli yaratildi.

Yuqorida ko'rsatilgan algoritmlarni ta'riflash uchun asosiy tushunchalari ifodalangan.

Bozor savatlarining yo'nalishlaridan amaliy misollar keltirilgan.

Cell va Intel protsessorlar bazasidagi hisoblash tizimi uchun moslashtirilgan bozor savatlarini tahlil qilish vazifalarini yechishning parallel algoritmi keltirilgan. Parallel algoritmlar to'plamni guruhlariga bo'lishi va hisoblash yadrolari bo'yicha bu guruhlarini tarqatish yo'li bilan erishilgan. Bunda savat to'plami har bir hisoblash yadrosiga uzatiladi.

Taklif qilingan algoritm samaradorligini ko'rsatuvchi hisoblash tajribalari natijalari keltirilgan.

Keltirilgan algoritm va model yordamida ko'pgina tajribalar o'tqazildi va tajriba natijalar asosida yakuniy xulosalar chiqarildi.

Faydalanilgan adabiyotlar ro'yxati

1. Wang, H. lay, and B. Liu. interestingness-Based Interval Merger for Numeric Association Rules. In *Proc. of the 4 th Intl. Conf. on Knowledge Discovery and Data Mining*, pages 121-128, New York, NY, August 2014.

2. I. Webb. Preliminary investigations into statistically valid exploratory rule discovery. In *Proc. of the Australasian Data Mining Workshop (AusDMOS)*, Canberra, Australia, December 2013.

3. Xiong, X. He, C. Ding, Y. Zhang, V. Kumar, and S. R. Holbrook. Identification of Functional Modules in Protein Complexes via Hyperclique Pattern Discovery. In *Proc. of the Pacific Symposium on Biocomputing, (PSB 2005)*, Maui, January 2010.

4. H. Xiong, S. Shekhar, P. N. Tan, and V. Kumar. Exploiting a Support-based Upper Bound of Pearson's Correlation Coefficient for Efficiently Identifying Strongly Correlated Pairs. In *Proc. of the 10th Intl. Conf. on Knowledge Discovery and Data Mining*, pages 334-343, Seattle, WA, August 2010.

5. H. Xiong, M. Steinbach, P. N. Tan, and V. Kumar. HICAP: Hierarchical Clustering with Pattern Preservation. In *Proc. of the SIAM Intl. Conf. on Data Mining*, pages 279-290, Orlando, FL, April 2011
6. H. Xiong, P. N. Tan, and V. Kumar. Mining Strong Affinity Association Patterns in Data Sets with Skewed Support Distribution. In *Proc. of the 2003 IEEE Intl. Conf. on Data Mining*, pages 387-394, Melbourne, FL, 2010
7. X. Yan and J. Han. gSpan: Graph-based Substructure Pattern Mining. In *Proc. of the 2002 IEEE Intl. Conf. on Data Mining*, pages 721-724, Maebashi City, Japan, December 2012.
8. C. Yang, U. M. Fayyad, and P. S. Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *Proc. of the 7th Intl. Conf. on Knowledge Discovery and Data Mining*, pages 194-203, San Francisco, CA, August 2011.