

MATNLI MA'LUMOTLAR UCHUN SINFLASHTIRISH MASALASINI YECHISH USUL VA ALGORITMLARI

Madrahimov Zafarbek Erkin o'g'li

Muhammad Al-Xorazmiy nomidagi TATU Urganch filiali

E-mail: Zafarbek0722@gmail.com

ANNOTATSIYA

Mazkur tezisda matnli ma'lumotlarni sinflashtirish usullari va algoritmlari tasvirlangan. Elektron hujjatlar bilan ishlashda matnli ma'lumotlarni sinflashtirish yordamida vaqtingiz tejaladi va ish samaradorligi oshadi. Xulosa asosida ushbu algoritmlarni qo'llashning o'ziga xos xususiyatlari keltirildi.

Kalit so'zlar: Matnli ma'lumotlar, algoritm, usul, yondashuv.

Matnli ma'lumotlarni sinflashtirishning bir necha yondashuvlari mavjud.

Birinchi yondashuv inson omili yordamida ya'ni matnli ma'lumotni insonlar yordamida o'qib olib uni sinflashtirilishi. Ushbu usul aniq, ammo asosiy kamchiliklaridan biri bu katta hajmdagi ma'lumotlarni maqbul vaqt ichida qayta ishlashning mumkin emasligi.

Ikkinchi yondashuv muntazam ifodalar asosida qoidalarni yozish. Ushbu yondashuvda matnni sinflash bo'yicha mutaxassis tomonidan qoidalar majmui tashkil etiladi. Natijada katta hajmdagi ma'lumotlarni qayta ishlashga imkon beruvchi muntazam ifodalarning asoslari shakllantiriladi. Biroq, bunday qoidalar yaratish uchun katta kuch va vaqt talab qilinadi. Shu bilan birga, qoidalarni aniqlashdan oldin mutaxassis barcha sinflardan turli xil ma'lumotlar namunalari bilan chuqur tanishishi kerak, bu juda ko'p vaqt talab qilishi mumkin.

Uchinchi yondashuv mashinali o'qitishga asoslangan. Ushbu yondashuvda namuna matnining sinfi avtomatik ravishda kompyuter tomonidan aniqlanadi. Ushbu yondashuv o'quv ma'lumotlarini oldindan qo'lda belgilashni talab qiladi, ammo bu barcha namunalarni sinflarga tegishli bo'lgan qoidalarni aniqlashdan ko'ra oddiy vazifadir. Hozirgi vaqtda ushbu yondashuv eng ko'p ishlatiladigan va istiqbolli hisoblanadi, chunki u insondan eng kam kuch talab qiladi va katta hajmdagi matnli ma'lumotlarni avtomatik qayta ishlash qobiliyatiga ega bo'ladi.[1]

Mashinani o'rganishga asoslangan matnni sinflash bir nechta asosiy algoritmlar bilan ifodalanadi.

Naif Bayes klassifikatori. Ushbu usul Bayes teoremasiga asoslangan ba'zi bir qo'shimchalar bilan ehtimollik tasnifi usuli hisoblanadi. Bayes teoremasi ikki

hodisaniing ehtimoli va ularning shartli ehtimoli o'rtasidagi munosabatni beradi. Naif Bayes klassifikatori sinfnig muayyan xususiyatlarining mavjudligi yoki yo'qligi boshqa xususiyatlarning mavjudligi yoki yo'qligi bilan bog'liq emasligini ta'kidlaydi. Aytaylik, ob'yekt rang, shakl va massa kabi xususiyatlarga ko'ra tasniflanishi mumkin.

Neyron to'rlari va uni o'qitish

Neyron tarmoq neyronlarning tizimi va ular orasidagi aloqa. Ta'lim jarayonida turli neyronlarni bog'laydigan aloqalarning vazni asta-sekin o'zgaradi. Neyron tarmoqni o'qitish natijasi bunday tarmoq bo'lib, uning aloqasi muammoning shartlarini qondiradigan koeffitsientga ega. Neyron tarmoqlari neyron aloqalar bilan o'rgatilgan qatlamlar soni bilan tavsiflanadi. Katta hajmdagi ma'lumotlar va ko'plab parametrlarga ega bo'lgan muammolarni hal qilishda ko'plab qatlamlarga ega bo'lgan neyron tarmoqlar qo'llaniladi. Bunday tarmoqlarni chuqur o'rganish deb ataladi. Chuqur o'rganish usullari ikkita muhim xususiyat bilan ajralib turadi: murakkab g'oyalarni qatlamli yaratish, shuningdek, oraligni batafsil o'rganish har bir qatlam yuqori qatlam vakillik ehtiyojlari va quyida qatlami ehtiyojlariga ko'ra yangilanadi, shunday qilib, taqdim etadi. Birgalikda, bu ikki xususiyat mashinani o'rganish uchun oldingi yondashuvlarga qaraganda chuqur o'rganishni ancha muvaffaqiyatli qiladi. Neyron tarmoqlarining afzalliklari shundaki odatda, boshqa usullarga nisbatan matnni tasniflashning yuqori aniqligi. Bundan tashqari, ushbu usulning aniqligi odatda ma'lumotlar sonining ko'payishi bilan kuchayadi, ya'ni ma'lumotlarni to'plash qobiliyatiga ega bo'lgan tizimlarda bu usul ko'proq mos keladi. Neyron tarmoqlarining kamchiliklari an'anaviy ravishda neyron tarmoqlardan foydalanishni murakkablashtiradigan hisoblash resurslariga bo'lgan ehtiyojning yuqori darajasi hisoblanadi.[2]

Xulosa matn ma'lumotlarini sinflashning mavjud usullarini taqqoslash natijasida quyidagi xulosaga kelish mumkin. Katta hajmdagi ma'lumotlarni qayta ishlashda chuqur o'rganish bilan ko'p qatlamli neyron tarmoqlardan foydalanish tavsiya etiladi. Bayes tasniflagichi usuli muqobil usul sifatida ishlatilishi mumkin, agar belgilar haqida qo'shimcha ma'lumot bo'lsa.

FOYDALANILGAN ADABIYOTLAR RO'YXATI: (REFERENCES)

1. **Srivastava N., Hinton G., Krizhevsky A., Sutskever I. and Salakhutdinov R.** Dropout: A Simple Way to Prevent Neural Networks from Overfitting Journal of Machine Learning Research. 2014[1]
2. **Чернодуб А.Н., Дзюба Д.А.** Обзор методов нейроруправления Проблемы программирования. – 2011.
3. **Обучение нейронных сетей.** — Горбань А.Н. - 1990. [2]

4. Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления. — 2002.
5. Ясницкий Л. Н. Введение в искусственный интеллект - 2005
6. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение – 2018
7. Круглов В. В., Борисов В. В. Искусственные нейронные сети. - 2001