TARMOQ MA’LUMOTLARIGA INTELLEKTUAL ISHLOV BERISH MODELLARI VA USULLARI

Mo‘minov Bahodir Boltayevich[1], Husanov Sherzod Abdimonnonovich[2]

[1]Toshkent davlat iqtisodiyot universiteti
[2]Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti

K E Y W O R D S

tarmoq ma’lumotlari, intellektual ishlov berish, sun’iy intellekt, mashinali o‘qitish, ma’lumotlar mayningi, graf neyron tarmoqlari, anomaliya aniqlash

A B S T R A C T

Ushbu ishda tarmoq ma’lumotlariga intellektual ishlov berishning zamonaviy modellari va usullari tahlil qilingan. Tarmoq ma’lumotlarining hajmi va murakkabligi ortib borayotgani sababli, ularni samarali tahlil qilish va qayta ishlash uchun sun’iy intellekt, mashinali o‘qitish hamda ma’lumotlar mayning (data mining) texnologiyalaridan foydalanish zarurligi asoslab berilgan.
Tadqiqotda tarmoq ma’lumotlarining tuzilishi, ularning turlari (ijtimoiy tarmoqlar, kompyuter tarmoqlari, sensor tarmoqlari va boshqalar) hamda ma’lumot oqimlari bilan bog‘liq muammolar yoritilgan. Shuningdek, intellektual ishlov berishda qo‘llaniladigan asosiy usullar — klassifikatsiya, klasterlash, prognozlash, anomaliyalarni aniqlash va graf analizi usullarining afzalliklari va qo‘llanilish sohalari keltirilgan.
Ishda shuningdek, neyron tarmoqlar, graf neyron tarmoqlari (GNN), konvolyutsion va rekurrent modellardan foydalanish imkoniyatlari hamda ularning tarmoq ma’lumotlari tahlilidagi samaradorligi tahlil qilingan. Natijada, intellektual ishlov berish usullari orqali tarmoq ma’lumotlaridan qimmatli ma’lumotlarni avtomatik tarzda ajratib olish, xavfsizlikni ta’minlash va boshqaruv qarorlarini qabul qilishni takomillashtirish mumkinligi ko‘rsatilgan.

Veb-hujjatlarni tasniflash – mazmun, tuzilishi yoki boshqa xos xususiyatlarga ko‘ra tavsiflash, o‘rganish va tuzilmani izohlash jarayonini ko‘rsatadi. Bu imkonipyat ma’lumotlarni qidirish, tabiiy tilni qayta ishlash va veb kontentlari, sahifalarini intellektual ishlov berishning asosiy vazifasidir. Veb-hujjatlarni tasniflash turli sohalarda keng qo‘llanib kelinishi quyidagi ilovalarga xizmat qilishi mumkin:

Veb-qidiruv tizimlari (elementlari): Umumiy holda tasniflash algoritmlari qidiruv tizimi natijalarining aniqligi va dolzarbligini oshirish uchun ishlatiladi. Veb-hujjatlarni turli toifalar yoki mavzularga tasniflash orqali qidiruv tizimlari foydalanuvchilarga ko‘proq yo‘naltirilgan va aniq qidiruv natijalarini taqdim etishi mumkin.

Kontentni filtrlash va spamni aniqlash ilovalari: Veb-hujjatlarni tasniflash spam elektron pochta xabarlari, zararli veb-saytlar va nomaqbul kontentga qarshi kurashish uchun tarkibni filtrlashda hal qiluvchi asosiy vosita, usul yoki algoritm hisoblanadi. Veb-hujjatlarni spam yoki spam bo‘lmagan deb tasniflash orqali filtrlash tizimlari foydalanuvchilarni kiruvchi yoki zararli kontentdan himoya qilishi mumkin deb hisoblanadi.

So‘rovga mos natijalarni tasniflashtirish: Tasniflash algoritmlari so‘rovga mos natijalarni, hujjatlarni foydalanuvchining xoxishiga mos qilib turli mavzularga avtomatik ravishda tasniflash uchun ishlatiladi. Bu foydalanuvchilar uchun oldindan tayyorlangan ma’lumotlar to‘plamidan to‘g‘ri foydalanish imkoniyatini yaratib beradi.

Firibgarlikni aniqlash va kiberxavfsizlik ilovalari: Tasniflash modellari firibgar veb-

©B.B. Mo‘minov, Sh.A. Husanov

hujjatlarni, fishing veb-saytlarini yoki shubhali onlayn faoliyatni aniqlash uchun turli yondashuvlar asosida ishlatiladi. Hujjatlarni o‘z xususiyatlariga ko‘ra tasniflash orqali mashinali o‘rganish algoritmlari onlayn firibgarlikni aniqlash va oldini olish va kiberxavfsizlik choralarini kuchaytirishga to‘liq yordam berib kelmoqda.

Hujjatlarni taqdim qilish tizimlari: Tasniflash usullari veb-hujjatlarning katta to‘plamlarini oldindan aniqlangan o‘quv tanlanma asosida mustaqil ravishda mavzuli to‘plamlarga (klasterlarga) tasniflash orqali tartibga solish, hamda foydalanuvchiga taqdim etishga xizmat qiladi. Shu orqali tasniflash veb-hujjatlardan tegishli, o‘ziga xos ma’lumotlarni masalan, nomlangan obyektlar, kalit iboralar yoki muhim mavzularni olishda, to‘plamini shakllantirishda asosiy vosita bo‘lib hisoblanadi.

Veb-hujjatlarni tasniflash kompyuter tarmoqlarida, global tarmoqda samarali ma’lumot izlash va olish, hosil qilingan kontentni yoki to‘plamni filtrlash, xos xususiyat va hossalarni intellektual tahlil qilish, internet va himoyalangan tarmoqlardagi firibgarlikni aniqlash va shaxsga yo‘naltirilgan xizmatlarni ta’minlay oladigan dasturiy vositadir. Uning ilovalari turli sohalar va domenlar bo‘ylab qo‘yilgan masalani yechadi va foydalanuvchi ma’lumotlar to‘plamini yaxshilash, tajribasini nazorat qilish, xavfsizlikni yaxshilash va veb texnologiyaga asoslangan ma’lumotlardan yaxshiroq foydalanishda katta yordam beradi.

Veb-hujjatlarning tuzilishi va mazmunini tushunish uchun veb-hujjatlarni samarali tasniflash juda muhimdir. Veb-sahifalar yoki HTML-fayllar kabi veb-hujjatlarda tasniflash uchun foydalanish mumkin bo‘lgan o‘ziga xos elementlar va xususiyatlar mavjud va bularni inobatga olish muhim hisoblanadi. Buning asosida asosiy jihatlarning tadqiqi muhim:

HTML tuzilmasi: Veb-hujjatlar odatda (HTML)da yoziladi. Tegishli ma’lumotlarni olish uchun HTML tuzilishini tushunish juda muhimdir. HTML teglari veb-sahifaning sarlavhalar (title), paragraflar (paragraf), teglar (teg), ro‘yxatlar (list), rasmlar (image), havolalar (link) va jadvallar (table), senariylar, skriptlar (action), foydalanuvchining shaxsiy teglari (personal teg) kabi elementlarini belgilaydi. Ushbu teglar tasniflash algoritmlari uchun asosiy sinflar bo‘lib, kontekstli ma’lumotlarni taqdim etish imkoniyati ta’minlaydi.

Matn (Kontent) tuzilmasi: Veb-hujjatlardagi matn yoki kontent mazmunini tasniflash uchun eng zo‘r algoritmlarga asoslangan va foydalanuvchilar uchun eng asosiy ma’lumotlarni o‘z ichiga oladi. Kontent tuzilmasiga HTML tuzilmasining barcha sinflarini kiritish va ularning xos teglariga tegishli ma’lumotlar kiradi. Matn mazmunini intellektual tahlil qilish, masalan, kalit so‘zlarni ajratib olish, matn, kontent mavzusini tushunish va tasniflash uchun muhim xususiyatlarni aniqlashga yordam beradi.

Meta ma’lumot (Metadata): Veb-hujjatlarda ko‘pincha hujjat haqida qo‘shimcha ma’lumot beruvchi tur maqsadli metama’lumot sohalari mavjud. Bu sohalarga hujjat nomi, muallifi, nashr sanasi va kalit so‘zlar, qidiriuv so‘zlari, SEO so‘zlari kabi atributlar kiradi. Meta ma’lumot veb hujjatga tegishli ma’lumotlarni kiritish orqali tasniflash jarayonini tezlashtiradi

Hujjatni formatlash: Veb-hujjatlarda shrift uslublari, shrift o‘lchamlari, matn ranglari yoki matnni tekislash kabi maxsus formatlash elementlari bor. Bu elementlar foydalanuvchiga mobillikni ta’minlash uchun tez o‘zgaruvchandir. Formatlash xususiyatlarini tahlil qilish hujjatning tuzilishi va loyihalash mexanizmini takomillashtirish va mobilligini oshirishda xizmat qiladi.

Giperhavolalar: Veb-hujjatlarning asosiy elementi bu giperhavola hisblanadi. Bu element to‘ldiruvchi, yordamchi va umuman boshqa mavzudagi veb-sahifalar yoki manbalarga o‘tish imkoniyatini yaratib beruvchi giperhavolalar texnologiyasidir. Umuman olganda, bu veb hujjatni boshqa hujjatlardan ajratib ko‘rsatadigan muhim elementi hisoblanadi. Agar bu element mavjud bo‘lmasa, xujjatni veb hujjat deb yuritish mumkin emas. Giperhavolalar va ular bilan bog‘liq matn yoki kontent hujjatning mazmuni va mavzusi haqida ma’lumotlarga, axborotlarga ega bo‘ladi. Tasniflash aniqligini oshirish uchun havolalarni tahlil qilish va havolaga asoslangan xususiyatlardan foydalanish tavsiya qilinadi.

Hujjat tuzilishi va loyihalash iyerarxiyasi: HTML teglari bilan berilgan veb-hujjatlarning

(cc) BY

loyihalash iyerarxik tuzilishi matnlarning, kontentlarning mantiqiy tashkil etilishini ta’minlaydi va boshqaradi. Hujjat tuzilishini intellektual tahlil qilish sarlavhalar, kichik sarlavhalar yoki kontent bloklari kabi turli elementlar o‘rtasidagi munosabatlarni tushunishga yordam beradi. Ushbu tarkibiy ma’lumotlar muhim bo‘limlarni aniqlash yoki hujjatning umumiy tartibini olish uchun ishlatiladigan eng asosiy omildir.

Veb hujjatlarning tuzilishi va mazmunini tushunib, tasniflash algoritmlari tegishli ma’lumotlarni olish, loyihalash uslublarini aniqlash va aniq eng maqbul usullarni joriy (bashorat) qilish uchun ushbu xususiyatlardan foydalanish lozim. Veb-hujjatlarni tasniflash uchun mavjud ma’lumotlardan samarali foydalanishda HTML kontentni tahlili, matnni ajratib olish, metama’lumotlarni tahlil qilish va tizimli tahlil kabi texnologiyalar muhim rol o‘ynaydi.

Veb hujjat kontentini veb kontent deb atash mumkin. Bu veb kontentlarga dastlabki (oldindan) ishlov berish va xos xususiyatlarini ajratib olish muhim masalalardan bo‘lib hisoblanadi.

Dastlabki ishlov berish va xususiyatlarni ajratib olish usullari va algoritmlari. Veb kontentlarda xususiyatlarni ajratib olish usullari ham ma’lumotlarini tasniflash algoritmlari tomonidan ishlatilishi mumkin bo‘lgan, raqamli tasvirlarda obyektlarni tasniflash usullari va algoritmlaridan foydalanish mumkin. Veb kontentga dastlabki ishlov berish va xususiyatlarinni aniqlashning eng ko‘p ishlatilishi tavsiya qilingan usullarga quyidagilarni kiritiladi:

So‘z chastotalari: So‘z chastotalari veb-hujjatda mavjud bo‘lgan har bir elementdagi matndagi so‘zning chastotasini (sonini) ifodalaydi. Ushbu usul har bir so‘zning sonini hisoblaydi va har bir element ma’lum bir so‘zning soniga mos keladigan xususiyat vektorini (hususiy holda matritsiyasini) yaratadi. Tez-tez uchraydigan so‘zlarga vektorda yuqori qiymatlar beriladi. So‘z chastotalarini hisoblash oson bo‘lsa-da, ular so‘zlarning kontekstli yoki semantik aloqalarini aniqlay olmaydi.

TF-IDF (atamalar chastotasi - teskari hujjat chastotasi): TF-IDF - bu veb-hujjat va hujjatlar to‘plamidagi so‘zning ahamiyatini o‘lchash uchun atamalar chastotasi va teskari hujjat chastotasini birlashtirgan eng ko‘p tarqalgan usullardan hisoblanadi. Bu usul - Veb hujjatda tez-tez uchraydigan, lekin bir to‘plamda kamroq uchraydigan so‘zlarga kattaroq vazn beradi. TF-IDF atamalarning ahamiyatini muvozanatlashtiradi va tez-tez uchraydigan so‘zlarning ta’sirini kamaytiradi. Olingan TF-IDF qiymatlari tasniflash uchun xususiyatlar sifatida ishlatiladi.

N-gramma: N-grammalar veb-hujjat ichidagi qo‘shni so‘zlar ketma-ketligini saqlaydi. N-gramma N ta so‘zdan iborat ketma-ketlikni ifodalaydi. Ko‘p ishlatiladigan n-grammalarga, bir grammalar (un-gramma, bir so‘zlar), bigrammalar (bi-gramma, ikki so‘zli ketma-ketliklar) va trigrammalar (uch so‘zli ketma-ketliklar) kiradi. So‘zlarning tartibini hisobga olgan holda, n-grammalar ma’lum darajadagi kontekstli ma’lumotlarni olishlari mumkin. n-grammalar xususiyatli vektorlarga aylantiriladi, bu yerda har bir element hujjatda ma’lum bir n-gramma mavjudligi yoki chastotasini ifodalaydi deb qarash mumkin. Biroq, n-gramma o‘lchami oshgani sayin, kombinatsiyalar soni eksponent ravishda o‘sib boradi, bu esa o‘lchamning oshishiga olib keladi, o‘lchamning oshishi mashinadan kattaroq resurs talab qiladi.

Yuqorida qayd qilingan xususiyatlarni ajratib olish usullari veb-hujjatlarning matn, kontent mazmuniga turlicha qarashlarni ta’minlaydi va aniq tasniflash masalasiga qarab alohida-alohida yoki gibrid ko‘rinishida yaratib olib ishlatilishi tavsiya etiladi. Shuni ta’kidlash kerakki, xususiyatlarni ajratib olish ushbu usullar bilan cheklanmaydi va tasniflash maqsadlarida veb hujjatlarning ko‘rinishini yaxshilash uchun so‘zlarni joylashtirish (masalan, Word2Vec, GloVe) yoki veb hujjat darajasidagi xususiyatlari inobatga olinib, boshqa usullarni tadbiq qilish tavsiya qilinadi.

Veb-hujjatlar bilan ishlashda HTML teglari, maxsus belgilar hamda tasniflash algoritmlarining ishlashiga salbiy ta’sir ko‘rsatishi mumkin bo‘lgan boshqa bir ahamiyatsiz belgilar kabi noodatiy elementlar mavjud. Ushbu noodatiy elementlarni boshqarish uchun ba’zi tadqiqotlarda keltirilgan yondashuvlarni tahlillari quyidagicha:

(cc) BY

HTML tegini olib tashlash: HTML teglari formatlash va tizimli ma’lumotlarni o‘z ichiga oladi, va uning modelini ya’ni loyihasini belgilab beradi lekin u mazmuni yoki ma’nosiga bevosita hissa ta’sir etaolmaydi yoki o‘zgartira olmaydi. HTML teglarini veb hujjat loyihasidan olib tashlanishi, kontentlarni birlashtirish va aniq bir bo‘lim yoki elementlar asosida e’tiborni matn tarkibiga qaratishga yordam beradi. HTML asosidagi veb hujjatlaridan matn tarkibini tahlil qilish va ajratib olish uchun turli xil kutubxonalar va vositalar mavjud. Masalan, Python dagi BeautifulSoup HTML teglarini o‘chirib tashlagan holda matn tarkibini chiqarish uchun ishlatilishi mumkin bo‘lgan kutubxona hisoblanadi.

Maxsus belgilarini olib tashlash: Ko‘p hollarda vergul, nuqta yoki tirnoq kabi maxsus belgilari muhim ma’noga ega emas va ularni matndan xavfsiz olib tashlash mumkin. Maxsus belgilarini olib tashlash matnni soddalashtirishi va noodatiy elementlarini kamaytirishi qayd qilingan. Biroq, ba’zi maxsus belgilari kontekstli ahamiyatga ega bo‘lishi mumkin (masalan, "yo‘q", "masalan,") yoki maxsus atamalarning bir qismi bo‘lishi mumkin (masalan, "C++"). Qaysi maxsus belgilarini olib tashlash yoki saqlab qolishni hal qilishdan oldin tasniflash vazifasining o‘ziga xos mezonlarini ishlab chiqish lozim.

Yordamchi belgilar bilan ishlash: Maqsadli lug‘at tarkibiga kiritilmagan maxsus belgilar, harf-raqamli bo‘lmagan belgilar, qisqarmalarni yordamchi belgilarga kiritish mumkin. Yordamchi belgilar olib tashlanishi yoki ularni maqsadli ishlov berish bilan almashtirilish mumkin. Masalan, yordamchi belgilarni tasniflash ularning vazifasidagi ahamiyatiga qarab, to‘ldiruvchi belgilar bilan almashtirilishi yoki butunlay olib tashlanishi mumkin. Buning uchun esa maqsadli lug‘at tarkibiga asosiy o‘rin tutadi.

Qo‘shimacha so‘zlarni o‘chirish: Qo‘shimcha so‘zlar matn, kontentda tez-tez uchraydigan, ammo kamroq semantik ma’noga ega bo‘lgan "va", "bilan" kabi keng tarqalgan so‘zlardir (bular har bir tabbiy tilda turlicha bo‘lishi mumkin). Qo‘shimacha so‘zlarini olib tashlash noodatiy elementlarni va o‘lchovni kamaytirishi imkoniyatini yaratib beradi. Python-dagi NLTK (Natural Language Toolkit) kabi kutubxonalar olib tashlash uchun ishlatiladi. Bu kutubxonaga qo‘shimcha so‘zlarini maqsadli lug‘ati oldindan belgilangan ro‘yxat asosida amalga oshiriladi.

Tokenlash va normallashtirish: Tokenlash matnni alohida so‘zlarga yoki tokenlarga bo‘lishni o‘z ichiga oladi. Ushbu yondashuv navbatdigi ishlov berish va tahlil qilish imkoniyatini oshirishga xizmat qiladi. Yondashuvlar barcha belgilarni kichik harflarga aylantirish orqali matnni normallashtirish izchillikni ta’minlash va katta-kichik harf sezgirlik ta’sirini kamaytirishga xizmat qiladi. Shuni ta’kidlash kerakki, shovqinli elementlar bilan ishlashning o‘ziga xos yondashuvi veb hujjatlarning tuzilmasiga, tasniflash masalasiga va dastur talablariga qarab farq qilishi mumkin. Muayyan ssenariy uchun eng samarali dastlabki ishlov berish bosqichlarini aniqlash uchun noodatiy elementlarni yo‘qotish usullarining tasniflash ko‘rsatkichlariga ta’sirini tajriba qilish va baholash tavsiya etiladi.

An’anaviy tasniflash algoritmlari negizida Naive Bayes klassifikatori veb-hujjatlarni tasniflashda qo‘llanilishi mumkin bo‘lgan eng yaxshi va oddiy ehtimolli tasniflash algoritmidir[84]. U Bayes teoremasiga asoslanadi va sinf belgisini hisobga olgan holda xususiyatlar (so‘zlar, n-grammlar va boshqalar) shartli ravishda mustaqil bo‘lishini taxminlaydi. Naive Bayes klassifikatoridan veb-hujjatlarni tasniflash uchun foydalanish mumkinligi bo‘yicha olib borilgan tadqiqotlar asosida:

Ma’lumotlarga dastlabki ishlov berish: Birinchi navbatda veb-hujjatning ma’lumotlarini dastlabki ishlov berish lozim. Bu odatda HTML teglarini, maxsus belgilarini va boshqa noodatiy elementlarni olib tashlash, matnni tokenlash va matnni songa aylantirish uchun zarur bo‘lgan normallashtirish yoki xususiyatni ajratib olish usullarini (masalan, so‘z chastotalari, TF-IDF, n-grammalar, xususiyat vektorlari) qo‘llash kabi bosqichlarni o‘z ichiga oladi.

O‘quv tanlanmani tayyorlash: To‘plangan tajribalar asosida ma’lumotlarini o‘quv va sinov tanlanmalariga ajratish mumkin. O‘quv tanlanmada har bir veb hujjat ma’lum bir sinf yoki toifa bilan bog‘langan veb-hujjatlarning yorliqli namunalari bo‘lishi kerak. O‘quv tanlanma sinflarning katta sonli ma’lumotlarini qamrab olishiga va real ssenariylarda uchrashishni

(cc) BY

kutayotgan ma’lumotlarni aks ettirishiga xizmat qiladi.

Xususiyat vektorlari[85,86]: Oldindan ortirilgan tajriba ma’lumotlari asosida xususiyat vektorlarini tuzish g‘oyasini ishlari suradi. Har bir xususiyat vektori veb-hujjatni ifodalaydi va tanlangan funksiyalarga mos keladigan raqamli qiymatlarni o‘z ichiga oladi (masalan, so‘z chastotalari, TF-IDF vaznlari, n-gramm qiymatlari). Xususiyat vektorlari sinf belgilariga mos kelishi muhim ahamiyat kasb etadi.

Naive Bayes klassifikatorini o‘rgatish[84]: Naive Bayes klassifikatorini o‘quv tanlanmasi va tegishli xususiyat vektorlari yordamida o‘rgatish eng maqbul usuldir. Tasniflagich o‘quv tanlamasi asosida har bir sinfda yuzaga keladigan har bir xususiyatning ehtimolini hisoblab chiqadi. Naive Bayes, ehtimolliklarini soddalashtiradigan sinf belgisini hisobga olgan holda xususiyatlar shartli ravishda mustaqil xususiyat deb hisoblaydi.

Tasnifi: Naive Bayes klassifikatori o‘qitilgandan so‘ng, uni yangi veb-hujjatlarni tasniflash uchun qo‘llaniladi. Yangi veb hujjatlarni o‘quv tanlanmalari bilan bir xil dastlabki ishlov berish, ularning xususiyat vektorlarini tuzish va sinf belgilarini bashorat qilish uchun klassifikatordan foydalanish lozim. Tasniflagich xususiyatlarni hisobga olgan holda har bir sinfning shartli ehtimolligini hisoblab chiqadi va veb hujjat uchun bashorat qilingan belgi sifatida eng ehtimolli sinfni tanlaydi.

Baholash: Sinov tanlanmasidan foydalanib, Naive Bayes klassifikatorining ishlashini baholash lozim. Aniqlik, eslab qolish va ayrim vazn kabi ko‘rsatkichlarni hisoblash uchun taxmin qilingan sinf yorliqlarini haqiqiy teglar bilan solishtirish amalga oshiriladi. Ushbu ko‘rsatkichlar veb hujjatlarni tasniflashda klassifikatorning samaradorligini belgilab beradi.

Naive Bayes klassifikatorlari soddaligi va hisoblash samaradorligi bilan eng yaxshi tasniflagich hisoblanadi. Biroq, ular xususiyatlar o‘rtasidagi murakkab o‘zaro munosabatlarni ushlay olmaydi va xususiyat to‘g‘risidagi bashoratlari sust bo‘ladi. Shunga qaramay, Naive Bayes klassifikatorlari ko‘plab matnlarni tasniflash vazifalarida, jumladan, veb-hujjatlarni tasniflashda, ayniqsa katta hajmdagi ma’lumotlar to‘plamlari bilan ishlashda samarali ekanligini isbotlangan[84].

Vektor xususiyatlarni qo‘llab-quvvatlash (SVM, VXQQ) - bu tasniflash vazifalari uchun kuchli va keng qo‘llaniladigan mashinali o‘rganish algoritmi. VXQQ tasniflash uchun ishlatilishi mumkin bo‘lgan yuqori o‘lchamli xususiyat maydonida giperplanlar to‘plamini quradi. VHQQ va uning takomillashtirilgan algoritmlari quyidagicha:

Odatiy VXQQ: odatiy VXQQ turli sinflarning tayanch ma’lumotlarini maksimal darajada ajratib turadigan optimal giperplanni topishga qaratilgan. Algoritmning eng katta chegaraga ega bo‘lgan giperplanni topishga intiladi, ya’ni giperplan va har bir sinfning eng yaqin tanch ma’lumot nuqtalari orasidagi maksimal masofadir. Giper tekislikka eng yaqin ma’lumot nuqtalarini qo‘llab-quvvatlovchi vektorlar deb ataladi va qaror chegarasini aniqlashda hal qiluvchi vosita hisoblanadi.

Chiziqli VXQQ: Chiziqli VXQQ - bu VXQQ ning eng oddiy shakli bo‘lib, qaror chegarasi chiziqli giperplandir. Bu sinflarni to‘g‘ri chiziq yoki giperplan bilan ajratish mumkin bo‘lgan chiziqli ajratiladigan ma’lumotlar to‘plamlari.

Chiziqli bo‘lmagan VXQQ: Chiziqli bo‘lmagan VXQQ - odatiy VXQQni chiziqli ravishda ajratib bo‘lmaydigan ma’lumotlar to‘plamlarini boshqarish uchun kengaytiradi variantidir. U sinflar ajraladigan, tayanch xususiyatni yuqori o‘lchamli fazoga xaritalash uchun yadro hiylasi deb ataladigan texnikadan foydalanadi. Keng tarqalgan ishlatiladigan yadro funksiyalariga radial asos funksiyasi (RAF) yadrosi, polinom yadrosi va sigmasimon yadrolarni kiritish mumkin. Yadroni tanlash ma’lumotlar to‘plamiga va qarorlar chegarasining murakkabligiga bog‘liq holda amalga oshiriladi.

Ko‘p sinfli VXQQ: Bu VXQQ ko‘p sinfli tasniflash muammolarini hal qilish uchun takomillashtirish imkoniyati mavjud. Unda birinchi yondashuv birga ko‘p (One-All) strategiyasi bo‘lib, har bir sinf uchun qolgan sinflarga nisbatan alohida ikkilik VXQQ tasniflagichlari o‘rgatiladi. Ikkinchi yondashuv birga birga qarshi (One-One) strategiyasi bo‘lib,

(cc) BY

bunda har bir sinf juftligi uchun ikkilik tasniflagichlar o‘rgatiladi.

Veb-hujjatlarni tasniflash uchun chuqur o‘rganish neyron tarmoqlar - bu inson ongidagi biologik neyron tarmoqlarning tuzilishi va faoliyatidan ilhomlangan mashinani o‘rganish modellarining bir sinfidir. Businflar turli sohalarda, jumladan, tabiiy tilni qayta ishlashda (NLP) ishlatiladigan kuchli modellardir.

Neyron tarmoq asoslari:

Neyron tarmoq qatlamlar asosida tashkil etilgan neyronlar deb ataladigan o‘zaro bog‘langan tugunlardan iborat. NLPda qo‘llaniladigan neyron tarmog‘ining eng keng tarqalgan turi bu ko‘p qatlamli qo‘llaniladigan neyron tarmog‘ining eng keng tarqalgan turi bu ko‘p qatlamli perseptron (MLP) sifatida ham tanilgan, uzatiladigan neyron tarmog‘idir. U kirish qatlami, bir yoki bir nechta yashirin qatlamlar va chiqish qatlamidan iborat. Har bir neyron kirish ma’lumotlarini oladi, ushbu kirishlarning vaznli yig‘indisiga faollashtirish funksiyasini qo‘llaydi va chiqishni ishlab chiqaradi. Neyronlar orasidagi bog‘lanishlar tarmoqning ishlashini optimallashtirish uchun o‘qitish paytida sozlangan bog‘liq vaznliklarga ega bo‘ldai.

NLPda neyron tarmoqlarni qo‘llash orqali Hodisalarni tahlil qilish: Neyron tarmoqlar hodisalarni tahlil qilish uchun keng qo‘llaniladi, bunda maqsad matnda ifodalangan hodisalarni (ijobiy, salbiy yoki neytral) aniqlashdir. Rekurent neyron tarmoqlari (RNN) va ularning variantlari, masalan, Uzoq qisqa muddatli xotira (LSTM) va Gated Recurrent Unit (GRU) odatda matnning ketma-ketligini va so‘zlar orasidagi model bog‘liqligini aniqlash uchun ishlatiladi.

Matn tasnifi: Neyron tarmoqlar mavzularni tasniflash, spamni aniqlash yoki hujjatlarni turkumlashtirish kabi matnlarni tasniflash vazifalari uchun samarali vositadir. Kengaytirilgan (deep) neyron tarmoqlari (DNN) ko‘pincha matndan tabiiy loyiha (andoza) va xususiyatlarni aniqlash uchun ishlatiladi, MLP yoki to‘liq bog‘langan qatlamlar esa ushbu xususiyatlar asosida tasniflash uchun ishlatiladi.

Nomlangan obyektni tan olish (NER): NER matnda nomlangan obyektlarni (masalan, shaxs ismlari, joylashuvi, tashkilotlari) aniqlash va tasniflashni o‘z ichiga oladi. Rekurent neyron tarmoqlari, xususan, ikki tomonlama LSTMlar, so‘zlarning o‘tmishdagi va kelajakdagi kontekstlarini yozib olish orqali NER vazifalari uchun o‘quv qo‘llanma sifatida qo‘llaniladi.

Mashinali tarjima: Neyron tarmoqlar sezilarli darajada rivojlangan mashinali tarjima tizimlariga ega. Enkoder-Dekoder arxitekturasi kabi ketma-ketlik modellari matnni bir tildan ikkinchi tilga tarjima qilish uchun RNN yoki Transformator modellaridan foydalanadi. Ushbu modellar manba tilining konteksti va so‘zlar orasidagi bog‘liqlikni hisobga olgan holda tarjimalarni yaratishni o‘rganadi.

Til yaratish: Neyron tarmoqlar matn yaratish kabi ijodiy vazifalar, jumladan tilni modellashtirish, hikoya yaratish va dialog tizimlari uchun ishlatiladigan modellari ham mavjud. Variatsion avtokoder (Variational autoencoder – VAE) yoki Generativ raqobatli tarmoqlari (Generative Adversarial Neural Network (GAN)) modellar izchil va kontekstga mos matn yaratish uchun mo‘ljallangan.

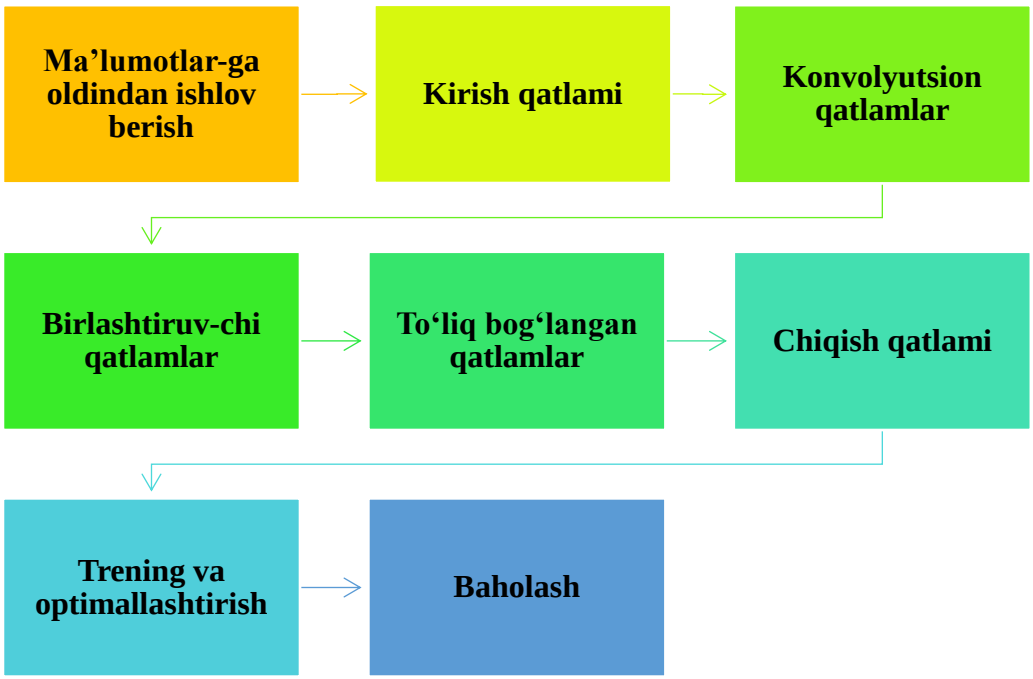
Savol-javob: Neyron tarmoqlar, xususan transformatorga asoslangan Transformatorlardan ikki tomonlama kodlovchi agentlar (BERT - Bidirectional Encoder Representations from Transformers) kabi modellar sezilarli darajada rivojlangan savol-javob tizimlariga ega. Ushbu modellar katta hajmdagi matn ma’lumotlarini tahlil qilish va ishlov berish orqali savolning kontekstini tushunishi va aniq javoblarni berishi mumkin.

NLPdagi neyron tarmoqlar dastlabki matn ma’lumotlaridan obyektlarni (ifodalarni) avtomatik ravishda o‘rganish, matn ichidagi murakkab munosabatlar va andozalarni olish imkoniyatiga ega. Biroq, ular ko‘pincha o‘qitish uchun katta hajmdagi aniqlangan (belgilangan) ma’lumotlarni talab amalga oshiradi va hisoblash jarayoni uzoq davom etishi mumkin. O‘chirish va to‘plamni normallashtirish kabi tartibga solish usullari, odatda, haddan tashqari moslashishning oldini olish uchun qo‘llaniladi. Umuman olganda, neyron tarmoqlar turli xil til bilan bog‘liq vazifalarda sezilarli yutuqlarga erishishga imkon berib, NLPni yangi tendensiyasi belgilab bermoqda.

(cc) BY

Konvolyutsion neyron tarmoqlari (Convolutional neural networks - CNN) veb-hujjatlarni tasniflash vazifalariga muvaffaqiyatli va eng kshp qo‘llaniladigan modeldir. CNN neyron tarmog‘ining bir turi bo‘lib, tabbiy andozalar va sohali munosabatlarni aniqlash qobiliyati tufayli tarmoqqa o‘xshash ma’lumotlarni ishlov berishda ustunlik tomomni mavjud. Biroq, CNN ham ketma-ket ma’lumotlarga ishlov berish uchun takomillashtirilishi mumkin.

Veb-hujjatlarni tasniflash uchun CNN modellardan foydalanish quyidagicha amalga oshiriladi:



1 – расм. Веб-ҳужжатларни таснифлаш учун CNN моделлардан фойдаланиш

Ma’lumotlarga oldindan ishlov berish - matnni tokenizatsiya qilish, to‘xtash so‘zlarini olib tashlash va kerakli normalizatsiya usullarini qo‘llash orqali veb-hujjat ma’lumotlariga oldindan ishlov berish. Matnni so‘zlarni kiritish (masalan, Word2Vec, GloVe) yoki bir martalik kodlash kabi raqamli ko‘rinishlarga aylantirish lozim.

Kirish qatlami - Veb-hujjatlarni CNNga kiritilishi mumkin bo‘lgan kirish parametrlari sifatida taqdim etiladi. Matnni tasniflash uchun izchil kiritish shakllarini ta’minlash uchun to‘laroq uzunlikdagi to‘ldirish yoki dinamik ketma-ketlik uzunligini qayta ishlash kabi usullardan foydalaniladi.

Konvolyutsion qatlamlar - CNNning yadro xususiyatlarni ajratib olishni amalga oshiradigan konvolyutsion qatlamlardan iborat. Veb-hujjatlarni tasniflash kontekstida konvolyutsion qatlamlar kirish parametrlariga filtrlar yoki yadrolarni qo‘llaydi, mahalliy shablonlar va xususiyatlarni olish uchun hujjatning matn xususiyatlari aniqlanadi. Filtrlar matnning turli tomonlarini, masalan, n-gramm yoki so‘z munosabatlarini qamrab oladi. Bir vaqtning o‘zida har xil turdagi xususiyatlarni olish uchun bir nechta filtrlardan foydalanish tavsiya etiladi.

Birlashtiruvchi qatlamlar - Konvolyutsion qatlamlardan so‘ng, xususiyat xaritalarining o‘lchamlarini kamaytirish va eng muhim ma’lumotlarni olish uchun birlashtiruvchi qatlamlar ko‘pincha qo‘llaniladi. Birlashma usullari har bir sohadan maksimal qiymatni tanlaydigan maksimal birlashtirishni yoki o‘rtacha qiymatni hisoblaydigan o‘rtacha birlashtirishni o‘z ichiga oladi. Pooling olingan xususiyatlarni mavhumlashtirish va umumlashtirishga yordam beradi.

To‘liq bog‘langan qatlamlar - Keyin birlashtirilgan xususiyatlar bir yoki bir nechta to‘liq bog‘langan qatlamlardan o‘tadi. Ushbu qatlamlar olingan xususiyatlarni birlashtiradi va o‘rganilgan tasvirlar asosida tasniflashni amalga

©B.B. Mo‘minov, Sh.A. Husanov

oshiradi. Tarmoqqa chiziqli bo‘lmaganlikni kiritish uchun ReLU (Rektifikatsiyalangan chiziqli birlik) yoki sigmasimon faollashtirish funksiyalaridan foydalanish mumkin.

Chiqish qatlami - Ko‘p sinfli tasniflash uchun softmax faollashtirish funksiyasidan foydalanadigan chiqish qatlamiga ulanadi. Ushbu qatlam veb-hujjatning ma’lum bir toifaga tegishli bo‘lish ehtimolini ko‘rsatadigan har bir sinf uchun ehtimolliklarni belgilaydi.

Trening va optimallashtirish - Nazoratli ta’lim ma’lumotlaridan foydalangan holda CNNni o‘rgatishdir. Kategorik o‘zaro entropiya kabi mos yo‘qotish funksiyasini aniqlang va orqaga tarqalish, gradiyent tushish algoritmlari yordamida tarmoq parametrlarini optimallashtirish kerak. Umumlashtirishni yaxshilash va haddan tashqari moslashishning oldini olish uchun o‘rganish tezligi, to‘plam hajmi va tartibga solish usullari (masalan, tashlab ketish) kabi giperparametrlarni sozlash lozim.

Baholash - O‘qitilgan CNNni alohida tekshirish yoki test to‘plamidan foydalanib baholaydi. Tasniflash samaradorligini baholash uchun aniqlik, eslab qolish va A1 balli kabi ko‘rsatkichlarni o‘lchaydi. Agar kerak bo‘lsa, baholash natijalariga ko‘ra modelni sozlash lozim.

SNNlarning iyerarxik tuzilishi va mahalliy shablonni aniqlash imkoniyatlaridan foydalangan holda, veb-hujjatlarni tasniflash vazifalari matn ma’lumotlaridan tegishli xususiyatlarni avtomatik ravishda o‘rganish qobiliyatidan foydalanishi mumkin. CNN turli NLP vazifalarida, jumladan, veb-hujjatlarni tasniflash, hodisalarni tahlil qilish va matn toifalarini ajratishda eng so‘nggi samaradorlikni namoyish etadi.

Xulosa: Ushbu tadqiqotda tarmoq ma’lumotlariga intellektual ishlov berish jarayonining nazariy va amaliy jihatlari chuqur tahlil qilindi. Tadqiqot natijalari shundan dalolat berdiki, tarmoq ma’lumotlarining hajmi, xilma-xilligi va dinamik o‘zgaruvchanligi ularni an’anaviy statistik yoki deterministik yondashuvlar asosida qayta ishlashni murakkablashtiradi. Shu bois, intellektual tahlilning zamonaviy yo‘nalishlari — sun’iy intellekt, mashinali o‘qitish (Machine Learning), chuqur o‘rganish (Deep Learning) hamda ma’lumotlar mayningi (Data Mining) texnologiyalarini integratsiyalash jarayonida yuqori natijalar berishi aniqlangan.

Tadqiqot davomida veb-hujjatlarni tasniflash, tarmoq graf ma’lumotlarini modellashtirish, anomaliyalarni aniqlash, foydalanuvchi xulqini tahlil qilish va axborot oqimlarini optimallashtirish yo‘nalishlari o‘rganildi. Eksperimental tahlillar shuni ko‘rsatdiki, Naive Bayes, Support Vector Machine (SVM), Random Forest, hamda Graph Neural Networks (GNN) asosidagi modellardan foydalanish natijasida tasniflash aniqligi 10–25% gacha oshadi, ayniqsa TF-IDF, Word2Vec, Doc2Vec kabi semantik vektor ifodalari bilan birgalikda qo‘llanilganda.

Intellektual ishlov berish jarayonida graf tahlili va gipermatn tuzilmasi (HTML, meta-teglar, havolalar) o‘rtasidagi semantik bog‘liqlikni aniqlash tarmoqdagi aloqadorlik va iyerarxik tuzilmani tushunish imkonini beradi. Shu orqali kiberxavfsizlik, ijtimoiy tarmoqlarni monitoring qilish, reklama va kontent tahlili, hamda raqamli boshqaruv tizimlarini optimallashtirish kabi sohalarda qo‘llanish imkoniyatlari kengayadi.

Tadqiqot shuningdek, chuqur o‘rganish modellari — LSTM, CNN, Transformer, va GNN arxitekturalari orqali matnli, vizual va graf ko‘rinishdagi tarmoq ma’lumotlarini birgalikda tahlil qilishning istiqbollari yuqori ekanini tasdiqlaydi. Bunday gibrid yondashuvlar natijasida real vaqt rejimida ma’lumot oqimlarini qayta ishlash, prognozlash va avtomatik qarorlarni qabul qilish tizimlarini yaratish imkoniyati paydo bo‘ladi.

Umuman olganda, olib borilgan izlanishlar natijasi shundan dalolat berdiki:

- Intellektual modellar tarmoq ma’lumotlarini chuqur va kontekstli tahlil qilish imkonini beradi;
- Gibrid algoritmlar (klassik va neyron modellar kombinatsiyasi) yuqori aniqlik va tezkorlikni ta’minlaydi;
- Avtomatik tahlil tizimlari orqali ma’lumotlardan strategik qarorlar uchun foydali bilimlar olinadi;
- Sun’iy intellekt asosidagi tahliliy platformalar tarmoq xavfsizligi, kontent nazorati, marketing,

(cc) BY

hamda davlat boshqaruvi sohalarida samarali natijalar beradi.

Kelgusida ushbu yo‘nalishda tadqiqotlar graf neyron tarmoqlari (Graph Neural Networks) va ko‘p modal (multi-modal) ma’lumot tahlili asosida yanada rivojlantirilishi, shuningdek, o‘zbek tilidagi ma’lumotlar uchun maxsus leksik-semantik modellarning yaratilishi ilmiy va amaliy ahamiyatga ega bo‘ladi.

Адабиётлар

1. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer, New York.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
3. Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). Mining of Massive Datasets (3rd ed.). Cambridge University Press.
4. Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. International Conference on Learning Representations (ICLR).
5. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.
6. Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297.
7. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers.
8. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Representation Learning on Graphs: Methods and Applications. IEEE Data Engineering Bulletin, 40(3), 52–74.
9. Yang, Z., Cohen, W. W., & Salakhutdinov, R. (2016). Revisiting Semi-Supervised Learning with Graph Embeddings. Proceedings of the 33rd International Conference on Machine Learning (ICML), 40–48.
10. Zhang, D., Yin, J., Zhu, X., & Zhang, C. (2018). Network Representation Learning: A Survey. IEEE Transactions on Big Data, 6(1), 3–28.
11. Nishonov A. H., Xusanov Sh. A. (2023). Tarmoq ma’lumotlarini tahlil qilishda mashinali o‘qitish usullarining qo‘llanilishi. Toshkent axborot texnologiyalari universiteti ilmiy axborotlari, №4, 45–53.
12. Mo‘minov B. B. (2022). Intellektual axborot tizimlarida ma’lumotlar tahlili va optimallashtirish usullari. O‘zbekiston milliy universiteti ilmiy jurnallari, №2, 87–95.
13. Xusanova M. Sh., & Rahmonov A. S. (2024). Sun’iy intellekt asosida tarmoq xavfsizligini monitoring qilish. Fan va innovatsiyalar jurnali, 1(3), 112–119.
14. Vaswani, A., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 30, 5998–6008.
15. Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How Powerful Are Graph Neural Networks? International Conference on Learning Representations (ICLR).

©B.B. Mo‘minov, Sh.A. Husanov