SUN’IY YARATILGAN MATNLARNI ANIQLASHDA INVARIANTLAR VA DASTURIY FILTRLASH JARAYONI

Muallif: Fozilov Murodjon Marifjon o‘g‘li
Tashkilot: Andijon davlat universiteti

Annotatsiya. Maqolada sun’iy yaratilgan matnlarni aniqlash algoritmini dasturiy filtrlash jarayoniga tatbiq etish masalasi ko‘rib chiqiladi. Yondashuv matnni dastlabki tayyorlash, lingvistik-statistik xususiyatlarni hisoblash, xususiyatlar vektorini sinf invariantlari bilan taqqoslash va qaror ishonchliligini baholash bosqichlariga ajratiladi. Manbada keltirilgan tajriba jadvalidagi I va II tur xatoliklar ko‘rsatkichlari tahlil qilinib, sinonimlash orqali yaratilgan matnlar uchun lug‘at tanlovi xatolik darajasiga ta’sir qilishi ko‘rsatiladi. Natijalar dasturiy filtrlashda faqat yakuniy yorliqni emas, balki qarorning ishonchlilik darajasi va xato xavfini ham hisobga olish kerakligini asoslaydi.

Kalit so‘zlar: sun’iy matnni aniqlash, invariantlar, tasniflash, dasturiy filtrlash, I tur xatolik, II tur xatolik, qaror ishonchliligi.

Kirish

Sun’iy matnni aniqlash amaliy tizimda faqat algoritm ishlab chiqish bilan tugamaydi. Algoritmni foydalanuvchi kiritgan matnga qo‘llash, xususiyatlarni hisoblash, natijani izohlash va ehtimoliy xatolarni nazorat qilish uchun ketma-ket dasturiy jarayon zarur.

Maqolaning asosiy masalasi — manbada taklif etilgan invariantlarga asoslangan aniqlash yondashuvini filtrlash jarayoni sifatida tavsiflash hamda manbada keltirilgan xatolik ko‘rsatkichlari orqali uning cheklovlarini muhokama qilish. Bu maqola lingvistik belgilarni batafsil solishtirishdan ko‘ra, tasniflash qarori va xatolarni boshqarishga e’tibor qaratadi.

Usulning dasturiy jarayondagi o‘rni

Manbadagi tavsifga muvofiq, tizimga kiruvchi matn dastlab tahlilga tayyorlanadi, so‘ng matnning tanlangan sifat va miqdoriy xususiyatlari hisoblanadi. Hisoblangan qiymatlar tegishli matn sinflari uchun shakllantirilgan invariantlar bilan solishtiriladi. Masofa yoki yaqinlik qiymatlari asosida matnning tabiiy yoki ma’lum usulda sun’iy yaratilgan sinfga mansubligi haqida xulosa chiqariladi.

Qaror qabul qilishda eng yaqin invariant bilan moslik va tegishlilik o‘lchovi muhim. Manbada qaror ishonchliligi R_A qiymati [0;1] oralig‘ida talqin qilinadi: qiymat kattalashgan sari matn xususiyatlari tanlangan invariantga yaqinroq bo‘ladi. Bu ko‘rsatkichni mutlaq ehtimollik deb talqin qilishdan oldin uni mustaqil tanlanmada kalibrlash zarur.

Dasturiy filtrlash nuqtayi nazaridan jarayonni to‘rt bosqichga ajratish mumkin: (1) kiruvchi matnni olish va me’yorlashtirish; (2) belgilangan xususiyatlarni hisoblash; (3) vektorni sinf invariantlari bilan taqqoslash; (4) tasnif natijasi, yaqinlik ko‘rsatkichi va noaniq holatlarni qaytarish. Ushbu bosqichlar manbadagi funksional tavsifning umumlashtirilgan ifodasidir.

Tajriba natijalari va xatoliklar tahlili

Manbada berilgan jadvalda sinonimlash yordamida yaratilgan matnlar uchun ikki xil lug‘at holati, shuningdek Markov zanjiri algoritmi bilan yaratilgan matnlar bo‘yicha I va II tur xatoliklar ko‘rsatilgan. I tur xatolik tabiiy matnning sun’iy deb noto‘g‘ri tasniflanishini, II tur xatolik esa sun’iy matnning tabiiy deb qabul qilinishini bildiradi.

Jadvaldagi qiymatlar quyidagicha: sinonimlash matni (havaskor lug‘at, o‘rtacha noyoblik 36%) uchun I tur xatolik 3,2%, II tur xatolik 6,1%; sinonimlash matni (ilova lug‘ati, o‘rtacha noyoblik 69%) uchun mos ravishda 3,2% va 2,8%; Markov zanjiri algoritmi yaratgan matnlar uchun 3,2% va 2,6%. Bu sonlar manba dissertatsiyasida keltirilgan tajriba natijalaridir.

Qiyoslashdan ko‘rinadiki, jadvaldagi uch holatda I tur xatolik 3,2% darajasida bir xil berilgan, II tur xatolik esa 6,1% dan 2,8% va 2,6% gacha farqlanadi. Sinonimlashda lug‘at tarkibi va uning noyoblik ko‘rsatkichi sun’iy matnni aniqlash natijasiga ta’sir qilishi mumkin. Shu sababli algoritmni faqat bitta lug‘at yoki bitta generator asosida baholash yetarli emas.

Xatoliklarning amaliy ahamiyati tizim ishlatiladigan vazifaga bog‘liq. Masalan, tabiiy matnni sun’iy deb belgilash noto‘g‘ri shubha uyg‘otishi mumkin; sun’iy matnni tabiiy deb qabul qilish esa filtrlash maqsadini susaytiradi. Shu bois dasturiy vosita natijani yakuniy va mutlaq hukm sifatida emas, balki ishonchlilik ko‘rsatkichi hamda zarur holatda qo‘shimcha tekshiruv talab qiladigan signal sifatida taqdim etishi maqsadga muvofiq.

Manba matnida TextOrigin tizimi va moderator konsoli tasvirlangan. Biroq taqdim etilgan materialda foydalanuvchilar soni, ishlov tezligi yoki server yuklamasi bo‘yicha raqamli sinov ko‘rsatkichlari yetarlicha berilmagan. Shu sababli ushbu maqolada bunday ko‘rsatkichlar to‘qib chiqarilmadi; dasturiy samaradorlikni to‘liq baholash uchun alohida yuklama sinovlari va tashqi test tanlanmasi talab qilinadi.

Xulosa

Invariantlarga asoslangan matnni aniqlash jarayoni xususiyatlarni hisoblash, sinf invariantlari bilan taqqoslash va qaror ishonchliligini baholashni birlashtiradi. Manbada keltirilgan tajriba jadvali II tur xatolik sinonimlashda qo‘llanilgan lug‘atga qarab sezilarli o‘zgarishini ko‘rsatadi; Markov zanjiri holatida esa II tur xatolik 2,6% deb berilgan.

Dasturiy filtrlash tizimida xatoliklarning ikki turini alohida kuzatish, qaror bilan birga ishonchlilik ko‘rsatkichini qaytarish va noaniq holatlarni inson tekshiruviga yo‘naltirish zarur. Keyingi bosqichda sinov tanlanmasini kengaytirish, matn janrlari va tillar bo‘yicha alohida baholash, shuningdek, tizimning tezligi va yuklama ostidagi barqarorligini real o‘lchash tavsiya etiladi.

Foydalanilgan adabiyotlar

1. Khmelev D. V. Распознавание автора текста с использованием цепей А. А. Маркова // Вестник Московского государственного университета. Серия 9: Филология. — 2000. — № 2. — С. 115–126.
2. Родненко В. Поиск нечетких дубликатов. Алгоритм шинглов для веб-документов. — 2009. — Elektron resurs.
3. Valgina N. S. Теория текста. — Москва: Логос, 2003. — 191 с.
4. Sun’iy yaratilgan matnlarni aniqlash usuli va dasturiy ta’minoti: foydalanuvchi taqdim etgan dissertatsiya materiali. — 2026. — 3-bob, dasturiy filtrlash va sinov natijalari.

Jadval 1. Manbada keltirilgan tasniflash xatoliklari

Matn yaratish holati	I tur xatolik	II tur xatolik
Sinonimlash, havaskor lug‘at (36% noyoblik)	3,2%	6,1%
Sinonimlash, ilova lug‘ati (69% noyoblik)	3,2%	2,8%
Markov zanjiri algoritmi	3,2%	2,6%