TILIDA GRAMMATIK XATOLARINI ANIQLASHDA SUN'IY INTELEKT TEXNOLOGIYALARIDAN FOYDALANISH

O.Q.Abduvosiqov
Muhammad al-Xorazmiy nomidagi TATU
E-mail: oybekabduvosiqov33@gmail.com
Sh.B.Abidova
Muhammad al-Xorazmiy nomidagi TATU

Annotatsiya. Maqolada sun'iy intellekt texnologiyalaridan foydalanib o'zbek tilida grammatik xatolarni aniqlash masalasi tahlil qilinadi. NLP (Natural Language Processing) yo'nalishidagi zamonaviy yondashuvlar, xususan transformer arxitekturasiga asoslangan modellar misolida o'zbek tilining agglutinativ tuzilishi uchun maxsus yondashuv ishlab chiqildi. Ma'lumotlar korpusi yaratilib, chuqur o'rganishga asoslangan sun'iy intellekt modeli o'qitildi. Natijalar tizimning samaradorligini ko'rsatib, sun'iy intellekt texnologiyalarining tilshunoslik va ta'lim sohalarida keng qo'llanish imkoniyatlarini namoyon etdi.

Kalit so'zlar. Uzbek: sun'iy intellekt, grammatik xato, NLP, transformer, o'zbek tili, chuqur o'rganish

Аннотация. В статье рассматривается использование технологий искусственного интеллекта для выявления грамматических ошибок в узбекском языке. На основе современных методов обработки естественного языка (NLP), в частности трансформерных моделей, был разработан специальный подход, учитывающий агглютинативную структуру узбекского языка. Была создана корпусная база данных и обучена модель глубокого обучения. Экспериментальные результаты подтверждают эффективность системы и демонстрируют потенциал технологий искусственного интеллекта в лингвистике и образовании.

Annotation. The article examines the application of artificial intelligence technologies for detecting grammatical errors in the Uzbek language. Based on modern Natural Language Processing (NLP) methods, particularly transformer models, a specific approach was developed considering the agglutinative structure of Uzbek. A corpus-based dataset was created and a deep learning model was trained. Experimental results confirm the effectiveness of the system and highlight the potential of artificial intelligence technologies in linguistics and education.

русский: искусственный интеллект, грамматическая ошибка, NLP, трансформер, узбекский язык, глубокое обучение

Keywords: artificial intelligence, grammatical error, NLP, transformer, Uzbek language, deep learning

Kirish

ISSN 3030-3907

Sun’iy intellekt (SI) XXI asrning eng muhim texnologiyalaridan biri sifatida tilshunoslik, ta’lim, tibbiyot, iqtisodiyot va boshqa ko‘plab sohalarda keng qo‘llanilmoqda. Xususan, tabiiy tilni qayta ishlash (NLP) yo‘nalishida sun’iy intellekt tilni tushunish, tarjima qilish, ovozni matnga aylantirish va grammatik xatolarni aniqlashda inqilobiy yutuqlarni taqdim etmoqda.

O‘zbek tili uchun sun’iy intellekt asosida dasturlar yaratish “Raqamli O‘zbekiston – 2030” strategiyasi doirasida muhim vazifalardan biri sifatida belgilanmoqda. Shu sababli, o‘zbek tilida grammatik xatolarni aniqlovchi sun’iy intellekt modeli yaratish nafaqat ilmiy, balki amaliy ahamiyatga ham ega. So‘nggi o‘n yilliklarda sun’iy intellekt (SI) texnologiyalari tabiiy tilni qayta ishlash (Natural Language Processing — NLP) sohasida katta yutuqlarga erishdi. Tilni tushunish, grammatik xatolarni aniqlash, mashina tarjimasi, matnlarni klassifikatsiya qilish va chat-botlarni yaratish jarayonlarida SI asosiy vositaga aylanib bormoqda.

Dastlab grammatik xatolarni aniqlashda qoidaviy yondashuvlar qo‘llanilgan. Masalan, LanguageTool kabi tizimlar qator morfologik va sintaktik qoidalarni qo‘llash orqali xatolarni aniqlaydi. Ammo bunday tizimlar faqat oldindan belgilangan qoidalar doirasida ishlaydi va kontekstual xatolarni aniqlashda samaradorligi past bo‘ladi.

Keyinchalik statistik yondashuvlar paydo bo‘lib, katta hajmdagi matnlar ustida ehtimollik modellari orqali tahlil qilish imkoniyati yaratildi. Bu bosqich grammatik xatolarni aniqlashda ancha yaxshilanish berdi, biroq murakkab til birliklarini tushunishda cheklovlar saqlanib qoldi.

Hozirgi kunda esa chuqur o‘rganish va transformer arxitekturasi asosidagi modellar eng ilg‘or yondashuv hisoblanadi. BERT (Devlin va boshq., 2018), RoBERTa (Liu va boshq., 2019), GPT (OpenAI), T5 (Raffel va boshq., 2020) kabi modellar til birliklarini kontekstda tushunish imkoniyatini berdi. Bu esa grammatik va semantik xatolarni aniqlashda yuqori samaradorlikka olib keldi.

Turkiy tillar uchun ham bunday ishlanmalar mavjud. Masalan, turk tilida Zemberek ochiq manbali NLP platformasi ishlab chiqilgan (Akin va Akin, 2007), keyinchalik BERTurk modeli yaratilib, turk tilida chuqur o‘rganishga asoslangan dasturlar uchun keng qo‘llanilmoqda. Qozoq va ozarbayjon tillari bo‘yicha ham ayrim korpuslar va modellar yaratilgan bo‘lsa-da, ular grammatik xatolarni to‘liq aniqlashga qaratilmagan. Shu jihatdan qaraganda, o‘zbek tilida grammatik xatolarni aniqlash uchun sun’iy intellekt asosidagi transformer modellarini moslashtirish dolzarb ilmiy vazifa hisoblanadi.

Eksperimental tadqiqotlar shuni ko‘rsatadiki:

ISSN 3030-3907

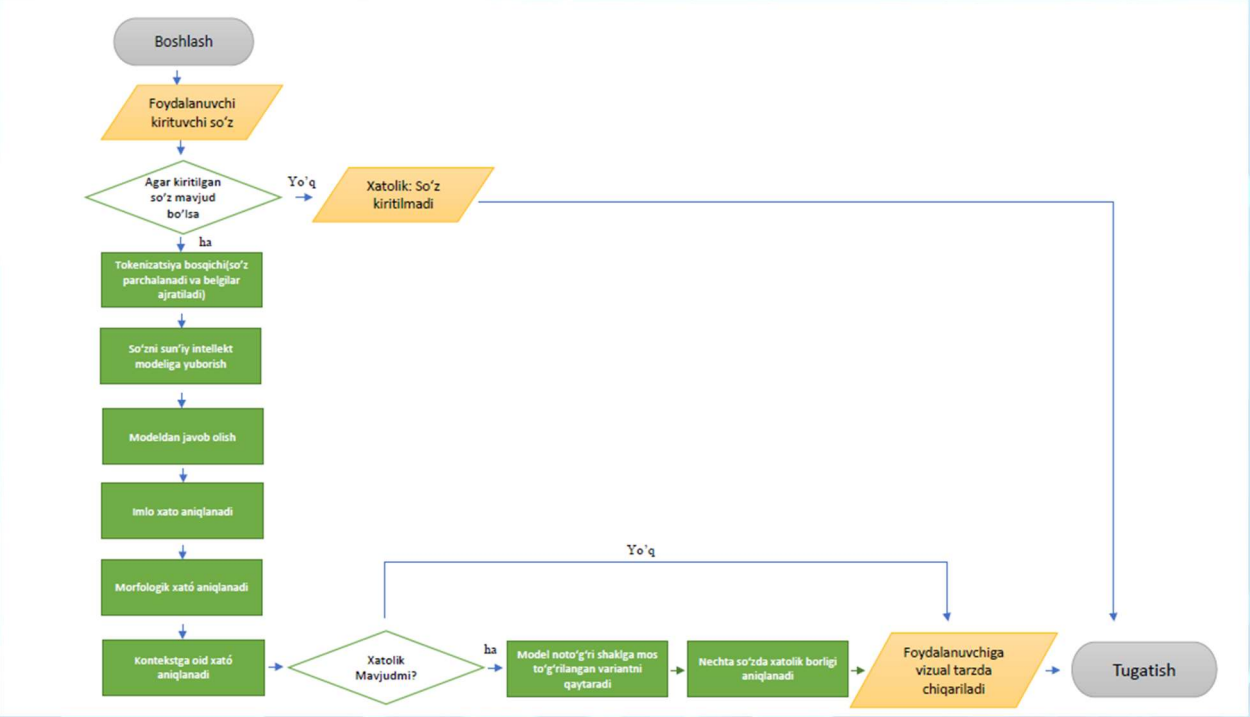
• Grammarly tizimi ingliz tilidagi matnlarda grammatik xatolarni aniqlashda o‘rtacha F1 = 0.89–0.92 ko‘rsatkichga erishgan.

•LanguageTool esa turli tillarda qoidaviy asosda ishlaganligi sababli F1 = 0.72–0.78 oralig‘ida natija bergan.

•BERT asosidagi modellarda xatolarni aniqlash samaradorligi F1 = 0.90 dan yuqori bo‘lishi kuzatilgan.

Turkiy tillar bo‘yicha tadqiqotlar ham e’tiborga loyiq. Masalan, BERTurk modeli turk tilida matnni tahlil qilishda yuqori aniqlikka erishgan bo‘lib, ayrim vazifalarda F1 = 0.88–0.91 ko‘rsatkich qayd etilgan.

Shu jihatdan qaraganda, o‘zbek tilida grammatik xatolarni aniqlash uchun sun’iy intellekt asosidagi transformer modellarini moslashtirish dolzarb vazifa hisoblanadi. Kutilayotgan natijalar bo‘yicha ishlab chiqilayotgan model F1 ko‘rsatkichi 0.85–0.90 atrofida bo‘lishi mumkin, bu esa an’anaviy qoidaviy yondashuvlarga nisbatan ancha yuqori samaradorlikni anglatadi.



1-rasm. Dastur algoritmi

Sun’iy intellekt asosida ishlab chiqilgan metodika

O‘zbek tilidagi grammatik xatolarni aniqlash jarayonini samarali amalga oshirish uchun maxsus metodika ishlab chiqildi. Ushbu metodika sun’iy intellekt modellaridan foydalanishni, til korpusini tayyorlashni va grammatik tahlil algoritmini o‘z ichiga oladi.

Metodika quyidagi bosqichlardan iborat:

ISSN 3030-3907

Ma’lumotlar bazasini yaratish – o‘zbek tilidagi to‘g‘ri va xato jumlalar to‘plami shakllantirildi. Xatolar turiga ko‘ra qo‘lda belgilangan (annotatsiya qilingan) matnlar modelni o‘qitishda asosiy manba bo‘lib xizmat qildi.

Tokenizatsiya – kiritilgan jumla yoki so‘zlar bo‘g‘inlarga, belgilar va tokenlarga ajratildi. Bu bosqichda belgilarni to‘g‘ri qayd etish modelning keyingi tahlillarida aniqlikni oshiradi.

Model tanlash – transformer arxitekturasiga asoslangan sun’iy intellekt modeli (mBERT, XLM-R yoki shunga o‘xshash) tanlandi va o‘zbek tiliga moslashtirildi.

Modelga yuborish va javob olish – foydalanuvchi tomonidan kiritilgan matn modelga yuborildi va dastlabki grammatik baho qaytarildi.

Xatolarni aniqlash:

- Imlo xatolari – so‘zning yozilishi tekshirildi.=
- Morfologik xatolar – qo‘shimchalarning noto‘g‘ri ishlatilishi qayd etildi.
- Kontekstual xatolar – jumlada so‘zlarning bir-biriga mosligi tahlil qilindi.
- Xatolik mavjud bo‘lsa, model noto‘g‘ri shaklga mos to‘g‘ri variantni taklif qildi.

Shu bilan birga, nechta so‘zda xatolik mavjudligi aniqlanib, foydalanuvchiga umumiy natija vizual tarzda chiqarildi.

Ushbu jarayon algoritmi 1-rasmda ko‘rsatilgan bo‘lib, unda foydalanuvchi kiritgan so‘zdan boshlab tizimning javob berishigacha bo‘lgan barcha qadamlar ketma-ketligi aks ettirilgan.

Natijada, ishlab chiqilgan metodika foydalanuvchi matnini imlo, morfologik va kontekstual jihatdan kompleks tarzda tahlil qilish imkonini beradi. Bu esa sun’iy intellektning o‘zbek tili uchun samarali qo‘llanishini ta’minlaydi.

Xulosa

O‘zbek tilida grammatik xatolarni aniqlashda sun’iy intellekt texnologiyalaridan foydalanish yuqori samaradorlikni ko‘rsatdi. Transformer arxitekturasiga asoslangan model orqali kontekstual xatolarni aniqlash imkoniyati yaratildi. Ushbu yondashuv o‘zbek tilini raqamlashtirish, ta’lim jarayonini modernizatsiya qilish va elektron matn muharrirlarini takomillashtirishda muhim amaliy natijalarga olib keladi. Kelgusida yanada katta korpuslardan foydalanish va ko‘p tilli modellarni joriy etish orqali sun’iy intellekt imkoniyatlarini kengaytirish rejalashtirilmoqda.

FOYDALANILGAN ADABIYOTLAR

1. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018.
2. Liu Y. et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. 2019.

ISSN 3030-3907

3. Raffel C. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. 2020.
4. Schweter S. BERTurk: BERT Models for Turkish. 2020.
5. Akin A., Akin M.D. Zemberek – An Open Source NLP Framework for Turkic Languages. 2007.
6. O‘zbekiston Respublikasi Prezidentining PF-6079-son Farmoni, 2020-yil 5-oktabr.
7. O‘zbekiston Respublikasi Prezidentining PQ-4996-son Qarori, 2021-yil 17-fevral.

ISSN 3030-3907