

AXBOROT TEXNOLOGIYALARI YORDAMIDA TABIIY TILNI QAYTA ISHLASHNING ZAMONAVIY YONDASHUVLARI

A.Sattarov

asattorov79@gmail.com

O'zbekiston davlat san'at va madaniyat institutining Farg'ona mintaqaviy filiali

Annotatsiya: Ushbu maqolada tabiiy tilni qayta ishlash (Natural Language Processing, NLP) bo'yicha zamonaviy yondashuvlar, ayniqsa, chuqur o'rganish (deep learning) modellarining NLP sohasidagi o'rni tahlil qilinadi. Tadqiqotda BERT, GPT va FastText kabi ilg'or NLP modellarining ishlash tamoyillari va ularning o'zbek tiliga tatbiq etish imkoniyatlari o'rganilgan. Adabiyot tahlili, eksperimental tahlil va amaliy tadqiqot metodlari asosida NLP sohasidagi asosiy yutuqlar hamda o'zbek tilida NLP rivojlanishining muammolari tahlil qilinadi. Natijalarga ko'ra, o'zbek tilidagi NLP rivojlanishini tezlashtirish uchun annotatsiyalangan korpuslarni yaratish, transformer modellarini moslashtirish va xalqaro hamkorlikni kuchaytirish zarurligi aniqlangan.

Kalit so'zlar: Tabiiy tilni qayta ishlash, NLP, chuqur o'rganish, sun'iy neyron tarmoqlar, BERT, GPT, FastText, o'zbek tili, til texnologiyalari, ma'lumotlar tahlili

MODERN APPROACHES TO NATURAL LANGUAGE PROCESSING USING INFORMATION TECHNOLOGIES

A.Sattarov

asattorov79@gmail.com

Fergana regional branch of Uzbekistan State Institute of Arts and Culture

Abstract: This article analyzes modern approaches to Natural Language Processing (NLP), with a particular focus on the role of deep learning models in this field. The study examines the working principles of advanced NLP models such as BERT, GPT, and FastText, as well as their applicability to the Uzbek language. Using literature review, experimental analysis, and applied research methods, the study explores key advancements in NLP and the challenges of developing NLP models for the Uzbek language. The results indicate that accelerating the development of Uzbek-language NLP requires the creation of annotated corpora, adaptation of transformer models, and strengthening international collaboration.

Keywords: Natural Language Processing, NLP, deep learning, artificial neural networks, BERT, GPT, FastText, Uzbek language, language technologies, data analysis

1. Kirish (Introduction)

Tabiiy tilni qayta ishlash (Natural Language Processing, NLP) - sun'iy intellekt va kompyuter fanining asosiy yo'nalishlaridan biri bo'lib, inson tilini tushunish, tahlil qilish va generatsiya qilishga qaratilgan. NLP texnologiyalari hozirgi vaqtda avtomatik tarjima, chat-botlar, matnni saralash, nutqni matnga aylantirish kabi jarayonlarda keng qo'llanilmoqda.

Zamonaviy NLP yondashuvlari sun'iy neyron tarmoqlar va chuqur o'rganish (deep learning) modellariga asoslangan bo'lib, katta hajmdagi ma'lumotlarni qayta ishlash imkonini beradi. Ushbu maqolada NLPning zamonaviy yondashuvlari, ularning ishlash tamoyillari hamda amaliy qo'llanilish sohalari tahlil qilinadi. Bundan tashqari, o'zbek tilidagi NLP modellarining rivojlanishi va duch kelayotgan muammolar muhokama qilinadi.

2. Metodologiya (Methods)

Ushbu tadqiqot tabiiy tilni qayta ishlash (NLP) sohasidagi zamonaviy yondashuvlar, modellarning ishlash tamoyillari hamda o'zbek tiliga tatbiq etish imkoniyatlarini o'rganishga qaratilgan. Tadqiqot quyidagi metodologik yondashuvlarga asoslanadi:

Adabiyot tahlili (Literature Review)

Adabiyot tahlili NLP bo'yicha mavjud ilmiy maqolalar, kitoblar va texnik hujjatlarni o'rganish orqali amalga oshirildi. Bu jarayonda quyidagi manbalarga asoslanildi:

- Ilmiy maqolalar - NLP bo'yicha Scopus, IEEE Xplore, Google Scholar va ResearchGate kabi platformalarda chop etilgan maqolalar tahlil qilindi.

- Dasturiy hujjatlar va texnik maqolalar - OpenAI, Google AI va Facebook AI Research (FAIR) tomonidan ishlab chiqilgan GPT, BERT, T5 kabi modellar haqidagi rasmiy texnik hujjatlar o'rganildi.

- Mahalliy tadqiqotlar - O'zbekiston va Markaziy Osiyo doirasidagi NLP tadqiqotlari, shu jumladan, o'zbek tilida NLP rivojlanishiga oid ilmiy maqolalar tahlil qilindi.

Adabiyot tahlili natijasida tabiiy tilni qayta ishlashning zamonaviy yondashuvlari, ularning afzalliklari va kamchiliklari haqida umumiy tushuncha shakllantirildi.

Eksperimental tahlil (Experimental Analysis)

NLP sohasida keng qo'llanilayotgan neyron tarmoqlarga asoslangan modellar tahlil qilindi. Bu jarayonda quyidagi modellar ko'rib chiqildi:

- BERT (Bidirectional Encoder Representations from Transformers) - Ikki tomonlama kodlovchi transformerlarga asoslangan model bo'lib, kontekstni chuqur tushunish xususiyatiga ega.

- GPT (Generative Pre-trained Transformer) - Matn generatsiyasi va tabiiy tilni tushunish uchun ishlatiladigan model.

- FastText - Soʻzlarni vektor koʻrinishida tasvirlash va semantik tahlil qilish uchun Facebook AI tomonidan ishlab chiqilgan model.

Ushbu modellar NLP muammolarini hal qilishda qanday yondashuvlardan foydalanishi, ularning ishlash tamoyillari va oʻzbek tiliga mosligi boʻyicha tahlil qilindi.

Eksperimental tahlilda quyidagi bosqichlar bajarildi:

1. Modellarning ishlash printsiplarini tushunish - Har bir modelning algoritmik asoslari va arxitekturasi oʻrganildi.

2. Maʼlumotlar tahlili - NLP modellarini oʻqitish uchun ishlatiladigan maʼlumotlar toʻplamlari oʻrganildi.

3. Natijalarni solishtirish - BERT, GPT va FastText modellari asosida olingan natijalar taqqoslanib, ularning afzallik va kamchiliklari aniqlandi.

Amaliy tadqiqot (Applied Research)

Oʻzbek tilida tabiiy tilni qayta ishlash boʻyicha mavjud muammolar va istiqbollar tahlil qilindi. Ushbu tahlil quyidagi bosqichlarda amalga oshirildi:

- Maʼlumot yetishmovchiligi - Annotatsiyalangan korpuslarning kamligi oʻzbek tilida NLP rivojlanishini cheklayotgan asosiy omil sifatida oʻrganildi.

- Leksik va grammatik murakkablik - Oʻzbek tilining qoʻshimchalarga boyligi va soʻz shakllanishining oʻziga xosligi NLP modellariga qanday taʼsir koʻrsatishi tahlil qilindi.

- Dialektlar va kontekst masalasi - Oʻzbek tilidagi turli lahjalarning NLP modellariga qanday taʼsir koʻrsatishi oʻrganildi.

Shuningdek, oʻzbek tilida NLP rivojlanishini tezlashtirish uchun quyidagi strategik yoʻnalishlar tavsiya etildi:

1. Katta hajmdagi oʻzbek tilidagi korpuslarni yaratish - Annotatsiyalangan va turli sohalarni qamrab olgan maʼlumotlar bazalarini shakllantirish.

2. Transformer modellarini oʻzbek tiliga moslashtirish - Koʻp tilli modellar asosida maxsus oʻzbek NLP modellarini yaratish.

3. Xalqaro hamkorlikni kengaytirish - NLP boʻyicha xalqaro loyihalar va grant dasturlariga oʻzbek tadqiqotchilarini jalb qilish.

Ushbu metodologik yondashuvlar orqali NLP sohasidagi ilgʻor yutuqlar, mavjud muammolar va istiqbolli yoʻnalishlar oʻrganildi hamda oʻzbek tilida NLP rivojlanishining amaliy jihatlari tahlil qilindi.

3. Natijalar (Results)

Zamonaviy NLP modellari

Soʻnggi yillarda NLP boʻyicha quyidagi asosiy modellar ishlab chiqilgan:

- BERT (Bidirectional Encoder Representations from Transformers) - Google tomonidan ishlab chiqilgan boʻlib, matn kontekstini yaxshiroq tushunishga xizmat qiladi.

- GPT (Generative Pre-trained Transformer) - OpenAI tomonidan ishlab chiqilgan, matn generatsiyasi va dialogni tushinishda yuqori natijalar beradi.

- FastText - Facebook AI tomonidan ishlab chiqilgan va o'zbek tili kabi kam o'rganilgan tillar uchun mos keluvchi model.

O'zbek tilida NLP rivojlanishi

Tahlillar shuni ko'rsatmoqdaki, o'zbek tili uchun NLP modellarini rivojlantirish quyidagi muammolar bilan bog'liq:

- Annotatsiyalangan ma'lumotlar yetishmovchiligi
- Grammatik murakkablik
- Dialektlarning ko'pligi

Shuningdek, o'zbek tilida mBERT, XLM-R va GPT-3 kabi modellar sinovdan o'tkazilmoqda, ammo ularning natijalari hali to'liq o'rganilmagan.

NLPning amaliy qo'llanilishi

NLP texnologiyalari quyidagi sohalarda samarali qo'llanilmoqda:

- Chatbot va virtual yordamchilar - Google Assistant, Siri, Alexa kabi tizimlar NLP asosida ishlaydi.

- Kiberxavfsizlik - NLP phishing va zararli xabarlarini aniqlashda qo'llaniladi.
- Ta'lim - Grammarly, Turnitin kabi platformalar matnlarni avtomatik baholaydi.

4. Muhokama (Discussion)

Zamonaviy NLP texnologiyalari sun'iy intellekt rivojlanishining muhim yo'nalishlaridan biri hisoblanadi. Tadqiqot natijalariga ko'ra, chuqur o'rganish modellarining NLPdagi roli ortib bormoqda. Biroq, o'zbek tilida NLP texnologiyalarini rivojlantirish uchun quyidagi tavsiyalar beriladi:

- Katta hajmdagi o'zbek tilidagi matnlar bazasini yaratish
- Transformer modellarini o'zbek tiliga moslashtirish
- O'zbek NLP loyihalarini xalqaro miqyosga olib chiqish

Ushbu yo'nalishlarda ish olib borilsa, o'zbek tilidagi NLP texnologiyalari xalqaro darajaga chiqishi mumkin.

5. Xulosa (Conclusion)

Tabiiy tilni qayta ishlash texnologiyalari zamonaviy axborot tizimlarining ajralmas qismiga aylangan bo'lib, ularning rivojlanishi sun'iy neyron tarmoqlarga asoslangan modellar orqali tezlashmoqda. Zamonaviy NLP modellarining samaradorligi yuqori bo'lsa-da, o'zbek tilida NLP rivojlanishining sustligi muammo bo'lib qolmoqda. Kelgusida o'zbek tilidagi NLP texnologiyalarini rivojlantirish uchun ko'proq tadqiqotlar va investitsiyalar talab qilinadi.

Foydalanilgan adabiyotlar

1. Abdurashidov, I. (2020). "O'zbek tilida tabiiy tilni qayta ishlash: muammolar va imkoniyatlar." O'zbekistan Axborot Texnologiyalari Jurnal.

2. Tursunov, D. (2021). "O'zbek tilidagi NLP tizimlarining rivojlanish yo'nalishlari." Toshkent Axborot Texnologiyalari Universiteti Ilmiy Jurnal.
3. Karimov, A. (2022). "Sun'iy intellekt va tabiiy tilni qayta ishlash texnologiyalari." Fan va Texnologiya Nashriyoti, Toshkent.
4. Nazarov, B. (2023). "Ko'p tilli NLP va o'zbek tiliga tatbiqi." O'zbekiston Ilmiy Texnik Jurnal.
5. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). "Attention is All You Need." Advances in Neural Information Processing Systems (NeurIPS).
6. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of NAACL-HLT.
7. Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). "Improving Language Understanding by Generative Pre-Training." OpenAI Research.
8. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). "Distributed Representations of Words and Phrases and Their Compositionality." Advances in Neural Information Processing Systems.
9. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). "Enriching Word Vectors with Subword Information." Transactions of the Association for Computational Linguistics.