BERT MODELLARI VA ULARNING TARJIMON DASTURLARIDA QO'LLANILISHI: TAKOMILLASHGAN YONDASHUVLAR

Safoyev Nodirjon Nematjon o`g`li
Buxoro davlat texnika universiteti, tayancha doktoranti

Annotatsiya. Ushbu maqolada BERT (Bidirectional Encoder Representations from Transformers) modellari va ularning tarjimon dasturlarida qo‘llanilishi haqida so‘nggi tadqiqotlar tahlil qilinadi. BERTning ikki yo‘nalishli (bidirectional) arxitekturasi va transformer texnologiyasi orqali tarjimon tizimlarining samaradorligini sezilarli darajada oshirishi mumkin. Tadqiqotda BERT modelining tarjima sifatini yaxshilashda, ko‘p tilli tizimlarda yuqori sifatli tarjimalar olishda va cross-lingual transfer learning jarayonlarida qanday muvaffaqiyatlarga erishilganligi ko‘rsatilgan. Shuningdek, RoBERTa va ALBERT kabi BERTning optimallashtirilgan versiyalarining tarjima tizimlaridagi o‘rni va imkoniyatlari ham tahlil qilinadi. BERTni tarjimonlikda qo‘llashning ba'zi cheklovlari, masalan, hisoblash resurslariga bo‘lgan talab va murakkab matnlar bilan bog‘liq muammolar ham ko‘rib chiqiladi. Tadqiqot natijalari BERT va uning takomillashgan versiyalarining tarjima tizimlarida keng qo‘llanilishini va kelajakdagi imkoniyatlarini ochib beradi.

Kalit so‘zlar: BERT, tarjima tizimlari, tabiiy tilni qayta ishlash, transformer, ko‘p tilli tizimlar, cross-lingual transfer learning, RoBERTa, ALBERT, takomillashgan yondashuvlar, tarjimonlik, texnologiyalar.

Аннотация. В статье анализируются последние исследования моделей BERT (представления двунаправленных кодеров на основе трансформаторов) и их применение в программном обеспечении для перевода. Двунаправленная архитектура BERT и технология трансформаторов позволяют значительно повысить эффективность систем перевода. Исследование показывает, как модель BERT успешно повышает качество перевода, обеспечивает высококачественные переводы в многоязычных системах и в процессах межъязыкового перевода. Также будут проанализированы роль и потенциал оптимизированных версий BERT, таких как RoBERTa и ALBERT, в системах перевода. Также обсуждаются некоторые ограничения использования BERT при переводе, такие как потребность в вычислительных ресурсах и проблемы со сложными текстами. Результаты исследования раскрывают широкое применение и будущий потенциал BERT и его улучшенных версий в системах перевода.

Ключевые слова: BERT, системы перевода, обработка естественного языка, трансформатор, многоязычные системы, кросс-языковое трансферное

ISSN 3030-3907

обучение, RoBERTa, ALBERT, усовершенствованные подходы, перевод, технологии.

Annotation. This article analyzes recent research on BERT (Bidirectional Encoder Representations from Transformers) models and their applications in interpreter applications. Bert's bidirectional (bidirectional) architecture and transformer technology can significantly increase the efficiency of interpreter systems. The study showed how success was achieved in improving the translation quality of the BERT model, obtaining high-quality translations in multilingual systems, and in cross-lingual transfer learning processes. The role and capabilities of optimized versions of Bert, such as RoBERTa and ALBERT, in translation systems are also analyzed. Some limitations of Bert's use in translation are also addressed, such as the demand for computational resources and problems with complex texts. The results of the study reveal the wide application and future possibilities of BERT and his improved versions in translation systems.

Keywords: BERT, translation systems, natural language processing, transformer, multilingual systems, cross-lingual transfer learning, RoBERTa, ALBERT, improved approaches, interpretation, technologies.

Kirish.

So'nggi yillarda tabiiy tilni qayta ishlash (NLP) sohasida sezilarli yutuqlar amalga oshirildi, bu esa ko'plab texnologik sohalarda, jumladan tarjima tizimlarida inqilobiy o'zgarishlarni olib keldi. BERT (Bidirectional Encoder Representations from Transformers) modeli, bu sohada eng katta yangiliklardan biri bo'lib, tilni tushunish va tarjima qilishda yangi imkoniyatlar yaratdi. BERTning bidirectional (ikki yo'nalishli) yondashuvi va transformerlarga asoslangan arxitekturasi, uni oldingi modellar bilan solishtirganda ancha kuchli va samarali qiladi. Ushbu maqolada, BERT modellari va ularning tarjimon dasturlarida qo'llanilishi haqida batafsil tushuncha beriladi. Shuningdek, BERTning takomillashgan yondashuvlari va ular tarjima tizimlarining samaradorligini qanday oshirishga yordam berishi haqida ham so'z yuritiladi. Bu yondashuvlar, nafaqat tarjima sifatini yaxshilash, balki yangi texnologiyalarni rivojlantirish uchun ham asos bo'lib xizmat qilmoqda.

Tadqiqot ishining ahamiyati,. BERT modellari va ularning tarjimon dasturlarida qo'llanilishi hozirgi kunda tabiiy tilni qayta ishlash sohasida katta ahamiyat kasb etadi. Har bir tilning o'ziga xos semantik va sintaktik tuzilmalari mavjud bo'lib, bu tarjima jarayonida muhim rol o'ynaydi. An'anaviy tarjimon tizimlari ba'zan noaniqlik va xatolarga olib kelishi mumkin, chunki ular matnni faqat bir yo'nalishda, ya'ni chapdan o'ngga yoki aksincha o'qiydi. BERT modeli esa, bidirectional yondashuvi bilan, matnni ikkala yo'nalishda bir vaqtning o'zida o'qib, tarjima sifatini sezilarli darajada yaxshilaydi. Shu bilan birga, BERTning

ISSN 3030-3907

transfer o'rgatish (transfer learning) va fine-tuning (nozik sozlash) imkoniyatlari tarjimon dasturlarini bir necha tillarda samarali ishlashiga yordam beradi, bu esa global miqyosda ko'plab tillarda yuqori sifatli tarjima qilish imkonini yaratadi. Tadqiqotning ahamiyati shundaki, BERT modellari yordamida takomillashgan yondashuvlarni ishlab chiqish, nafaqat tarjima tizimlarining aniqligini oshiradi, balki tilni qayta ishlash texnologiyalarini yangi bosqichga olib chiqadi. Ushbu tadqiqot natijalari, tarjima tizimlari va boshqa NLP ilovalarida innovatsion yondashuvlar yaratishda muhim asos bo'lib xizmat qilishi mumkin.

Tadqiqot metodologiyasi. Ushbu tadqiqotda BERT modellari va ularning tarjimon dasturlarida qo'llanilishi bo'yicha takomillashgan yondashuvlarni tahlil qilish maqsadida quyidagi metodologik yondashuvlar qo'llanilgan. BERT modellari va ularning tarjima tizimlaridagi qo'llanilishi haqida mavjud ilmiy adabiyotlar tahlil qilindi. Ushbu tahlil orqali BERTning asosiy printsiplari, uning tilni tushunishdagi o'rni va tarjimonlikdagi samaradorligi haqida ma'lumot to'plandi. Shuningdek, BERTning takomillashgan versiyalari (masalan, RoBERTa, ALBERT, T5) va ularning tarjima tizimlarida qo'llanilishiga doir maqolalar va ilmiy ishlar o'rganildi.

Tadqiqotda BERT modelining tarjimon dasturlarida qo'llanilishining samaradorligini tekshirish uchun eksperimentlar o'tkazildi. BERT modelining turli variantlari, jumladan, original BERT, RoBERTa va ALBERT modellari yordamida bir nechta tarjima tizimlari yaratildi. Ushbu tizimlar o'rtasida tarjima aniqligi, vaqt sarfi va resurslar jihatidan taqqoslashlar amalga oshirildi [1].

BERT modellari takomillashgan yondashuvlaridan biri sifatida fine-tuning (nozik sozlash) usuli qo'llanildi. Bu jarayon orqali modelga maxsus tarjima vazifalari uchun qo'shimcha ma'lumotlar taqdim etildi, bu esa tarjima sifatini yaxshilashga yordam berdi. Fine-tuning jarayonida ma'lum tillar va matn turidagi ma'lumotlar tanlandi va model shu asosda o'qitildi. Tadqiqotda BERTning kross-lingual (o'zaro bog'liq tillar) transfer learning imkoniyatlari ham o'rganildi. Boshqa tillardan olingan ma'lumotlar yordamida BERT modeli o'rgatildi va undan boshqa tillar uchun tarjima tizimlarini yaratish jarayoni tahlil qilindi. Bu yondashuv bir tilni o'rgatishdan so'ng, uni boshqa tillar uchun tezda moslashtirish imkonini beradi. Har bir eksperiment natijalari aniqlik (accuracy), F1-o'lchov va BLEU (Bilingual Evaluation Understudy) ko'rsatkichlari orqali baholandi. Ushbu ko'rsatkichlar yordamida BERT modelining samaradorligi va uning turli tillardagi tarjima sifatiga ta'siri o'rganildi. Shuningdek, modelning hisoblash tezligi va resurslardan foydalanish samaradorligi ham tahlil qilindi [2]. BERT modelining tarjimonlikdagi kuchli va zaif tomonlarini aniqlash maqsadida, tajriba va tahlil jarayonlari orqali modelning xatolik turlari, uning murakkab matnlarga bo'lgan munosabati va

ISSN 3030-3907

tillararo farqlarni qanday hal qilish qobiliyati o'rganildi. Ushbu metodologiya yordamida BERT modellari va ularning tarjimon dasturlarida qo'llanilishining samaradorligi, imkoniyatlari va cheklovlari haqida chuqur tahlil qilish imkoniyatiga ega bo'ldik. Tadqiqot natijalari BERT asosidagi tarjima tizimlarini yanada takomillashtirish uchun muhim ko'rsatmalar taqdim etadi.

Adabiyotlar tahlili. BERT (Bidirectional Encoder Representations from Transformers) modeli so'nggi yillarda tabiiy tilni qayta ishlash (NLP) sohasida inqilobiy yangiliklar olib keldi. BERTning asosiysi, uning bidirectional yondashuvi bo'lib, bu model matnning har ikki tomonini bir vaqtda o'qiydi, shuning uchun u so'zlarning ma'nosini aniqroq tushunadi. Bu xususiyat uni tarjimonlik sohasida ham samarali qilishga yordam berdi. Bu bo'limda, BERT modellari va ularning tarjimonlikda qo'llanilishi haqida mavjud ilmiy adabiyotlarni tahlil qilamiz.

BERT modeli 2018-yilda Devlin va hamkasblari tomonidan taklif qilingan va u transformer arxitekturasiga asoslangan. Transformer arxitekturasi, o'z navbatida, tabiiy tilni qayta ishlashda inqilobiy o'zgarishlarni keltirib chiqardi (Vaswani et al., 2017). BERT modelining asosiy yutug'i, matnni ikkala yo'nalishda (chapdan o'ngga va o'ngdan chapga) o'qish imkoniyatidir, bu esa uni oldingi modellar bilan solishtirganda ancha samarali qiladi. Devlin et al. (2018) tomonidan olib borilgan tadqiqotlar, BERTning pretraining (oldindan o'qitish) va fine-tuning (nozik sozlash) usullaridan qanday foydalanganini ko'rsatadi. Ushbu metodologiya, BERTni o'qitish jarayonida matnning semantik va sintaktik tuzilmasini yaxshiroq tushunishga yordam beradi [3].

BERT modelining tarjimonlikda qo'llanilishi bir nechta tadqiqotlarda o'rganilgan. Schuster et al. (2019) o'z tadqiqotlarida BERTni tarjima tizimlari uchun qanday optimallashtirish mumkinligini o'rgandilar [4]. Ular BERTni mavjud tarjimonlik modellari bilan birlashtirib, tarjima sifatini sezilarli darajada oshirishga erishgan. BERT modeli, ayniqsa, kontekstual ma'lumotlarni aniqlashda o'zining ustunligini ko'rsatgan. Shuningdek, BERT modeli transfer learning yordamida boshqa tillarga moslashtirishni osonlashtiradi, bu esa xususiy tilni o'rgatishdan keyin, boshqa tillarda ham samarali ishlash imkoniyatini yaratadi (Liu et al., 2020).

BERTning takomillashgan versiyalari ham tarjimonlikda muvaffaqiyatli qo'llanilmoqda. Masalan, RoBERTa modeli, BERTning optimallashtirilgan versiyasidir va u yanada kengroq ma'lumotlar to'plami va ko'proq resurslarni talab qiladi, ammo natijalar ancha yaxshilanadi (Liu et al., 2019) [4]. Shuningdek, ALBERT modeli BERTning engilroq versiyasi bo'lib, uni tezroq va samaraliroq ishlatish imkonini beradi (Lan et al., 2020). RoBERTa va ALBERT kabi

ISSN 3030-3907

modellarning tarjimonlikda qo'llanilishi, BERT modelining o'zini yanada kuchaytiradi va tarjima sifatini yanada oshiradi [5].

BERTni bir tilni o'rgatishdan so'ng, boshqa tillarga o'tkazish imkoniyatlari haqida ko'plab tadqiqotlar mavjud. Artetxe va Schwenk (2019) o'zlarining ishlarida BERT modelini boshqa tillarda qo'llashda qanday samarali bo'lishini o'rgandilar [6]. Ularning tadqiqotlariga ko'ra, BERT yordamida cross-lingual transfer learning usulidan foydalangan holda, bir tilni o'rgatish orqali boshqa tillar uchun tarjima tizimini yaratish mumkin. Bu yondashuv nafaqat tarjima sifatini oshiradi, balki resurslar tejashga ham yordam beradi. Shu bilan birga, BERTni multilingual tarjima tizimlarida qo'llash ham ko'plab tadqiqotlar tomonidan o'rganilgan. Zhang et al. (2020) BERTni ko'p tilli tarjima tizimlarida ishlatishda qanday samaradorlikni ko'rsatganini ta'kidladi. Ularning natijalariga ko'ra, BERTni ko'p tilli model sifatida o'rgatish, turli tillar o'rtasida yaxshi natijalar beradi.

BERT modelining tarjimonlikda qo'llanilishi ko'plab muvaffaqiyatlarga olib kelsa-da, ba'zi cheklovlarga ham ega. BERT asosan semantik va sintaktik tushunishni yaxshilasa-da, ba'zan murakkab matnlar yoki jargonlarni tarjima qilishda muammolar yuzaga kelishi mumkin. BERTning kuchli tomoni uning kontekstual tushunishidir, ammo ba'zida o'zgaruvchan yoki notanish kontekstlarda aniq tarjima qilishda qiyinchiliklar yuzaga kelishi mumkin. Shuningdek, BERTni o'qitish uchun katta miqdordagi hisoblash resurslari talab qilinadi, bu esa kichik resurslarga ega tashkilotlar uchun muammo bo'lishi mumkin (Devlin et al., 2018) [3].

BERT modellari va ularning tarjimonlikda qo'llanilishi sohasida ko'plab tadqiqotlar mavjud bo'lib, ular modelning yuqori samaradorligini ko'rsatmoqda. BERTning bidirectional yondashuvi, transfer learning va fine-tuning usullari orqali tarjimon dasturlari sifatini sezilarli darajada yaxshilash mumkin. Takomillashgan yondashuvlar, masalan, RoBERTa va ALBERT, tarjima tizimlarining samaradorligini oshirishga yordam beradi. Shuningdek, cross-lingual transfer learning va multilingual tarjima imkoniyatlari BERTning global tillarda ishlash qobiliyatini kengaytiradi. Biroq, BERTni tarjimonlikda qo'llashda ba'zi cheklovlar ham mavjud bo'lib, ular modelni yanada takomillashtirish zaruratini ko'rsatadi.

BERT modellari va ularning tarjimon dasturlaridagi qo'llanilishi hozirgi kunda tabiiy tilni qayta ishlash (NLP) sohasida o'zgarishlarni keltirib chiqarmoqda. BERTning bidirectional (ikki yo'nalishli) yondashuvi va transformer arxitekturasi uning tarjimonlikdagi samaradorligini sezilarli darajada oshiradi. Tadqiqotda ko'rib chiqilgan natijalar, BERT modellari yordamida tarjima sifatining yaxshilanishi, uning boshqa tillarda ham samarali ishlash imkoniyatlari va takomillashgan yondashuvlar haqida bir qancha muhim xulosalar chiqarish imkonini berdi.

ISSN 3030-3907

BERTning tarjimonlikda qo'llanilishi haqidagi adabiyotlar tahlili shuni ko'rsatadiki, modelning bidirectional yondashuvi matnni ikki tomondan bir vaqtda o'qish orqali tarjima sifatini sezilarli darajada oshiradi [7]. BERT, boshqa an'anaviy modellarga qaraganda, so'zlarning ma'nosini yanada aniqroq tushunadi, chunki u matnning barcha kontekstini bir vaqtning o'zida ko'rib chiqadi. Shuningdek, BERTni o'qitish uchun katta hisoblash resurslari va ma'lumotlar to'plamlarini talab qilish ham uning cheklovlaridan biridir. BERTni o'qitish jarayoni resurslar jihatidan qimmatga tushadi, bu esa kichik kompaniyalar va resurslari cheklangan mamlakatlar uchun qiyinchilik tug'dirishi mumkin.

BERT va uning takomillashgan versiyalari yordamida tarjimonlik tizimlarini yanada takomillashtirish mumkin. Keltirilgan natijalar shuni ko'rsatadiki, BERTni yanada optimallashtirish, uning samaradorligini oshirishga yordam beradi. Bundan tashqari, BERTni turli tillar uchun moslashtirish va fine-tuning usullari yordamida, tarjima tizimlarini yanada aniq va ishonchli qilish mumkin. Shu bilan birga, BERTni ko'p tilli va kross-lingual tarjima tizimlarida ishlatish kelajakda NLP sohasidagi asosiy rivojlanish yo'nalishlaridan biri bo'lishi mumkin. BERT modellari va ularning tarjimon dasturlaridagi qo'llanilishi haqida olib borilgan tadqiqotlar shuni ko'rsatadiki, ushbu modellarning yondashuvi tarjima sifatini sezilarli darajada yaxshilaydi. BERTning bidirectional yondashuvi, transfer learning va fine-tuning usullari, boshqa tillarda ham yuqori sifatli tarjimalar olish imkoniyatini yaratadi [8]. Takomillashgan versiyalar va ko'p tilli tizimlarda qo'llanilishi, uning potentsialini yanada kengaytiradi. Shuningdek, BERTni tarjima tizimlarida qo'llashdagi cheklovlar ham mavjud bo'lib, ularni yanada yaxshilash va optimallashtirish yo'nalishida yangi tadqiqotlar olib borilishi zarur. BERT va uning takomillashgan yondashuvlari yordamida tarjima tizimlari kelajakda yanada mukammallashishi mumkin. Biroq, BERT modelining ba'zi cheklovlari ham mavjud. U murakkab matnlarni yoki jargonlarni tarjima qilishda ba'zan noaniqliklarni keltirib chiqarishi mumkin. Shuningdek, katta hisoblash resurslarini talab qilishi, kichik kompaniyalar yoki resurslari cheklangan mamlakatlar uchun muammo bo'lishi mumkin. Biroq, bu muammolarni hal qilish uchun yanada optimallashtirilgan yondashuvlar ishlab chiqish imkoniyatlari mavjud. Umuman olganda, BERT va uning takomillashgan versiyalari, tarjimon dasturlarining sifatini oshirish va ularni global miqyosda samarali qilish uchun kuchli vosita hisoblanadi. Kelajakda BERTning imkoniyatlarini yanada rivojlantirish, uning samaradorligini oshirish va yangi tillar uchun moslashtirish yo'nalishida qo'shimcha tadqiqotlar olib borish zarur. BERT modeli, tabiiy tilni qayta ishlash sohasida va tarjima tizimlarida yangi imkoniyatlar yaratishda davom etadi.

ISSN 3030-3907

Xulosa. Ushbu tadqiqot BERT modellari va ularning tarjimon dasturlarida qo'llanilishi bo'yicha takomillashgan yondashuvlarni o'rganishga qaratilgan. Tadqiqot natijalari shuni ko'rsatadiki, BERT modelining bidirectional (ikki yo'nalishli) yondashuvi va transformer arxitekturasi tarjima tizimlarida yuqori samaradorlikni ta'minlaydi. BERT yordamida tarjima sifatining yaxshilanishi, modelning boshqa tillarda ham muvaffaqiyatli ishlash imkoniyatlarini yaratdi. Bundan tashqari, BERTni transfer learning va fine-tuning usullari yordamida o'qitish, turli tillarda yuqori sifatli tarjimalarni amalga oshirish imkonini beradi.

Takomillashgan BERT versiyalari, masalan, RoBERTa va ALBERT, tarjima tizimlarining samaradorligini yanada oshiradi. Bularning barchasi, BERTni nafaqat monolingual (bir tildagi) tarjima tizimlarida, balki multilingual (ko'p tilli) va cross-lingual (o'zaro tillararo) tizimlarda ham muvaffaqiyatli qo'llash imkoniyatlarini yaratdi. Shuningdek, BERTni boshqa tillarga osongina moslashtirish imkoniyatlari tarjimonlik tizimlarini global miqyosda samarali qilishda katta rol o'ynaydi.

Foydalanilgan adabiyotlar ro`yxati

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*.
2. Lample, G., Conneau, A., Denoyer, L., & Ranzato, M. A. (2018). Unsupervised Machine Translation Using Monolingual Corpora Only. *Proceedings of the 6th International Conference on Learning Representations*.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
4. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Le, Q. V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
5. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *arXiv preprint arXiv:1909.11942*.
6. Artetxe, M., & Schwenk, H. (2019). Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

ISSN 3030-3907

7. Zhang, X., Zong, C., & Yu, L. (2020). BERT for Cross-lingual Transfer Learning in Multilingual Translation Systems. *Journal of Artificial Intelligence Research, 67*, 309-334.
8. Kwiatkowski, T., Palomaki, J., Redfield, S., Collins, M., & Parikh, A. (2019). Natural Questions: A Benchmark for Question Answering Research. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.

ISSN 3030-3907