

TABIY TILNI QAYTA ISHLASHDA BERT MODELINING AMALIY QO‘LLANILISHI

Rasulov Omadbek Rustam o‘g‘li

omadbekrasulov171@gmail.com

Tel: +998772860600

Farg‘ona Davlat

Texnika Universiteti talabasi

Annotatsiya: Ushbu maqolada tabiiy tilni qayta ishlash (NLP) jarayonida BERT (Bidirectional Encoder Representations from Transformers) modelining nazariy asoslari va amaliy qo‘llanilishi keng tahlil qilinadi. Maqolada modelning arxitekturasini, uning oldingi tahlil usullaridan farqi, mashinaviy o‘qitishdagi afzalliklari hamda BERT asosida yaratilgan amaliy tizimlar haqida batafsil ma’lumot beriladi. Shuningdek, maqolada O‘zbek tili uchun BERT modelini moslashtirish bo‘yicha ilmiy istiqbollari, ma’lumotlarga o‘plamlari bilan ishlashda yuzaga keladigan muammolar va ularni hal etish strategiyalari yoritilgan.

Аннотация: В данной статье широко анализируются теоретические основы и практическое применение модели BERT (Bidirectional Encoder Representations from Transformers) в процессе обработки естественного языка (NLP). В статье подробно рассматриваются архитектура модели, её отличие от предыдущих методов анализа, преимущества в машинном обучении, а также практические системы, созданные на основе BERT. Кроме того, освещаются научные перспективы адаптации модели BERT для узбекского языка, проблемы, возникающие при работе с наборами данных, и стратегии их решения.

Annotation: This article provides a comprehensive analysis of the theoretical foundations and practical applications of the BERT (Bidirectional Encoder Representations from Transformers) model in the field of Natural Language Processing (NLP). The paper examines the model’s architecture, its differences from previous analysis methods, advantages in machine learning, and practical systems developed based on BERT. In addition, the article discusses scientific prospects for adapting the BERT model to the Uzbek language, challenges encountered when working with datasets, and strategies for addressing these issues.

Kalit so‘zlar: BERT, NLP, sun’iy intellekt, transformer, tabiiy tilni qayta ishlash, semantik tahlil, mashinaviy o‘qitish, deep learning, contextual embedding, O‘zbek tili, AI model.

Tabiiy tilni qayta ishlash (Natural Language Processing — NLP) so‘nggi yillarda sun’iy intellektning eng tez rivojlanayotgan yo‘nalishlaridan biriga aylandi.

Bugungi kunda chatbotlar, avtomatik tarjimonlar, tovushni matnga aylantirish tizimlari, sentiment tahlili kabi amaliy sohalarda NLP algoritmlarining samaradorligi inson bilan mashina o'rtasidagi muloqot nitubdan o'zgartirmoqda.

Shunday yutuqlar orasida Google AI tomonidan 2018-yilda ishlab chiqilgan BERT modeli tabiiy tilni qayta ishlash sohasida inqilobiy yangilik bo'ldi. U tilning kontekstual (kontekstga bog'liq) ma'nosini ikki tomondan —chapdan va o'ngdan — tahlil qilish imkonini beruvchi birinchi yirik transformer arxitekturasiga asoslangan model sifatida tanilgan.

BERT modeli bugungi kunda Google Search, ChatGPT, HuggingFace, Amazon Comprehend va boshqa tizimlarning intellektual yadrosiga aylangan.

BERT modelining yaratilish tarixi va ilmiy asoslari

BERT (Bidirectional Encoder Representations from Transformers) modeli Devlin va hamkasblari (2018) tomonidan "Attention is All You Need" tamoyiliga asoslangan Transformer arxitekturasidan foydalangan holda ishlab chiqilgan.

BERT'ning asosiy g'oyasi — so'zlarning ma'nosini atrofidagi kontekst orqali to'liq anglash va uni matematik ifoda shaklida kodlashdir.

An'anaviy NLP modellari (masalan, Word2Vec, GloVe) har bir so'zni alohida vektor sifatida ko'rgan bo'lsa, BERT har bir so'zning ma'nosini gapdagi barcha so'zlar bilan o'zaro bog'liq holda o'rganadi. Bu esa tilni yanada chuqurroq semantik tahlil qilish imkonini beradi.

BERT modelining nazariy asosi:

Transformer Encoder: o'z ichiga ko'p qatlamli (multi-head) "attention" mexanizmini oladi.

Bidirectional Learning: har bir so'zning chap va o'ng kontekstini bir vaqtning o'zida tahlil qiladi.

Pre-training va Fine-tuning: model avval yirik matn korpusida o'qitiladi (pre-training), so'ngra ma'lum topshiriqlarga moslashtiriladi (fine-tuning).

BERT arxitekturasida: Transformer asosidagi yangilik

Transformer arxitekturasida — bu BERT modelining "yuragi"dir. U quyidagi asosiy komponentlardan tashkil topgan:

Self-Attention (o'ziga e'tibor mexanizmi) — har bir so'z boshqa so'zlar bilan o'zaro semantik aloqani hisoblaydi.

Multi-Head Attention — bir vaqtning o'zida bir nechta kontekst qatlamlarini o'rganadi.

Positional Encoding — matndagi so'zlarning joylashuvini eslab qoladi.

Feed-Forward Layers — ma'lumotni chuqur neyron tarmoq orqali qayta ishlaydi.

Bu yondashuv tufayli BERT nafaqat sintaktik, balki semantic bog‘liqliklarni ham chuqur o‘rganadi.

BERT modelining amaliy qo‘llanilishi

Bugungi kunda BERT modeli ko‘plab real tizimlarda ishlatiladi:

Google Search: so‘rovning haqiqiy ma‘nosini tahlil qilib, foydalanuvchining niyatiga mos natijani beradi.

Chatbotlar: foydalanuvchi savolini semantik jihatdan tushunadi va to‘g‘ri javob shakllantiradi.

Tibbiy NLP tizimlari: klinik hujjatlarni avtomatik tahlil qiladi.

Hujjatlarni tasniflash va sentiment tahlili: ijtimoiy tarmoqlardagi fikrlarni musbat yoki salbiy sifatida ajratadi.

Avtomatik tarjima tizimlari: matnning kontekstual tarjimasini amalga oshiradi.

Masalan, “BERT-base” modeli yordamida o‘zbek tilidagi tweetlarni ijobiy /salbiy deb tasniflovchi tizimlar ishlab chiqilmoqda.

O‘zbek tili uchun BERT modelini moslashtirish

O‘zbek tili agglutinatив (qo‘shimchalar orqali so‘z yasaluvchi) til bo‘lgani uchun, BERT modelini to‘g‘ridan-to‘g‘ri qo‘llashda bir qator muammolar yuzaga keladi. Shu sababli tadqiqotchilar “UzBERT”, “XLM- RoBERTa” va “mBERT” kabi ko‘p tilli versiyalardan foydalanmoqda. O‘zbek tili uchun BERT modelini tayyorlashda quyidagi bosqichlar amalga oshiriladi:

Korpus tayyorlash — 10–20 millionta so‘zli matn to‘plami yig‘iladi.

Tokenizatsiya — morfologik murakkablikni inobatga olgan holda so‘zlarni bo‘lish.

Modelni pre-train qilish — GPU yordamida minglab epoxalar davomida o‘qitish.

Fine-tuning — ma‘lum vazifalarga (sentiment tahlili, tarjima, savol-javob tizimi) moslashtirish.

Natijada “UzBERT” modeli O‘zbek tilidagi matnlarni kontekst asosidatan tahlil qilishda an‘anaviy modellarni 20–30% ga ortda qoldirgan.

BERT va boshqa NLP modellarining taqqoslanishi

Model	Arxitektura	Kontekst	Samaradorlik yo‘nalishi (%)
Word2Vec	Skip-gram CBOW	Bir yo‘nalishli	70
GloVe	Matritsa faktorizatsiyasi	Bir yo‘nalishli	75
ELMo	RNN asosida	Ikki yo‘nalishli	82

BERT	Transformer Encoder	To'liq yo'nalishli	ikki	93
------	------------------------	-----------------------	------	----

BERT nafaqat aniqroq natija beradi, balki u o'qitishdan so'ng turli topshiriqlarga moslashuvchan tarzda qayta ishlatilishi mumkin.

BERT modelining afzalliklari

Kontekstni chuqur anglash qobiliyati;

Turli NLP vazifalariga moslashuvchanlik;

Transfer learning imkoniyati;

Katta miqdordagi ma'lumotlar bilan ishlash;

Multitasking (bir nechta vazifani bir modelda bajarish).

Kamchiliklar va muammolar

Katta hisoblash quvvatini talab qiladi (GPU/TPU).

O'qitish jarayoni vaqt jihatdan uzoq.

Agglutativ tillarda (masalan, o'zbek) tokenizatsiya qiyin.

Model "katta ma'lumot"ga haddan tashqari bog'liq.

Shunga qaramay, bu muammolarni hal etish uchun "DistilBERT", "ALBERT" va "TinyBERT" kabi yengillashtirilgan versiyalar ishlab chiqilgan.

BERT asosidagi kelajak istiqbollari Kelajakda O'zbekiston ilmiy markazlarida BERT modeli asosida:

Avtomatik tarjima tizimlari,

Ta'lim platformalarida intellektual suhbat tizimlari,

O'zbek tilidagi sun'iy intellekt agentlari,

Hujjat tahlilchi modellar yaratilishi kutilmoqda.

Bu yo'nalishdagi ishlar "O'zAI", "AI Uzbekistan" va "Digital Education" kabi tashabbuslarda amalga oshirilmoqda.

Xulosa: BERT modeli tabiiy tilni qayta ishlash sohasida tub burilish yasadi. U kontekstual tahlilga asoslangan yangi paradigma — bidirectional understandingni yaratdi.

Bugungi kunda BERT va uning variantlari inson tilini chuqur tahlil qilish, avtomatik tushunish va sun'iy intellekt tizimlarida muloqot sifatini yaxshilashda asosiy vosita bo'lib xizmat qilmoqda.

O'zbek tili uchun BERT modelini moslashtirish — bu nafaqat ilmiy, balki milliy texnologik mustaqillik sari tashlangan muhim qadamlardan biridir.

FOYDALANILGAN ADABIYOTLAR:

1. Devlin, J. et al. (2018). BERT: Pre-training of Deep Bidirectional
1. Transformers for Language Understanding.
2. Vaswani, A. et al. (2017). Attention is All You Need. NeurIPS
3. Conference.

4. Wolf, T. et al. (2020). Transformers: State-of-the-Art NLP Models.
5. Rahmatov A. (2023). O‘zbek tilida NLP modellarining rivoji.
6. Toshkent: Innovatsiya nashriyoti.
7. 5. HuggingFace Research Center. (2024). UzBERT: Multilingual
8. Adaptation.

