IQRO JURNALI , IQRO JOURNAL, ИКРО ЖУРНАЛИ
ISSN:2181-4341| wordlyknowledge.uz/index.php/iqro
Volume 23. Issue 02. Sentabr ,2026

NEYRON TARMOQLAR ASOSIDA O'ZBEK TILIDAGI MATNLARNING SEMANTIK TAHLILI VA HISSIYOTLARNI AVTOMATIK ANIQLASH MODELLARINI YARATISH

Buriyeva Madina Aliyorovna
Webster universiteti
MA Tesol Magistratura talabasi

Mamatkulova Saidaxon Shukurjonovna
O'zbekiston Jahon tillari universiteti
Dotsent, filalogiya fanlari doktori

Annotatsiya

Mazkur maqolada o'zbek tilidagi matnlarni chuqur o'rganish (Deep Learning) va zamonaviy neyron tarmoqlar arxitekturasi asosida semantik tahlil qilish hamda hissiyotlarni (Sentiment Analysis) avtomatik aniqlash masalalari atroflicha o'rganilgan. O'zbek tilining aglutinativ xususiyatlari, morfo-sintaktik murakkabliklari va resurslarning cheklanganligi sharoitida matn vektorlash usullari (FastText, Uzbek-BERT) hamda LSTM, BiLSTM, Transformer kabi neyron tarmoq modellarining samaradorligi qiyosiy tahlil qilingan. Tadqiqot natijasida o'zbek tilidagi ijtimoiy tarmoq sharhlari va yangiliklar matnlarining hissiyotlarini aniqlovchi optimallashtirilgan model taklif etilib, uning aniqlik ko'rsatkichlari (***Accuracy***, ***Precision***, ***Recall***, ***F1-score***) bo'yicha eksperimental natijalar keltirilgan.

Kalit so'zlar: O'zbek tili, tabiiy tilni qayta ishlash (NLP), hissiyotlarni tahlil qilish (Sentiment Analysis), chuqur o'rganish, LSTM, BiLSTM, Uzbek-BERT, semantik tahlil, neyron tarmoqlar.

Kirish

Axborot texnologiyalarining jadal rivojlanishi va raqamli iqtisodiyotning shakllanishi davrida internet tarmog'ida o'zbek tilidagi kontent hajmi keskin ortib bormoqda. Ijtimoiy tarmoqlar, yangiliklar portallari, elektron tijorat platformalari va rasmiy veb-resurslarda har kuni millionlab matnli xabarlar yaratilmoqda. Ushbu ulkan matnli ma'lumotlar oqimini inson omilisiz tezkor tahlil qilish, foydalanuvchilarning mahsulotlar, xizmatlar yoki ijtimoiy-siyosiy jarayonlarga bo'lgan munosabatini (hissiyotini) aniqlash tabiiy tilni qayta ishlash (Natural Language Processing – NLP) fanining eng dolzarb vazifalaridan biridir.

O'zbek tili aglutinativ tillar guruhiga kirib, so'z yasalishi va turlanishining o'ta boy tizimiga ega ekanligi bilan ajralib turadi. Bitta o'zakdan affikslar zanjiri yordamida yuzlab yangi ma'nodagi so'zlarni hosil qilish mumkin. Bu holat an'anaviy statistik va qoidaga asoslangan (rule-based) matn tahlili usullarining samaradorligini sezilarli darajada pasaytiradi. Shu sababli, oxirgi yillarda sun'iy intellekt va neyron tarmoqlar, xususan, chuqur o'rganish (Deep Learning) modellariga asoslangan semantik tahlil usullarini o'zbek tiliga tatbiq etish zarurati yuzaga keldi.

Hissiyotlarni tahlil qilish (Sentiment Analysis yoki Opinion Mining) — matndagi muallifning hissiy bo'yog'ini (ijobiy, salbiy yoki neytral) avtomatik ravishda aniqlash jarayonidir. Ushbu maqolaning asosiy maqsadi — neyron tarmoqlar yordamida o'zbek tilidagi matnlarning semantik xususiyatlarini chuqur ochib berish va hissiyotlarni aniqlashning yuqori aniqlikdagi modellarini yaratish hamda ularning amaliy samaradorligini baholashdan iborat.

Adabiyotlar tahlili va muammoning o'rganilganlik darajasi

Tabiiy tilni qayta ishlash sohasida so'nggi yillarda ulkan muvaffaqiyatlarga erishildi. Ingliz va boshqa resurs boy tillar uchun Word2Vec (Mikolov va boshq., 2013), GloVe (Pennington va

boshq., 2014), FastText (Bojanowski va boshq., 2017) kabi so'z vektorlash modellari hamda Transformer arxitekturasiga asoslangan BERT (Devlin va boshq., 2018), GPT seriyali modellar standart vositaga aylandi.

O'zbek tilida NLP yo'nalishidagi tadqiqotlar nisbatan keyinroq boshlangan bo'lsada, oxirgi yillarda sezilarli siljishlar kuzatilmoqda. Jumladan, xorijiy va mahalliy tadqiqotchilar tomonidan o'zbek tili uchun korpuslar yaratish, mashinaviy tarjima va matnni tokenizatsiya qilish ishlari olib borildi. Xususan, o'zbek tili uchun FastText vektorlari va dastlabki korpuslar yaratildi. Shuningdek, huggingface platformasida o'zbek tili uchun maxsus pre-trained qilingan ***Uzbek-BERT*** modellari paydo bo'ldi.

Biroq, mavjud ilmiy ishlar tahlili shuni ko'rsatadiki, o'zbek tilining o'ziga xos morfo-sintaktik xususiyatlarini (jumladan, so'z tartibining erkinligi SOV sxemasi, ko'p qo'shimchalilik, omonimiya va sinonimiya) to'laqonli hisobga oluvchi, ijtimoiy tarmoqlar va norasmiy matnlardagi leksik xatoliklarni (slenglar, imlo xatolari, arab/kirill/lotin alifbolarining aralash qo'llanilishi) bartaraf etuvchi barqaror neyron tarmoq arxitekturalari yetarlicha ishlab chiqilmagan. Ushbu bo'shliqni to'ldirish mazkur tadqiqotning dolzarbligini belgilaydi.

Metodologiya va ma'lumotlar bazasini tayyorlash

Hissiyotlarni aniqlash modeli uchun dastlab sifatli matnli ma'lumotlar bazasini (korpus) yig'ish va tozalash talab etiladi. Ushbu tadqiqot doirasida Telegram kanallari, Facebook ijtimoiy tarmog'i guruhlari, Kun.uz, Daryo.uz kabi yirik axborot portallari sharhlari hamda elektron tijorat (Uzum Market va boshqalar) mahsulot sharhlaridan iborat umumiy hajmda ***50,000*** ta matndan iborat maxsus dataset shakllantirildi.

Ma'lumotlarni oldindan qayta ishlash (Preprocessing)

O'zbek tilidagi xom matnlar shovqindan tozalash bosqichidan o'tkazildi:

- Lotin va kirill alifbosidagi matnlarni yagona standartga (lotin alifbosiga) o'tkazish;
- URL havolalar, maxsus belgilar, emojilar va tinish belgilarini olib tashlash (emojilar ba'zi bosqichlarda hissiyot belgisi sifatida saqlab qolindi);
- Barcha harflarni kichik harflarga o'tkazish (***lowercase***);
- Stop-so'zlarni (bog'lamlar, ko'makchilar) olib tashlash yoki kontekstga qarab saqlash;
- Lematizatsiya va stemming jarayonlari: o'zbek tilining aglutinativ tabiatidan kelib chiqqan holda, PyUzb va maxsus qoidalar asosida so'zlarning o'zak va qo'shimchalarini ajratish.

Vektorlash usullari

Matnli ma'lumotlarni neyron tarmoqqa uzatish uchun ularni raqamli ko'rinishga (vektorlarga) o'tkazish zarur. Tadqiqotda quyidagi usullar sinovdan o'tkazildi:

1. **FastText:** O'zbek tili uchun o'rgatilgan, so'zlarning ichki morfologik tuzilishini (n-grammalarni) hisobga oluvchi vektorlar. Bu o'zbek tilidagi qo'shimchali so'zlar uchun juda muhimdir.
2. **Uzbek-BERT Tokenizer:** Transformer modellari uchun maxsus sub-word tokenizatsiya usuli.

Neyron tarmoq arxitekturalari

Matnlar semantikasini o'rganish va hissiyotlarni klassifikatsiya qilish uchun turli darajadagi neyron tarmoq modellarining arxitekturasi qurildi va o'qitildi:

LSTM va BiLSTM tarmoqlari

Ketma-ket ma'lumotlarni qayta ishlashda Rekurrent neyron tarmoqlarning (RNN) yo'qolib boruvchi gradient muammosini hal qilish uchun Long Short-Term Memory (LSTM) va uning

ikki tomonlama versiyasi bo'lgan Bi-directional LSTM (BiLSTM) tarmoqlaridan foydalanildi. BiLSTM tarmoqi matnni ham oldinga, ham orqaga qarab o'qib, har bir so'zning oldingi va keyingi kontekstini birdek hisobga oladi.

$$h = f(W h + W x + b) t h t-1 x t$$

Bu yerda x — t-qadamdagi kirish vektori, h — yashirin qatlam holati, W va b — o'rganiluvchi $t t$

parametrlar.

CNN-LSTM Gibrid modeli

Matndagi mahalliy xususiyatlarni (local features) tezkor ajratib olish uchun bir o'lchovli konvolyutsiya qatlami (1D CNN) va ketma-ketlikni bog'lash uchun LSTM qatlami birlashtirilib, gibrid arxitektura yaratildi.

Transformer (Uzbek-BERT) modeli

Diqqat mexanizmiga (Self-Attention) asoslangan Transformer arxitekturasi matndagi so'zlar orasidagi uzoq masofali bog'liqliklarni mukammal ochib beradi. Tadqiqotda ***multilingual-BERT*** va maxsus o'zbekcha pre-trained qilingan ***Uzbek-BERT*** modellari fine-tuning (nozik sozlash) qilindi.

Eksperimental natijalar va tahlil

Yaratilgan modellarni baholash uchun ma'lumotlar bazasi 80% o'qitish (training), 10% validatsiya (validation) va 10% test to'plamlariga bo'lindi. Modellarning samaradorligini baholashda quyidagi metrikalardan foydalanildi: Aniqilik (***Accuracy***), Aniqlik (***Precision***), To'liqlik (***Recall***) va F1-o'lchov (***F1-score***).

1-jadval. Turli neyron tarmod modellarining qiyosiy natijalari

Model arxitekturasi	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Baseline (TF-IDF + Logistic Regression)	78.4	77.9	78.1	78.0
FastText + LSTM	84.6	84.1	84.5	84.3
FastText + BiLSTM	87.2	86.8	87.0	86.9
1D CNN + BiLSTM Gibrid	88.5	88.1	88.3	88.2
Uzbek-BERT (Fine-tuned)	**93.7**	**93.4**	**93.6**	**93.5**

Jadval ma'lumotlaridan ko'rinib turibdiki, an'anaviy mashinaviy o'qitish usullari pastroq natija ko'rsatgan bo'lsa, chuqur o'rganish modellari, ayniqsa ***Uzbek-BERT*** modeli ***93.7%*** aniqlik ko'rsatkichiga erishdi. BiLSTM asosidagi tarmoqlar ham yetarlicha yuqori natija (***87.2%***) berib, hisoblash resurslari cheklangan tizimlar uchun maqbul yechim ekanligini ko'rsatdi.

Xulosa

Ushbu tadqiqot doirasida o'zbek tilidagi matnlarning semantik tahlili va hissiyotlarni avtomatik aniqlash bo'yicha mukammal ilmiy-uslubiy yondashuvlar ishlab chiqildi. O'zbek tilining morfo-sintaktik xususiyatlarini inobatga olgan holda maxsus matnlar korpusi shakllantirildi va oldindan qayta ishlash konveyeri yaratildi.

O'tkazilgan eksperimentlar natijasida quyidagi xulosalarga kelindi:

• O'zbek tili uchun FastText vektorlari so'zlarning ichki morfologiyasini to'laqonli ochib berib, rekurrent tarmoqlar samaradorligini oshiradi;

• BiLSTM va 1D CNN gibrid arxitekturasi matn semantikasini o'rganishda tezkor va yuqori aniqlik (***88.5%***) ta'minlaydi;

• Transformer arxitekturasiga asoslangan ***Uzbek-BERT*** modeli esa o'zbek tilidagi murakkab sintaktik tuzilmalar va kontekstlarni eng yuqori darajada (***93.7% F1-score***) tahlil qilish imkonini berdi.

Kelgusidagi tadqiqotlarda o'zbek tili uchun kattaroq hajmli pre-trained LLM (Large Language Models) modellarini yaratish va ularni real vaqt rejimida (real-time) ijtimoiy tarmoqlar monitoringida qo'llash rejalashtirilgan.

Foydalanilgan adabiyotlar

1. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 3111-3119.
2. 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*.
3. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information.
4. *Transactions of the Association for Computational Linguistics*, 5, 135-146.
5. Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.
6. Юсупов, Б. Р., & Рахимов, Ш. У. (2022). Ўзбек тилидаги матнларни табиий тилни айта ишлаш (NLP) усуллари ёрдамида талил илиш муаммолари. *Ахборот технологиялари ва телекоммуникациялар муаммолари*, 4(2), 45-52.
7. Karimova, M. A. (2023). Sentiment analysis in Turkic languages: Challenges and solutions for the Uzbek language. *Journal of Intelligent Systems and Data Science*, 11(1), 18-29.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I.
9. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5000-5010.
10. Gensig, R., & O'Brien, S. (2021). Deep Learning for Natural Language Processing: Text Classification and Sentiment Analysis. *Springer Nature*, 312 p.