



Sabohat ALLANAZAROVA,

Toshkent davlat o'zbek tili va adabiyoti universiteti tayanch doktoranti

E-mail: allanazarovasabohat@gmail.com

Dilrabo ELOVA,

Toshkent davlat o'zbek tili va adabiyoti universiteti katta o'qituvchisi, (PhD)

E-mail: elova@navoiy-uni.uz

ToshDO'TAU dotsenti, DSc Sh.Xamroyeva taqrizi asosida

EXPERIENCE OF SENTIMENT ANALYSIS IN NLP

Annotation

Sentiment analysis is an area of NLP that helps identify sentiments in text. In this analysis, the comments left by users are the object of research, and they are classified in terms of emotional expressiveness. Although resource-rich languages already have dozens of linguistic resources for analysis, there are no such resources for Uzbek. This is a hindrance to the development of many areas. Owing to the fact that sentiment analysis is used in the decision-making process. This article analyzes the research and modern approaches in this field.

Key words: Sentiment analysis, NLP, linguistic database, reviews, social network, public opinion.

ОПЫТ СЕНТИМЕНТ АНАЛИЗА В НЛП

Аннотация

Анализ тональности – это сфера НЛП, которая помогает выявить настроения в тексте. В данном анализе объектом исследования являются комментарии, оставленные пользователями, которые классифицируются по эмоциональной выразительности. Хотя богатые ресурсами языки уже имеют десятки лингвистических ресурсов для анализа, для узбекского языка таких ресурсов нет. Это препятствует развитию многих направлений. Это связано с тем, что анализ настроений используется в процессе принятия решений. В данной статье анализируются исследования и современные подходы в этой области.

Ключевые слова: Анализ тональности, НЛП, лингвистическая база, отзывы, социальная сеть, общественное мнение.

NLPDA SENTIMENT TAHLILI TAJRIBASI

Annotatsiya

Sentiment tahlili – matndagi his-tuyg'ularni aniqlashga yordam beradigan NLP sohasi hisoblanadi. Bu tahlilda foydalanuvchilar qoldirgan sharhlar tadqiqot obyekti bo'lib, ular emotsional-ekspressivlik jihatdan tasniflanadi. Resurslarga boy tillarda tahlil uchun allaqachon o'nlab lingvistik ta'minot manbalari mavjud bo'lsa-da, o'zbek tili uchun bunday resurslar mavjud emas. Bu esa ko'plab sohalar rivoji uchun to'sqinlik qiladi. Sababi sentiment tahlili qaror qabul qilish jarayonida qo'llaniladi. Ushbu maqolada mazkur sohada amalga oshirilgan tadqiqotlar va zamonaviy yondashuvlar tahlil qilinadi.

Kalit so'zlar: Sentiment tahlili, NLP, lingvistik ta'minot, sharhlar, ijtimoiy tarmoq, jamoatchilik fikri.

Kirish. So'nggi yillarda matnni hissiy tahlil qilishga ko'proq e'tibor qaratilib, asta-sekin axborot olish, tabiiy tilni qayta ishlash (NLP) va boshqa sohalarining tadqiqot nuqtasiga aylandi [1].

Ma'lumki, biror mahsulot yoki muammoga qiziqish dastlab uni o'rganishdan boshlanadi. Ijtimoiy tarmoqlarda esa turli mavzularda istalgancha ma'lumot olish imkoni mavjud. Axborotlar shunchalik ko'pki, bu ma'lumotlarni to'liq o'rganib chiqish uchun kishilarning bir kunlik vaqti kifoya qilmaydi. Sentiment tahlilining maqsadi ijtimoiy tarmoqlarning kuchidan turli xil sohalaridagi kayfiyat va jamoatchilik fikrini o'rganishdan iborat. So'nggi yillarda his-tuyg'ularni tahlil qilish bo'yicha katta ishlar amalga oshirildi [2].

Mavzuga oid adabiyotlarning tahlili. ST bo'yicha dunyo miqyosidagi olimlar turli tadqiqotlar bilan shug'ullanib kelganlar va hozirda ham shug'ullanmoqdalar. Ingliz tilidagi matnlarni fikrlar asosida tahlil qilish juda mashhur va yaxshi o'rganilgan mavzu hisoblanadi. Ingliz tili uchun ko'plab tasniflash manbalari mavjud. Masalan: SentiWordNet [3], SenticNet [4] va NRC Emotion Lecikon va boshqalar [5].

O'zbek tili uchun Sentiment tahlilida qo'llaniladigan lingvistik ta'minot ishlab chiqish bo'yicha yetarlicha ma'lumotlar mavjud emas va to'laqonli o'rganilmagan. Shunga qaramasdan o'zbekcha matnlar ustida bir nechta

tadqiqotlar o'tkazilgan. Xususan, S.Matlatipov va boshqalar tomonidan [6,7] o'zbek tilidagi matnlarning sentiment tahlili uchun annotatsiyali korpus qilingan. Ushbu qurilgan ikkala modelning natijasi mos ravishda 88.89% va 89.56% foiz aniqlik ko'rsatgan. I. Rabbimov boshchiligidagi tadqiqotchilar [8] guruhi filmlariga qoldirilgan sharhlarni emojiilar asosida o'rganishdi. Sababi ko'pchilik fikr bildirishdan ko'ra emojiini afzal bilishadi. Tadqiqot ishi uchun YouTube videoxostingdagi o'zbek filmlariga qoldirilgan sharhlar obyekt qilib olingan. Tadqiqot ishi 85.25% ga teng aniqlikda natija ko'rsatgan.

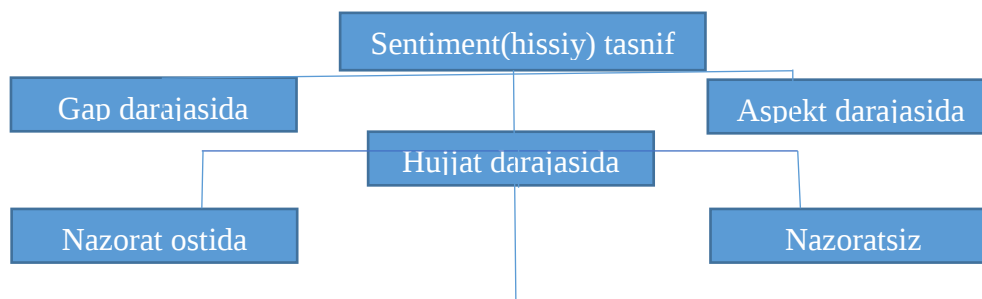
Turkiy tillar oilasiga mansub bo'lgan qo'shni Turk va Qozoq tillari bu sohada ancha taraqqiy etdi. Masalan: Turk tilidagi kino sharhlariga o'tkazilgan hissiy tahlil qilish tizimi 79.06% aniqlik ko'rsatdi [9]. Qozoq tilidagi terrorizm tahdidlari bo'lgan matnlar ustida ishlab, morfologik, sintaktik va hissiy jihatdan tahlil qilish usullari ishlab chiqilgan [10]. Ushbu ishda morfologik qoidalar va ontologik modelga asoslanib, qozoq tilidagi matnlarni hissiy tahlil qilish uchun emotsional so'zlar lug'ati yordamida qoidalarga asoslangan usul 83% aniqlik ko'rsatgan.

Xu va Liu [11] o'z tadqiqotlarini mijozlar sharhlarini tasniflashga qaratdilar, ya'ni ular hissiyotlarni o'z ichiga olgan mahsulot xususiyatlarini ajratib oldilar, so'ngra ushbu

xususiyatlar asosida fikrlarni tasnifladilar va natijada mahsulot sharhlarining qisqacha mazmuni yaratildi. Jumlagi ijobiy yoki salbiy belgi qo'yish uchun tadqiqotchilar birinchi navbatda har bir sharhdan qutbli so'z (sifat)larni ajratib olishgan va o'rtacha aniqlik 84% ni ko'rsatgan. Kim va Hovy [12] berilgan mavzudagi matn va uning ma'ruzachisining kayfiyatini o'rganib chiqdilar. Tadqiqot mualliflari bir nechta

tasniflagichlardan foydalanganlar. Birinchi klassifikator jumladagi har bir so'zning qutbligini olish uchun qo'llanildi. Ikkinchi tasniflagich fikr tashuvchisi tomonidan ifodalangan butun jumlaning qutbligini aniqladi.

Tadqiqot metodologiyasi. Sentiment tahlil qilish darajasiga ko'ra uch turga bo'linadi: hujjat darajasidagi, gap darajasidagi va aspekt darajasidagi tahlil [13].



1-rasm. Darajasiga ko'ra Sentiment tasnifi

Hozirgi kunda sentiment tahlilida qo'llanilishi mumkin bo'lgan uch xil: leksikaga, mashinaviy o'qitishga asoslangan va gibril yondashuvlar ko'rsatiladi.

Leksikaga asoslangan yondashuv. Leksikaga asoslangan yondashuvda atamalarining emotsional-eksperimental jihatdan bo'yoqdorligi bor yoki yo'qligiga qarab tuyg'u belgilanadi.

Jumlagi ijobiy yoki salbiy baho qo'yish uchun birinchi navbatda har bir sharhdan hissiy baho ifodalovchi so'zlar ajratib olinadi. Bunda sifatlar asos qilib olinadi. Tasnif sifatning sinonimlari bilan bir xil qutbligini va antonimlariga qarama-qarshi qutbligini ko'rsatadi. Emotsional-ekspressiv bo'yoqdor so'zlar WordNetdan ma'lum yo'nalishga ega bo'lgan sinonim va antonimlari qidiriladi. So'ngra sharhda berilgan qutbli so'zlarning yo'nalishi aniqlanadi. Tadqiqotchilarning ta'kidlashicha, ba'zi sinonimlar/antonimlar betaraf yoki hatto qarama-qarshi yo'nalishga ega, bu ularni ishlatish uchun nomaqbul hisoblanadi [14]. Biroq, taklif qilingan yondashuv bir qator kamchiliklarga ega. Asosiy muammo shundaki, sinonim va antonimlarni yig'ish ko'p vaqt talab etadi.

Park va Kim [15] tanlangan ma'lumotlar to'plamidan sinonim va antonimlarni olish uchun uch xil lug'atdan (an'anaviy ravishda faqat bittasi qo'llaniladi) foydalanadigan usulni ishlab chiqdilar. Keyinchalik, sharhlarni tasniflash uchun kengaytirilgan lug'at ishlatilgan. Biroq, taklif qilingan yondashuv bir qator kamchiliklarga ega. Asosiy muammo shundaki, sinonim va antonimlarni yig'ish ko'p vaqt talab etadi. Shuningdek, lug'atlar odatda rasmiy so'zlarni o'z ichiga oladi, aksincha sharhlar norasmiy so'zlarga to'la bo'ladi.

Rus olimi F.P. Filin «Rus adabiy tilining lug'ati» 7-jildini tahlil qilib, undagi 15530 so'zdan 3925 so'f stilistik pozitsiyada ekanligini aniqladi [16]. Demak, 75% so'z neytral bo'lib, 25% so'z stilistik bo'yoqli ekan.

1.2. Mashinaviy o'qitishga asoslangan yondashuv. ML alqoritmlari ko'pincha nazorat ostida yoki nazoratsiz ML texnikasi yordamida amalga oshiriladi. Matla va Badugu tomonidan [17] tvitlar CNN (55,50%) va Naïve Bayes (72,06%) alqoritmlarining ishlash samaradorligi solishtirilganda, NB samaradorligini ko'rsatgan. Mazkur ishning kamchiligi jumladagi kontekstual ma'lumotlarning kamligi hisoblanadi. Mozafari va Tahayori tomonidan 7 xil hissiyot toifalari SVM asosida tasniflangan [18]. Ishning kamchiligi kontekstual ma'lumotlarni tahlil qilishdagi zaiflik. Huang va boshqalar [19] matnga asoslangan his-tuyg'ularni aniqlash uchun HRLCE va BERT modelini quvnoq, g'azablanish, va qayg'uli hissiyot toifalarini tasniflashda foydalanishgan, ammo noto'g'ri tasniflarning ko'pligi ishning

sifatini kamaytirgan. Jayakrishnan va boshqalar [20] «Malayalam» romanidagi jummalarni quvnoq, qayg'uli, qo'rquv, g'azab va hayrat hissiyot toifalariga SVM yordamida tasniflashgan va mos ravishda 0,94, 0,92, 0,93, 0,90 va 0,90 aniqlikka erishganiga qaramasdan matnlardagi semantik ma'lumotlarga e'tibor bermaslik; jumladagi so'zlar qanday kontekstual ma'no ifodalayotganini hisobga olmaslik kabi kamchiliklari mavjud. Wikarsa va Thahir [21] NB usuli yordamida Twitter foydalanuvchilarining his-tuyg'ularini aniqlash uchun dastur ishlab chiqdi. Twitter API yordamida 105 ta sharh ajratib olingandan so'ng qayta ishlandi. Keyin, NB tasniflash usulidan foydalanib, matnlarni baxt, qayg'u, qo'rquv, g'azab, ajablanish va jirkanchlik deb tasniflandi va 83% aniqlikka erishgan.

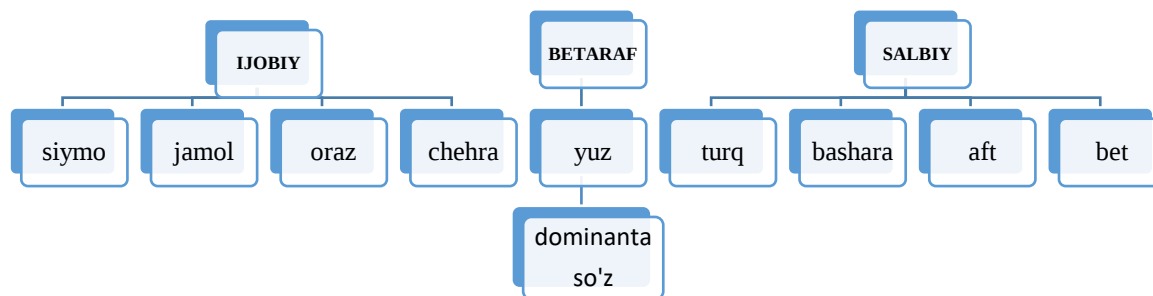
Gibril yondashuv. Yuqoridagi ikkala yondashuvdagi kamchiliklarni bartaraf etish uchun gibril yondashuvdan foydalaniladi. Bu kombinatsiyali yechim ikkala yondashuvdan natijalar aniqligi va tizimning kuchliligi bilan ustun turadi. Grover va Verma [22] panjabi matnlarida his-tuyg'ularni aniqlash uchun gibril yondashuvdan foydalanganlar. Ular ba'zi qoidalaridan foydalangan holda his-tuyg'ularning mavjudligini aniqladilar va ularni Ekmaning olti hissiyot toifalaridan biriga, ya'ni SVM va NB, ML tasniflagichlari yordamida baxtli, qayg'u, qo'rquv, g'azab, ajablanish va jirkanish kabi tasnifladilar. O'z ishlarida ular yanada mustahkam ML texnikasi ularning ish faoliyatini yaxshilashi mumkinligini ta'kidlashdi. Perikos va Hatzilygeroudis [23] matnlardagi hissiyotlarni avtomatik aniqlash uchun tasniflagichlar kombinatsiyasidan foydalanishdi. Natijalar shuni ko'rsatdiki, klassifikatorlar jamlanmasi yagona klassifikatorlar qo'llanilgan holatlarga qaraganda yaxshiroq ishlaydi.

Ghanbari-Adivi va Mosleh [24] ikkita ma'lumotlar to'plamidan (ya'ni, ISEAR va OANC) his-tuyg'ularni aniqlash uchun NLP vositalaridan TPE asosidagi gibril tasniflagichlardan foydalanilgan. Ularning natijalari Twitterdagi tartibsiz ma'lumotlar uchun 88,49% va ISEAR ma'lumotlar to'plami uchun 99,49% aniqlik ko'rsatdi, ammo taklif etilgan tizim juda murakkab.

Tahlil va natijalar. Garchi his-tuyg'ularni tahlil qilish juda faol tadqiqot bo'lgan bugungi kunda ushbu sohaning bir qator murakkab muammolari hali ham qolmoqda. Birinchidan, sarkazm muammosi; Sarkazm - nutq birliklarining murakkab shakli bo'lib, unda sharh muallifi yoki yozuvchi o'zi nazarda tutgan narsaning aksini aytadi yoki yozadi. Tuyg'ularni tahlil qilishda istehzoli jummalarni tahlil qilish juda qiyin, chunki fikr aniq va to'g'ridan-to'g'ri ifodalangan. Tilning leksikasi eng nozik ma'noga va stilistik bo'yoqdorlik bilan axborot berishda

juda katta imkoniyatlarga ega. Ular ichida, ayniqsa, sinonimlar, antonimlar stilistik imkoniyatga boy vositalar

hisoblanadi. Stilistik bo'yoqli so'zlar, ayniqsa, sinonimik qatorda o'z ifodasini topadi.



2-rasm. O'zbek tilidagi ko'p qatorli sinonimik sinset na'munasi.

Hozirgi o'zbek tilida nechta bo'yoqdor so'z bor? Bunga javob berish juda qiyin. Bu vazifani amalga oshirish uchun kuchli lingvistik bilim va lug'atlar ustida tadqiqot olib borish talab qilinadi. Dastlabki muammo o'zbek tili uchun sentiment tahlilida qo'llash mumkin bo'lgan lingvistik ta'min ishlab chiqilmagan. Bunday holatda yaxshi yechim qo'shni turk tilidagi Turkish WordNet bo'lishi mumkin [25]. Sababi turk va o'zbek tillari turkiy oilaga mansub va grammatik jihatdan agglyutinativ xususiyatga ega.

Ba'zi so'zlarda mintaqa va ma'naviy qadriyatlar jihatdan farq qilishi mumkin. Sababi, insonning madaniy

mansubligi ularning vaziyatlarga nisbatan ifodalangan his-tuyg'ulariga katta ta'sir ko'rsatadi.

Xulosa va takliflar. Ushbu maqolada sentiment tahlili bo'yicha yondashuvlar va shu kungacha amalga oshirilgan tadqiqot ishlari ko'rib chiqildi. O'zbek tili uchun ushbu sohada qo'llanilishi mumkin bo'lgan lingvistik ta'minotini ishlab chiqish uchun metod va yondashuvlar tanlandi. Ingliz, fransuz, ispan, hind, xitoy, rus, turk kabi tillardagi boy resurslarning mavjudligi o'zbek tili uchun lingvistik ta'minot ishlab chiqish va turli tillarda bajarilgan ishlarni o'rganish bo'yicha qilinadigan tadqiqotlarni sezilarli darajada rag'batlantiradi.

ADABIYOTLAR

1. Pei, Y.; Chen, S.; Ke, Z.; Silamu, W.; Guo, Q. AB-LaBSE: Uyghur Sentiment Analysis via the Pre-Training Model with BiLSTM. Appl. Sci. 2022, 12, 1182.
2. Gautam, Geetika, and Divakar Yadav. "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis." 2014 Seventh International Conference on Contemporary Computing (IC3), 2014. doi:10.1109/IC3.2014.6897213.
3. Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani//SEN TI WORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining Conference: Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010.
4. Erik Cambria, Daniel Olsher, Dheeraj Rajagopal (2015) SenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis
5. Saif M. Mohammad and Peter D. Turney (2013) NRC Emotion Lexicon. National Research Council Canada Ottawa, Ontario, Canada, K1A 0R6 <https://doi.org/10.4224/21270984>
6. Matlatipov S., Kuryozov E., Miguel A. A., Corlos-Rodriguez. Deep learning vs. classic models on a new uzbek sentiment analysis dataset. Conference: 9th language & rechnology conference: Human language technologies as a challenge for computer science and linguistics Poznan, -Poland- 2019. -P. 258-262
7. Kuriyozov E., Matlatipov S. Building a new Sentiment Analysis Dataset for Uzbek Language and Creating Baseline Models.// Multidisciplinary Digital Publishing Institute Proceedings. 2019. -№1. Pages 37.
8. Rabbimov I., Mporas I., Kobilov S. Investigating the effect of emoji in opinion classification of uzbek movie review comments. International Conference on Speech and Computer Science. -2020. -P. 435-445.
9. Vural, A. Gural, B. Barla Cambazoglu, Pinar Senkul, and Z. Ozge Tokgoz, 2012. A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish.
10. Yergesh, Banu, Gulmira Bekmanova, Altynbek Sharipbay, and Manas Yergesh, 2017. Ontology-based sentiment analysis of Kazakh sentences. In International Conference on Computational Science and Its Applications. Springer.
11. Ding X., Liu, B., & Yu, P. S. (2008, February). A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 international conference on web search and data mining (pp. 231-240). ACM.
12. Kim, S. M., & Hovy, E. (2004, August). Determining the sentiment of opinions. In Proceedings of the 20th international conference on Computational Linguistics (p. 1367). Association for Computational Linguistics.
13. Devika M D, Sunitha C, Amal Ganesh //Sentiment Analysis:A Comparative Study On Different Approaches. Procedia Computer Science 87 (2016) 44 – 49
14. Hu M., & Liu B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 168-177). ACM.
15. Park S., & Kim Y. (2016, June). Building thesaurus lexicon using dictionary-based approach for sentiment classification. In Software Engineering Research, Management and Applications (SERA), 2016 IEEE 14th International Conference on (pp. 39-44). IEEE.
16. Филин Ф.П. Историческая лексикология русского языка. Изд.2. УРСС. 2008.
17. Suhasini M., Srinivasu B.Emotion detection framework for Twitter Data Using Supervised Classifiers. New York, NY: Springer; 2020 (pp.565-576).
18. Mozafari F., Tahayori H. Emotion detection by using similarity techniques. Paper presented at: Proceedings of the 2019 7thIranian Joint Congress on Fuzzy and Intelligent systems; 2019:1-5; IEEE.
19. Huang C., Trabelsi A., Zaiane OR.ANA at SemEval-2019 Task 3: contextual emotion detection in conversations through hierarchical LSTMs and BERT; 2019.arXiv preprint arXiv: 1904.00132.

20. Jayakrishnan R., Gopal Greeshma N., Santhikrishna MS. Multi-class emotion detection in malayalan novels. Paper presented at: Proceedings of the 2018 International Conference on Computer Communication and Informatics; 2018: 1-5; IEEE.
21. Wikarsa L, Thahir SN. A text mining application of emotion classifications of Twitter's users using Naive Bayes method. Paper presented at: Proceedings of the 2015 1st International Conference on Wireless and Telematics (ICWT); 2015:1-6; IEEE.
22. Grover S, Verma A. Design for emotion detection of punjabi text using hybrid approach. Paper presented at: Proceedings of the International Conference on Inventive Computation Technologies; 2016:1-6; IEEE
23. Perikos I, Hatzilygeroudis I. Recognizing emotions in text using ensemble of classifiers. Eng Appl Artif Intell. 2016; 51: 191- 201.
24. Ghanbari-Adivi F, Mosleh M. Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). Neural Comput Appl. 2019; 31(12): 8971- 8983.
25. Orhan Bilgin, Ozlem C., Etino'glu, Kemal Oflazer. Building a WordNet for Turkish. Romanian journal of information science and technology.// 2004. V.7. № 1-2. P. 163-172.