

O'ZBEKCHA MATNLARNI HISSIY TAHLIL QILISH UCHUN LEKSIKONGA ASOSLANGAN YONDASHUV

Allanazarova Sabohat Yusupboyevna

allanazarova.sabosh@gmail.com

ToshDO'TAU tayanch doktoranti

Annotatsiya. Axborot asrida ijtimoiy tarmoqlardan foydalanish orqali mijozlarning xizmat sifati, brend yoki mahsulotlar va hodisalarga bo'lgan munosabatlarini o'rganishga asoslangan tadqiqotlar ustunlik qilmoqda. Bunday onlayn sharhlar yordami bilan turli xil sohalarda mijozlarning fikrlarini o'rganish maqsadida qo'llanilishi mumkin. Hissiy tahlil qilishda qo'llaniladigan bir necha yondashuvlar mavud. Ushbu maqolada emotsional-ekspressivlik jihatidan hissiy baho ifodalovchi so'zlar lug'atidan foydalangan holda leksikonga asoslangan hissiy tahlil usuli ko'rib chiqiladi.

Kalit so'zlar: hissiy tahlil, ijtimoiy tarmoq, sharhlar, mijozlar fikri.

Abstract. In the information age, research based on the study of customer attitudes toward service quality, brand or products, and events through the use of social networks dominates. With the help of such online reviews, it can be used for the aim of studying the opinions of costumers in various fields. There are several approaches used in sentiment analysis. In this article, we will considered the method of lexicon based approach, using a dictionary of words expressing emotional-expressiveness.

Key words: sentiment analysis, social network, reviews, customer's opinion.

Аннотация. В век информации преобладают исследования, основанные на изучении отношения клиентов к качеству услуг, бренду или товарам и событиям посредством использования социальных сетей. С помощью таких онлайн-обзоров можно узнать мнение клиентов в различных отраслях. В анализе настроений используется несколько подходов. В данной статье рассматривается метод эмоционального анализа на основе лексики с использованием словаря слов, выражающих эмоциональную оценку с точки зрения эмоционально-экспрессивности.

Ключевые слова: анализ тональность, социальная сеть, отзывы, мнение клиентов.

Mijozlarning mahsulot yoki xizmatlar haqida qanday fikrda ekanligini tushunish biznesda har doim eng muhimlaridan biri bo'lib kelgan. Ijtimoiy tarmoq foydaluvchilarining xizmatlar va mahsulotlar haqidagi o'zgalarning fikri bilan qiziqishga o'chligi va bu fikrlarga asoslangan holda xizmatlar sifatini yaxshilashga qaratilgan biznes talabi NLP sohasi bo'lgan hissiy tahlil qilishni dolzarb tadqiqotlar

qatoriga qo'shdi. Katta ma'lumotlar davrida millionlab iste'molchilarning sharhlari va munozaralari internetda kuniga behisob aylanadi [1]. Bunday katta hajmli ma'lumotlarni qo'lda saralash va mavuviy tasniflash esa tasavvur qilganimizdek, murakkab vazifa. Ijtimoiy tarmoqdagi tizimga solinmagan matnlardagi hissiyotlarni avtomatik ravishda tahlil qilish uchun sentiment tahliliga ehtiyoj bor. Sentiment tahlili (ST) - yozma matnda ifodalangan fikrlarni miqdoriy jihatdan baholash va turkumlashtirish jarayonidir.

ST uchun qo'llanilishi mumkin bo'lgan ilk yondashuv – leksikaga asoslangan yondashuv bo'lib, u ikki xil ko'rinishda namoyon bo'ladi. Birinchisi **lug'atga asoslangan**, ikkinchisi esa, **korpusga asoslangan yondashuv**dir.

Lug'at asoslangan yondashuv. Lug'atga asoslangan yondashuvning asosiy g'oyasi matndan his-tuyg'u ifodalovchi so'zlarni ajratish uchun leksik ma'lumotlar bazasidan foydalanishdir. Har qanday tilda ST tizimini ishlash chiqish uchun asosiy vazifa bu hissiyot leksikonidir. Boshlang'ich kayfiyatdagi so'zlar to'plamiga (masalan: **yaxshi, yomon, ajoyib, qimmat**) asoslanib ularning qutblari qo'lda yig'iladi [2]. Dastlab, bu boshlang'ich to'plam katta bo'lishi shart emas, 30 ta fikr so'zlari yetarli [3]. Keyingi qadam leksik ma'lumotlar bazasidan mos sinonim va antonimlarni izlash orqali to'plamni boyitish uchun qutbli so'zlardan foydalanishdir. Resurslarga boy ingliz tilida o'nlab hissiyot leksikonlari (masalan, **SentiWordnet** [4], **Senticnet** [5], **NRC Emotion Lexicon** [6], **General Inquirer** [7], **AFIN** [8], **Wordnet-Affect** [9], **Opinion Lexicon** [10], **Subjectivity Lexicon** [11], **So-Cal** [12], **Micro-WNOp** [13], **SentiSense** [14]) mavjud. Boshqa tomondan, o'zbek tili resurslari kam til bo'lib, hissiyot leksikonlari hozircha ishlab chiqilmagan. Bunday lug'atlar qo'lda va yarim avtomat yoki avtomatik ravishda yaratilishi mumkin.

Lug'atlarni qo'lda ishlab chiqishda vazifa mahalliy tilda so'zlashuvchi tilshunos mutaxassislardan tarkib topgan guruh a'zolari tomonidan taqsimlanadi. Har bir nomzod so'z (bunda ot, fe'l, sifat, ravish, iboralar tahlilga tortilishi mumkin) a'zo etiketidan o'tgach, barcha tomonidan bir xil baholangan so'z o'zgarishsiz qoladi. Ba'zi birliklar esa o'zgarishi mumkin, sababi chunki har kim dunyoni turlicha qabul qiladi va tasavvur doirasi turlicha bo'ladi.

Mavjud leksikonlar orasida SentiWordNet (SWN) eng keng tarqalgan bo'lib, chunki u har bir sinsetga tayinlangan obyektiv, subyektiv va salbiy uchta belgining birlashishi hisoblanadi. Bundan tashqari, har bir sinset raqamli ballar yordamida belgilanadi, bu endi his-tuyg'u to'liq so'zning o'zi emas, balki so'zning ma'nosi bilan bog'liqligini ko'rsatadi. Bundan tashqari, bir vaqtning o'zida ijobiy va salbiy sinset ham hisobga olinadi. Bu muhim xususiyatlar SWN ni boshqa leksik manbalardan farqi hisoblanadi.

Har bir soha (xizmat ko'rsatish, media, fond bozorlari, siyosat va boshqalar) ixtisoslashgan lug'atlarga muhtoj bo'lsa-da, umumiy lug'atlar ham foydalidir, chunki ular ish jarayonida aniqlanishi va to'ldirilishi mumkin bo'lgan boshlang'ich material bo'lib xizmat qilishi mumkin. Buning sababi shundaki, har bir hissiyot atamalariga

tegishli his-tuyg'ular toifasi (**IJOBIY**: sevinch, mamnuniyat, ishonch; **SALBIY**: norozilik, nafrat, qo'rquv) va ball beriladi, bu so'z, jumla va hujjat darajasi kabi turli darajalarda ballni hisoblashda yordam beradi.

Korpusga asoslangan yondashuv sintaktik va semantik belgilarni ko'rib chiqishni o'z ichiga oladi. O'z-o'zidan korpusga asoslangan usul korpus tarkibidagi so'zlarning cheklanganligi sababli lug'atga asoslangan usuldan samaradorligi kamroq. Biroq ushbu yondashuvdan foydalanish turli sohalar va kontekst mavzusiga xos lug'atlarni shakllantirishga yordam beradi [15]. Ushbu usul uch qismdan iborat:

- Pos tegging usuli (har bir so'z ajratib olinadi);
- Har bir tokenga hissiy ohangiga qarab alohida ball beriladi;
- Gapning umumiy qutbliligi hisoblanadi.

Birinchidan, sharhlar oldindan qayta ishlanadi, ular tuzatiladigan jummalarga bo'linadi va jumladagi har bir so'zni belgilash va saqlash uchun “Matnni ajratish” (POS) usuli qo'llaniladi. Ikkinchidan, kontekstual ma'lumotlarga va jumla tuzilishiga asoslanib, gapning qutbliligini aniqlashga imkon beruvchi so'zni ajratib olish bosqichiga o'tiladi. Bu gap tarkibidagi iboralarni tanib olish imkonini beruvchi aspekt hisoblanadi. Oxirgi qism qoidaga asoslangan modul yordamida jumalarni obyektiv yoki subyektiv tasniflashdan iborat.

Leksikaga asoslangan yondashuv asosida subyektivlik tahlili ham amalga oshiriladi. Bu tahlilda har bir matn tarkibidagi gap, har bir gap tarkibidagi so'zlar subyektiv yoki obyektiv birlik ekanligi aniqlanadi. Subyektiv gapda hissiyotni idodalovchi so'z qatnashgan, obyektiv gapda esa hech qanday fikr so'zlari yo'q bo'ladi.

ID	Gap	Hissiyot so'zlari	Subyektiv/Obyektiv
S1	Yer quyosh atrofida aylanadi	-	obyektiv
S2	Men kecha telefon sotib oldim	-	obyektiv
S3	Juda shinam joy. Avvalo, uning atmosferasi yoqimli.	shinam, yoqimli	subyektiv

1-rasm. Subyektivlik tahlili na'munasi

Yuqoridagi 3-gapda **shinam** va **yoqimli** so'zlari bu gapni subyektiv deb topilishiga asos bo'lgan.

Ta'sirchan ma'no bo'yicha dastlabki ishlar [Osgood va boshq.] 1957-yilda ta'sirchan so'zlarning ma'nosini uchta muhim o'lchov (salbiy, ijobiy va betaraf) bo'yicha o'zgarishini aniqlashdan boshlangan.

Sentiment (hissiy) tahlilda tadqiqot obyekti sifatida restoran sharhlari, mahsulot yoki brend sifati, xizmat ko'rsatish, siyosat yoki mavjud vaziyat haqidagi sharhlar olinishi mumkin. Ijtimoiy tarmoqdagi sharhlar kulgich (emoji) va jargonlar bilan to'la bo'lganligi sababli, ularning barchasini boshqarish uchun ixtisoslashgan leksika kerak bo'ladi. Sentiment tahlil qilishda hissiyot so'zlari turli xil sohalarga sezgir va tasniflashda turlicha talqin qilinadi. Shu sababli, tasniflashda ixtisoslashgan lug'atlardan foydalanish maqsadga muvofiq. Restoranlar sharhlarida mayonez va yog'li sochiq kabi so'zlar salbiy ma'noga ega. Agar bu so'zlar sharhda paydo bo'lsa,

u odatda bu joyda salbiy baho sifatida baholanadi. Shubhasiz, matnlardagi bunday yashirin jihatni tahlil qilish alohida ahamiyatga ega.

Hissiyotlarni tahlil qilishda turli darajadagi vazifalar mavjud. Asosiy vazifa – berilgan matnda bildirilgan fikr ijobiy/salbiy yoki betaraf qutbliligini aniqlash. Qutblilik – berilgan gapning ijobiy yoki salbiy ekanligini ko'rsatadigan o'lchovdir. Biroz murakkabroq vazifa esa 1 dan 5 gacha bo'lgan shkala yordamida matndan hissiyot darajasini hisoblashdir, bu yerda 1 qiymati eng salbiy, aksincha, 5 qiymati eng ijobiyni anglatadi. Yana bir muhim masala - ma'lum bir jumlada qutblanish modifikatorining doirasini aniqlash (masalan, inkor ma'noni). Masalan, **bu yangi telefonning dizayni menga yoqadigan emas, lekin u ba'zi qiziqarli funksiyalarga ega** jumlasida **emas** faqat yoqadigan so'ziga tegishli bo'lib, **qiziqarli** so'zining qutblanishini o'zgartirmaydi.

Google Map ilovasida o'zbek restoran va oshxonalariga sharhlar berilgan va tegishli yulduzchalar bilan belgilangan. Bunda biz 5 yulduzli sharhlarda qanday ijobiy so'zlar, 1 yulduzchali sharhda qanday salbiy so'zlar uchrashini bilishimiz mumkin. Ilovaning kamchilik tomoni, Twitterda bo'lgani kabi avtomatik ijobiy/salbiy sharhlarga ajratilmagan.

Modifikator	Qutblilik
cheksiz	+ ijobiy
juda	+ ijobiy
bag'oyatda	+ ijobiy
mukammal	+ ijobiy
to'liq	+ ijobiy
ko'p	+ ijobiy
benihoya	+ ijobiy
behad	+ ijobiy
zo'rg'a	- salbiy
bazo'r	- salbiy
arang	- salbiy
deyarli	- salbiy
biroz	- salbiy
picha	- salbiy

2-rasm. O'zbek tilida uchrashi mumkin bo'lgan modifikatorlarning qisqacha ro'yxati

Norasmiy til birliklari, shu jumladan kulgichlar va jargonlar, ko'pincha ijtimoiy tarmoq foydalanuvchilari tomonidan o'z xabarlarini yetkazish uchun qo'llaniladi. Ingliz tili kulgichlar va jargonlar va hattoki haqorat so'zlari uchun leksik manbalar bilan yaxshi ta'minlangan bo'lsa-da, o'zbek tilidagi ST uchun bunday manbalarga muhtoj va o'zbekcha sharhlardagi norasmiy tilni tahlil qilishga to'sqinlik qiladi. Va aksincha, ijtimoiy tarmoqdagi sharhlarda bunday lingvistik manbalar ko'p uchraydi.

Xulosa. Internetning paydo bo'lishi mahsulotlar, xizmatlar, voqealar va siyosiy fikrlar kabi mavzular bo'yicha ma'lumotlarni oson va keng almashish uchun eshiklarni ochdi. O'rgatilgan resurslari kam o'zbek tilida hissiyot tahlili uchun lingvistik ta'minot manbalariga muhtoj.

Foydalanilgan adabiyotlar:

1. Dang, Yan, Yulei Zhang, and Hsinchun Chen. 2010. A lexicon enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, 25(4):46–53.
2. H. Han, Y. Zhang, J. Zhang, J. Yang, and X. Zou, “Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias”, pp. 1–11, 2018
3. B. Liu, “Sentiment analysis: Mining opinions, sentiments, and emotions,” *Cambridge University Press*, 2015.
4. Baccianella S, Esuli A, Sebastiani F. Sentiwordnet3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the 7th International Conference on Language Resources and Evaluation. 2010, 2200–2204
5. Cambria E, Olsher D, Rajagopal D. SenticNet3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2014, 1515–1521
6. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465
7. Stone P J, Bales R F, Namenwirth J Z, Ogilvie D M. The general inquirer: a computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, 1962, 7(4): 484–498
8. Nielsen F. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In: Proceedings of Workshop on Making Sense of Microposts: Big Things Come in Small Packages. 2011, 93–98
9. Strapparava C, Strapparava C, Valitutti A. WordNet-affect: an affective extension of WordNet. In: Proceedings of the 4th International Conference on Language Resources and Evaluation. 2004, 1083
10. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, 2015, 89: 14–46
11. Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 2005, 39(2): 165–210
12. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 2011, 37(2): 267–307
13. Cerini S, Compagnoni V, Demontis A, Formentelli M, Gandini G. Micro-WNOp: a gold standard for the evaluation of automatically compiled lexical resources for opinion mining. In: Proceedings of Language Resources and Linguistic



Theory: Typology, Second Language Acquisition, English Linguistics. 2007, 200–210

14. De Albornoz J C, Plaza L, Gervais P. SentiSense: an easily scalable concept-based affective lexicon for sentiment analysis. In: Proceedings of the 8th International Conference on Language Resources and Evaluation. 2012, 23–25

15. Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In LREC, 2010.

16. Elov B., Hamroyeva Sh., Xusainova Z. NLP (tabiiy tilga ishlov berish)ning vazifalari va zamonaviy yondashuvlar. TerDU, Filologik tadqiqotlar: til, adabiyot, ta'lim. 2022, 5-6.