

MATNLARNING ASPEKTGA ASOSLANGAN HISSIY TAHLILI

Allanazarova Saboxat Yusupboyevna

allanazarova.sabosh@gmail.com

ToshDO‘TAU tayanch doktoranti

Annotatsiya. Hissiyotlarni tahlil qilish biznes va jamiyat uchun muhim ahamiyati tufayli Tabiiy tilni qayta ishlash (NLP) ning eng faol sohalaridan biri bo‘lib kelgan. Bu fikrning (ijobiy, salbiy, betaraf) bahosini aniqlashga qaratilgan tadqiqot sohasi. Hissiy (sentiment) tahlil qilishning turli darajalari bo‘lib, biz ushbu maqolada eng qiyin turi bo‘lgan aspekt darajasidagi tahlilga e’tibor qaratamiz.

Abstract. Sentiment analysis has been one of the most active areas of Natural Language Processing (NLP) due to its importance to business and society. This is a field of research aimed at determining the value of an opinion (positive, negative, neutral). There are different levels of sentiment analysis, and in this article we will focus on the most difficult type, aspect level analysis.

Аннотация. Анализ тональности был одной из наиболее активных областей обработки естественного языка (НЛП) из-за его важности для бизнеса и общества. Это область исследований, направленная на определение ценности мнения (положительного, отрицательного, нейтрального). Существуют разные уровни анализа настроений, и в этой статье мы сосредоточимся на самом сложном типе — анализе на уровне аспектов.

Kalit so‘zlar: *Tabiiy tilni qayta ishlash, ijobiylik, betaraf, salbiylik, hissiy tasnif, fikrlarni o‘rganish.*

Hissiyot tahlili katta hajmdagi umumiy fikrlarni avtomatik saralab, tushunish, tezkor xulosa chiqarish imkonini beradi. Ya’ni qiyinchiliklarni oddiy misollar orqali ko‘rsatish mumkin bo‘ladi. Sentiment (hissiy) tahlil - matndagi emotsional rangdagi so‘z boyligini avtomatlashtirilgan tarzda aniqlash uchun ishlatiladi. Hissiy tahlil quyidagi darajalarda amalga oshiriladi:

1. *Hujjat darajasidagi tahlil.* Matnda ifodalangan butun fikrlarni tasniflashdir. Masalan, biror bir mahsulot yoki xizmat haqida qoldirilgan sharhni tahlil qilish. Biror matnni boshqa matnga qiyoslamasdan to‘liq matn darajasida tahlil qiladi, bitta hissiy baho bitta sharh/matn uchun tegishli bo‘ladi.

Yangi versiyasi juda ham qulay ekan. Mb kam ketishi bilan boshqa applardan farqi bor ekan. Ishingizga omad sizlardan yangidan yangi xabarlar kutamiz. (ijobiy)

2. *Gap darajasidagi tahlil.* Har bitta gapga alohida hissiy baho beriladi. Bu tahlil turi subyektivlik tahlili uchun amalga oshiriladi.

1) *Men telefon sotib oldim.* (obyektiv)

2) *Restoranning ovqatlari ajoyib ekan.* (Subyektiv)

3. *Aspektga darajasidagi tahlil.* Bir xil so‘zning turli gaplardagi ma’nosi qarama-qarshi bo‘lishi mumkin yoki bir xil jumladagi bir xil shaxsning qutbliligi

turli jihatlarida farqlanadi. Shu sababli, matndan ma'lum obyektlar va ularning jihatlarini haqidagi fikrlarni aniqlash uchun aspekt darajasidagi hissiyotlarni tasniflash taklif etiladi.

Mehmonxonaning ovqatlari shirin ammo, narxi qimmat.

Aspekt darajasidagi hissiyotlarni tahlil qilishda muhim muammolardan biri jumlar o'rtasidagi semantik munosabatlarni qanday modellashtirishdir. Muammoni hal qilish uchun ko'plab usullar taklif qilingan [Rahman, 2013: 6]. So'nggi yillardagi tadqiqotlarda aspekt darajasidagi hissiyotlarni tahlil qilishda neyron tarmoqlar (NN) ustunlik qilmoqda. Aspekt darajasidagi hissiyotlarni tasniflash ko'pincha adabiyotlarda tasniflash muammosi sifatida qaraladi. Mashinaviy o'qitish ham hissiyotlarni tasniflashning asosiy usuli hisoblanadi. Bu usullar, asosan, matn ko'rinishini va xususiyatlarni ajratib olishni masalan, so'zlar to'plami modellari va hissiyot leksikasining xususiyatlari, so'ngra qo'llab-quvvatlash vektor mashinasi (SVM) va logistik regressiya (LR) kabi hissiyotlar tasniflagichini o'rgatishni o'z ichiga oladi. Bunda aspekt atamaları va aspekt toifalarini aniqlash, shuningdek, mijozlar sharhlarida atamalar va tomonlar toifalariga nisbatan bildirilgan his-tuyg'ularni aniqlash uchun boshqariladigan mashinaviy o'qitish usullari tavsiflanadi. Biroq, an'anaviy mashinaviy o'rganish usullari odatda ko'plab muhim xususiyatlar kontekstini modellashtira olmaydi.

Hissiyotlarni aspekt bo'yicha tahlil qilish muammosida odatda matn bitta obyektning baholashni o'z ichiga oladi deb taxmin qilinadi, aslida obyektning turli qismlari VA / YOKI atributlari bilan bog'liq holda ko'rib chiqilishi tushuniladi. Aspekt bo'yicha his-tuyg'ularni tahlil qilish uchun odatiy matndagi (bu ba'zi mahsulot yoki xizmatlar haqida foydalanuvchilar qoldirgan sharhlar bo'lishi mumkin) aspektlarni toifalarga guruhlash mumkin. Aspektga asoslangan tahlilda obyektning biror bir jihatiga qaratilgan bo'ladi. Masalan, oshxonalarining xizmat ko'rsatish sifati, yoki ovqatlari tanlab olinishi mumkin. Shuningdek, sharhlarda siz obyektning umumiy bahosini topishingiz ham mumkin: ajoyib restoran. Matndagi biror jihatga ishora qilish uchun ishlatiladigan so'z va iboralar aspekt atamaları deyiladi. Restoranlar uchun bu odatda oshxona, ovqat, likobcha, xizmat ko'rsatish, interyer (mebellar) kabilar.

Tovarlar yoki xizmatlar uchun belgi/xususiyatlarni aniqlashni avtomatlashtirish. Aspekt terminlari sifatida ko'pincha otlar va otli birikmalari ko'rib chiqiladi [Asuncion, 2009: 27]. Ismlar guruhining uzunligi 3-4 so'zdan oshmasligi kerak. Shu bilan birga, faqat individual otlar aspekt sifatida ajratilsa, ular ko'pincha noaniq bo'lishi mumkin, bu mutaxassislar tomonidan xato talqinga olib keladi [Боронцов, 2014: 676]. Matnlardan aspekt atamalarini ajratib olishni avtomatlashtirishning to'rtta asosiy yondashuvi mavjud:

1. Otlar va ot guruhlariga asoslangan yondashuv;
2. Baholovchi iboralar va aspekt terminlari o'rtasidagi munosabatlardan foydalanadigan yondashuv;
3. Nazorat ostida mashinaviy o'qitish yondashuvi;

4. Statistik modellarga asoslangan yondashuv.

Shunday qilib, aspektlar bo'yicha hissiyotlarni tahlil qilish vazifasi quyidagi kichik vazifalarni o'z ichiga oladi:

- Aspektni ifodalovchi atamalarni tanlash;
- Aspekt atamaları orqali toifalarga ajratish;
- Tanlangan toifalarga nisbatan aspektlarni avtomatik aniqlash.

Mavzu sohasidagi aspekt atamalarini bir necha jihatdan tasniflash mumkin. Aspekt atamalarining eng keng tarqalgan turi - obyekt, uning qismlarini yoki xususiyatlarini aniq nomlaydigan, matn muallifi tomonidan baholanadigan, masalan, osh, xizmat, zal, restoran sharhlarida aniq aspekt terminlaridir. Aniq aspekt terminlari ko'pincha otlar yoki ot guruhiga mansub so'zlar bilan ifodalanadi, lekin ba'zi jihatlar fe'llar bilan ham ifodalanishi mumkin, masalan, restoranlarda xizmat ko'rsatish sifatini baholashda kutish, kelmaslik, kuttirish kabi birliklar uchraydi.

Aspekt atamalarini aniqlashning ikkinchi turiga yashirin aspekt atamaları kiradi, ular bir vaqtning o'zida muhokama qilinayotgan jihatni (odatda ancha umumlashgan jihat kategoriyasi) ko'rsatadigan ma'noning aniq baholovchi komponentiga ega bo'lgan so'zlardir, masalan: **mazali** (ijobiy + taom) restoran sharhlarida), **qulay** (avtomobil sharhlarida ijobiy + qulaylik). Boshqa baholovchi so'zlar singari, noaniq jihat atamalar deb ataladigan so'zlar bilan birlashtirilishi mumkin. O'z bahosini o'zgartiradigan yoki kuchaytiradigan baholash modifikatorlari: **juda** mazali emas, **uncha** qulay emas.

Avtomatik his-tuyg'ularni tahlil qilish tizimlarining lug'atlari uchun bunday aspektli atamalarining ahamiyati shundaki, ular tan olinmagan holatlarda aniq atamalar muallifi (matbaa xatolari, yangi lug'at, murakkab havolalar) yashirin aspekt shartlari foydalanuvchining ba'zi jihatlar toifasiga nisbatan pozitsiyasini ajratib olish imkonini beradi.

Ma'lum bir obyektning ba'zi bir xususiyati haqida o'z fikrini ifodalashning uchinchi turi - bu sodir bo'lgan salbiy yoki ijobiy fakt to'g'risidagi xabar, bu bir vaqtning o'zida ham aspekt toifasini, ham uning foydalanuvchi tomonidan baholanishini (baholash faktlari) ko'rsatadi. Baholash faktlarining turlaridan biri bu sharhlarda qayd etilgan texnik muammolardir [Girolami, 2003: 433]. Ingliz tilidagi texnik muammolarga havolalar ko'pincha quyidagilarni o'z ichiga oladi:

- *Biror narsa sodir bo'lganligini bildiruvchi maxsus fe'llar to'plami* (fail-muvaffaqiyatsizlik, halokat-crash, ortiqcha yuk-overload, sayohat-trip, tuzatish-fix, tartibsizlik-mess, buzish-break, ortiqcha zaryadlash-overcharge, vayron qilish-distrupt);

- *Biror narsa sodir bo'lmaganligini bildiruvchi fe'llar to'plami* va ko'pincha bu fe'llar inkorlar bilan, shuningdek, fe'llar bilan (to'xtash, rad etish, to'xtatish va boshqalar) bilan tilga olinadi;

- *shuningdek, ot va iboralar.*

Ma'no jihatdan yaqin bo'lgan baholovchi faktlar matnda turli yo'llar bilan ifodalanishi mumkin, bu esa ularni aniqlashni qiyinlashtiradi. Matnda tez-tez

uchraydigan lekin baholanmagan soʻzlar, ijobiy yoki salbiy konnatativ naʼno ifodalashi mumkin. Ijtimoiy-siyosiy matnlarda salbiy maʼnoga ega boʻlgan bunday soʻzlarga **ishsizlik, inflyatsiya, turgʻunlik** soʻzlari misol boʻla oladi. Restoran sharhlari sohasida **soch, mayonez** soʻzlari salbiy maʼnoga ega, yaʼni matnlarda bunday soʻzlarning paydo boʻlishi matnning ohangi ancha salbiy boʻlishidan dalolat beradi.

Texnik sohada bunday soʻzlar [Lee, 2011: 884] da koʻrsatilganidek, buzilishlarni bildiruvchi soʻzlar (qobiliyatsiz, avariya, ortiqcha yuk, safar, tuzatish, tartibsizlik, tanaffus) hisoblanadi. Shuningdek, “resurs muammosi” deb ataydigan yashirin baholar va jihatlarning yana bir toifasi mavjudligi koʻrsatilgan: **Bu kir yuvish mashinasi juda koʻp suv sarflaydi**. Shunday qilib, suv isteʼmoli bu yerda jihat, suv esa resurs atamasi boʻlib, uni ortiqcha isteʼmol qilish salbiy faktdir.

Baʼzi tadqiqotlarda da resurs atamalarini koʻp sonli miqdor koʻrsatkichlari bilan, shuningdek, isteʼmol feʼllari asosida ajratib olish kerakligi koʻrsatilgan [Wang, 2009: 1982]. Maqolada maʼlum isteʼmol feʼllari, shuningdek, tabiiy resurslar: gaz, suv, elektr, gaz maʼlum sonli algoritm asosida koʻrib chiqiladi.

Mikolov maqolasida restoranlar va avtomobillar haqidagi sharhlar sohasida eng koʻp uchraydigan baholovchi faktlarni koʻrib chiqadi [Mikolov, 2013: 4546]. Restoran sharhlaridagi manbalarga quyidagilar kiradi:

- mehmonlarni qabul qilish vaqti;
- ofitsiantlarning eʼtibori;
- oziq-ovqat bilan bogʻliq boʻlgan uch turdagi resurslar: likobchadagi oziq-ovqat miqdori, menyuda tanlovning kengligi va idishlarning tozaligi;
- restorandagi joy va bepul oʻrindiqlar;
- xizmat koʻrsatish narxi.

Aspekt terminlari sifatida nafaqat otlar, balki otli birikma ham ajratilishi, keyin ot soʻz birikmasining uzunligini toʻgʻri aniqlash uchun qoʻshimcha belgilardan foydalanish kerak.

Xulosa. Matnlarning avtomatik hissiy tahlil qilish vazifasi murakkab muammodir. Hissiyotlarni tahlil qilishning mashhur va ommalashgan turi bu – aspektga asoslangan tahlildir. Matnlarni avtomatik tahlil qilishning dastlabki natijalarini muhandis-linvistik yondashuv asosida ham, mashinani oʻrganish usullari asosida ham tezda olish mumkin.

Foydalanilgan adabiyotlar:

1. Asuncion A., Welling M., Smyth P., Teh Y. W. On smoothing and inference for topic models // Proceedings of the International Conference on Uncertainty in Artificial Intelligence. — 2009. — Pp. 27–34.
2. Girolami M., Kabʼan A. On an equivalence between PLSI and LDA // SIGIRʼ03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. — 2003. — Pp. 433–434.



3. Lee S. S., Chung T., McLeod D. Dynamic item recommendation by topic modeling for social networks // Information Technology: New Generations (ITNG), 2011 Eighth International Conference on. — IEEE, 2011. — Pp. 884–889.
4. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed representations of words and phrases and their compositionality // CoRR. — 2013. — Vol. abs/1310.4546.
5. Rahman M. Search engines going beyond keyword search: A survey // International Journal of Computer Applications. — August 2013. — Vol. 75, no. 17. — Pp. 1–8.
6. Wang C., Blei D. M. Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process // NIPS. — Curran Associates, Inc., 2009. — Pp. 1982–1989.
7. Воронцов К. В., Потапенко А. А. Регуляризация вероятностных тематических моделей для повышения интерпретируемости и определения числа тем // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2014 г.). — Вып. 13 (20). — М: Изд-во РГГУ, 2014. — С. 676–687.