

PARALLEL KORPUSLARDA SENTIMENT TAHLIL QILISHDA “ALIGNER” VOSITASINING O'RNI

Matyakubova Noila Shakirjanovna,
tayanch doktoranti
matyakubovanoila@navoiy-uni.uz
ToshDO'TAU

Annotatsiya. Parallel korpuslarda sentiment tahlilini amalga oshirishda “Aligner” vositalari muhim rol o'ynaydi. Ular matnlarni to'g'ri moslashtirish orqali SLda berilgan gapning hissiy holatini saqlashda, tarjimaning sifatini oshirishda va semantik xatolarni kamaytirishda yordam beradi. “Aligner” vositalarining samarali ishlatilishi parallel korpuslardan to'liq foydalanishga, shuningdek, turli tillarda kayfiyatni aniq aks ettirishga imkon yaratadi. Bu jarayonni mukammal bajarish, parallel korpuslar asosida yaratilgan tarjima tizimlarining sifatini oshirishga yordam beradi va umumiy tilni qayta ishlash sohasiga katta hissa qo'shadi.

Annotation. Aligner software tools play a crucial role in conducting sentiment analysis on parallel corpora. By accurately aligning texts, they help preserve the emotional tone of a sentence in the source language (SL), improve translation quality, and reduce semantic errors. The effective use of aligner tools enables full utilization of parallel corpora and allows for accurate representation of sentiment across different languages. Proper implementation of this process contributes to enhancing the quality of translation systems built on parallel corpora and makes a significant contribution to the broader field of natural language processing.

Аннотация. Программные инструменты Aligner играют ключевую роль в проведении сентимент-анализа на параллельных корпусах. За счёт точного выравнивания текстов они помогают сохранить эмоциональную окраску предложений, заданных на языке-источнике (SL), повысить качество перевода и уменьшить количество семантических ошибок. Эффективное использование инструментов выравнивания позволяет полноценно использовать параллельные корпуса, а также точно передавать настроение на разных языках. Корректное выполнение данного процесса способствует повышению качества систем перевода, основанных на параллельных корпусах, и вносит значительный вклад в развитие области обработки естественного языка.

Kalit so'zlar. *Parallel korpus, Aligner vositasi, asliyat til (SL), tarjima tili (TL), sentiment tahlil, ijobiy, salbiy, neytral.*

Kirish. Parallel korpuslar – bu ikki yoki undan ortiq tillarda o'xshash mazmundagi matnlar to'plamidir. Tabiiy tilni qayta ishlash (NLP) sohasida parallel korpuslar tarjima tizimlarini o'rgatishda, til o'rtasidagi so'z va frazalarni moslashtirishda keng qo'llaniladi. Biroq, parallel korpuslardan samarali foydalanish uchun ularni to'g'ri moslashtirish zarur. “Aligner” vositalari parallel matnlar orasidagi to'g'ri so'z yoki ibora moslashuvlarini topish uchun ishlatiladi. Sentiment tahlili, matndagi kayfiyatni aniqlashda yordam beruvchi jarayon sifatida, parallel korpuslar orqali samarali o'rganilishi mumkin.

Parallel korpuslarni moslashtirishda “Aligner” vositalarining ishlatilishi, til o'rtasidagi sintaktik va semantik farqlarni kamaytirishga yordam beradi, ya'ni bir tildagi so'z yoki iborani boshqa tilga to'g'ri moslashtirishda ishlatiladi. Bu vositalar yordamida sentimentni to'g'ri aks ettiruvchi so'zlar va iboralar aniqlanadi, salbiy yoki ijobiy kayfiyatini aks ettiruvchi so'zlari, boshqa tilda ham to'g'ri tarjima qilinganligini aniqlashga yordam beradi. “Aligner” vositalari parallel korpuslarning sifatini oshiradi va asliyat til (SL) matnidagi kayfiyatni tarjima til (TL) matnida to'g'ri saqlashga yordam beradi.

Sentiment tahlili va “Aligner” vositalarining birgalikda ishlatilishi, parallel korpuslarda kayfiyatni aniq aks ettirishga va tarjima tizimlarining sifatini oshirishga imkon yaratadi. Shu bilan birga, bu yondashuv mashina tarjimasi tizimlarining samarasini oshiradi va ko'p tilli NLP modellarini kuchaytiradi. Maqola parallel korpuslarni to'g'ri moslashtirish orqali sentiment tahlilining aniqligini ta'minlashning ilmiy va amaliy ahamiyatini ko'rsatadi.

Adabiyotlar tahlili.

So'nggi yillarda sentiment tahlili (sentiment analysis/opinion mining) yo'nalishlarida xalqaro va milliy miqyosda ko'plab ilmiy tadqiqotlar olib borilmoqda. Sentiment tahlili ijtimoiy tarmoqlar, mahsulot sharhlari, yangiliklar, fikr va munosabatni ifodalovchi matnlarda keng qo'llanilmoqda [1].

Ko'p tilli sentiment tahlili (Multilingual Sentiment Analysis, MSA) doirasida, turli tillar o'rtasida semantik moslikni aniqlash muammosi dolzarb hisoblanadi[2]. Ayniqsa, kam resursli tillar (low-resource languages), jumladan, o'zbek tili uchun bu masala yana ham keskinroq bo'lib, parallel korpuslarga bo'lgan ehtiyoj ortib bormoqda [3]. O'zbek tilida sentimentga oid belgilangan korpuslar mavjud emasligi sababli, parallel matnlar asosida tahlil yuritish strategiyasi afzallik kasb etadi.

Bu borada “Aligner” vositalari ko'p tilli sentiment tahlilida asosiy vositalardan biri sifatida e'tirof etiladi. IBM Model 1–5, GIZA++, hunAlign kabi an'anaviy moslashtirish vositalari uzoq yillar davomida parallel korpuslar uchun ishlatilib kelinmoqda [4/5]. So'nggi yillarda esa neyron tarmoqlariga asoslangan “Aligner”lar (semantik moslikni chuqur kontekstual model yordamida aniqlash

imkonini yaratdi [6]. Ayniqsa, BERT, mBERT, XLM-R kabi pre-trained til modellari yordamida kontekstual semantika chuqurroq tahlil qilinmoqda.

O'zbekistonlik olimlar ham so'nggi yillarda matnlar semantikasini tahlil qilish, mashinaviy tarjima va korpus lingvistikasi yo'nalishlarida tadqiqotlar olib bormoqda. Jumladan, N.Abdurahmonova, Sh.Xamroyeva, N.Matyakubova, S.Avezov, R.Karimov, A.Axmedova kabi tadqiqotchilar tomonidan o'zbek tilida avtomatik tarjima, parallel korpuslar tuzish va lingvistik birliklarni moslashtirish usullari bo'yicha tadqiqotlar e'lon qilingan[7/8]. Ushbu ishlar o'zbek tili uchun til juftligi bo'yicha parallel korpuslar yaratish va tahlil qilishning dastlabki bosqichlarini shakllantirgan.

Bundan tashqari, E.Kuriyozov, S.Allanazarova, U.Salaev, S.Matlatipov, N.Raximov, va U.Bekmuradov tomonidan o'zbek tilidagi matnlar uchun NLP modellari, klassifikatsiya usullari, shuningdek, elektron korpuslar asosida sentiment tahlil uslublarini ishlab chiqish bo'yicha amaliy tadqiqotlar olib borilgan[9/10]. Ularning ishlarida asosiy e'tibor matn segmentatsiyasi, belgilash va sentiment tahlil qilish imkoniyatlarini yaratishga qaratilgan.

Parallel korpuslar va sentiment tahlili

Parallel korpuslarda SL va TL matnlari orasidagi semantik va sintaktik o'xshashliklarni o'rganish, tilni qayta ishlashning muhim jihatidir. Sentiment tahlili – matnda ijobiy, salbiy yoki neytral kayfiyatni aniqlash jarayonidir. Parallel korpuslar yordamida, bir tilning matni boshqa tilga tarjima qilinayotganda, SL matn ifodalayotgan kayfiyat saqlanishi yoki o'zgarishi mumkin. Shuning uchun parallel korpuslarda sentiment tahlilini amalga oshirishda asosiy e'tibor, kayfiyatni to'g'ri aks ettirishga qaratilishi lozim.

Sentiment tahlilining parallel korpuslarda o'rganilishining ahamiyati shundaki, bu metod bir tilning hissiy holatini boshqa tildagi mos keluvchi elementlar bilan to'g'ri tarjima qilishga yordam beradi. Bunda, qayta tarjima jarayonida ijobiy yoki salbiy kayfiyatni o'zgartirishga yo'l qo'ymaslik uchun sentiment tahlilidan foydalanish zarur.

Aligner vositalari va parallel korpuslarni moslashtirish

“Aligner” vositalari parallel korpuslarda SL va TL orasidagi moslikni aniqlashga xizmat qiladi. Ular matnlarni moslashtirish, ya'ni bir tildagi so'zlarni yoki iboralarni boshqa tilga to'g'ri tarjima qilish uchun ishlatiladi. Aligner vositalari yordamida parallel matnlarning har bir elementini o'zaro bog'lash (align) mumkin. Bu vositalar so'zlar va gaplar o'rtasidagi o'zaro bog'lanishlarni aniqlashga imkon beradi, shuningdek, tarjimadagi noaniqliklarni va kayfiyatdagi farqlarni kamaytirishga yordam beradi.

Bugungi kunda asosan statistik va neyron tarmoqlariga asoslangan “Aligner”lar NLPning ko‘plab sohalarida keng qo‘llanilmoqda. Statistik “Aligner”lar matnlar orasidagi moslik statistik metodlar asosida aniqlaydi. Giza++, FastAlign kabi moslashtiruvchi vositalar bunga misol bo‘ladi. Neyron tarmoqlariga asoslangan “Aligner”lar so‘ngi avlod moslashtiruvchi vositalardan bo‘lib, matnlarni moslashtirishda yuqori aniqlik darajasidagi natijalar taqdim etadi va mashina o‘rganish modellari yordamida matnlarni samarali tarzda moslashtiradi.

“Aligner” vositalarining sentiment tahlilidagi o‘rni

Sentiment tahlilining muvaffaqiyatli amalga oshirilishi uchun, parallel korpuslarda hissiy holatni to‘g‘ri aks ettiruvchi so‘z va iboralarning to‘g‘ri moslashuvi zarur. Aligning (moslashtirish) jarayoni, so‘zlarning yoki iboralar orasidagi bog‘lanishni aniq belgilashga yordam beradi, bu esa sentiment tahlilining to‘g‘ri bajarilishi uchun zarur.

“Aligner” vositalari yordamida parallel korpuslarda ijobiy yoki salbiy kayfiyatga ega bo‘lgan so‘zlar to‘g‘ri moslashtiriladi. Masalan, ingliz tilidagi “*brilliant*” so‘zi o‘zbek tilida “*a‘lo*” deb tarjima qilinadi, va ijobiy hissiy holatni saqlab qoladi. Agar bu so‘z noto‘g‘ri moslashtirilsa, hissiy holatni o‘zgarishiga olib kelishi mumkin. “Aligner” vositalari bu jarayonni aniq va to‘g‘ri amalga oshirishga yordam beradi.

Ba‘zan til o‘rtasidagi semantik farqlar natijasida sentiment tahlili xatolari yuzaga kelishi mumkin. Misol uchun, ba‘zi tillarda salbiy hissiy holatni ifodalovchi so‘zlar boshqa tildagi so‘zlarga mos kelmasligi mumkin. “Aligner” vositalari bu kabi xatolarni aniqlashda yordam beradi, chunki ular matndagi ijobiy yoki salbiy elementlarni to‘g‘ri bog‘lashda aniq va sifatli moslashuvni ta‘minlaydi.

Sentiment tahlilining parallel korpuslarda samarali amalga oshirilishi uchun, sentimentga mos tarjima modellari ishlab chiqish zarur. Aligner vositalari bu modellarga kerakli parallel matnlarni moslashtirishda yordam beradi, bu esa ijobiy yoki salbiy kayfiyatni saqlab qolgan holda tarjima qilish imkonini beradi.

O‘zbek-ingliz parallel korpusida sentiment tahlilini amalga oshirish.

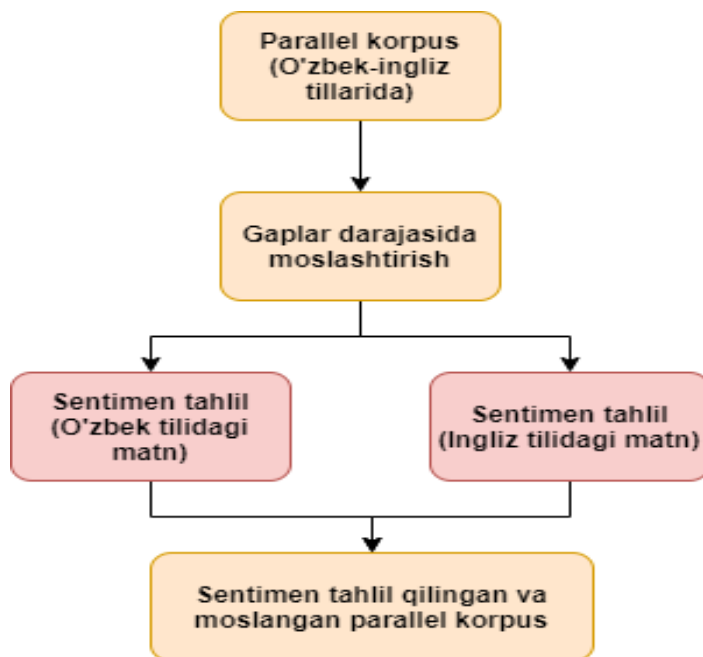
O‘zbek-ingliz paralell korpusida sentiment tahlil jarayonini amalga oshirish bir nechta bosqichda amalga oshiriladi (Qarang: 1-rasmga).

Dastlab parallel korpus shakllantirilishi lozim. Bunda SL va TLda fikrni ifodalovchi gaplar yoki qisqa matnlar jamlanadi. Matnlarga ishlov berish bosqichidan keyin ular alohida SL matnlar va TL matnlar shaklida alohida fayllar ko‘rinishida shakllantiriladi.

Korpus uchun matnlar jamlangach moslashtirish bosqichiga o‘tiladi. Bu bosqich o‘zbek-ingliz “Aligner” vositasi yordamida gaplar darajasida moslikni

aniqlash bosqichi amalga oshiriladi. Bu qadam til o‘zaro farqlarni yengillashtirish va keyingi sentiment tahlilida to‘g‘ri ma‘nolarni aniqlash uchun muhimdir.

“Aligner”dan olingan moslangan ma‘lumot asosida sentiment tahlili amalga oshiriladi. Bu yerda klassifikator yoki sentiment leksikonidan foydalangan holda, har bir segmentning ijobiy, salbiy yoki neytral ekanligi aniqlanadi.



1-rasm. Parallel korpusni moslashtirish va sentiment tahlil qilish bosqichi.

Moslashtirilgan va tahlil qilingan korpus natijasi jadval ko‘rinishida taqdim etiladi. Bunda ijobiy mazmunga ega gaplar *positive*, salbiy mazmundagilar *negative* va neytral mazmundagi gaplar *neutral* belgisi asosida belgilanadi.

1-jadval. Moslangan parallel korpus va sentiment tahlil natijasi.

ID	Inglizcha matn	O‘zbekcha matn	Sentiment tahlil natijasi
1	The book elicited a highly favorable impression from the reader.	Kitob o‘quvchida nihoyatda ijobiy taassurot uyg‘otdi.	positive
2	The quality of service at the restaurant was critically inadequate.	Restorandagi xizmat sifati juda past darajada bo‘lgan.	negative
3	Meteorological conditions appear to be stable and pleasant today.	Bugungi meteorologik sharoitlar barqaror va yoqimli kechmoqda.	neutral
4	The user expressed complete satisfaction with the provided service.	Foydalanuvchi ko‘rsatilgan xizmatdan to‘liq mamnunligini bildirdi.	positive

5	The recurrent tardiness of the individual provoked irritation.	Shaxsning muntazam kechikishi g'azabni keltirib chiqardi.	negative
6	The cinematic experience was acceptable, yet lacked distinctiveness.	Kinematografik tajriba qoniqarli bo'lsa-da, o'ziga xoslikdan yiroq edi.	neutral
7	The previous day's meeting was perceived as a pleasant encounter.	Kechagi uchrashuv yoqimli muloqot sifatida baholandi.	positive
8	The culinary offerings at this venue failed to meet expectations.	Ushbu maskandagi taomlar kutilgan darajada emas edi.	negative
9	The accommodation maintained high standards of cleanliness and comfort.	Turar joy tozalik va qulaylik mezonlariga to'liq javob berdi.	positive
10	The application encountered frequent crashes, causing frustration.	Ilova tez-tez nosozlikka uchrab, foydalanuvchida noqulaylik tug'dirdi.	negative

Xulosa.

Sentiment tahlilini kam resursli tillarda amalga oshirishda parallel korpuslar va ularni aniqlik bilan moslashtiruvchi “Aligner”lar asosiy vosita bo'lib xizmat qiladi. Aynan o'zbek tilida bu yo'nalishdagi tadqiqotlar hali boshlang'ich bosqichda bo'lsa-da, parallel korpus tuzish, ularni yuqori aniqlikda moslashtirish va bu orqali sentimentni tahlil qilish istiqbolli va dolzarb ilmiy yo'nalish hisoblanadi.

Foydalanilgan adabiyotlar ro'yhati.

1. Liu B. Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers. 2012.
2. Barnes J., Klinger R., Schulte im Walde S. Projecting Embeddings for Domain Adaptation: Joint Modeling of Sentiment Analysis in Diverse Languages. In Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018), 2018. – P. 818–830.
3. Bojar O., Chatterjee R., Federmann C., Graham Y., Haddow B., Huck M., Birch, A. Findings of the 2016 Conference on Machine Translation (WMT16). In Proceedings of the First Conference on Machine Translation, Berlin, Germany. 2016.
4. Och F., Ney H. A systematic comparison of various statistical alignment models. Computational Linguistics, 2003, 29(1), – P. 19–51.
5. Varga D., Halácsy P., Kornai A., Nagy V., Németh L., Trón V. Parallel corpora for medium density languages. In Proceedings of RANLP 2007, Borovets, Bulgaria, 2007. – P. 1–8.



6. Dou Z., Neubig G. Word Alignment by Fine-tuning BERT with Multilingual Objectives. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021), 2021. – P. 8502–8512.
7. Abdurahmonova N. Mashina tarjimasining lingvistik asoslari. Akademnashr. Toshkent, 2012. – 30 b.
8. Xamroyeva Sh., Matyakubova N. Korpus lingvistikasida “Aligner” dasturlarning o‘rni / “Fan va Jamiyat” ilmiy-uslubiy jurnal. Nukus, 2023. – B. 74-76.
9. Matlatipov S., Rahimboeva H., Rajabov J., Kuriyozov E. Uzbek Sentiment Analysis Based on Local Restaurant Reviews. CEUR Workshop Proceedings, 3315, 2022. – P. 126–136.
10. Allanazarova, S. Korpus va lug‘atga asoslangan matnning sentiment tahlili. Xorazm Ma‘mun akademiyasi axborotnomasi, 2024. 7(4), – P. 28-31.