

NLPDA KOREFERENSIYA MASALASI

Jumaqulova Nasiba Qurbonovna

nasiba.qurbonovna@gmail.com

ToshDO‘TAU magistranti

Annotatsiya. Ushbu tezisda NLPda koreferensiya muammosini hal qilishning ahamiyati va qoidaga asoslangan usullar, Hobbs algoritmi hamda koreferensiya yechimidagi qiyinchiliklar haqida fikr yuritiladi.

Abstract. In this thesis, the importance of resolving coreference issues in NLP and the methods based on rules, the Hobbs algorithm, and the difficulties in coreference resolution are discussed.

Аннотация. В данной диссертации обсуждается важность разрешения кореференции в NLP, методы, основанные на правилах, алгоритм Хоббса и трудности в разрешении кореференции.

Kalit so‘zlar: *NLP, koreferensiya, qoidaga asoslangan usullar, Hobbs algoritmi.*

Tabiiy tilni qayta ishlash (NLP) – tilshunoslik va sun’iy intellekt kesishmasida joylashgan soha bo‘lib, u tabiiy tilning rasmiy axborot uslubini yaratish bilan shug‘ullanadi. NLPdagi asosiy muammolardan biri bu koreferensiya muammosini hal qilishdir, bu jarayon matndagi turli xil so‘z yoki ifodalar bir xil obyekt yoki tushunchaga ishora qilishini avtomatik aniq aniqlashni o‘z ichiga oladi. Bu vazifa matnni umumlashtirish va savollarga javob berishdan tortib chatbotlar va ma’lumot qidirish tizimlarigacha bo‘lgan turli ilovalar uchun juda muhimdir.

Azim issiq yoz kuni ko‘cha chetida qovun sotardi, u har o‘tgan piyodalarga uning mazzasini ta’riflashdan charchamasdi.

Azim → u

Qovun → uning

Koreferensiya muammosining ahamiyati

Matn ma’lumotlarining tuzilishi va semantikasini tushunishda asosiy rol o‘ynaydi. Bu olmoshlar, otlar va boshqa ishora qiluvchi ifodalarni ajratishda yordam beradi va natijada matnni umumiy tushunishni yaxshilaydi. Koreferensiya muammosini hal qilish muhim bo‘lgan bir nechta asosiy sabablar mavjud [1]:

1. **Matndagi koreferensiya:** Koreferensiya yechimini hal qilish matnning uyg‘unligi va o‘qilishini oshiradi. Bu o‘quvchilar yoki NLP tizimlari matnning turli qismlari orasidagi mos yozuvlar aloqalarini chalkashmasdan kuzatishini ta’minlaydi.

2. **Ma’lumot olish:** Ma’lumot olish vazifalarida, masalan, nomli obyektни tanib olish yoki hodisani ajratib olishda, koreferensiya yechimi matnning turli qismlaridan ma’lum bir obyekt yoki hodisa haqidagi ma’lumotlarni jamlashda yordam beradi.

3. **Matnni umumlashtirish:** Avtomatik matnni umumlashtirish uchun matndagi qaysi ifodalar yoki atoqli otlar obyekt bir obyektga tegishli ekanligini tushunish juda muhimdir. Bu qisqa, izchil xulosalar yaratish imkonini beradi.

4. **Savol-javob:** Savollarga javob berish tizimlari uchun koreferensiya yechimi zarur. Savol berilganda, asosiy ma'lumotlarni hal qilish savolga javob berish uchun zarur bo'lgan tegishli ma'lumotlarni aniqlashi mumkin.

5. **Chatbotlar va virtual yordamchilar:** Virtual yordamchilar, chatbotlar va mijozlarni qo'llab-quvvatlash tizimlari foydalanuvchilar bilan tabiiy tilda suhbatlar o'tkazyotganda koreferensiya yechimidan foyda ko'radi. Bu ularga kontekstni saqlab qolish va aniqroq javob berish imkonini beradi.

Koreferensiya muammosini hal qilish usullari

Koreferensiya muammosini hal qilishga turli usullar orqali erishish mumkin, ularning har biri o'zining afzalliklari va cheklovlariga ega:

1. **Qoidalarga asoslangan usullar** [2]: Qoidalarga asoslangan tizimlar asosiy murojaatlarni aniqlash uchun oldindan belgilangan lingvistik qoidalaridan foydalanadi. Ushbu qoidalar sintaktik tuzilma, shaxs-son moslashuvi va jins kabi omillarni hisobga oladi. Qoidaga asoslangan usullar koreferensiyani hal etishni to'liq qamrab olmaydi. Tabiiy til ijtimoiylik xususiyatiga ega murakkab hodisa bo'lganligi uchun, qoidalarga bo'ysinmaydigan holatlar ham uchraydi.

Ushbu modellar qaysi eslatmalar bir obyektga tegishli ekanligini aniqlash uchun oldindan belgilangan qoidalar va usullar to'plamiga tayanadi. Qoidalarga asoslangan keng tarqalgan modellardan biri, Hobbsning olmoshlarni aniqlash algoritmidir. Biroq, boshqa qoidalarga asoslangan modellar mavjud bo'lib, ular ot ifodalari, tegishli otlar va boshqa turdagi ifodalar o'rtasidagi bog'liqlikni hal qilishga qaratilgan. Ushbu qoidaga asoslangan modellar odatda matndagi eslatmalarning sintaktik va semantik xususiyatlarini tahlil qilib, ularning asosiy munosabatlarini aniqlaydi. Ular asosiy qarorlarni qabul qilish uchun grammatik shaxs-son kelishuvi, miqdor, jins kelishuvi va semantik muvofiqlik kabi omillarni hisobga oladi.

Qoidalarga asoslangan modellar ma'lum kontekstlarda samarali bo'lgan bo'lsa-da, ular ham cheklovlarga ega. Ular noaniq yoki murakkab lingvistik konstruktsiyalarga duch kelishi va koreferensiya muammosini hal qilishga ta'sir qiluvchi lingvistik hodisalarni to'liqligicha qamrab ololmasligi mumkin. Garchi, qoidaga asoslangan modellar usulining cheklovlari mavjud bo'lsa-da, ular tabiiy tilni qayta ishlashda asosiy koreferensiya yechimlarini aniqlash usullarining muhim tarkibiy qismi bo'lib qolmoqda.

Hobbs algoritmi [3] gapdagi olmoshning ishorasini aniqlash uchun tabiiy tilni qayta ishlashda qo'llaniladigan usuldir. Bu olmosh va uning atrofidagi so'zlarning kontekstini tahlil qilish uchun bir qator qoidalar va metoddan foydalanadi. Ushbu algoritmi, ayniqsa, so'zlar va ularning referentlari o'rtasidagi munosabatlarni tushunish va to'g'ri talqin qilish uchun juda muhim bo'lgan mashina tarjimasini, ma'lumotni qidirish va matnni umumlashtirish kabi vazifalarda foydalidir.

Algoritm 1978-yilda Wiilliam A.Woods tomonidan taklif qilingan va keyinchalik 1979-yilda Jerry R.Hobbs tomonidan takomillashtirilgan. O'shandan beri u tabiiy tilni qayta ishlashda asosiy texnikaga aylandi va turli ilovalarda keng qo'llanildi.

Xobbs algoritmining asosiy xususiyatlaridan biri uning olmoshlarni yechishda ham sintaktik, ham semantik axborotni hisobga olish qobiliyatidir. Unda potentsial oldingi gapning grammatik roli, olmoshga yaqinligi, ikkalasi o'rtasidagi semantik moslik kabi omillar ko'rib chiqiladi.

Algoritmni qo'llash uchun odatda jumlaning sintaktik tuzilishini tahlil qilish va olmoshning potentsial oldingilarini aniqlash uchun tahlilchi ishlatiladi. Keyin algoritm potentsial oldingilarni referent bo'lish ehtimoli asosida tartiblash uchun o'z qoidalari va evristikasini qo'llaydi.

Xobbsning algoritmi tabiiy tilni qayta ishlash vazifalarida olmoshlarning aniqligini oshirishda samarali ekanligi isbotlangan. Biroq, bu cheklovlardan holi emas va tadqiqotchilar murakkab lingvistik kontekstlarda olmoshlarning yechimini yanada yaxshilash uchun yangi yondashuvlarni o'rganishda davom etmoqdalar.

Umuman olganda, Xobbsning olmoshlarni aniqlash algoritmi tabiiy tilni qayta ishlashda muhim vosita bo'lib, matndagi olmoshlarning murojaatlarini aniqlashning tizimli va ishonchli usulini ta'minlaydi. Uning doimiy rivojlanishi va takomillashtirilishi turli NLP ilovalaridagi yutuqlarga hissa qo'shishi mumkin.

2. Mashinali o'rganish yondashuvlari: Nazorat qilinadigan mashinani o'rganish modellari asosiy aloqalarni bashorat qilish uchun etiketli ma'lumotlarga o'rgatilgan bo'lishi mumkin. Ushbu modellar haqida ma'lumot berish uchun sintaktik va semantik o'xshashlik, so'zlarni joylashtirish va jins kelishuvi kabi xususiyatlardan foydalaniladi. Mashinali o'rganish algoritmlariga misollar orasida qarorlar to'plami, qo'llab-quvvatlovchi vektor mashinalari va logistik regressiya kiradi.

3. Chuqur o'rganish modellari: Chuqur o'rganish, xususan, neyron tarmoqlar, koreferensiya yechim rezolyutsiyasida inqilob qildi. Neyron tarmoqlar, shu jumladan takroriy neyron tarmoqlari (RNNs) va transformatorga asoslangan modellar matndagi murakkab naqshlar va bog'liqliklarni olishlari mumkin. BERT va uning variantlari kabi zamonaviy modellar koreferensiya yechimida ajoyib natijalarga erishdi.

4. Gibrid yondashuvlar: Ba'zi tizimlar ikkalasining kuchli tomonlaridan foydalanish uchun qoidalarga asoslangan va mashina o'rganish yoki chuqur o'rganish usullarini birlashtiradi. Misol uchun, qoidaga asoslangan usullardan dastlabki yadroviy klasterlarni yaratish uchun foydalanish mumkin, ular keyinchalik mashinani o'rganish yoki chuqur o'rganish modellari yordamida takomillashtiriladi.

Koreferensiya yechimidagi qiyinchiliklar

Koreferensiya yechimi quyidagi omillar tufayli qiyin NLP vazifasidir[1]:

1. Noaniqlik: Til tabiatan noaniq va bir nechta obyektlar yoki tushunchalar o'xshash atributlarga ega bo'lsa, asosiy ma'lumotlarni hal qilish qiyin bo'ladi.

2. Ichki ma'lumotnomalar: Matn ko'pincha ichki o'rnatilgan ko'rsatmalarni o'z ichiga oladi, bu yerda bir obyekt boshqa obyektga ishora qiladi, bu esa o'z navbatida boshqa obyektga ishora qiladi. Bunday ichki o'zaro munosabatlarni boshqarish ilg'or texnikani talab qiladi.

3. Uzoq masofaga bog'liqliklar: Bir nechta jumalarni yoki hatto paragraflarni qamrab oluvchi asosiy ma'lumotlarni hal qilish murakkab vazifa bo'lib, kengaytirilgan nutq orqali kontekstni tushunishni talab qiladi.

4. Jahon bilimlari: Ma'lumotnomalarni tushunish uchun ko'pincha dunyo bilimlari talab qilinadi, ular trening ma'lumotlarida mavjud bo'lmasligi mumkin. Masalan, “president” ma'lum bir shaxsga tegishli ekanligini bilish tashqi bilimlarni talab qiladi.

5. Pronominal va Null Anaphora: Ba'zi tillarda olmoshlar o'tkazib yuborilgan qo'zg'atuvchi hodisalar mavjud va koreferensiya yechimi aniq bo'lmasligi mumkin. Null anaforani aniqlash ayniqsa qiyin.

Xulosa

Xulosa qilib aytadigan bo'lsak, koreferensiya yechimi tabiiy tilni qayta ishlashning muhim tarkibiy qismi bo'lib, mashinalarga inson tilini yanada samaraliroq tushunish va talqin qilish imkonini beradi. U matnning uyg'unligini osonlashtiradi, ma'lumot olishda yordam beradi va NLP vazifalarining keng doirasi uchun ilovalarga ega. NLP tadqiqotlari va texnologiyasi rivojlanishda davom etar ekan, koreferensiya yechim o'lchamlari mashinalarga inson tilini aniqroq va ravonroq tushunish va yaratishga imkon beradigan hal qiluvchi jihat bo'lib qolmoqda.

Foydalanilgan adabiyotlar:

1. <https://medium.com/aimonks/coreference-resolution-in-natural-language-processing-nlp-5ba4f570bffe>
2. <https://medium.com/analytics-vidhya/hobbs-algorithm-pronoun-resolution-7620aa1af538>
3. <https://galhever.medium.com/a-review-to-co-reference-resolution-models-f44b4360a00>