

SUN'IY INTELLEKT YORDAMIDA TABIIY TILNI TUSHUNISH USULLARI VA ALGORITMLARI TAHLILI

Sulaymonova Fozila Mo'minjon qizi

Muhammad al-Xorazmiy nomidagi

Toshkent Axborot Texnologiyalari universiteti magistranti

sulaymonovaf24@gmail.com

Annotatsiya. Ushbu maqolada sun'iy intellekt yordamida tabiiy tilni tushunish usullari va algoritmlari haqida fikr yuritilgan. Jumladan, qoidalarga asoslangan tizimlar (Rule based systems) va statistik usullar (xususan, mashinali o'qitish) ning qiyosiy tahlili keltirilgan.

Kalit so'zlar: sun'iy intellekt (AI), mashinali o'qitish algoritmlari, neyron tarmoqlar, tabiiy tilni tushunish (NLU – natural language understanding), nomli obyektning tanib olish (NER), Rule based systems, til xususiyatlari, hisoblash tilshunosligi.

Annotation. This article discusses methods and algorithms for understanding natural language with the help of artificial intelligence. In particular, a comparative analysis of rule-based systems and statistical methods (particularly, machine learning) is presented.

Keywords: artificial intelligence (AI), machine learning algorithms, neural networks, natural language understanding (NLU), named object recognition, Rule based systems, language features, computational linguistics.

KIRISH. NLU tabiiy tilni tushunishni anglatadi. Bu sun'iy intellekt (AI) va hisoblash tilshunosligi (computational linguistics) ning kichik sohasi bo'lib, u tabiiy tildan foydalangan holda kompyuterlar va odamlar o'rtasidagi o'zaro ta'sirga qaratilgan. NLU ning maqsadi kompyuterlarga inson tilini mazmunli va kontekstga mos keladigan tarzda tushunish, talqin qilish va javob berish imkonini berishdir. NLU sintaksis, semantika va pragmatika kabi tilning turli jihatlarini tushuna oladigan algoritmlar va modellarni ishlab chiqishni o'z ichiga oladi. Bu matnni tushunish (text comprehension), his-tuyg'ularni tahlil qilish (sentiment analysis), nomli obyektning aniqlash (named entity recognition) va tilni tarjima qilish (language translation) kabi vazifalarni o'z ichiga oladi. NLU virtual yordamchilar, chatbotlar va til tarjimasini xizmatlari kabi inson tilini tushunish va qayta ishlash qobiliyatini talab qiluvchi ilovalar va tizimlarning muhim tarkibiy qismidir.

MATERIALLAR VA USULLAR

Tabiiy tilni tushunishda (NLU) turli xil algoritmlar va yondashuvlar qo'llaniladi. Qaysi algoritmi yoki usulni tanlash ko'pincha muayyan vazifa yoki dasturga bog'liq. Quyida bir nechta umumiy NLU algoritmlari va texnologiyalarini ko'rib chiqamiz:

1. Qoidalarga asoslangan tizimlar (Rule-Based Systems): Bu tizimlar matndan ma'lumot olish uchun oldindan belgilangan qoidalar va namuna(shakl)lardan foydalanadi. Qoidalar odatda tilshunoslar yoki domen mutaxassislari tomonidan muayyan lingvistik namuna(shakl)larni qo'lga kiritish uchun ishlab chiqariladi.

Sintaksis qoidalar: Gapnlarni tahlil qilish uchun grammatik qoidalar va sintaktik tuzilmalarni aniqlash.

Semantik qoidalar: so'zlar va iboralar o'rtasidagi ma'no va munosabatlarni aniqlash uchun qoidalarni o'rnatish.

Namuna (shakl)ni moslashtirish: obyektlni, munosabatlarni yoki muayyan til konstruktsiyalarini aniqlash uchun oldindan belgilangan namunalardan foydalanish.

2. Statistlik usullar:

Mashinali o'qitish: Ma'lumotlardagi shakl va munosabatlarni o'rganadigan algoritmlardan foydalanish.

N-gramm modellari: Keyingi so'zni bashorat qilish yoki shakllarni aniqlash uchun n ta so'z ketma-ketligini tahlil qilish.

Yashirin Markov modellari (HMM): Nutqni aniqlash va nutqning bir qismini belgilash kabi vazifalar uchun foydali bo'lgan ketma-ket ma'lumotlarni modellashtirish.

3. Mashinali o'qitish algoritmlari (Machine Learning Algorithms):

Nazorat ostidagi o'qitish (Supervised learning): matnni tasniflash, hissiyotlarni tahlil qilish va nomli ob'yektni tanib olish kabi vazifalar uchun yorliqli ma'lumotlar to'plamida modellarni o'qitish. Quyidagi algoritmlar Supervised learning ga misol bo'ladi:

- Naive Bayes: Matnlarni tasniflash va hissiyotlarni tahlil qilish uchun keng tarqalgan ehtimollik algoritmi.

- Vektorli mashinalarni qo'llab-quvvatlash (Support Vector Machines - SVM): Matnni tasniflash va nomli ob'yektlarni tanib olish kabi vazifalar uchun samarali.

- Qaror daraxtlari va tasodifiy o'rmonlar (Decision Trees and Random Forests): Tasniflash yoki ierarxik qarorlar qabul qilishni talab qiladigan vazifalar uchun foydalidir.

Nazoratsiz o'qitish (Unsupervised learning): yorliqsiz ma'lumotlardagi shakllarni aniqlash uchun klasterlash va mavzuni modellashtirish.

Yarim nazorat ostida o'qitish (Semi-supervised learning): modellarni o'qitish uchun yorliqli va yorliqsiz ma'lumotlarni birlashtirish.

4. Neyron tarmoqlari (Neural Networks): Qaytalanuvchi neyron tarmoqlari (Recurrent neural network - RNN) va Transformatorlar kabi chuqur o'qitish modellari turli NLU vazifalarida, jumladan, tilni modellashtirish, his-tuyg'ularni tahlil qilish va nomlarni tanib olishda muvaffaqiyatli bo'ladi.

Takroriy neyron tarmoqlari (RNN): Tilni modellashtirish va hissiyotlarni tahlil qilish kabi vazifalar uchun mos keladigan ketma-ket ma'lumotlarni qayta ishlash.

Uzoq qisqa muddatli xotira (Long short-term memory LSTM): RNN turi ma'lumotlarning uzoq muddatli bog'liqligini qo'lga kiritish uchun mo'ljallangan.

Transformatorlar: BERT va GPT kabi e'tiborga asoslangan modellar turli NLU vazifalarida eng so'nggi natijalarga erishdi.

5. Named Entity Recognition (NER): NER matndagi obyektlarni (masalan, odamlar, tashkilotlar, joylar nomlari) identifikatsiyalash va tasniflashni o'z ichiga olgan NLUdagi o'ziga xos vazifadir. NER uchun odatda shartli tasodifiy maydonlar (Conditional random fields - CRF) va ikki tomonlama (bidirectional) LSTMlar qo'llaniladi.

6. So'zlarni joylashtirish (Word Embeddings): Word2Vec, GloVe va fastText kabi texnikalar vektor bo'shliqlarida so'zlarni ifodalab, so'zlar orasidagi semantik munosabatlarni aks ettiradi. Ushbu o'rnatishlar ko'pincha quyi oqim (downstream) NLU vazifalari uchun xususiyatlar sifatida ishlatiladi.

7. Semantik rol yorlig'i (Semantic Role Labeling - SRL): Bu gapdagi so'zlar o'rtasidagi munosabatlarni aniqlash va subyekt, obyekt yoki predikat kabi rollarni belgilashni o'z ichiga oladi. Grafikga asoslangan modellar va chuqur o'qitish yondashuvlarida ko'pincha SRL qo'llaniladi.

8. Asosiy rezolyutsiya (Coreference Resolution): Bu vazifa matndagi ikki yoki undan ortiq iboralar qachon bir xil obyektga tegishli ekanligini aniqlashni o'z ichiga oladi. Bunga ko'pincha neyron tarmoqlar kabi mashinali o'qitish modellari yordamida yondashadi.

9. Transfer Learning: Katta hajmdagi ma'lumotlar bo'yicha katta til modellarini oldindan o'qitish va ularni muayyan vazifalar uchun nozik sozlash odatiy amaliyotga aylangan. BERT (Bidirectional Encoder Representations from Transformers - Transformatorlardan ikki tomonlama kodlovchi vakilliklari) va GPT (Generative Pre-trained Transformer - generativ oldindan tayyorlangan transformator) kabi modellar bir qator NLU vazifalarida juda yaxshi natijani namoyish etdi.

10. Ansambl usullari (Ensemble Methods): Umumiy ishlash va mustahkamlikni yaxshilash uchun bir nechta modellarni birlashtirish.

11. Tushuntirish va talqin qilish usullari: Diqqat mexanizmlari va xususiyat ahamiyatini tahlil qilish kabi NLU modellari tomonidan qabul qilingan qarorlarni tushunish va izohlash usullari. Transformator modellari tomonidan ommalashgan

diqqat mexanizmlari modellarga kirish ketma-ketligining turli qismlariga e'tibor qaratish imkonini beradi va tabiiy tilda kontekstni tushunishni yaxshilaydi.

Bu faqat bir nechta misollar va NLU sohasi dinamik bo'lib, davom etayotgan tadqiqot va ishlanmalar yangi texnika va takomillashtirishga olib keladi. Algoritmni tanlash ko'pincha NLU vazifasining o'ziga xos talablari va xususiyatlariga ko'ra amalga oshiriladi.

Tabiiy tilni tushunish (NLU) kontekstida qoidalarga asoslangan tizimlar (Rule based systems) va statistik usullar (jumladan, mashinali o'qitish) o'rtasidagi tanlov turli omillarga asoslanadi va ularning taqqoslanishi quyidagicha:

	Qoidalarga asoslangan tizimlar	Statistik usullar (mashinali o'qitish)
Ijobiy tomonlari	Talqin qilinish qobiliyati: Qoidalarga asoslangan tizimlar domen mutaxassislari tomonidan oson tushunilishi va o'zgartirilishi mumkin bo'lgan aniq, izohlanadigan qoidalarni ta'minlaydi.	Umumlashtirish: Statistik usullar, xususan, mashinali o'qitish modellari ma'lumotlarni umumlashtirishda ustunlik qiladi, bu ularga keng ko'lamli lingvistik shakllarga moslashish imkonini beradi.
	Domenning o'ziga xosligi: Qoidalarga asoslangan tizimlar aniq bilimlar mavjud bo'lgan va qoidalarga kodlanishi mumkin bo'lgan domenga xos ilovalar uchun juda mos keladi.	Moslashuvchanlik: Mashinali o'qitish modellari yangi ma'lumotlarga avtomatik ravishda moslasha oladi, bu ularni tildan foydalanish yoki turli domenlar bilan bog'liq vazifalar uchun mos qiladi.
	Aniqlik va nazorat: Ishlab chiquvchilar tizim xatti-harakatlarini aniq nazorat qiladi, bu esa tizimning muayyan talablarga rioya	Avtomatik xususiyatni o'rganish: Mashinali o'qitish modellari qoidalarga asoslangan tizimlarda ko'rinmasligi mumkin bo'lgan murakkab munosabatlarni yozib olgan holda ma'lumotlardan

	qilishini osonlashtiradi.	tegishli xususiyatlarni avtomatik o'rganishi mumkin.
Kamchiliklari	Cheklangan umumlashtirish: Qoidalarga asoslangan tizimlar lingvistik o'zgaruvchanlikni boshqarishda qiynalishi va qoidalar bilan aniq qamrab olinmagan yangi, ko'rinmaydigan shakllarni yaxshi umumlashtirishi mumkin.	Qora quti tabiati: Ba'zi ilg'or mashinali o'qitish modellari qora qutilar sifatida ko'rilishi mumkin, bu esa ularning qaror qabul qilish jarayonini izohlashni qiyinlashtiradi, bu ma'lum ilovalarda hal qiluvchi ahamiyatga ega.
	Xizmat ko'rsatish uchun qo'shimcha xarajatlar: Qoidalarga asoslangan tizimlarga texnik xizmat ko'rsatish ko'p mehnat talab qilishi mumkin, ayniqsa qoidalar soni ortib borayotganligi sababli yoki tizim rivojlanayotgan til shakllariga moslashishi kerak.	Ma'lumotlarga bog'liqlilik: Statistik modellar ko'p jihatdan o'quv ma'lumotlarining sifati va representativligiga bog'liq. Noto'g'ri yoki representativ bo'lmagan ma'lumotlar noto'g'ri modellarga olib kelishi mumkin.
		Resurs intensivligi: Murakkab mashinali o'qitish modellarini o'qitish va qo'llash hisoblash jihatidan qimmat bo'lishi va katta resurslarni talab qilishi mumkin.

Ko'pgina hollarda ikkalasining kuchli tomonlarini ishlatish uchun qoidalarga asoslangan tizimlarni statistik usullar bilan birlashtirgan gibrid yondashuv qo'llaniladi. Misol uchun, qoidalarga asoslangan tizim qoidalar aniq bo'lgan muayyan holatlarni ko'rib chiqishi mumkin, mashinali o'qitish modeli esa yanada moslashuvchan va mustahkam NLU tizimini yaratishga imkon beruvchi murakkab va noaniq holatlarni ko'rib chiqishi mumkin.

XULOSA. Xulosa qilib aytadigan bo'lsak, tabiiy tilni tushunishda qoidalarga asoslangan tizimlardan til shakllari yaxshi tushunilgan va aniq qoidalar belgilanishi mumkin bo'lgan holatlarda foydalangan ma'qul. Bu yuridik yoki tibbiy sohalar kabi ekspert bilimlari muhim bo'lgan vazifalarni hal qilishda qo'llaniladi. Murakkab til shakllari bilan ishlashda statistik usullarni tanlash o'rinlidir, va bunday tizim rivojlanayotgan turli til shakllariga moslashuvchanlik xususiyatiga ega. Bu ko'pincha his-tuyg'ularni tahlil qilish, niyatni aniqlash va obyektini tanib olish kabi vazifalar uchun xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR RO'YXATI:

1. Uzbek Speech Synthesis Using Deep Learning Algorithms. MI Abdullaeva, DB Juraev, MM Ochilov, MF Rakhimov. *Intelligent Human Computer Interaction: 14th International Conference, IHCI 2022, Tashkent, Uzbekistan, October 20–22, 2022.*
2. Agglutinativ tillar uchun pos teglash va stemming masalasi (turk, uyg'ur, o'zbek tillari misolida). B Elov, Sh Hamroyeva, O Abdullayeva, Z Husainova, N Xudayberganov. *O'ZBEKISTON: TIL VA MADANIYAT. (ISSN 2181-922X) 2023 Vol. 2*
3. Sun'iy intellekt algoritmlari yordamida O'zbek tabiiy tilini tushunishning ahamiyati. F Sulaymonova. *Educational Research in Universal Sciences. (ISSN: 2181-3515) VOLUME 2 | SPECIAL ISSUE 2 | 2023*
4. <https://monkeylearn.com/blog/natural-language-understanding>