O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi O'zbek tilini rivojlantirish jamg'armasi

O'zbekiston Respublikasi Raqamli ta'lim texnologiyalar vazirligi

O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi

Sankt-Peterburg davlat universiteti

Istanbul texnika universiteti

Toshkent davlat o'zbek tili va adabiyoti universiteti

Samarqand davlat universiteti

Toshkent axborot texnologiyalari universiteti Samarqand fiiliali

MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARIUNIVERSITETI SAMARQAND FILIALI

“O‘ZBEK TILINING MILLIY KORPUSI: MUAMMOLAR VA VAZIFALAR”

MAVZUSIDAGI XALQARO
ILMIY-AMALIY ANJUMANI MA’RUZALAR TO‘PLAMI
2023 yil 25 mart

PROCEEDINGS
of the International Scientific and Practical Conference

“THE NATIONAL CORPUS OF THE UZBEK LANGUAGE: PROBLEMS AND TASKS”

March 25, 2023

O‘ZBEKISTON RESPUBLIKASI
RAQAMLI TEXNOLOGIYALAR VAZIRLIGI
OLIY TA’LIM, FAN VA INNOVATSIYALAR VAZIRLIGI

MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI SAMARQAND FILIALI

“O‘ZBEK TILINING MILLIY KORPUSI: MUAMMOLAR VA VAZIFALAR” MAVZUSIDAGI XALQARO ILMIY-AMALIY ANJUMANI MA’RUZALAR TO‘PLAMI

2023 yil 25 mart

PROCEEDINGS

of the International Scientific and Practical Conference
“THE NATIONAL CORPUS OF THE UZBEK LANGUAGE: PROBLEMS AND TASKS”
March 25, 2023

SAMARQAND 2023

O‘zbek tilining milliy korpusi: muammolar va vazifalar / Xalqaro ilmiy-amaliy konferensiya to‘plami. – Samarqand:
TATU Samarqand filiali, 25.03.2023. – 394 b.

Mazkur to‘plamdan kompyuter lingvistikasi sohasiga doir tadqiqot natijalari, filologik dasturiy ta’minotlar, ularning imkoniyatlari va afzalliklari, tabiiy tilni qayta ishlashga oid masalalar, semantik tizim bazalari, til va adabiy korpuslar, lingvistik ta’minot manbalari, mashina tarjimasida lingvistik muammolar, lingvodidaktik dasturlar va platformalar, terminologik lug‘atlarni yaratish masalalariga oid materiallar joy olgan.

Ushbu ilmiy to‘plamdan kompyuter lingvistikasi sohasi vakillari, tadqiqotchilar, mustaqil izlanuvchilar, magistrantlar va kelajak sohalariga qiziquvchi talabalar foydalanishlari mumkin.

Mas’ul muharrir:
Karimov S.A. Sharof Rashidov nomidagi Samarqand davlat universiteti, professor, f.f.d.

Taqrizchilar:
Xujayorov I.Sh. Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti Samarqand filiali, dotsent, PhD.
Mahmadiyev Sh.S. Sharof Rashidov nomidagi Samarqand davlat universiteti, dotsent, f.f.n.

TURLI SO‘Z TURKUMLARI ORASIDAGI OMONIMIYANI NAÏVE BAYES KALSSIFIKTORI YORDAMIDA ANIQLASH
DETERMINING HOMONY BETWEEN DIFFERENT WORD GROUPS USING THE NAÏVE BAYES CLASSIFIER

*Elov Botir Boltayevich, **Axmedova Xolisxon Ilxomovna
*Texnika fanlari bo‘yicha (PhD), dotsent, Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti, Toshkent, O‘zbekiston
elov@navoiy-uni.uz
**Tayanch doktorant, Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti, Toshkent, O‘zbekiston
xolisa0629@gmail.com

Annotatsiya: Tabiiy tilni qayta ishlash jarayonlarining dolzar masalalaridan biri bu omonim so‘zlarni semantik farqlash. Omonim so‘zlarni semantik farqlashda Qoidalarga asoslangan, Stoxastik, Mashinali o‘qitishga asoslangan hamda Chuqur o‘qitish usullaridan foydalanilanish mumkin. Naïve Bayes klassifikatori ana shunday usullardan biri. O‘zbek tilidagi turli va grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlari orasidagi omonimiyani bartaraf etishda Naïve Bayes klassifikatori soddaligi va tezkorligi bilan boshqa usullardan ajralib turadi. Naïve Bayes klassifikatoridan foydalanish uchun dastlab so‘z turkumlarini tasniflovchi parametrlarni ajratib olish va kuzatuvlar olib borish zarur.

Annotation: One of the important issues of natural language processing is the semantic differentiation of homonyms. Rule-based, Stochastic, Machine learning and Deep learning methods can be used for semantic differentiation of homonyms. Naïve Bayes classifier is one such method. In eliminating homonymy between different and grammatically similar word groups in the Uzbek language, the Naïve Bayes classifier is distinguished from other methods by its simplicity and speed. To use the Naïve Bayes classifier, it is first necessary to extract the parameters that classify word groups and make observations.

Kalit so‘zlar: Tabiiy tilni qayta ishlash jarayonlari, omonimiya, Naïve Bayes klassifikatori, aprior ehtimollik, aposterior ehtimollik, semantik tahlil.

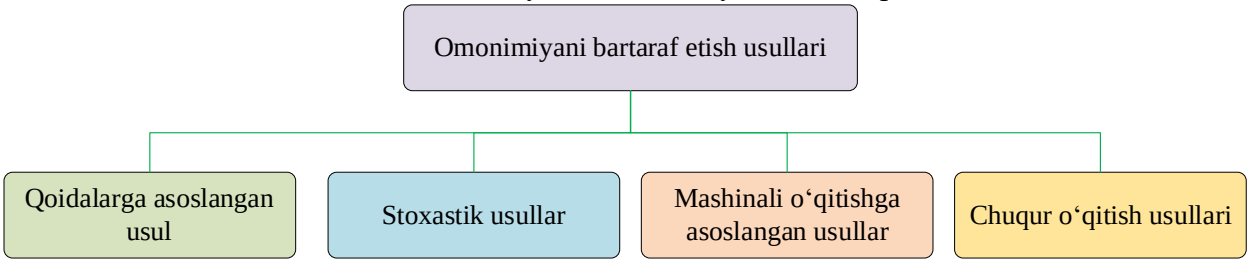
Keywords: Natural language processing, homonymy, Naïve Bayes classifier, a priori probability, a posteriori probability, semantic analysis.

Matnlarni semantik tahlil qilish tabiiy tilni qayta ishlash sohasining eng muhim masalalaridan biri, ya’ni matndagi so‘zlarni ma’nosini aniqlash orqali gapni ma’nosini, gap orqali matnning ma’nosini aniqlash dolzarb mavzulardan biri hisoblanadi. Matnlarni ma’nosini aniqlash orqali ularning qisqacha mazmunini anglatuvchi matn (annotatsiya) hosil qilinadi. Matnlarni semantik tahlil qilish ikki qismga bo‘linadi:

- So‘z ma’nosini aniqlash (Word sense disambiguation (WSD))
  - Sinonim
  - Omonim
  - Polisemantik
  - Polifunksional
  - Antonim
  - Paronim
  - Giponim
  - Meronim
  - Giperonim
- Sentiment tahlil qilishi
  - Matndagi his-tuyg‘uni aniqlash

- Matnning maqsadini aniqlash
- Matnning sohasini aniqlash

So‘z ma’nosini aniqlashning muhim masalalaridan biri omonimiyani bartaraf etish masalasidir. Shu o‘ringa savol paydo bo‘ladi. Nega aynan omonim so‘zlarni aniqlash dolzarb mavzu? Omonim so‘z bir so‘z turkumi doirasida ham turli so‘z turkumlari doirasida ham turli ma’nolarni anglatishi mumkin. Shu omonim so‘zning aynan qaysi ma’noni anglatayotganligini aniqlash semantik tahlil uchun juda muhim. Jahon tajribasi shuni ko‘rsatadiki, omonimiyani bartaraf etish masalasi bir nechta usullar yordamida o‘z yechimini topishi mumkin.



1-rasm. *Omonimiyani bartaraf etish usullari*

Qoidalarga asoslangan usul– so‘zlarni til grammatikasi qoidalari yordamida semantik tahlil qilish nazarda tutilgan.

Stoxastik usul– masalani yechishda so‘zlarning grammatik parametrlarini tasniflash orqali foydalaniladi. Bu parametrlar turli tabiiy tillarda turlicha tanlab olinadi.

Mashinali o‘qitishga asoslangan usul –sun’iy intellektga asoslangan. Ushbu usul statistic usullarga qo‘shimcha bo‘lib, jahon hozirda omonimiyani aniqlashda keng qo‘llanilmoqda.

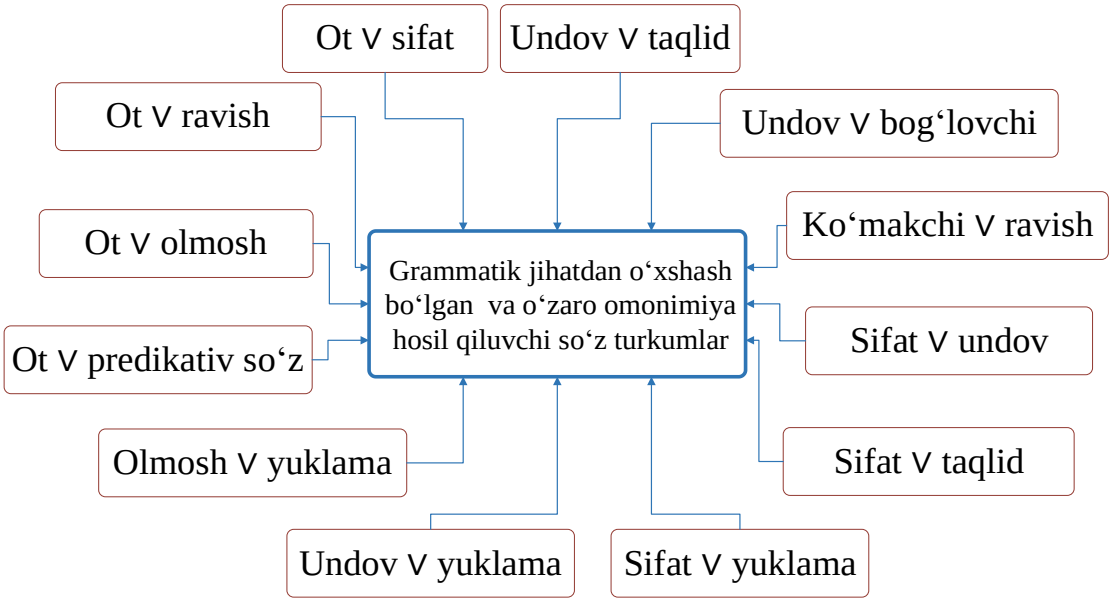
Chuqur o‘qitish usullari – (shuningdek, chuqur tuzilgan o‘rganish deb ham ataladi) sun’iy neyron tarmoqlarga asoslangan mashinali o‘rganish usullarining kengroq oilasining bir qismidir.

Naïve Bayes klassifikatori Mashinali o‘qitishga asoslangan usullarning tarkibiy qismi bo‘lib tabiiy tilni qayta ishlash jarayonlaridagi ko‘plab masalalarini yechishga yordam beradi. Masalan, o‘sha mashhur, kommentariyalarni sharhlash masalasi [1], Fake ma’lumotlarni aniqlash [2], matnlarni sentiment tahlil qilish [3,4,5,6] kabi masalalar shular jumlasidandir. Tabiiy tilni qayta ishlash jarayonlarining yana bir muhim vazifalaridan biri so‘z ma’nosini aniqlash bo‘lib, bu masalani yechishda ham Naïve Bayes klassifikatori alohida o‘rin tutadi. Pal, A. R. va boshqalar tomonidan Bengal tilida so‘z ma’nolarini aniqlashda mashinali o‘qitishning 4 ta Qarorlar daraxti (Decision Tree-DT) usuli, Support Vector Machine (SVM) usuli, Sun’iy neyron tarmoqlari (Artificial Neural Network -ANN) usuli va Naïve Bayes (NB) usullari qo‘llanilgan. Ushbu asosiy strategiyalar mos ravishda 63,84%, 76,9%, 76,23% va 80,23% aniq natijalarni berdi [7]. Undan tashqari jinoyat ishlari bo‘yicha hisobotlar tayyorlashda ham ushbu klassifikatordan keng foydalanilgan va 87- 93,74% gacha aniqlikka erishilganligini ko‘rish mumkin [8].

O‘zbek tilidagi omonim so‘zlarni semantik farqlashda ham ushbu usullardan foydalanish maqsadga muvofiq. O‘zbek tilidagi omonim so‘zlarni semantik farqlashda ularni so‘z turkumlari doirasida uchrashiga ko‘ra ikkita guruhga ajratildi: *bir so‘z turkumi doirasidagi va turli so‘z turkumi doirasidagi omonim so‘zlar.*

Bir so‘z turkumi doirasidagi omonimiyani aniqlashda *mashinali o‘qitishga asoslangan* usuldan foydalanish qulayligi oldingi maqolalarimizda keltirib o‘tgandik [9]. Turli so‘z turkumlari doirasidagi omonim so‘zlar ham o‘z navbatida 2, 3 va 4 so‘z turkumlari doirasida omonimlik hosil qilishiga qarab 3 ta guruhga ajratish mumkin [10,11]. Turli so‘z turkumlari doirasidagi omonim so‘zlarni *Qoidalarga asoslangan usullar, statistik ma’lumotlarga asoslangan* usullar va Mashinali o‘qitishga asoslangan usullar yordamida farqlash mumkin. So‘z turkumlari orasida grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlari borki, ularni farqlovchi qoidalar mavjud emas. Agar bor bo‘lsa ham qat’iy emas. Masalan, *ot va sifat* so‘z turkumlari grammatik jihatdan o‘xshash so‘z

turkumlari deyish mumkin. Ya’ni, sifat so‘z turkumi otlashish xususiyatiga ega, bu degani ikkala so‘z turkumiga oid so‘zga bir xil qo‘shimcha qo‘shilishi mumkin. O‘zbek tilida grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlari orasida omonimiya hosil qiluvchi so‘zlarni 2-rasmdagi kabi guruhlarga ajratish mumkin.



2-rasm. *Grammatik jihatdan o‘xshash bo‘lgan va o‘zaro omonimiya hosil qiluvchi so‘z turkumlari*

Ma’lumki,2-rasmda keltirilgan so‘z turkumlari orasidagi omonimiya hosil qiluvchi so‘zlarni semantik farqlashda statistik ma’lumotlardan olinadi. [11,12] maqolada Chastotali usul yordamida aniqlashni keltirgan edik. Omonimiyani chastotali usul yordamida aniqlanish uchun katta hajmdagi ma’lumotlarga ehtiyoj paydo bo‘ladi. Huddi shunday bigdata ustida sodda va tezkor amallar bajarishni ta’minlovchi usullardan biri Naïve Bayes klassifikatori qulay.

Naive Bayes algortimi – Bayes teoremasiga asoslangan statistik tasniflash usuli deya ta’riflash ham mumkin [13]. *Statistik tasniflash*– tasniflangan matnlar ustida o‘tkazilgan kuzatuv natijalaridir.

Naive Bayes klassifikatori sinfdagi ma’lum bir xususiyatning ta’siri boshqa xususiyatlardan mustaqil ekanligini taxmin qiladi. Bu xususiyatlar o‘zaro bog‘liq bo‘lsa-da, ular mustaqil ravishda ham ko‘rib chiqiladi. Naive Bayes klassıfikatorı har bir xususıyat uchun ehtimolliklarni hisoblab chiqadi va eng yuqori ehtimollik bilan natijani tanlaydi.

$$P(class|data) = \frac{P(data|class)\cdot P(class)}{P(data)} \qquad (1)$$

– **P(class):** *class gipotezasining rost bo‘lish ehtimoli (ma’lumotlardan qat’iy nazar). Bu class ning aprior ehtimoli sifatida qo‘llaniladi.*

– **P(data):** *ma’lumotlarning ehtimolligi (gipotezadan qat’iy nazar). Bu aprior ehtimollik sifatida qo‘llaniladi.*

– **P(class|data):** *data ma’lumotlar asosida class gipoteza ehtimoli. Bu aposterior ehtimollik sifatida qo‘llaniladi.*

– **P(data|class):** *class gipotezasi to‘g‘ri bo‘lganligi sharti asosida ma’lumotlarning data ehtimoli. Bu aposterior ehtimollik sifatida qo‘llaniladi.*

Naive Bayes klassifikatori.Matnli ma’lumotlarni klassifikator orqali diskret belgilashlarga tasniflashimiz sababli, Naive Bayes algoritmi uchun *kirish funksiyalari to‘plami* hamda *ularga mos tegishli chiqish sinfiga* ega bo‘lamiz. Naive Bayes klassifikatori quyidagi formula yordamida ehtimollikni hisoblaydi:

$$P(y_1|x_1,x_2,x_3) = \frac{P(y_1)\cdot P(x_1|y_1)\cdot P(x_2|y_1)\cdot P(x_3|y_1)}{P(x_1)\cdot P(x_2)\cdot P(x_3)} \quad (2)$$

Ushbu tenglamadagi $P(y_1|x_1,x_2,x_3)$ qabul qilingan {x_1, x_2, x_3} ma’lumotlar asosida, chiqish sifatida y_1 bo‘lish ehtimolini anglatadi [14,15].

Agar bizning NLP masalamiz jami 2 ta sinfga ega bo‘lsin, ya’ni {y_1, y_2}. Endi yuqoridagi formuladan avval y_1 sodir bo‘lish ehtimolini, keyin esa y_2 sodir bo‘lish ehtimolini hisoblash lozim bo‘ladi [16, 17]. Qaysi bir ehtimoli yuqori bo‘lsa, bizning taxmin qilingan sinfimiz shu hisoblanadi. Naïve Bayes klassifikatori yordamida grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlari orasida omonimiya hosil qiluvchi so‘zlarni farqlash masalasini yechilishini ko‘rib chiqsak. Buning bizga so‘z turkumlarni tasniflash xususiyatlar zarur. Buni so‘z turkumlarini grammatik xususiyatlaridan kelib chiqib aniqlash mumkin.

Faraz qilaylik, ot yoki sifat so‘z turkumlari orasidagi omonimiyani aniqlash masalasi qo‘yilgan bo‘lsin. Dastlab ot yoki sifat so‘z turkumlari uchun tasniflash parametrlari aniqlanishi lozim. Kuzatishlar shuni ko‘rsatadiki, ot va sifat so‘z turkumlari asosan o‘zak va tarkibida qo‘shimcha mavjud bo‘lgan holatlarda bo‘lishi mumkin, ya’ni *o‘zak va o‘zak +aff (affix).* Misol tariqasida, *ot yoki sifat* so‘z turkumlari doirasida omonimiya hosil qiluvchi *issiq* so‘zini keltiramiz.

O‘zbek tili korpusi ma’lumotlari orasida *issiq* so‘zini izlaganimizda 30782 ta o‘rinda ishtiroki kuzatildi. Kuzatuvlarda issiq so‘zi

- *Uy **issiq** bo’lgani uchun tutun yuqoriga ko’tarilmaydi.*
- *Onamning issiqqina bag‘rini, mehrini baribir hech narsa bosolmaydi-da.*
- *Jazirama issiqda sun’iy suv havzalarida cho‘milish bolalarga bir olam zavq berishi shubhasiz.*
- *Charchadim o‘qishdanmi, o‘ylovlardanmi, issiqdanmi, sovuqdanmi bilmayman.*

kabi so‘zshakllarda uchrash holatlari kuzatildi. Kuzatuvlar natijasida ot yoki sifat so‘z turkumlari orasida omonimlik hosil qiluvchi so‘zlarni tasniflovchi parametrlar ularni *o‘zak* va *o‘zak+aff* ekanligi aniqlandi. Demak x_1= *o‘zak*, x_2= *o‘zak+aff, y_1=sifat va y_2= ot*. Dastlab, o‘zak shaklida uchrash holatlari tahlil qilindi. Tahlillarga ko‘ra issiq so‘zi o‘zak shaklida asosan sifat so‘z turkumiga oid bo‘lishi kuzatildi. Kuzatuvlar natijasidan olingan statistik ma’lumotlar asosida *issiq* so‘zining aprior va aposterior ehtimolliklarni hisoblash jarayonini soddalashtirish uchun *chastota* va *ehtimollik jadvali*dan foydalanish mumkin. Ushbu jadvallarning ikkalasi ham aprior va aposterior ehtimolliklarni hisoblashda yordam beradi [18]. Chastotalar jadvali barcha xususiyatlar uchun teglarni o‘z ichiga oladi. Shuningdek, ikkita ehtimollik jadvali ham mavjud bo‘lib, 1-jadvalda yorliqlarning aprior ehtimoli, 2-jadvalda esa aposterior ehtimoli ko‘rsatilgan.

	Parametrlar	So‘z turkumi
1	O‘zak	Sifat
2	O‘zak +aff	Ot
3	O‘zak	Sifat
4	O‘zak	Sifat
5	O‘zak	Ot
6	O‘zak	Sifat
7	O‘zak +aff	Ot
8	O‘zak	Sifat
9	O‘zak +aff	Ot
10	O‘zak	Sifat
11	O‘zak +aff	Ot
12	O‘zak +aff	Ot
13	O‘zak	Sifat
14	O‘zak	Ot
…	…	

Parametrlar	Ot	Sifat
O‘zak	2	7
O‘zak +aff	4	1
Umumiy	6	8

1-Ehtimollik jadvali

Parametrlar	Ot	Sifat		
O‘zak	2	7	=9/14	0.64
O‘zak +aff	4	1	=5/14	0.36
umumiy	6	8		
	=6/14	=8/14		
	0.43	0.47		

2-Ehtimollik jadvali

Parametrlar	Ot	Sifat	Bo’lish ehtimoli (ot)	Bo’lmaslik ehtimoli (sifat)
O‘zak	2	7	2/9=0.22	7/9=078
O‘zak +aff	4	1	4/5=0.8	1/7=0.14
umumiy	6	8		

3-rasm. *Chastota va ehtimollik jadvallari*

Gap tarkibida uchragan ot yoki sifat so‘z turkumlari doirasiga omonimlik hosil qiluvchi so‘zlarning sifat so‘z turkumiga doir bo‘lish ehtimolini hisoblash jarayonini ko‘rib chiqamiz. Bu hisoblashlarni quyidagi jumla misolida ko‘rib chiqsak

*Uy **issiq** bo‘lganligi sababli tutun yuqoriga ko‘tarilmaydi.*

Demak, jumlada qatnashgan *issiq* so‘zi o‘zak so‘z, bu so‘z uchun x_1 va x_2 xususiyatlariga moslik ehtimolligini hisoblaymiz.

So‘zning sifat so‘z turkumiga oid bo‘lish ehtimolini hisoblash:

$$P(Sifat|O'zak) = \frac{P(O'zak|Sifat) \cdot P(Sifat)}{P(O'zak)} \qquad (3)$$

Aprior ehtimolliklarni hisoblash:

$P(\text{O'zak}) = 7/14 = 0.5$

$P(\text{sifat}) = 8/14 = 0.57$

Aposterior ehtimolliklarni hisoblash:

$P(O'zak|\text{Sifat}) = 7/9 = 0.78$

Aprior va Aposterior ehtimolliklar orqali **sifat so‘z turkumiga doir ekanlik ehtimoli**ni hisoblash:

$$P(Sifat|\text{O'zak}) = \frac{0.78 \cdot 0.57}{0.5} = 0.89\ (Yuqori)$$

Ot so‘z turkumiga oid bo‘lish ehtimolini hisoblash:

$$P(Ot|O'zak) = \frac{P(O'zak|Ot) \cdot P(Ot)}{P(O'zak)} \qquad (4)$$

Aprior ehtimolliklarni hisoblash:

$P(Ot) = 6/14 = 0.43$

$P(O'zak) = 9/14 = 0.64$

Aposterior ehtimolliklarni hisoblash:

$P(O'zak|Ot) = 2/9 = 0.22$

Aprior va Aposterior ehtimolliklar orqali so‘zning sifat bo‘lmaslik (ot bo‘lish)ehtimolini hisoblash :

$$P(Ot|O'zak) = \frac{0.22 \cdot 0.64}{0.43} = 0.33$$

Demak,

$P(Sifat|\text{O'zak}) = 0.89$

$P(Ot|O'zak)$=0.33

Hisoblashlar natijasiga ko‘ra “Uy ***issiq*** bo‘lganligi sababli tutun yuqoriga ko‘tarilmaydi” gapida keltirilgan *issiq* so‘zi 88% aniqlik bilan sifat so‘z turkumiga oid deya baholaymiz. Kuzatuvlar shuni ko‘rsatadiki *issiq* so‘zi x_2 *=o‘zak+aff* shaklida ham uchrashi mumkin.

Masalan,

***Issiqda** terlab-pishib sport bilan shug‘ullanish nafaqat yoqimsiz, balki sog‘liq uchun xavflidir.*

Jumlasidagi *issiqda* so‘zini qaysi so‘z turkumiga oid ekanlik ehtimolligini hisoblaganimizda $P(Sifat|\text{O'zak} + \text{aff}) = 0.35$ *va* $P(Ot|\text{O'zak} + \text{aff}) = 0.93$ qiymatlar hosil bo‘ldi. Bundan kelib chiqadiki “***Issiqda** terlab-pishib sport bilan shug‘ullanish nafaqat yoqimsiz, balki sog‘liq uchun xavflidir*” gapidagi *issiq* omonim so‘zi 93% ehtimollik bilan ot so‘z turkumiga oid ekanligi aniqlandi. Olib borilgan kuzatuvlar shuni ko‘rsatadiki, sifat yoki ot so‘z turkumi doirasidagi omonim so‘zning ot so‘z turkumi va sifat so‘z turkumiga oidlik hodisalari birgalikda bo‘lmagan hosidalar. Ya’ni x_1 va x_2 xususiyatlardan bir vaqtda (3) formuladan foydalanib bo‘lmaydi.

Xulosa o‘rnida shuni aytish mumkinki, o‘zbek tilida omonimiyani aniqlash masalasi murakkab jarayon. Omonimiyani aniqlashda ularni turli guruhlarga ajratish va ularning farqlovchi omillarga asosan turli usullardan foydalanish maqsadga muvofiq. Naïve Bayes klassifikatori ana shunday usullardan biri. Naïve Bayes klassifikatoridan o‘zbek tilidagi grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlari orasidagi omonimiyani bartaraf etishda foydalanildi. Turli so‘z turkumlari doirasidagi omonimiyani aniqlashda Naïve Bayes klassifikatoridan foydalanish uchun dastlab grammatik jihatdan o‘xshash bo‘lgan so‘z turkumlarini tasniflash parametrlari aniqlandi. Tasniflash parametrlarini inobatga olgan holda statistik ma’lumotlar olindi. Statistik ma’lumotlar asosida Naïve Bayes hisoblashlari amalga oshirildi va natijalar baholandi.

Foydalanilgan adabiyotlar ro‘yxati

1. Sourav Kunal, Arijit Saha, Aman Varma, Vivek Tiwari. Textual Dissection Of Live Twitter Reviews Using Naive Bayes. International Conference on Computational Intelligence and Data Science (ICCIDS 2018). Procedia Computer Science 132 (2018) 307–313.
2. Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. In 2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings (pp. 900–903). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/UKRCON.2017.8100379.
3. Bahri, S., Saputra, R. A., & Wajhillah, R. (2017). Analisa sentimen berbasis Natural Languange Processing (NLP) dengan Naïve-Bayes clasifier. Konferensi Nasional Ilmu Social & Technology, 1(1), 176–180. Retrieved from https://www.researchgate.net/
4. Rusli, N. L. I., Amir, A., Zahri, N. A. H., & Ahmad, R. B. (2019). Snake species identification by using natural language processing. Indonesian Journal of Electrical Engineering and Computer Science, 13(3), 999–1006. https://doi.org/10.11591/ijeecs.v13.i3.pp999-1006
5. Kaur, C. (2020). Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach. International Journal of Advanced Trends in Computer Science and Engineering, 9(4), 6303–6311. https://doi.org/10.30534/ijatcse/2020/310942020
6. Siddiqui, S., Rehman, M. A., Daudpota, S. M., & Waqas, A. (2019). Opinion mining: An approach to feature engineering. International Journal of Advanced Computer Science and Applications, 10(3), 159–165. https://doi.org/10.14569/IJACSA.2019.0100320
7. Pal, A. R., Saha, D., Naskar, S. K., & Dash, N. S. (2021). In search of a suitable method for disambiguation of word senses in Bengali. International Journal of Speech Technology,24(2), 439–454.https://doi.org/10.1007/s10772-020-09787-8
8. Ku, C. H., & Leroy, G. (2014). A decision support system: Automated crime report analysis and classification for e-government. Government Information Quarterly, 31(4), 534–544. https://doi.org/10.1016/j.giq.2014.08.003
9. Elov B.B., Axmedova X.I. Uchta so‘z turkumi doirasidagi omonimiyani farqlovchi biznes jarayonni modellashtirish// O‘zbekiston respublikasi innovatsion rivojlanish vazirligining, Ilm-fan va innovasion rivojlanish ilmiy jurnal 2022 / 1, 150-162-b.
10. Axmedova X. I. Turli so‘z turkumlari orasidagi omonimiyani aniqlovchi matematik modellar// Science and innovation international scientific journal volume 1 issue 7 uif-2022: 8.2 | issn: 2181-3337. https://doi.org/10.5281/zenodo.7238546
11. Axmedova X.I. Chastotali usul yordamida omonimiyani aniqlash// “O‘ZBEK AMALIY FILOLOGIYASI ISTIQBOLLARI” Respublika ilmiy-amaliy konferensiyasi Toshkent: 2022.-164-170 b.
12. Elov B.B., Axmedova X.I. Determining homonymy using statistical methods. // “Hisoblash modellari va texnologiyalari (HMT 2022)” O‘zbekiston-Malayziya ikkinchi xalqaro konferensiyasi materiallari-Toshkent,2022 16-17 sentabr,-106 b.
13. Anggraeni, M., Syafrullah, M., & Damanik, H. A. (2019). Literation Hearing Impairment (I-Chat Bot): Natural Language Processing (NLP) and Naïve Bayes Method.

In Journal of Physics: Conference Series (Vol. 1201). Institute of Physics Publishing. https://doi.org/10.1088/1742-6596/1201/1/012057

14. Putong, M. W., & Suharjito. (2020). Classification model of contact center customers emails using machine learning. Advances in Science, Technology and Engineering Systems, 5(1), 174–182. https://doi.org/10.25046/aj050123

15. Bogery, R., Babtain, N. A., Aslam, N., Alkabour, N., Hashim, Y. A., & Khan, I. U. (2019). Automatic semantic categorization of news headlines using ensemble machine learning: A comparative study. International Journal of Advanced Computer Science and Applications, 10(11), 689–696. https://doi.org/10.14569/IJACSA.2019.0101190

16. Nahar, K. M. O., Jaradat, A., Atoum, M. S., & Ibrahim, F. (2020). Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning. Jordanian Journal of Computers and Information Technology, 6(3), 247–262. https://doi.org/10.5455/jjcit.71-1586289399

17. Taheri, S., & Mammadov, M. (2013). Learning the naive bayes classifier with optimization models. International Journal of Applied Mathematics and Computer Science, 23(4), 787–795. https://doi.org/10.2478/amcs-2013-0059

18. Foster, J., & Wagner, J. (2021). Naive Bayes versus BERT: Jupyter notebook assignments for an introductory NLP course. In Teaching NLP 2021 - Proceedings of the 5th Workshop on Teaching Natural Language Processing (pp. 112–114). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/2021.teachingnlp-1.20

ЛИНГВИСТИЧЕСКАЯ СТЕГОСИСТЕМА ДЛЯ СОКРЫТИЯ КОРОТКИХ СООБЩЕНИЙ В УЗБЕКСКОМ ТЕКСТЕ
LINGUISTIC STEGOSYSTEM FOR HIDING SHORT MESSAGES IN UZBEK TEXT

Вафаев Мирабид Алимович, Толибов Самохиддин Вохид угли,
Норбеков Рустам Исмат угли
Самаркандский филиал Ташкентского университета информационных технологий имени Мухаммада Ал-Хоразмий, Самарканд, Узбекистан
samohiddin2004@gmail.com, norbekovrustam12001@gmail.com

Аннотация. В этой статье мы представляем простой и новый подход к стеганографии на базе узбекского языка. Мы анализируем характеристики узбекского языка и предлагаем некоторые эффективные методы преобразования сообщения в бит. Мы находим, что одно узбекское текстовое сообщение может быть преобразовано в несколько секретных битов. Результаты показывают, что предложенный метод очень важен для успешной стеганографической техники. Кроме того, мы обнаруживаем, что встроенные биты могут быть правильно и легко извлечены людьми-наблюдателями без использования специального декодера. Таким образом, предлагаемая схема является практичной и эффективной для скрытия сообщения и является новым подходом для узбекского языка.

Abstract. In this article, we present a simple and novel approach to steganography based on the uzbek language. We analyze the characteristics of the uzbek language and offer some efficient methods for converting a message into a bit. We find that one uzbek text message can be converted into several secret bits. The results show that the proposed method is very important for a successful steganographic technique. In addition, we find that embedded bits can be correctly and easily extracted by human observers without the use of a special decoder. Thus, the proposed scheme is practical and effective for hiding the message and is a novel approach for the uzbek language.