MATNLARNI VEKTORLASHNI KVANT ILHOMLANGAN WORD2VEC ALGORITMI

Niyozmatova Nilufar A'loxanovna
"Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti" Milliy tadqiqot universiteti, Raqamli texnologiyalar va sun'iy intellekt kafedrasi dotsenti
E-mail: n_niyozmatova@mail.ru
Turg'unova Nafisaxon Maxammadjon qizi
"Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti" Milliy tadqiqot universiteti, Raqamli texnologiyalar va sun'iy intellekt kafedrasi assistenti
E-mail: nafisa_turgunova@tiiame.uz
Mamatov Abduvali Abduvohid o'g'li
"Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti" Milliy tadqiqot universiteti, Raqamli texnologiyalar va sun'iy intellekt kafedrasi assistenti
E-mail: abduvali_mamatov@mail.ru

Annotatsiya: Mazkur maqola moliyaviy xizmatlar bo'yicha foydalanuvchilar shikoyatlarini avtomatik tasniflashga bag'ishlangan bo'lib, unda belgilarni shakllantirishning kvant Word2vec algoritmi taklif etilgan. Algoritm har bir so'zni kompleks vektor ko'rinishida ifodalab, o'rtacha vektorga o'tkazish orqali mashinali o'qitish usullariga kiruvchi ma'lumot sifatida uzatadi. Klassik va kvant Word2vec algoritmi asosida shakllantirilgan belgilardan Decision Tree, Random Forest, Naive Bayes va Passive Aggressive usullar orqali tanib olishda foydalanilgan. Tadqiqot natijasida kvant Word2vec algoritmi klassik algoritmiga nisbatan o'rtacha 28% yuqori aniqlikni ko'rsatdi. Bu kvant Word2vec algoritmini klassik algoritmga nisbatan yuqori samaradorlikka ega ekanligini bildiradi. Bunda Decision Tree, Random Forest, va Passive Aggressive usullari 90% dan yuqori aniqlikni ta'minladi. Bu kvant Word2vec algoritmi yordamida shakllantirilgan belgilar vektorlari shikoyat matnlarini ma'no va kontekstini yetarli darajada ifodalashini ko'rsatadi.

Kalit so'zlar: kvant Word2vec, matn tasniflash, iste'molchi shikoyatlari, mashinali o'qitish, tabiiy tilni qayta ishlash, vektor, Decision Tree, Random Forest, Naive Bayes, Passive Aggressive.

Аннотация: Статья посвящена автоматической классификации жалоб пользователей в сфере финансовых услуг, в которой предлагается квантовый алгоритм Word2vec для формирования символов. Алгоритм представляет каждое слово в виде сложного вектора и переносит его в усредненный вектор

ISSN 3030-3907

в качестве информации для методов машинного обучения. Символы, сформированные на основе классического и квантового алгоритмов Word2vec, использовались для распознавания с использованием методов Decision Tree, Random Forest, Naive Bayes и Passive Aggressive. В результате исследования квантовый алгоритм Word2vec показал среднюю точность на 28% выше, чем классический алгоритм. Это говорит о том, что квантовый алгоритм Word2vec имеет более высокую эффективность, чем классический алгоритм. При этом методы Decision Tree, Random Forest и Passive Aggressive обеспечили точность более 90%. Это показывает, что векторы символов, сформированные с использованием квантового алгоритма Word2vec, в достаточной степени отражают смысл и контекст текстов жалоб.

Ключевые слова: квантовый Word2vec, классификация текста, жалобы потребителей, машинное обучение, обработка естественного языка, вектор, дерево решений, случайный лес, Наивный Байесовский алгоритм, пассивно-агрессивный.

Abstract: This article is devoted to the automatic classification of user complaints in financial services, in which the quantum Word2vec algorithm for character formation is proposed. The algorithm represents each word in the form of a complex vector and transfers it to the average vector as information for machine learning methods. Characters formed based on the classical and quantum Word2vec algorithms were used for recognition using Decision Tree, Random Forest, Naive Bayes, and Passive Aggressive methods. As a result of the study, the quantum Word2vec algorithm showed an average accuracy of 28% higher than the classical algorithm. This indicates that the quantum Word2vec algorithm has a higher efficiency than the classical algorithm. In this case, the Decision Tree, Random Forest, and Passive Aggressive methods provided an accuracy of more than 90%. This shows that the character vectors formed using the quantum Word2vec algorithm sufficiently represent the meaning and context of the complaint texts.

Keywords: quantum Word2vec, text classification, consumer complaints, machine learning, natural language processing, vector, Decision Tree, Random Forest, Naive Bayes, Passive Aggressive.

KIRISH

Bugungi raqamli iqtisodiyotda mijozlardan kelgan shikoyat matnlarini avtomatik tasniflash muhim masalaga aylanib bormoqda. Ayniqsa, moliyaviy xizmatlar yo‘nalishida iste’molchi mulohazalari, shikoyat va fikrlarini tahlil qilish bilan xizmat ko‘rsatish sifatini oshirish, muammoli hududlarni aniqlash va yechimlar ishlab chiqish bugungi kundagi eng dolzarb masalalar hisoblanadi.

ISSN 3030-3907

Ushbu masalalarni yechishda tabiiy tilni qayta ishlash (NLP) va mashinali o‘qitish (ML) texnologiyalari muhim ahamiyatga egadir [1-3].

Ko‘plab tadqiqotchilar matnli ma’lumotlarni vektorizatsiyasi va tasniflashda TF_IDF, Word2vec, GloVe modellaridan foydalanadi [3-6]. Biroq, bu yondashuvlarda katta hajmli ma’lumotlar bilan ishlashda, semantik bog‘liqliklarni ifodalashda turli murakkabliklarni yuzaga keltirib chiqaradi. Shuning uchun kvant hisoblashlarni klassik usullar bilan birgalikda qo‘llash hozirgi kunda ilmiy izlanuvchilarda katta qiziqish uyg‘otmoqda. Chunki ular klassik algoritmlarga nisbatan murakkablik va umumlashtirish nuqtai nazardan ko‘plab afzalliklarga ega hisoblanadi [5-7].

Hongying va boshqalar tadqiqotlarida kvant hisoblash orqali Word2vec va kvant konvolyutsion neyron tarmog‘i (Quantum Convolutional Neural Network - QCNN) bilan elektr energetikasi sohasida tezaurusni qurishni taklif etishgan. U Word2vec modeli bo‘yicha olingan vektorlarni kvant konvolyutsion neyron tarmog‘i bilan chuqur semantik ma’nolarni olgan va yaxshilangan kvant zarralar to‘dasini optimallashtirish algoritmi bilan qidiruv imkoniyatini oshirishda foydalangan. Natijada QCNN klassik CNNga nisbatan 3% yuqori aniqlikni ko‘rsatgan [6-11]. Bundan tashqari, Qiuchi va boshqalar kvant ilxomlangan hamda so‘zlarni vektorlashni yaxshilashni taklif etishgan. Unda asosiy maqsad so‘zlarni kompleks sonli vektor sifatida ifodalash, fazalar interferensiyasi orqali ma’no jixatdan yaqin so‘zlarni belgilash va matematik model orqali gaplar semantikasini yaxshiroq ifodalash hamda ularni tasniflash qilib olingan [12-16].

Mazkur maqolada kvant hisoblashlar asosida ishlab chiqilgan matnlarni vektorlash algoritmi taklif etilgan bo‘lib, u orqali foydalanuvchilar shikoyat matnlarini moliyaviy xizmatlar bo‘yicha tasniflash amalga oshirilgan. Tadqiqotda Qaror daraxti (Decision Tree - DT), Tasodifiy o‘rmon (Random Forest - RF), Naive Bayes (Naive Bayes - NB) va Passiv tajovuzkor (Passive Aggressive - PA) mashinali o‘qitish usullari asosida tasniflash amalga oshirilib, ular aniqlik ko‘rsatkichlari bo‘yicha taqqoslangan. Shuningdek, kvant ilhomlangan Word2vec yordamida so‘zlar orasidagi bog‘liqliklar kompleks fazoda ifodalanib, tasniflash natijalariga qanday ta’sir qilishi tahlil qilindi.

Mazkur ilmiy ish taklif etilgan matn tasniflash yondashuvlariga nisbatan yangi va amaliy yondashuv bo‘lib, kelgusida kvant hisoblash asosidagi NLP modellarini qayta ishlash va tahlil qilishda muhim qadam bo‘lishi mumkin.

METODOLOGIYA

Ushbu tadqiqotda iste’molchi shikoyat matnlarini moliyaviy xizmat turlari bo‘yicha tasniflash uchun kvant ilhomlangan Word2vec yondashuvi asosida

ISSN 3030-3907

so‘zlarni vektorlashtirish va klassik mashinali o‘qitish usullaridan foydalanildi. U quyidagicha bosqichlardan iborat:

1-qadam. Ma’lumotlarni tayyorlash. Dastlab iste’molchilar shikoyatlarini o‘z ichiga olgan real bank xizmatlari ma’lumotlar bazasi https://github.com/Nafisaxon/shikoyat_dataset/tree/main manzilidan olingan bo‘lib, u har xil turdagi iqtisodiy xizmatlarga oid shikoyatlarni o‘z ichiga qamrab olgan. Unda 20 ga yaqin moliyaviy xizmatlar mavjud. Ma’lumotlar to‘plamini yagona to‘plam sifatida ifodalab olindi:

$$D = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\},$$

bu yerda x_i - matn, y_i - tegishli sinfi.

Xizmat turlarini raqamli formatga o‘tkaziladi:

$$y_i = \{0,1,2,3,4\}.$$

2-qadam. Lug‘at yaratish. Har bir matn so‘zlarga ajratiladi va umumiy so‘zlar to‘plami yaratiladi:

$$V = \bigcup_{i=1}^{n} \{w_{i1}, w_{i2}, ..., w_{in}\},$$

bunda w_{in} - so‘z, lug‘atdagi har bir so‘z uchun murakkab sonli vektor yaratiladi.

3-qadam. Kvant ilhomlangan Word2vec vektorlarini yaratish.

Har bir so‘z uchun murakkab vektor quyidagicha yaratiladi:

$$\vec{\upsilon}_w = \vec{r}_w + i \cdot \vec{i}_w, \qquad \vec{r}_w, \vec{i}_w \square U(-1,1)^d$$

bu yerda U - barcha qiymatlar teng ehtimollikda tanlanadigan taqsimot, $d = 2 \times |V|$ - har bir vektor o‘lchami.

Normallashtirish uchun kvadrat normalizatsiyasi qo‘llaniladi:

$$\vec{\upsilon}_w \leftarrow \frac{\vec{\upsilon}_w}{\sqrt{\sum_{j=1}^{d} |\vec{\upsilon}_{w,j}|^2}}.$$

So‘zlar vektorlarini birlik sfera ustida joylashtirishni ta’minlaydi va bu kvant holat uchun zarur shartni ifodalaydi.

4-qadam. Matnni o‘rtacha vektorga aylantirish. Har bir shikoyat matni uchun so‘zlarning o‘rtacha vektori olinadi:

$$x_i = \{w_1, w_2, ..., w_k\},$$

$$\vec{s}_i = \frac{1}{k} \sum_{j=1}^{k} R\left(\vec{\upsilon}_{w_j}\right),$$

bunda natijani haqiqiy qismi olinadi, chunki mashinali o‘qitish usuli haqiqiy sonlar bilan ishlaydi.

5-qadam. Kvant holat simulyatsiyasi. O‘rtacha vektorlar istalgan holatda kvant holatga o‘tkazilishi mumkin:

ISSN 3030-3907

$$|\psi_i\rangle = \sum_{j=0}^{2^n-1} a_j |j\rangle,$$

bunda $a_j = s_{i,j}$ bo‘lsa, $\sum |a_j|^2 = 1$ ga teng.

Bu bosqichda hosil qilingan murakkab sonli vektorlar qiskit kutubxonasi yordamida kvant holat sifatida ifodalanadi. Buning uchun QuantumCircuit.initialize() funksiyasi qo‘llaniladi, u orqali vektor kvant zanjiriga yuklanadi. Undan maqsad matnni klassik vektorlar fazosida emas, balki ko‘p o‘lchovli va murakkab fazoda (Gilbert fazosida) ifodalab, so‘zlar orasidagi yashirin bog‘liqliklarni yanada aniqroq qilishdir.

6-qadam. Kosinus o‘xshashlikni hisoblash. Ikkita matn vektori $\vec{a}, \vec{b}$ orasidagi o‘xshashlik:

$$sim(\vec{a}, \vec{b}) = R(\langle \psi_a | \psi_b \rangle),$$

bu kvant ichki ko‘paytma sifatida ishlaydi va kompleks fazodagi yaqinlikni ko‘rsatadi. Bu bosqich faqat o‘quv va nazoratlarda yaxshi natijalarga erishish uchun qo‘llaniladi.

7-qadam. Klassifikatsiya uchun mashinali o‘qitish usullaridan foydalanish. Hosil bo‘lgan vektorlar har bir matn uchun quyidagicha belgilanadi:

$$X = \{\vec{s}_1, \vec{s}_2, ..., \vec{s}_N\} \in R^{N \times d}.$$

Mashinali o‘qitish usullari uchun ma’lumotlarning 70% o‘quv va 30% nazorat to‘plamlariga ajratildi hamda DT, RF, NB, PA usullar asosida o‘qitildi.

8-qadam. Baholash va vizualizatsiya. Har bir model quyidagicha baholanadi:

$$Accuracy = \frac{\sum_{i=1}^{N} I(\hat{y}_i = y_i)}{N}.$$

Bashoratlar va haqiqiy sinflar orasidagi farqlar ajratilib, noto‘g‘ri tasniflangan matnli ma’lumotlarni alohida tahlil qilindi.

NATIJALAR

Tadqiqotda kvant ilhomlangan Word2vec asosida vektorlashtirilgan shikoyat matnlari DT, RF, NB va PA mashinali o‘qitish usullari yordamida tasniflandi. Usullar samaradorligini baholash uchun aniqlik ko‘rsatkichi asos qilib olindi.

1-jadval

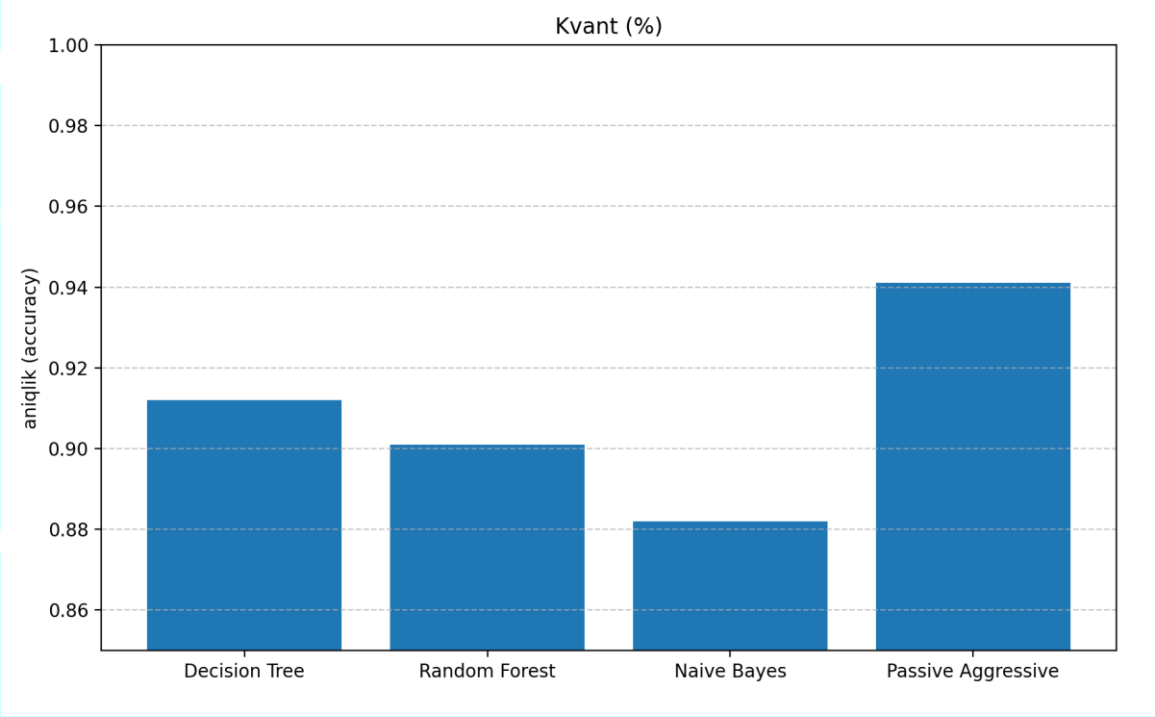
Har bir usul aniqlik natijalari

№	Usul nomi	Klassik (%)	Kvant (%)
1	DT	0.629	0.912%

ISSN 3030-3907

2	RF	0.571	0.901%
3	NB	0.600	0.882%
4	PA	0.714	0.941%

Word2vec modelni kvant usuli klassik usulga nisbatan yuqori natijalarni ko‘rsatdi. Ayniqsa, RF usuli uchun farq sezilarli bo‘lib, aniqlik 0.571 dan 0.901 gacha oshdi. Bu kvant ilhomlangan Word2vec algoritmi ma’lumotlarni yanada yaxshi ifodalash va usul o‘rganishini samaraliroq ta’minlash qobiliyatini ko‘rsatadi. PA usulida eng yuqori aniqlik kuzatildi. Kvant Wor2vec vektorlash orqali yaratilgan matn vektorlarini muvaffaqiyatli o‘rganib, nazorat to‘plamidagi namunalarni 94% aniqlik bilan to‘g‘ri tasniflay oldi. Natijalar quyidagi rasmda keltirib o‘tilgan.



1-rasm. Mashinali o‘qitish modellari bo‘yicha tahlili

Natijada kvant ilhomlangan Word2vec algoritmi yordamida yaratilgan vektorlar ko‘plab tasniflash modellarida yuqori aniqlikni ta’minladi. Kvant usullar DT, RF va PA usullarni barchasida klassik yondashuvga nisbatan yuqori aniqlikka erishdi. Bu kvant ilhomlangan Word2vec algoritmi tomonidan yaratilgan kompleks vektorlar shikoyat matni semantik mazmunini ancha yaxshi ifodalashini va vektorlar mashinali o‘qitish usullariga aniq hamda farqli xususiyatni belgilashini ko‘rsatadi.

Naive Bayes modeli 88.2% aniqlikni ko‘rsatgani uni sodda ehtimolli tuzilmasi bilan bog‘liq bo‘lishi mumkin. Ushbu usul so‘zlar orasidagi bog‘liqliklarni mustaqil deb hisoblaydi, bu kvant ilhomlangan Word2vec singari

ISSN 3030-3907

kompleks tuzilmalarda samarasi 90%dan yuqori bo‘lmasligi mumkin. Shunga qaramay, klassik natijalarga nisbatan aniqligi 28%ga oshdi. 90%dan yuqori aniqlik bilan ishlagan DT, RF va PA usullari umumiy xususiyati ularni vektordagi o‘zgarishlarga sezgirligi deyish mumkin. Aynan kvant ilhomlangan Word2vec vektorlari yuqori o‘lchovli murakkab tuzilmasi bunday usullar uchun xos deb hisoblash mumkin. Kvant ilhomlangan Word2vec yondashuvi matnni chuqurroq va aniqroq ko‘rinishda ifodaladi hamda tasniflash jarayonida yuqori aniqlikni ko‘rsatdi.

XULOSA

Ushbu maqolada iste’molchi shikoyatlarini moliyaviy xizmatlar bo‘yicha avtomatik tasniflash vazifasi uchun kvant ilhomlangan Word2vec algoritmiga asoslangan yondashuv taklif etildi. Modelda har bir so‘z kompleks vektor orqali ifodalanib, matnlar real fazodagi o‘rtacha vektorga aylantirildi. Shikoyat matnlariga boy semantik ifoda beruvchi ushbu vektorlar turli mashinali o‘qitish usullarida sinovdan o‘tkazildi.

Kvant ilhomlangan Word2vec yondashuvi asosida yaratilgan vektorlar klassissik Word2vec algoritmiga nisbatan o‘rtacha 28% yuqori aniqlikni ta’minladi. DT, RF va PA usullari esa 90% dan yuqori aniqlikni ko‘rsatdi.

Tadqiqot natijalari kvant ilhomlangan Word2vec yondashuvlarni NLP sohasida amaliy qo‘llashda yuqori natijalar olish mumkinligini ko‘rsatadi va ushbu yo‘nalishda keyingi ishlanmalar uchun amaliy hamda nazariy asos bo‘lib xizmat qiladi. Kelgusida ushbu modelni yanada rivojlantirish, real kvant qurilmalarida sinovdan o‘tkazish maqsadga muvofiq deb hisoblaymiz.

Foydalanilgan adabiётlar

1. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space // arXiv preprint arXiv:1301.3781. – 2013.

2. N. A. Niyozmatova, N. S. Mamatov, B. I. Otaxonova, A. N. Samijonov and K. K. Erejepov, "Classification Based On Decision Trees And Neural Networks," *2021 International Conference on Information Science and Communications Technologies (ICISCT)*, Tashkent, Uzbekistan, 2021, pp. 01-04, doi: 10.1109/ICISCT52966.2021.9670345.

3. Babomuradov O. O. J. et al. Text documents classification in Uzbek language //International journal of recent technology and engineering. – 2019. – T. 8. – №. 2S11. – S. 3787-3789.

4. Pennington J., Socher R., Manning C.D. GloVe: Global Vectors for Word Representation // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2014. – P. 1532–1543.

ISSN 3030-3907

5. Schuld M., Sinayskiy I., Petruccione F. An Introduction to Quantum Machine Learning // Contemporary Physics. – 2015. – Vol. 56, No. 2. – P. 172–185.

6. Babomuradov, O. J., Mamatov, N. S., Boboyev, L. B., & Otaxonova, B. I. (2019). Classification of texts using decision trees algorithms. *Descendants of Muhammad al-Khwarizmi*, *4*(10).

7. Babomuradov O. J. i dr. Qaror daraxti algoritmidan foydalanib matnlarni tasniflash //Muxammad al-Xorazmiy avlodlari. – 2019. – №. 4 (10). – S. 22.

8. Mamatov, N. S., Turg'unova, N. M., Turg'unov, B. X., & Samijonov, B. N. (2025). Algorithm for analysis of texts in Uzbek language. In *Artificial Intelligence and Information Technologies* (pp. 511-515). CRC Press.

9. Zhang Y., Wallace B.C. A Sensitivity Analysis of CNN for Sentence Classification // arXiv preprint arXiv:1510.03820. – 2015.

10. Sebastiani F. Machine Learning in Automated Text Categorization // ACM Computing Surveys. – 2002. – Vol. 34, No. 1. – P. 1–47.

11. He, Hongying and Liu, Nan and Zhu, Jizhong and Luo, Diansheng and Hong, Runli, A Thesaurus Constructing Method in Electric Power Domain Based on Word2vec and Quantum Convolutional Neural Network. Available at SSRN: https://ssrn.com/abstract=4840583 or http://dx.doi.org/10.2139/ssrn.4840583

12. Li Q. et al. Quantum-inspired complex word embedding //arXiv preprint arXiv:1805.11351. – 2018.

13. Raschka S., Mirjalili V. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. – Birmingham: Packt Publishing, 2017. – 620 p.

14. Muhamediyeva D., Niyozmatova N., Turgunova N., & Mamatov A. (2025) TEXT VECTORIZATION BASED ON THE QUANTUM BAG-OF-WORDS ALGORITHM. Electronic Journal of Actual Problems of Modern Science, Education and Training. April, (4), ISSN 2181-9750, 33–38.

15. Ниёзматова Н., Маматов Н., & Тургунова Н. (2025). КВАНТОВЫЙ АЛГОРИТМ ONE-HOT ВЕКТОРИЗАЦИИ ТЕКСТОВ. *Jeleznodorojnыy transport: aktualnые voprosы i innovaцii*, (1), 119–124.

16. Тургунов, Б. Х. У., Тургунова, Н. М. К., Иброхимова, М. И. К., & Абдукодирова, Д. А. К. (2023). АЛГОРИТМЫ ОТДЕЛЕНИЯ ФАКТОВ ОТ ИНФОРМАЦИИ НА УЗБЕКСКОМ ЯЗЫКЕ ИЗ ЗАЯВЛЕНИЙ,

ISSN 3030-3907

ПОЛУЧЕННЫХ ОТ ВИРТУАЛЬНЫХ ПРИЕМНЫХ. *Universum: технические науки*, (12-2 (117)), 9-11.

ISSN 3030-3907