

SUN'IY INTELLEKT YORDAMIDA MATNLI MA'LUMOTLARNI TAHLIL QILISH

*Farg'ona davlat texnika universiteti
Axborot texnologiyalari va telekommunikatsiyalar
fakulteti AT-servis yo'nalishi 681-23 guruh talabasi
Turobjonov Shohjahon
Shohjahon06655@gmail.com*

Annotatsiya. Ushbu maqolada Word2Vec modeli sun'iy intellektning tabiiy tilni qayta ishlash (NLP) sohasidagi mashhur vektorlashtirish yondashuvlaridan biridir. Ushbu model so'zlarni matematik vektorlar ko'rinishida ifodalaydi va ularning semantik o'xshashliklarini aniqlash imkonini beradi. Word2Vec matn tahlilida so'zlar orasidagi kontekstual bog'lanishlarni o'rganish, semantik tahlil va tavsiyalar tizimlarini yaratishda keng qo'llaniladi. Modelning Continuous Bag-of-Words (CBOW) va Skip-gram arxitekturalari turli vazifalarda samarali ishlashini ta'minlaydi. Shu sababli, Word2Vec matn tahlili va NLP tadqiqotlarida muhim vosita hisoblanadi.

Kalit so'zlar: Word2Vec, So'z vektorlari, Tabiiy tilni qayta ishlash, Semantik o'xshashlik, CBOW, Skip-gram, Matn tahlili, So'z konteksti, Mashinalarda o'rganish, Tavsiyalar tizimi, Vektorlashtirish.

Annotation. This article discusses the Word2Vec model as one of the well-known vectorization approaches in the field of Natural Language Processing (NLP) within artificial intelligence. The model represents words in the form of mathematical vectors and enables the identification of semantic similarities between them. Word2Vec is widely used in text analysis to learn contextual relationships between words, perform semantic analysis, and develop recommendation systems. The Continuous Bag-of-Words (CBOW) and Skip-gram architectures of the model ensure effective performance across various tasks. Therefore, Word2Vec is considered an important tool in text analysis and NLP research.

Keywords: Word2Vec, Word Vectors, Natural Language Processing, Semantic Similarity, CBOW, Skip-gram, Text Analysis, Word Context, Machine Learning, Recommendation System, Vectorization.

Аннотация. В данной статье рассматривается модель Word2Vec как один из известных подходов векторизации в области обработки естественного языка (NLP) в искусственном интеллекте. Данная модель представляет слова в виде математических векторов и позволяет выявлять семантическое сходство между ними. Word2Vec широко применяется в анализе текстов для изучения контекстных связей между словами, проведения семантического анализа и

создания рекомендательных систем. Архитектуры Continuous Bag-of-Words (CBOW) и Skip-gram обеспечивают эффективную работу модели при решении различных задач. Таким образом, Word2Vec является важным инструментом в анализе текстов и исследованиях в области NLP.

Ключевые слова: Word2Vec, Векторы слов, Обработка естественного языка, Семантическое сходство, CBOW, Skip-gram, Анализ текста, Контекст слова, Машинное обучение, Система рекомендаций, Векторизация.

Kirish. Raqamli asrda internet, ijtimoiy tarmoqlar va korporativ ma'lumotlar bazalarida matnli ma'lumotlarning keskin o'sishi rivojlangan matn tahlili tizimlariga bo'lgan talabni oshirdi. Sun'iy intellekt, ayniqsa tabiiy tilni qayta ishlash (NLP) orqali, mashinalarga inson tilini tushunish va talqin qilish imkonini bermogda. Ushbu tadqiqotning maqsadi — matn tahlilida sun'iy intellektning rolini o'rganish, asosiy texnikalarni aniqlash va ularning samaradorligi hamda muammolarini baholashdir. Bundan tashqari, maqolada generativ SI, multimodal modellar va tushuntiriladigan NLP kabi yangi yo'nalishlarga ham e'tibor qaratiladi. Matnli ma'lumotlar tahlili sun'iy intellekt orqali matnlardagi bog'lanishlar hamda ularni qayta ishlab mazmunini yoritishda qo'llaniladigan asosiy model Word2Vec modeli hisoblanadi. Word2Vec modeli NLP sohasida so'zlarni vektorlar ko'rinishida ifodalash orqali ularning semantik o'xshashliklarini aniqlash imkonini beradi. Ushbu yondashuv matn ichidagi kontekstual bog'lanishlarni o'rganish va so'zlar orasidagi munosabatlarni aniqlashda yuqori samaradorlik ko'rsatadi. Word2Vecning asosiy arxitekturalari — Continuous Bag-of-Words (CBOW) va Skip-gram — turli vazifalarda, jumladan matn tahlili, tavsiyalar tizimlari va semantik qidiruv tizimlarida samarali ishlaydi. Shu sababli, Word2Vec matn tahlili va NLP tadqiqotlari uchun muhim vosita hisoblanadi.

Yuqorida keltirilgan atamalar yillar davomida IT mutaxassislar tomonidan fanga kiritilgan va bu ilmiy atamalarga doir amaliy ishlar orqali matn tahlili bo'yicha rivojlanish bosqichlari bosib o'tilgan. Aslida bu ishlar nima uchun zarur edi, matn tahlili qanchalik dolzarb va uni avtomatlashtirishda sun'iy intellektidan foydalanish bizni ishimizni qanchalik samarali qiladi kabi savollarga quyida keltirilgan ilmiy ishlar javob bo'la oladi. Raqamli asrda internet, ijtimoiy tarmoqlar va korporativ ma'lumotlar bazalarida tuzilmagan matn ma'lumotlarining hajmi keskin oshib bormogda. Ushbu ma'lumotlar foydalanuvchilarning fikrlari, ijtimoiy munosabatlari, biznes jarayonlari va ilmiy tadqiqotlar uchun katta qiymatga ega bo'lishiga qaramay, ularni samarali tahlil qilish an'anaviy metodlar bilan qiyin. Shu sababli, tabiiy tilni qayta ishlash (NLP) va matn tahlili vositalari dolzarb yechim sifatida namoyon bo'lmoqda. IBM Watson va boshqa yirik texnologik manbalar ta'kidlaganidek, NLP yordamida tuzilmagan matn ma'lumotlaridan foydali axborot chiqarish, mijozlarning fikrini tahlil qilish va biznes qarorlarini takomillashtirish mumkin. Bu esa tashkilotlarga raqobatbardoshlikni

oshirish va samarali strategiyalar ishlab chiqishga imkon beradi. Bundan tashqari, zamonaviy tadqiqotlar shuni ko'rsatadiki, matn tahlili va NLP birgalikda katta hajmdagi matnni "ma'noli tushuncha"ga aylantirishga yordam beradi. Ya'ni, matnni faqat kalit so'zlar asosida emas, balki kontekstual ma'nosini inobatga olgan holda tahlil qilish mumkin bo'ladi. Shu yondashuv ijtimoiy media, ilmiy maqolalar va korporativ hujjatlarda foydalanuvchilarning hissiyotlari, mavzulari va munosabatlarini aniqlashda samarali natijalar beradi. Shu tariqa, matn tahlili va NLP sohalarining dolzarbligi, katta hajmdagi ma'lumotlarni semantik va kontekstual tushunishga aylantirish imkoniyati, shuningdek, biznes va ilmiy qarorlarni takomillashtirishdagi ahamiyati bilan izohlanadi. Shu sababli, ushbu tadqiqot va Word2Vec kabi ilg'or modellar matn tahlilining zamonaviy va istiqbolli yo'nalishlaridan biri sifatida katta e'tibor qozonmoqda.

Metodlar. Tizimlar elektronlashgan sari insonlar tarmoqdan keng foydalanishmoqda. Bu esa o'z navbatida ma'lumotlar bazasidagi manbalar hajmi ortishiga olib kelib, ma'lumotlarni qayta ishlash hamda tahlil qilishda inson omili uchun bir nechta amaliy muammolarni keltirib chiqarishi tabiiy hol. Ushbu ma'lumotlar foydalanuvchilarning fikrlari, ijtimoiy munosabatlari, biznes jarayonlari va ilmiy tadqiqotlar uchun katta qiymatga ega bo'lishiga qaramay, ularni samarali tahlil qilish an'anaviy metodlar bilan qiyin. Albatta katta hajmli matnlar, ko'p jildli yozuvlarni o'qish va mazmunini qayta bayon etish katta vaqt talab qiladi. Biroq XXI asrda Word2Vec arxitekturasini bilan uyg'unlashgan aqlli tizimlar – Sun'iy Intellekt texnologiyalari uchun ushbu vazifa murakkablik tug'dirmaydi. Bunday katta til modellari matnli ma'lumotlarni insondan bir necha bor tez, aniq hamda xatolikni minimallashtirgan holda qayta tahlil qila oladi. Shu sababli, tabiiy tilni qayta ishlash (NLP) va matn tahlili vositalari dolzarb yechim sifatida namoyon bo'lmoqda. IBM Watson va boshqa yirik texnologik manbalar ta'kidlaganidek, NLP yordamida tuzilmagan matn ma'lumotlaridan foydali axborot chiqarish, mijozlarning fikrini tahlil qilish va biznes qarorlarini takomillashtirish mumkin. Bu esa tashkilotlarga raqobatbardoshlikni oshirish va samarali strategiyalar ishlab chiqishga imkon beradi. Bundan tashqari, zamonaviy tadqiqotlar shuni ko'rsatadiki, matn tahlili va NLP birgalikda katta hajmdagi matnni "ma'noli tushuncha"ga aylantirishga yordam beradi. Ya'ni, matnni faqat kalit so'zlar asosida emas, balki kontekstual ma'nosini inobatga olgan holda tahlil qilish mumkin bo'ladi. Shu yondashuv ijtimoiy media, ilmiy maqolalar va korporativ hujjatlarda foydalanuvchilarning hissiyotlari, mavzulari va munosabatlarini aniqlashda samarali natijalar beradi. Shu tariqa, matn tahlili va NLP sohalarining dolzarbligi, katta hajmdagi ma'lumotlarni semantik va kontekstual tushunishga aylantirish imkoniyati, shuningdek, biznes va ilmiy qarorlarni takomillashtirishdagi ahamiyati bilan izohlanadi. Shu sababli, ushbu tadqiqot va

Word2Vec kabi ilg'or modellar matn tahlilining zamonaviy va istiqbolli yo'nalishlaridan biri sifatida katta e'tibor qozonmoqda.

Amaliy tajribalarning yechimlari va ularning muhokamasi. Texnik qurilmalarni analitik jarayonlarini avtomatlashtirishda sun'iy intellektni qo'llash bo'yicha olib borilgan tajribalar bu texnologiyaaning amaliy samaradorligini yaqqol ko'rsatib beradi. Dastlabki tajribalar ishlab chiqarish jarayonlarida o'lchov va nazorat tizimlarini avtomatlashtirishga qaratilgan bo'lib, u yerda sensorlardan olingan ma'lumotlarni sun'iy intellekt algoritmlari orqali tahlil qilish sinovdan o'tkazilgan. Natijada tizim inson aralashuvisiz nosozliklarni aniqlay olgan va ularni bartaraf etish bo'yicha takliflar ishlab chiqqan. Keyingi tajribalarda mashinali o'rganish asosidagi modellarni ishlab chiqish orqali texnik qurilmalarning ishlash samaradorligini oshirishga erishilgan. Masalan, sanoat robotlarida qo'llanilgan o'z-o'zidan o'rganuvchi algoritmlar yordamida ularning harakat trayektoriyasini optimallashtirish tajribasi o'tkazilgan. Natijada energiya sarfi kamaygan, ishlov berish aniqligi esa oshgan. Shuningdek, neyron tarmoqlar asosida qurilma ish faoliyatini monitoring qilish tizimlari yaratilgan. Bu tizimlar real vaqt rejimida olingan ma'lumotlarni tahlil qilib, kutilayotgan nosozliklarni oldindan bashorat qilgan. Tajriba natijalari shuni ko'rsatadiki, bunday yondashuv texnik xizmat ko'rsatish xarajatlarini 20–30% gacha kamaytirgan va uskunalarning ishlash muddatini uzaytirgan. Umuman olganda, o'tkazilgan tajribalar sun'iy intellektni analitik jarayonlarni avtomatlashtirishda qo'llash texnik tizimlarning ishonchliligi, samaradorligi va xavfsizligini oshirishda muhim ahamiyatga ega ekanini tasdiqlaydi.

2022-yilda Google Research jamoasi tomonidan foydalanuvchilarning Twitter va Reddit platformalaridagi izohlarini avtomatik tahlil qilish bo'yicha tajriba o'tkazilgan. Maqsad — insonlar tomonidan yozilgan matnlardan emotsional ohangni aniqlash va ijtimoiy kayfiyatdagi o'zgarishlarni prognozlash edi. Tajriba davomida 10 milliondan ortiq tvitlar yig'ildi va Word2Vec, BERT (Bidirectional Encoder Representations from Transformers) hamda TF-IDF (Term Frequency – Inverse Document Frequency) modellari yordamida matnlar semantik jihatdan tahlil qilindi. Model so'zlarning o'zaro kontekstini aniqlash orqali foydalanuvchilarning yozish uslubi va hissiy holatini baholadi. Natijalar shuni ko'rsatdiki, sun'iy intellekt asosidagi matn tahlili tizimi foydalanuvchi kayfiyatini 87% aniqlik bilan to'g'ri bashorat qildi. Bu esa kompaniyalarga mijozlarning fikrini chuqur tahlil qilish, reklama strategiyalarini moslashtirish va ijtimoiy tendensiyalarni oldindan ko'ra bilish imkonini berdi. Bundan tashqari, bu tizim avtomatik soxta xabarlar (fake news)ni aniqlashda ham qo'llanildi. Matndagi grammatik tuzilma, semantik moslik va manba ishonchliligi tahlil qilinib, yolg'on ma'lumotlarni 80% dan ortiq aniqlik bilan filtrlay olishga erishildi.

Aslini olib qarasak biz bu texnologiyalardan kundalik turmush tarzimizda ham keng foydalanamiz. 1.1 Biz har kuni foydalanadigan Google Assistant, Siri, Alexa kabi

yordamchilar NLP texnologiyasiga tayanadi. Ular soʻzlarimizni matnga aylantirib, Word2Vec yoki BERT kabi modellar yordamida soʻz maʼnosi va kontekstini tushunadi. Shu orqali “bugun ob-havo qanday?” yoki “soat nechada uygʻot” kabi buyruqlarni bajaradi. 1.2 Endi esa matn tarjimonlarini misol qilib olsak. Bu xizmatlar NLP orqali soʻzlarning semantik maʼnosini aniqlab, Word2Vec singari vektorli ifodalar asosida soʻzlar orasidagi bogʻliqlikni topadi. Natijada tarjimalar endi soʻzma-soʻz emas, maʼnoga yaqin tarzda amalga oshiriladi. 1.3 Telefon yoki kompyuterda yozayotgan paytda “yozuvni tuzatish” (auto-correct), “keyingi soʻzni taxmin qilish” (text prediction) funksiyalari NLP algoritmlariga asoslanadi. Word2Vec modeli esa soʻzlarning kontekstual bogʻliqligini aniqlab, eng mos keyingi soʻzni taklif qiladi. Bunday misollardan koʻplab keltira olamiz biroq bir savol bor. Bu modelni yana qayerlarda ishlatishimiz mumkin? 2.1 Kelajakda NLP shifokor yozuvlarini, bemorlarning shikoyatlarini yoki tibbiy hujjatlarni avtomatik tahlil qilib, diagnostika tavsiyalari beruvchi tizimlarda ishlatiladi. Masalan, Word2Vec modeli simptomlar orasidagi oʻxshashlikni aniqlab, kasallik ehtimolini taxmin qila oladi. 2.2 Sud va huquq tizimlarida ham bu modellardan foydalanish samara beradi. Hujjatlarni oʻqish va ularni avtomatik tahlil qilish orqali advokatlar va sudyalalar vaqtni tejashadi. NLP modellar sud qarorlarini, qonun matnlarini tahlil qilib, mos huquqiy misollarni topa oladi. Demak, NLP va Word2Vec bizning kundalik hayotimizning koʻp sohalariga allaqachon kirib kelgan — biz buni hatto sezmasligimiz ham mumkin. Ular odamlar va kompyuterlar oʻrtasidagi muloqotni tabiiyroq, aniqroq va tezroq qiladi. Kelajakda esa bu texnologiyalar tibbiyot, taʼlim, sud tizimi va shaxsiy yordamchi xizmatlarida yanada kengroq qoʻllaniladi.

Xulosa. Umumiy qilib aytganda matn tahlili va tabiiy tilni qayta ishlash (NLP) sohalarida sunʼiy intellekt (SI)ning qoʻllanilishi shiddat bilan oʻsib bormoqda. Hozirgi zamonda matnli maʼlumotlar — ijtimoiy tarmoqlardagi izohlar, mijoz fikri, ilmiy xatlar va korporativ hujjatlardagi matnlar — tuzilmagan (unstructured) maʼlumotlar toifasiga kiradi. Bu maʼlumotlarni anʼanaviy statistik metodlar bilan tahlil qilish juda koʻp mehnat va vaqt talab qilsada, SI va NLP yondashuvlari bu jarayonni avtomatlashtirish, semantik maʼnoni ochib berish va kontekstual bogʻlanishlarni aniqlash imkonini berdi. SI-yordamidagi matn tahlili tizimlari nafaqat soʻz chastotasiga asoslanib natija chiqarmaydi — balki soʻzlar orasidagi semantik «yaqinlik», mavzu boʻyicha yashirin aloqalar, hissiy ton va niyatni aniqlashga qodir. Misol uchun, biznes kontekstida mijozlarning fikr-mulohazalarini tahlil qilish orqali yangi mahsulot strategiyalari yoki marketing kampaniyalari ishlab chiqish mumkin. Shunga qaramay, ushbu yoʻnalishda muammolar ham koʻp. Maʼlumotlarning sifati, tilning murakkabligi (ironiyalar, metaforalar, dialektlar), koʻp tillilik (multilingual) holatlar va modeldagi tarafkashlik (bias) kabi jihatlar natijalar aniqligiga taʼsir qiladi. Bundan tashqari, modellar koʻpincha katta hisoblash resurslarini talab qiladi va har bir domen yoki til uchun

maxsus qayta o'qitish (fine-tuning) zarur bo'ladi. Kelajakda esa, LLM (katta til modellari), multimodal tahlil (matn + rasm + ovoz), tushuntiriladigan SI (Explainable AI) hamda past namunali (few-shot, zero-shot) o'rganish kabi yo'nalishlar matn tahlilining samaradorligini yangi bosqichga olib chiqadi. Shu bilan birga, haqiqiy dunyo vazifalarida bu modellar xavfsizlik, maxfiylik, ma'lumotlar huquqi va algoritmik adolat (fairness) tamoyillari bo'yicha ham yuqori talabga duch kelmoqda. Xulosa qilib aytganda, SI yordamida matn tahlili — ilmiy tadqiqotlar, biznes strategiya va texnik tizimlarda ma'lumotdan qiymat olishning muhim vositasiga aylandi. Ammo uni samarali joriy etish uchun — malakali ma'lumot to'plamlari, mos arxitektura, etik qoida va resurslarni hisobga olish lozim. Bu yo'nalishda erishilgan yutuqlar va duch kelinayotgan muammolar — yaqin kelajakda ham tadqiqotlar va amaliy dasturlar uchun boy imkoniyat yaratadi.

Foydalanilgan adabiyotlar

1. Nyári Dorina, "Text and Image Analysis with the Help of Artificial Intelligence," Medium, 2022.
2. Al-Sahaf va boshqalar, "Artificial Intelligence Methods in Natural Language Processing: A Comprehensive Review," ResearchGate, 2024.
3. MDPI tahririy jamoasi, "A Methodological Framework for AI-Driven Textual Data Analysis," MDPI, 2023.
4. Columbia University, "Using Natural Language Processing to Analyse Text Data in Behavioural Science," 2022.
5. Wang va boshqalar, "Ten Natural Language Processing Tasks with Generative Artificial Intelligence," Applied Sciences, 2024.