Proceedings of the 4th International Scientific and Practical Conference «Scientific Progressive Methods and Tools»

(October 6-8, 2024). Riga, Latvia

No 219

PHILOLOGY AND LINGUISTICS

Effective machine learning in linguistics

Krasniuk Svitlana[1]

[1] Senior Lecturer;
Department of philology and translation,
Kyiv National University of Technologies and Design; Ukraine

Abstract. Machine learning (ML) in modern linguistics is extremely relevant and effective due to its ability to automate complex processes of natural language processing, text analysis, and linguistic data processing. Modern language models and machine learning algorithms are able to perform tasks that previously required significant human resources or were difficult to achieve. Machine learning is critical to modern linguistics, as it allows automating the processing of linguistic data, greatly increasing the efficiency and accuracy of linguistic research and practical applications. Its relevance lies in the need to process large volumes of textual information, and efficiency is ensured by the speed and adaptability of algorithms. The future of machine learning in linguistics is closely related to the development of computational methods, access to qualitative data, and the development of new models to better understand language.
Keywords: *linguistics, machine learning, deep machine learning.*

Introduction.

Machine learning is a key technology in the framework of data mining, which allows automating the processes of analyzing large volumes of data, making accurate predictions, classifying and finding hidden structures in data [1, 2, 3]. Together, they create a powerful tool for working with information in various fields — from linguistics to business analytics, medicine, and social sciences [4, 5].

Machine Learning is a subfield of artificial intelligence that aims to develop algorithms and models that allow computer systems to learn and improve their results on the basis of experience, that is, on the basis of data, without explicit programming for each specific task [6, 7, 8 , 9].

The relevance and importance of machine learning in linguistics is extremely high, since this approach opens up new opportunities for automating the analysis and processing of natural language, and also allows to significantly increase the efficiency of working with large arrays of text data.

Machine learning in linguistics is the application of artificial intelligence methods and algorithms for the

PHILOLOGY AND LINGUISTICS

analysis of language data and the automation of natural language processing processes. This direction of research is a component of computational linguistics and includes the use of mathematical models to solve problems of a linguistic nature. Machine learning allows computer systems to learn on their own from existing data, doing so by identifying patterns and regularities in language structures.

The main part.

The main areas of application of machine learning in linguistics.

1. Natural Language Processing (NLP)

Machine learning is the basis for NLP, a field that aims to create systems that can understand, analyze and generate natural language. Examples of problems:

- Morphological analysis (detection of parts of speech).
- Syntactic analysis (determining the grammatical structure of sentences).
- Semantic analysis (recognizing the meaning of words and phrases in context).
- Automatic summarization of texts.

2. Machine translation.

Modern machine translation systems, such as Google Translate or DeepL, use neural networks to translate texts between different languages. Neural models are trained on huge text corpora of bilingual data, allowing them to translate texts more accurately and efficiently compared to previous methods.

3. Sentiment Analysis.

This technology is used to automatically determine emotional shades in texts. Thanks to machine learning, the models can recognize the positive, negative or neutral mood of the text. It has wide applications in marketing, sociology, and public opinion analysis.

4. Speech recognition.

Machine learning allows you to convert spoken speech into text (speech-to-text), which is the basis for voice assistants such as Siri or Google Assistant. Neural networks, especially recurrent and deep networks, learn to recognize and transform audio data into textual representations [10, 11].

5. Identification and classification of text categories.

Machine learning techniques are widely used to automatically classify texts such as news, e-mails, scientific articles, etc. Systems learn from examples of texts with certain labels and can independently categorize new

PHILOLOGY AND LINGUISTICS

documents.

Types of machine learning algorithms in linguistics.

1. Classification based on frequency features

This is one of the basic approaches, which consists in representing texts in the form of frequency feature vectors (bag-of-words, TF-IDF), and then classifying them using machine learning algorithms, such as naive Bayes, SVM or logistic regression.

2. Neural networks.

Deep learning using neural networks, in particular recurrent neural networks (RNN) and their variants (LSTM, GRU), are among the most common approaches for processing text sequences. Transformer models (such as BERT, GPT) are used for translation, text generation, and context analysis tasks.

3. Word Embeddings.

Vector word representation technology, such as Word2Vec, GloVe or FastText, transforms words into multidimensional vectors that take into account the semantic similarity between words. This allows the model to better understand the meaning of words in different contexts.

4. Thematic modeling.

Algorithms such as Latent Dirichlet Allocation (LDA) allow the discovery of themes in large text corpora by grouping words that occur together most often. This is useful for analyzing large volumes of textual data and identifying their underlying themes.

Challenges of machine learning in linguistics.

1. Linguistic heterogeneity

Different languages have their own unique structures and features, which makes creating universal models a difficult task. Some languages have a rich morphology that requires additional consideration of grammatical changes.

2. Lack of data for rare languages

For many little-studied or rare languages, there is often a lack of sufficient training data, which limits the application of machine learning methods.

3. Contextuality and ambiguity

Languages often have polysemous words that have different meanings depending on the context. This is a challenge for machine learning, as the system must be able to distinguish the values correctly.

Conclusions.

Machine learning is a powerful tool for automating

decisions and analyzing large volumes of data. Due to the application of various methods and algorithms, it can be used in many fields, from science to business. However, challenges related to data, computational resources, and model interpretability must be considered for successful application.

Machine learning is an important component of modern linguistics and plays a key role in the development of such areas as natural language processing, machine translation, speech recognition and sentiment analysis. It allows you to significantly improve the quality and efficiency of linguistic tasks, opening up new opportunities for automation and expanding the scope of language research.

Machine learning in linguistics is developing at a rapid pace due to the improvement of algorithms and the increase of data volumes. Modern models are becoming more and more context-oriented, which makes it possible to understand language more effectively and solve more complex tasks. Further development in this area may lead to the creation of even more advanced translation systems, text analysis and automation of linguistic tasks.

Discussion and prospects for further research.

Despite significant progress, machine learning in linguistics faces a number of challenges, such as the ambiguity of words, cultural contexts and nuances, and the lack of sufficiently large and high-quality data corpora for some languages. However, the development of deep learning methods and an increase in the amount of available language data promises to overcome these difficulties and further expand the possibilities of automating language processes.

Considering the above, it is hybrid machine learning in linguistics that represents a promising direction of development, which combines various methods and approaches with the aim of improving the results of natural language processing (NLP), text analysis, and solving complex linguistic tasks. Hybrid methods include combining traditional data processing approaches with innovative Deep Learning algorithms, machine learning methods, and statistical analysis [12, 13, 14, 15].

Moreover, based on the above, the author emphasizes that the relevance of Big Data processing, analysis and analytics is constantly growing due to the exponential increase in the amount of textual information that needs to be analyzed, processed and used in various contexts [16, 17, 18, 19]. It

PHILOLOGY AND LINGUISTICS

is Big Data in linguistics that makes it possible to solve complex language tasks, automate natural language processing (NLP) processes, and carry out a more accurate analysis of language structures, meanings, and regularities. That is, Big Data in linguistics is a necessary tool for the automation of natural language processing, the study of changes in languages, and the adaptation of language systems to different contexts and cultures. Their relevance is related to the ability to process huge amounts of textual information, which allows to significantly improve the quality of research, create new tools for automation and ensure the development of technologies that work with language.

These directions of perspective scientific research of the author will be reflected in the following publications.

References:

[1] Kulynych Y., Krasnyuk M., Krasniuk S. Knowledge discovery and data mining of structured and unstructured business data: problems and prospects of implementation and adaptation in crisis conditions. *Grail of Science*. 2022. (12-13). pp. 63-70.

[2] Krasnyuk M.T., Hrashchenko I.S., Kustarovskiy O.D., Krasniuk S.O. (2018) Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. *Economies' Horizons.* 2018. No. 3(6). pp. 121-136

[3] Tetiana Tsalko, Svitlana Nevmerzhytska, Svitlana Krasniuk, Svitlana Goncharenko, Liubymova Natalia «Features, problems and prospects of data mining and data science application in educational management» // Bulletin of Science and Education, № 5(23), 2024. pp.637-657

[4] Науменко, М. (2024). Інтелектуальний аналіз бізнесових даних як фактор посилення конкурентної позиції підприємства. *Успіхи і досягнення у науці*, 2024, 5 (5). https://doi.org/10.52058/3041-1254-2024-5(5)-746-762

[5] Краснюк Світлана. (2024) Data Science у освітньому менеджменті // *Діалог культур у Європейському освітньому просторі* [Електронний ресурс] : Матеріали IV Міжнародної конференції, м. Київ, 10 травня 2024р. Київський національний університет технологій та дизайну / упор. С. Є. Дворянчикова. – К. : КНУТД, 2024. – С. 119-124.

[6] Науменко, М. (2024). Ефективне застосування класичних алгоритмів машинного навчання при прийнятті адаптивних управлінських рішень. *Наукові перспективи*, 2024, 5 (47). https://doi.org/10.52058/2708-7530-2024-5(47)-855-875

[7] Krasnyuk M., Tkalenko A., Krasniuk S. Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. *ΛΟΓΟΣ*. 2021.

[8] Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. *ΛΟΓΟΣ*. 2021.

[9] Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. *ΛΟΓΟΣ*. 2020. P. 55-57.

PHILOLOGY AND LINGUISTICS

[10] Науменко, М. (2024). Оптимальне використання алгоритмів глибокого машинного навчання в ефективному управлінні підприємством. *Успіхи і досягнення у науці*, 2024, 4 (4). https://doi.org/10.52058/3041-1254-2024-4(4)-776-794

[11] Maxim Krasnyuk, Svitlana Krasniuk, Svitlana Goncharenko, Liudmyla Roienko, Vitalina Denysenko, Liubymova Natalia. Features, problems and prospects of the application of deep machine learning in linguistics // *Bulletin of Science and Education*, № 11(17), 2023. pp.19-34. http://perspectives.pp.ua/index.php/vno/article/view/7746/7791

[12] Krasnyuk M., Hrashchenko I., Goncharenko S., Krasniuk S. Hybrid application of decision trees, fuzzy logic and production rules for supporting investment decision making (on the example of an oil and gas producing company). *Access to science, business, innovation in digital economy*. ACCESS Press. 2022. 3(3). P. 278-291.

[13] Краснюк М.Т. Гібридизація інтелектуальних методів аналізу бізнесових даних (режим виявлення аномалій) як складовий інструмент корпоративного аудиту. *Стан і перспективи розвитку обліково-інформаційної системи в Україні*: матеріали III Міжнар. наук.-практ. конф. (м. Тернопіль, 10-11 жовт. 2014 р.). Тернопіль: ТНЕУ, 2014. С. 211-212.

[14] Krasnyuk M., Goncharenko S., Krasniuk S. Intelligent technologies in hybrid corporate DSS. *Інноваційно-інвестиційний механізм забезпечення конкурентоспроможності країни:* колективна монографія / за заг. ред. О. Л. Гальцової. Львів-Торунь: Ліга-Прес, 2022. С. 194-211.

[15] Гращенко І.С., Краснюк М.Т., Краснюк С.О. Гібридно-сценарне застосування інтелектуальних, орієнтованих на знання технологій, як важливий антикризовий інструмент логістичних компаній в Україні. *Вчені записки Таврійського Національного Університету імені В. І. Вернадського*. Серія: Економіка і управління. 2019. Т. 30 (69). С. 121-129.

[16] Maxim Krasnyuk, Dmytro Elishys (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. *Science and technology today*, 4 (32) 2024. pp. 833-857

[17] Krasnyuk M., Krasnuik I. Big data analysis and analytics for marketing and retail. *Штучний інтелект у науці та освіті:* збірник тез Міжнародної наукової конференції (*AISE*) (1-2.03.2024 р.), Київ, 2024.

[18] Maxim Krasnyuk, Svitlana Nevmerzhytska, Tetiana Tsalko. (2024). Processing, analysis & analytics of big data for the innovative management. *Grail of Science*, 38, April 2024. pp. 75-83. https://www.journal-grail.science/issue38.pdf

[19] Науменко , М. (2024). Аналіз та аналітика великих даних в маркетингу та торгівлі конкурентного підприємства. *Grail of Science*, (40), 117-128. https://doi.org/10.36074/grail-of-science.07.06.2024.013