

DOI 10.36074/logos-18.10.2024.058

EFFECTIVE DATA MINING IN PHILOLOGY

Svitlana Krasniuk¹

1. Senior Lecturer

Department of philology and translation

*Kyiv National University of Technologies and Design, UKRAINE***ORCID ID: 0000-0002-5987-8681**

Introduction.

Data Mining is the process of discovering hidden patterns, relationships, and new knowledge from large amounts of data using machine learning, statistics, artificial intelligence, and databases [1, 2, 3]. This technology is key to data analysis in various fields, including philology, medicine, economics, marketing, and others [4, 5].

Data mining in philology is the process of identifying hidden patterns, structure and information in textual data using machine learning algorithms [6, 7, 8, 9], statistical methods and tools for processing large volumes of textual information. In modern philology, the analysis of texts is becoming more automated due to the use of computer technologies, which allows the study of linguistic phenomena on large text corpora.

Deep data mining in philology is an interdisciplinary approach that combines modern methods of data analysis and deep learning for the study of texts, linguistic phenomena and linguistic regularities [10, 11]. This process involves the automated processing and interpretation of large volumes of textual information in order to obtain new knowledge, structured data or to identify hidden patterns in texts. Philology, as a science of language and literature, increasingly uses the capabilities of artificial intelligence and computing methods for a deeper understanding of language processes and phenomena.

The main part.

1. Main aspects of data mining in philology

Data mining in philology includes the analysis of large volumes of textual data to solve tasks such as:

Identification of lexical and syntactic regularities in different types of texts.

The study of the evolution of language or vocabulary through the analysis of texts from different periods.

SECTION 22.

PHILOLOGY AND JOURNALISM

Study of connections between styles, authors and genres.

Automation of text processing for classification of genres, authorship or thematic features.

2. Data mining methods in philology

Philological data mining uses several main approaches and methods:

Topic modeling: Models like LDA (Latent Dirichlet Allocation) are used to automatically detect topics in texts. This makes it possible to determine the main thematic directions in a large corpus of documents.

Frequency analysis: the study of the frequency of use of words and phrases to identify stylistic or lexical features of a certain author or period.

Text Classification: Using machine learning algorithms (eg, Naive Bayesian, SVM) to automatically classify texts by genre, topic, or style.

Sentiment Analysis: automatic determination of the emotional coloring of the text to study texts from the point of view of positive, negative or neutral emotions.

Tokenization and vectorization: techniques for dividing text into individual words or phrases and turning them into numerical vectors that facilitate their analysis by machine learning algorithms.

3. Examples of the use of data mining in philology

Analysis of literary works: automated study of literary texts to identify main themes, stylistic features or evolution of the author's style.

Cultural-historical research: analysis of large corpora of texts to investigate the development of vocabulary and the use of changing words through cultural, political and social events.

Comparative language study: identifying differences between texts written in different languages or dialects, which helps in the study of linguistic diversity.

Analysis of media texts: the study of trends in the use of language in the media to analyze public sentiment, political discourse or reactions to events.

4. Challenges of data mining in philology

Unstructured textual data: a large part of philological data is unstructured (literary texts, articles, manuscripts), which complicates their automated analysis.

Language diversity: analyzing texts in different languages requires the creation of separate models and tools for each language or group of languages.

Large amounts of data: working with large text corpora requires significant computing resources and efficient algorithms.

5. Advantages of data mining in philology

Scalability: data mining allows you to process huge volumes of textual information, which is unattainable with manual analysis.

Automation of research: data mining significantly accelerates the research process by providing automated approaches to the analysis of large corpora of texts.



Detection of hidden regularities: data mining methods make it possible to find regularities and connections that may be invisible during traditional philological analysis.

Conclusions.

Machine learning and data mining are closely related, as both approaches are used to identify patterns in data, but have some key differences in methods and applications. Machine learning is one of the main tools for data mining. Machine learning algorithms make it possible to identify patterns in data without the need for prior knowledge of all the rules that govern this data. Machine learning provides the ability to automate the process of data analysis and make more accurate predictions, while data mining provides a structured process for extracting useful knowledge/patterns from this data.

In general, data mining in philology opens up new opportunities for the study of textual data, allowing to automate the analysis of linguistic phenomena and quickly explore large volumes of literary and cultural texts. This makes possible a more accurate and quick study of language evolution, stylistic features, as well as the analysis of sociocultural phenomena through the lens of language.

Discussion and prospects for further research.

Analysis and analytics of big data in philology is a set of methods and techniques for researching large arrays of textual information in order to identify regularities, trends, new meanings and contexts in big semi-structured data [12, 13, 14, 15]. The use of big data in philology is part of an interdisciplinary approach that combines classical humanities with modern technologies of information processing and analysis. Thus, the author separately emphasizes that it is the analysis and analytics of big data in philology that significantly expands the possibilities for researching linguistic phenomena and text structures. Modern natural language processing technologies make it possible to automate the research processes of big text corpora, which makes it possible to solve both classic and new tasks in philological science.

The author emphasizes that it is the hybrid methods and algorithms of data mining in philology that combine several approaches and technologies to achieve a more effective analysis of big volumes of textual data [16, 17]. These methods integrate various machine learning algorithms, linguistic rules and statistical models, which allows to optimize the processing of linguistic data and provide a deeper and more accurate analysis of philological information [18, 19].

This two directions of scientific research of the author will be reflected in the following publications.

SECTION 22.

PHILOLOGY AND JOURNALISM

REFERENCES:

- [1] 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](https://doi.org/10.52058/3041-1254-2024-5(5)-746-762)
- [2] 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.
- [3] 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
- [4] 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
- [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](https://doi.org/10.52058/2708-7530-2024-5(47)-855-875)
- [7] Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. *Λ'ΟΓΟΣ*. 2020. P. 55-57.
- [8] Krasnyuk M., Tkalenko A., Krasniuk S. Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. *Λ'ΟΓΟΣ*. 2021.
- [9] Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. *Λ'ΟΓΟΣ*. 2021.
- [10] 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>
- [11] 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](https://doi.org/10.52058/3041-1254-2024-4(4)-776-794)
- [12] 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
- [13] Krasnyuk M., Krasniuk 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.
- [14] 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>
- [15] 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>

- [16] 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.
- [17] 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.
- [18] 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: THEU, 2014. С. 211-212.
- [19] 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.