

SECTION 11.

PHILOLOGY AND JOURNALISM

Svitlana Goncharenko 

Senior Lecturer

Institute of Law and Modern Technologies

Kyiv National University of Technologies and Design, Ukraine

Svitlana Krasniuk 

Senior Lecturer

Department of philology and translation

Kyiv National University of Technologies and Design, Ukraine

MODERN DATA MINING IN PHILOLOGY

Introduction.

Modern data mining is a set of methods and technologies for automated detection of hidden patterns, trends, and relationships in large data sets [1, 2, 3]. It combines elements of statistics, machine learning, artificial intelligence, and databases to analyze data from various fields — from business to medicine, from finance to hydrocarbon extraction, etc. [4].

Machine learning as a component of data mining is an important component for analyzing large volumes of data and identifying hidden patterns, connections and knowledge [5, 6, 7]. Machine learning (ML) and Data Mining have a common goal — automatically finding valuable information (regularities, patterns) from data sets [8, 9], however, machine learning is part of the data mining process, responsible for building models and algorithms that allow making predictions, classification and other tasks [10]. Machine learning is a powerful tool in data mining, helping to automatically find patterns and trends in large data sets [11, 12, 13]. This allows business and academics to improve the decision-making process, increase the efficiency of operations and find new knowledge that was not available with traditional methods of analysis.

Considering the direction of the author's experience and author's research, it is also worth highlighting such a direction of data mining as Educational Data Mining (EDM) — it is the use of data analysis methods to study educational processes, identify patterns and improve educational results. It helps educational institutions analyze student behavior, identify problems in educational processes, and make decisions based on data to improve the quality of education [14, 15].

The Main Part.

Data Mining in philology is the application of methods and algorithms for the automatic analysis of large volumes of textual information in order to identify new patterns, trends and gain knowledge from textual data [16]. In philological research, this tool is becoming increasingly popular due to its ability to analyze large arrays of texts, literary works, language corpora, etc. Key aspects of the use of data mining in philology are discussed below:

1. Corpus linguistics

- Analysis of large text corpora: Data mining helps in the automated analysis of large language corpora, which can contain millions of words and thousands of texts from different authors, genres and eras. Using clustering algorithms, frequency analysis, phrase search, it is possible to investigate language features and trends in literature.

- Statistical analysis of linguistic phenomena: Data mining algorithms allow to analyze the frequency of use of words, grammatical constructions, semantic structures in different languages or texts, which helps to study changes in language and stylistics.

2. Stylometry

- Identification of authorship: Data mining helps to analyze the stylistic features of the text, such as the length of sentences, frequency of word usage, syntactic constructions, rhythm, which allows to draw conclusions about the authorship of the texts. For example, the analysis of literary works can help establish who is the author of anonymous or controversial works.

- Comparative analysis of style: Stylometry also allows you to compare the styles of different authors or to determine the changes in the style of one author during his creative journey, using machine learning to classify texts according to stylistic features.

3. Analysis of topics and semantics

- Topic Modeling: Topic modeling algorithms such as Latent Dirichlet Allocation (LDA) are used to automatically detect topics in large text arrays. It helps to identify the main themes discussed in certain texts or genres and their relationships.

- Semantic analysis: Data mining allows defining lexical fields, analyzing semantic connections between words and phrases, which can be useful for in-depth analysis of the content of literary works or scientific texts.

4. Sentiment analysis

- Mood analysis in literature: Data mining helps to analyze the emotional coloring of the text, revealing positive, negative or neutral moods in literary works.

This allows researchers to analyze how authors express emotions in their works or how the tone of the text changes depending on the context.

- Social media and modern language: Sentiment analysis is also used to analyze speech in social networks, blogs and online forums, which helps to study modern language trends, understand changes in lexicon and stylistics at the level of everyday communication.

5. Machine translation and lexicography

- Analysis of multilingual texts: Data mining is used to improve the quality of machine translation by analyzing large bilingual or multilingual text corpora. It helps improve translation matching by automatically learning equivalent structures between languages.

- Automatic creation of dictionaries: Lexicography also uses data mining to automatically create or update dictionaries by analyzing new words, phrases and language structures in different texts.

6. Analysis of language evolution

- Diachronic analysis: Data mining allows studying changes in language and stylistics throughout history. By analyzing ancient texts and comparing them with modern works, it is possible to trace how language norms, vocabulary and grammar changed.

- Changes in cultural ideas: By analyzing texts from different historical periods, it is possible to determine how cultural, social and political ideas reflected in literature or journalism have changed.

7. Study of folklore and ethnolinguistics

- Analysis of fairy tales and folk texts: Data mining allows automatic analysis of a large number of folklore texts, revealing common motifs, structures and themes in folk works. This helps researchers better understand the cultural and linguistic characteristics of different peoples.

- Ethnolinguistic analysis: By analyzing texts from different regions and language communities, data mining can help identify ethnolinguistic features such as specific terms, cultural realities and unique forms of expression.

Conclusions.

Modern data mining provides companies and researchers with powerful tools to extract valuable information from huge amounts of data. Using a variety of algorithms, methods, and tools, you can find new business opportunities, improve customer service, predict market trends, and make important scientific discoveries.

Data mining in education opens up new opportunities for improving the quality and efficiency of education. The use of data makes it possible to adapt educational

programs to the individual needs of students, predict their success, improve educational methods and management processes. This helps make education more personalized, flexible and result-oriented.

Data mining in philology opens up new opportunities for the study of linguistic phenomena, literary works and textual data in general. By automating the analysis of large volumes of text, researchers can discover patterns and trends that would be difficult or impossible to detect using traditional methods. This makes data mining a powerful tool for modern philologists and linguists, contributing to the development of science and expanding the understanding of language and literature.

Deep analysis and analytics of big data is a complex process of collecting, processing, analyzing and interpreting large volumes of data that are constantly generated from various sources [17, 18]. Big Data Analytics is aimed at identifying patterns, trends, and relationships in data to help organizations make more informed decisions [19, 20]. Deep analysis and big data analytics are powerful tools for obtaining valuable information that allows organizations and enterprises to gain a competitive advantage, optimize processes and make informed decisions. Big data is changing the way business and science work with information, opening new horizons for innovation and development.

The authors claim that in-depth intellectual analysis and analytics of big data continues to gain applied relevance in philological studies of linguistic and literary phenomena [21]. Thanks to the development of digital technologies, philologists gained access to large-scale text corpora that were previously impossible to process manually [22]. The use of big data makes it possible to analyze millions of texts, study language patterns, identify themes, styles, and even make predictions about language development.

Big data is revolutionizing philological research, opening new horizons for the analysis of texts, language and literature. This allows researchers to work with huge amounts of information, find new patterns, explore language changes and make new discoveries. Thanks to the development of tools for processing and analyzing big data, philology is becoming a more technologically advanced and interdisciplinary science.

The authors separately emphasize that deep machine learning in philology opens up new opportunities for automation and improvement of many processes in natural language processing (NLP), text analysis, linguistic research and translation [23]. Thanks to the ability of deep neural networks to detect complex patterns [24] in large text corpora, this approach allows solving tasks that were previously inaccessible or required significant human resources.

Discussion and prospects for further research.

The authors propose that the application of soft computing, in particular, using genetic algorithms, is an interesting and promising application of artificial intelligence methods for solving problems in linguistics, text analysis, translation, semantic analysis, and other philological tasks. Genetic algorithms are one of the methods of evolutionary programming that simulate the process of natural selection with the aim of finding optimal solutions for complex tasks [25, 26, 27]. In other words, the authors put forward the thesis that genetic algorithms in philology have great potential for solving complex linguistic problems that are difficult to deal with traditional approaches. They can be effective in the tasks of automatic translation, semantic analysis, stylistic analysis and many other aspects of philological analysis, helping specialists to find new approaches to the study and processing of textual data.

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. 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. 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.
4. Krasnyuk, M., & Krasniuk, S. (2020). Application of artificial neural networks for reducing dimensions of geological-geophysical data set's for the identification of perspective oil and gas deposits. *Збірник наукових праць ЛОГОС*, 18-19. <https://doi.org/10.36074/24.04.2020.v2.05> DOI: <https://doi.org/10.36074/24.04.2020.v2.05>
5. Краснюк М.Т. Проблеми застосування систем управління корпоративними знаннями та їх таксономія. *Моделювання та інформаційні системи в економіці: міжвід. наук. зб. Відп. ред. В.К. Галіцин. К.: КНЕУ, 2006. Вип. 73. С. 34-44.*
6. Tuhaienko V., Krasniuk S. Effective application of knowledge management in current crisis conditions. *International scientific journal "Grail of Science"*. 2022. № 16. pp. 348-358.
7. Кулинич Ю.М., Краснюк М.Т. (2022) Ефективна інтеграція орієнтованої на знання концепції бізнес-правил в рамках управляючої інтелектуальної корпоративної інформаційної системи // *Current state and prospects for the development of enterprises in Ukraine : theory, methodology, practice* : a collective monograph / Collective of authors. Poltava: РС «Astraya», 2022. – С. 73-79
8. Науменко М. А. Моделі бізнесових знань в системах штучного інтелекту для ефективного конкурентного підприємства // *Міжнародний науковий журнал "Інтернаука"*. Серія: "Економічні науки". – 2024. – №6. <https://doi.org/10.25313/2520-2294-2024-6-10010>
9. Ситник В.Ф., Краснюк М.Т. Політика управління знаннями нафтогазової компанії як ключовий фактор підвищення її ефективності. *Проблеми формування ринкової економіки: міжвідомчий науковий збірник*. Відп. ред. О.О. Беляєв. К.: КНЕУ, 2002. Вип. 10. 326 с.
10. Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. *ЛОГОС*. 2020. P. 55-57.
11. Krasnyuk M., Tkalenko A., Krasniuk S. Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. *ЛОГОС*. 2021.
12. Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. *ЛОГОС*. 2021.
13. 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)
14. Tetiana Tsalko, Svitlana Nevmerzhytska, Svitlana Krasniuk, Svitlana Goncharenko, Liubymova Natalia (2024). 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

15. Краснюк Світлана. (2024) Data Science у освітньому менеджменті // *Діалог культур у Європейському освітньому просторі*: 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.
16. Krasniuk, S., & Goncharenko, S. (2024). MODERN MATHEMATICAL LINGUISTICS. *Collection of Scientific Papers «SCIENTIA»*, (September 13, 2024; Pisa, Italia), 115–119. Retrieved from <https://previous.scientia.report/index.php/archive/article/view/2060>
17. 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
18. 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>
19. 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.
20. 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>
21. Goncharenko, S., & Krasniuk, S. (2024). Innovative architecture of large language models. *In Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування*. Національний авіаційний університет.
22. Krasniuk, S., & Goncharenko, S. (2024). Ethics of using large language models in machine linguistics. *In Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування*. Національний авіаційний університет.
23. 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)
24. Maxim Krasnyuk, Svitlana Krasniuk, Svitlana Goncharenko, Liudmyla Roienko, Vitalina Denysenko, Liubymova Natalia (2023). 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>
25. Krasnyuk, M., & Krasniuk, S. (2024). Evolutionary technologies and genetic algorithms in machine translation. *European Science*, 3(sge30-03), 91–98. <https://doi.org/10.30890/2709-2313.2024-30-00-025>
26. Науменко, М., & Краснюк, М. (2024). Ефективне застосування генетичних алгоритмів у вирішенні багатоекстремумних оптимізаційних задач в менеджменті конкурентного підприємства. *Grail of Science*, (41), 65–73. <https://doi.org/10.36074/grail-of-science.05.07.2024.008>
27. Kulynych, Y. ., Krasnyuk, M. ., & Krasniuk, S. . (2022). Efficiency of evolutionary algorithms in solving optimization problems on the example of the fintech industry. *Grail of Science*, (14-15), 77–84. <https://doi.org/10.36074/grail-of-science.27.05.2022.010>