SECTION 11.
PHILOLOGY AND JOURNALISM

Svitlana Krasniuk
Senior Lecturer
Department of philology and translation
Kyiv National University of Technologies and Design, Ukraine

ADVANCED TEXT MINING IN PHILOLOGY

Introduction.

Big data in philology is an interaction between the latest technologies of Big Data processing and analysis [1, 2, 3, 4] and traditional philological research. Thanks to the processing of large text arrays, researchers can conduct in-depth analysis of linguistic, literary and cultural phenomena. This integration provides new opportunities for the study of texts, while at the same time posing new challenges for philologists in the use of modern technologies.

Deep text intellectual analysis of philological data (usually Big Text Data) is the use of advanced methods of machine learning [5, 6, 7, 8], in particular deep learning, to analyze large volumes of textual information in philological research. The purpose of such an analysis is not only the processing of texts, but also the detection of hidden semantic, syntactic and pragmatic regularities [9, 10], which contribute to a deeper understanding of linguistic phenomena and literary texts.

1. The importance of deep text analysis in philology.

Philological studies include the study of languages, literature, the cultural context of texts and their impact on society. Deep text analysis allows:

Identify linguistic structures that remain invisible to traditional methods of analysis.

To study the evolution of languages through the analysis of large corpora of texts from different eras.

Analyze the style and tone of authors and literary works, revealing hidden motives or narrative structures.

Find intertextual connections between different texts or authors.

2. Features of in-depth text analysis

2.1. Deep text mining uses a variety of techniques that allow not only to analyze the surface characteristics of the text (such as word frequency or basic classification), but also to understand more complex aspects such as:

2.2. Semantic understanding: analysis of text content based on deep context and connections between words and phrases.

2.3. Pragmatics and contextuality: evaluating the influence of context on the meaning of words and phrases, which is important in the study of fiction and interlanguage communication.

2.4. Identifying irony and sarcasm: More sophisticated models can recognize the indirect or hidden meaning of utterances, which is a challenge for traditional algorithms.

3. Technologies and methods.

The main methods used for deep text analysis are based on the latest advances in deep learning:

3.1. Transformers: These are deep learning architectures such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) that provide a powerful mechanism for understanding context, allowing systems to consider the meaning of words in different contexts. These models take into account the context both at the level of individual sentences and at the level of the entire text. This allows for a better understanding of the relationships between words and phrases, particularly in literary or scientific texts.

3.2. Sequential models, i.e. Recurrent Neural Networks (RNN) and LSTM (Long Short-Term Memory): are used to analyze texts with long dependencies, where it is important to consider the previous context to understand the current text.

3.3. Topic modeling models (eg, Latent Dirichlet Allocation - LDA): enable the extraction of underlying themes or motifs in large collections of texts.

3.4. Semantic networks: used to create models that represent the relationships between different concepts in a text and allow the identification of interactions between terms.

4. Main areas of application

In-depth intellectual analysis of texts has numerous areas of very effective application (especially for large arrays of semi-structured text):

4.1. Analysis of literary works to determine author's styles, themes, emotional components and hidden meanings in literary texts. This allows for a detailed analysis of both classical and modern works. Thanks to deep learning, it is possible to conduct a detailed analysis of the authors' style. For example, models can reveal characteristic features of style, such as lexical choice, syntactic structures or tonality, which allows determining the authorship of texts or exploring the development of a writer's style during his work.

4.2. Social media: Monitoring and analyzing emotional state, public opinion or even political sentiment based on texts in social networks.

4.3. Automatic text summarization: Using algorithms to generate short summaries of large documents, useful in journalism, law or scientific research.

4.4. Information retrieval: Improving the accuracy of search engines by better understanding the text query and the relevance of the documents found.

4.5. Semantic analysis and meaning: deep models allow studying the content of texts at the level of semantics. This includes identifying ambiguity, the complex shades of meaning of words and phrases, which is critical to literary criticism, especially in poetry or metaphorical speech.

4.6. Pragmatic and contextual analysis: By analyzing a text in its broader context, deep learning models can detect complex pragmatic elements such as tone, irony, sarcasm, or other rhetorical devices. This is especially useful for studying literature or political texts.

4.7. Thematic and lexical analysis of literature: In-depth analysis can automatically identify main themes or lexical fields in large corpora of literary texts. This helps philologists explore the structure and content of works in a new way, especially in comparative literary studies.

5. Advantages and challenges of deep textual analysis

5.1. Advantages:

Automation and Scalability: The ability to analyze large sets of texts faster than traditional methods.

Improved understanding of text: Deep learning models are capable of interpreting complex relationships in text that cannot be achieved by standard methods.

New opportunities for interdisciplinary research: the integration of philological research with other disciplines such as cognitive linguistics, cultural studies or history.

5.2. Challenges:

Linguistic ambiguity: deep models are sometimes wrong in understanding complex linguistic constructions, especially those that depend on context or sociocultural features [11, 12]. Even modern algorithms sometimes get it wrong in cases where the meaning of words or phrases is highly dependent on complex social or cultural contexts.

Context sensitivity: Some literary texts require a deep cultural or historical context that is difficult to embed in machine models.

The need for large corpora: Quality training of models requires large amounts of textual data, which can be a problem for rare or understudied languages.

Polysemy: Ambiguity remains a problem, as some words can have multiple meanings, and determining the correct one requires additional knowledge of the context.

Multilingualism: Developing models capable of analyzing text in multiple languages simultaneously remains a technical challenge.

6. Examples of the application of deep textual analysis

6.1. Analysis of fiction

Using deep learning to study narrative structures, stylistic features, symbolism, and themes in classic and contemporary literary works. For example, automatic analysis of novels allows you to identify thematic changes and variations in the development of the plot.

6.2. Corpus linguistics

Deep models can analyze large language corpora to study the evolution of linguistic phenomena, new words, stylistic changes, or social and cultural influences on language.

6.3. Lexicography and linguistic research

Deep learning can help create new dictionaries and databases that reflect modern linguistic realities, as well as identify changes in the usage of words and expressions.

Conclusions.

Deep textual intellectual analysis of philological data opens up new opportunities for the study of language, literature and culture. The use of modern deep learning technologies allows obtaining new insights from large text arrays, which makes this method indispensable for modern philology and interdisciplinary research.

Discussion and prospects for further research.

Hybrid models in deep text mining combine different approaches and algorithms for the analysis of large text data, which allows to increase the accuracy and flexibility of the analysis [13, 14]. The use of such models is important in situations where the same type of approaches do not provide adequate efficiency or depth of analysis [15, 16]. Hybrid models in deep text mining are a promising approach that combines the best features of various text processing and data analysis methods. They make it possible to increase the efficiency and accuracy of the analysis of philological data, opening up new opportunities for research in the field of linguistics and literary studies.

This direction of scientific research of the authors will be reflected in the following publications.

References:

1. 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
2. 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.

3. 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
4. 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
5. 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
6. Krasnyuk M., Krasniuk S. Modern practice of machine learning in the aviation transport industry. *ΛΟΓΟΣ*. 2021.
7. Krasnyuk M., Krasniuk S. Comparative characteristics of machine learning for predicative financial modelling. *ΛΟΓΟΣ*. 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. *ΛΟΓΟΣ*. 2021.
9. 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
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. 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.
12. Goncharenko, S., & Krasniuk, S. (2024). Innovative architecture of large language models. *In Лінгвістичні та методологічні аспекти викладання іноземних мов професійного спрямування*. Національний авіаційний університет.
13. 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.
14. 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.
15. 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.
16. 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.