Proceedings of the 17th International Scientific and Practical Conference «International Forum: Problems and Scientific Solutions»

(September 26-28, 2026). Melbourne, Australia

No 311

LITERARY STUDIES

Innovative Trends and Prospects of Evolutionary Computation for Complex Authorship Attribution: Navigating Uncertainty and High-Dimensional Linguistic Environments

Svitlana Krasniuk[1]

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

Abstract. The article examines innovative approaches, trends, and prospects regarding the application of evolutionary computation to solve complex problems in authorship attribution and textology. It substantiates the effectiveness of these methods for the intelligent selection and combination of stylometric features, which enable the identification of individual markers of authorial style. Furthermore, the study explores the potential for hybridizing evolutionary algorithms with machine learning and natural language processing technologies amidst turbulence, uncertainty, and systemic risks.

Keywords: *modern linguistics, computational linguistics, evolutionary computation, evolutionary algorithms, authorship attribution, textology, stylometry, authorial style, linguistic features, text mining, machine learning, natural language processing, feature optimization, uncertainty, turbulence, systemic risks.*

INTRODUCTION. Modern linguistics is evolving amidst the rapid growth of digital text data, increasingly complex communication environments, the emergence of new genres and forms of written communication, and constantly shifting linguistic practices [1-3]. While the digitalization of humanities research significantly expands the potential for automated text analysis, it simultaneously creates new methodological challenges-such as high-dimensional feature spaces, heterogeneous text corpora, variations in authorial style, and data uncertainty [4-6]. Tasks related to authorship attribution and textology are particularly complex in this context, as identifying a text's true author often requires analyzing a vast array of diverse linguistic features with unknown interrelationships.

LITERARY STUDIES

Classical authorship attribution methods generally rely on the statistical comparison of texts based on a predefined set of characteristics. However, when analyzing anonymous, disputed, stylistically heterogeneous, or historically evolved texts, a fixed set of features may prove insufficient. An author's individual style is shaped by a multitude of interconnected characteristics: word and function-word frequencies, preferred syntactic structures, sentence length, parts-of-speech distribution, punctuation patterns, the use of set phrases, lexical diversity, and other parameters. Consequently, the challenge lies not only in text classification but also in the intelligent identification of the most informative combination of features capable of distinguishing an author's unique style from the general characteristics of the language, genre, and era. Evolutionary computation-based on the principles of search, adaptation, and solution selection-is of particular interest in this context [7]. Genetic algorithms, genetic programming, evolutionary strategies, and other forms of evolutionary optimization enable the exploration of an immense space of potential feature combinations and the gradual development of models that ensure high discriminative power in text attribution [8]. Their fundamental advantage lies in the ability to perform not only automatic feature selection but also the discovery of feature combinations, interactions, and structural dependencies [9].

The relevance of applying evolutionary computation is further heightened by the dynamic and uncertain nature of the modern linguistic landscape [10-12]. Shifts in linguistic norms, the emergence of new digital genres, the transformation of communicative practices, the growth of data volumes, the presence of noise and corpus incompleteness, and the proliferation of automated text generation and editing technologies create a situation where the robustness of traditional authorship attribution models may diminish. Consequently, the development of adaptive intelligent methods capable of accounting for data variability and re-optimizing the feature space represents a promising direction.

Therefore, the aim of this article is to substantiate innovative approaches, current trends, and future prospects regarding the use of evolutionary computation to address complex tasks in authorship attribution and textology-specifically through the intelligent identification of an author's unique stylometric markers amidst the high

LITERARY STUDIES

dimensionality, uncertainty, and dynamic variability of the linguistic environment.

Main Part. Attribution of authorship is the task of establishing the likely author of a text based on its linguistic and stylistic characteristics. In the most complex version, such a task involves the presence of an anonymous or controversial text and a set of texts by well-known authors, which are used as a reference sample. In this, the researcher must determine to what extent the structure of the unknown text corresponds to the individual characteristics of one of the potential authors.

An important feature of this task is that the author's style does not exist in the form of one separate feature. It is a complex system regarding the persistent preferences of the author, which are revealed at different levels of the text. Potentially significant signs include the frequency of certain words, especially functional ones; length of words and sentences; distribution of punctuation marks; frequency of parts of speech; syntactic constructions; lexical diversity; the nature of the use of conjunctions, prepositions and pronouns; repetition of certain grammatical patterns; paragraph structure; persistent phrases and other characteristics.

Thus, the task of attribution can be represented as a search for such a configuration of features that provides the most reliable distinction of authors while minimizing the influence of random and contextually determined characteristics of the text. The problem is that the number of possible combinations of features is extremely large. Even a relatively small set of features with different methods of their selection and combination forms a huge space of potential models. It is practically impossible to go through all the options, and the use of a single pre-installed model may not take into account the individual specifics of a particular case.

It is here that the field of application of evolutionary computing is formed. Their logic allows you to consider different sets and combinations of stylometric features as competing solutions, evaluate them according to a given quality criterion, and with the help of iterative selection, obtain more and more effective configurations. In this case, the "evolving" objects can be both individual features, but also whole sets of features, rules for their association, parameters of classifiers or structures of mathematical models.

LITERARY STUDIES

One of the most promising approaches is the use of genetic algorithms for automated trait selection. In a simplified representation, each possible combination of stylometric characteristics can be coded as a separate individual of the population. The presence or absence of a specific feature is determined by the corresponding element of the chromosome. For each combination, a fitness function is calculated that reflects the quality of the attribution. After that, the best solutions are used to form new combinations using selection, crossover and mutation operators.

Such a mechanism is fundamentally different from the simplest selection of the most frequent or statistically significant features. The evolutionary algorithm is able to detect combinations of features, which individually may have limited distinguishing ability, but collectively form a rather specific author profile. For example, a certain frequency of service words as such may be characteristic of a large number of authors. However, its combination with a certain distribution of sentence length, features of punctuation and the frequency of specific syntactic constructions can form a much more individualized stylometric marker.

This highlights one of the key advantages of the evolutionary approach: the search is conducted not only within the space of individual features but also within the space of their interactions. Consequently, evolutionary computation can be used to construct a kind of "evolutionary portrait" of an author's style, wherein individuality is defined by a combination of relatively weak yet complementary characteristics.

Another promising area involves the application of genetic programming. Unlike the classical genetic algorithm, which focuses primarily on optimizing a fixed solution structure, genetic programming allows the structures of analytical expressions and rules themselves to evolve. For authorship attribution tasks, this opens up the possibility of automatically generating interpretable classification rules that combine various stylometric metrics.

For instance, the system might search not merely for a set of features but for a specific structure of their interaction-such as the relative frequencies of certain function words, average sentence length, punctuation variability, and lexical diversity indices. This can result in the formation of an individualized analytical construct

LITERARY STUDIES

that describes the differences between authors not only through a final classification category but also through the set of characteristics that influenced the decision.

The issue of the fitness function becomes particularly critical here. In authorship attribution tasks, it should not be limited solely to a classification accuracy metric. High accuracy on a specific corpus might stem from the genre, time period, subject matter, or length of the texts, rather than from the author's actual style. Therefore, multi-criteria evaluation functions-which simultaneously account for attribution accuracy, result stability, feature set compactness, the model's ability to generalize to new texts, and noise robustness-offer greater promise.

Adopting the principle of multi-criteria optimization allows for a shift from seeking a merely "accurate" model to finding a balanced solution. Consequently, an optimal configuration can simultaneously ensure high discriminative power and a relatively small number of features. This is particularly significant in textology, where a researcher needs not only to obtain an automated result but also to interpret the grounds upon which that result was generated.

An important avenue of development involves hybridizing evolutionary computation with machine learning methods. An evolutionary algorithm can serve as an intelligent feature selector and parameter optimizer, while the machine learning classifier performs the actual attribution. Such an architecture enables the separation of two interdependent tasks: identifying an informative feature space and constructing a classification model.

In the future, such hybrid systems could integrate evolutionary algorithms, statistical analysis methods, machine learning, natural language processing, and big text data analysis technologies. In this setup, the evolutionary component performs adaptive search, while other intelligent methods handle the extraction, representation, and interpretation of linguistic information.

The application of evolutionary calculations in textology acquires special importance. If classical attribution mainly solves the question "to whom does the text belong?", textological research can contain a wider set of tasks: establishing the origin of the text, comparing redactions, identifying later interventions, analyzing variants, reconstructing the text and determining the degree of proximity of different text versions.

LITERARY STUDIES

In these tasks, evolutionary methods can be used to find the optimal correspondence between text variants, identify stable and changing elements, cluster texts and identify hidden structural regularities. The use of evolutionary mechanisms for the analysis of large corpora, in which the manual comparison of all text variants becomes time-consuming or practically impossible, is especially promising.

It is also important to take into account the dynamic nature of the author's style. The author does not necessarily use the same language means throughout the entire creative period. The style may change depending on the author's age, genre, subject, historical period, addressee and communicative situation. Therefore, a model built for one time period may show a decrease in efficiency when analyzing texts from another period.

In this context, evolutionary computing provides the possibility of periodic adaptation of the model. At least the arrival of new data of the decision population can be subjected to additional optimization, as a result of which the system is able to change the set of the most significant features. Such adaptability is particularly important for digital linguistics, where the linguistic environment changes much more rapidly than traditional historical corpora.

Additional complexity is created by the systemic risks of digital textology. These include incompleteness of corpora, digitization errors, differences in text presentation standards, metadata of inadequate quality, genre biases of the sample, insufficient volume of reference texts, and the presence of texts that have undergone editorial or automatic transformation. If such factors are not taken into account, the algorithm is able to accept random features of the corpus as stable authorial features.

Therefore, a promising system of evolutionary attribution should be oriented not only to the maximization of quality, but also to the management of uncertainty. In the evaluation function, it is advisable to take into account the sensitivity of the result to changes in the composition of the training sample, resistance to noise, variability of signs and the degree of confidence in the obtained decision. Because of the war, it is advisable to present the result of attribution not as an absolute statement, but as a reasonable probable conclusion with an indication of the stability of the result.

This approach becomes especially relevant under the conditions of the spread of generative artificial

LITERARY STUDIES

intelligence. Automatically generated or significantly human-edited texts may violate traditional notions of the author's style. Blending of human and machine text, automatic redaction, paraphrasing, and translation can alter individual stylometric characteristics. Therefore, future attribution systems should take into account both traditional authorship markers and the probability of transformation of the original text.

Evolutionary computations can be useful here as well, as they allow adaptive search for stable features in the conditions of a changed feature space. At the same time, the joint use of surface stylistic signs and deeper linguistic representations, including syntactic, semantic and discursive characteristics, becomes promising.

Thus, the modern trend involves a shift from static attribution to adaptive, intelligent attribution. In the traditional model, the set of features and system parameters are predefined by the researcher. In the evolutionary model, the analytical configuration itself becomes the object of optimization. The system does not merely answer the question of whether a text belongs to a specific author; rather, it seeks the most informative representation of the author's style for a given corpus and research task.

Looking ahead, the development of multi-level evolutionary linguistic systems is a possibility. The first level involves the collection and normalization of text data; the second, the automatic extraction of a wide range of potential features; the third, the evolutionary selection and combination of features; the fourth, machine classification; the fifth, the assessment of result stability and uncertainty; and the sixth, the researcher's interpretation of the identified authorial markers.

Human-machine interaction takes on particular importance in this process. Despite the high computational efficiency of the algorithms, the final textological interpretation must account for historical, cultural, genre-related, and biographical contexts. Evolutionary computing should be viewed not as a replacement for the textology specialist, but as a tool to expand their analytical capabilities. The machine is capable of exploring a vast space of feature combinations, while the researcher provides a meaningful interpretation of the patterns discovered.

A promising avenue is the creation of self-adapting evolutionary systems capable of adjusting search parameters

LITERARY STUDIES

based on the nature of the corpus. For historical texts, certain groups of features may take priority, whereas for modern digital communication, others do. Similarly, the attribution of fiction, scholarly texts, journalism, social media content, and business communication requires distinct models of authorial profiles.

Evolutionary computation may gradually evolve from a supplementary optimization tool into an independent intelligent mechanism for generating linguistic hypotheses. Its potential lies in the ability to handle vast numbers of candidate solutions, uncover latent feature combinations, adapt to changing data, and identify robust patterns amidst uncertainty.

Conclusions.

The analysis demonstrates that evolutionary computation holds significant potential for addressing complex tasks related to authorship attribution and textology in modern linguistics. Its fundamental value lies in the ability to conduct an intelligent search within the high-dimensional space of potential stylometric features and to identify combinations of these features that most comprehensively characterize the unique nature of an author's style.

A key application of evolutionary computation is the automated selection and combination of features characterizing word frequency, sentence length, punctuation patterns, syntactic structures, lexical diversity, and other textual parameters. In this context, an individual authorial marker is viewed not as an isolated characteristic, but as the result of a complex interplay among a multitude of relatively stable features.

The development of hybrid evolutionary-intelligent systems-combining genetic algorithms, genetic programming, machine learning methods, statistical analysis, and natural language processing technologies-represents a promising direction. Such integration enables the simultaneous optimization of feature spaces, model parameters, and classification procedures, thereby enhancing the analytical system's adaptability.

Amidst turbulence, uncertainty, and systemic risks, the shift from static attribution models to adaptive systems is becoming increasingly important. Changes in linguistic practices, the structure of digital corpora, genres, editing methods, and the emergence of texts created or transformed by artificial intelligence necessitate the continuous

LITERARY STUDIES

verification of authorial marker stability and the capacity for model re-optimization. For textology, the potential of evolutionary computing lies not only in authorship attribution but also in the automated comparison of textual variants, the identification of transformation patterns, corpus clustering, and support for reconstructing textual history. This expands the scope of evolutionary methods from the narrow task of classification to comprehensive intelligent support for textological research.

Thus, the future development of evolutionary computing in modern linguistics should be linked to the creation of adaptive, hybrid, and interpretable intelligent systems for authorship attribution. Their prospective architecture must ensure not only high classification accuracy but also robustness against noise and incomplete data, uncertainty assessment, the interpretability of identified authorship markers, and adaptability to new text types.

In a broader scientific context, evolutionary computing facilitates a promising shift from the traditional search for predefined authorship features to the intelligent discovery of optimal configurations. It is precisely the ability of evolutionary algorithms to explore vast numbers of potential feature combinations-and to identify stable, individual patterns within them-that lays the foundation for a new generation of authorship attribution systems and digital textology, designed to operate within the dynamic, uncertain, and rapidly evolving environment of modern linguistics.

References:

[1] Zhang, Z., Song, W., & Liu, P. (2024). Making and interpreting: Digital humanities as embodied action. Humanities and Social Sciences Communications, 11, Article 13. https://doi.org/10.1057/s41599-023-02548-3

[2] De Luca, E. W., Fallucchi, F., Ghattas, B., & Spielhaus, R. (2024). The digital transformation processes for supporting digital humanities researchers in text analysis. Journal of Documentation, 80(2), 378-391. https://doi.org/10.1108/JD-07-2022-0143

[3] Goncharenko S. (2025). Methodological bases of innovative information systems in the field of computer linguistics. Trends, Issues, and Challenges in Modern Science : Proceedings of the 2nd International Scientific Conference (Cambridge, United Kingdom, 5 September 2025). - Lulu Press, Inc., 2025. - P. 100-103. URL: https://er.knutd.edu.ua/handle/123456789/31071

[4] Smits, T., & Wevers, M. (2023). A multimodal turn in Digital Humanities: Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. Digital

LITERARY STUDIES

Scholarship in the Humanities, 38(3), 1267-1280. https://doi.org/10.1093/llc/fqad008

[5] Rettberg, J. W. (2022). Algorithmic failure as a humanities methodology: Machine learning's mispredictions identify rich cases for qualitative analysis. Big Data & Society, 9(2). https://doi.org/10.1177/20539517221131290

[6] Grieve, J., Bartl, S., Fuoli, M., Grafmiller, J., Huang, W., Jawerbaum, A., Murakami, A., Perlman, M., Roemling, D., & Winter, B. (2024). The sociolinguistic foundations of language modeling. Frontiers in Artificial Intelligence, 7, 1472411. https://doi.org/10.3389/frai.2024.1472411

[7] Krasnyuk M.T., Naumenko M.A. (2024) Efektyvne zastosuvannia henetychnykh alhorytmiv u vyrishenni bahatoekstremumnykh optymizatsiinykh zadach v menedzhmenti konkurentnoho pidpryiemstva [Effective application of genetic algorithms in solving multi-extremum optimization problems in the management of a competitive enterprise]. Grail of Science, №41, June 2024, pp. 65-73. URL: https://archive.journal-grail.science/index.php/2710-3056/issue/view/05.07.2024/29 [in Ukrainian].

[8] Goncharenko S. (2025). Mathematical optimization as element of innovative educational management in modern complex conditions. Science and Global Challenges in the Modern World : Proceedings of the 2nd International Scientific Conference (Leicester, United Kingdom, 5 October 2025). - Lulu Press, Inc., 2025. - P. 90-94. URL: https://er.knutd.edu.ua/handle/123456789/31667

[9] 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, May 2022, (pp. 63-70) DOI: https://doi.org/10.36074/grail-of-science.27.05.2022

[10] Araujo, L. (2009). How evolutionary algorithms are applied to statistical natural language processing. Artificial Intelligence Review, 28(4), 275-303. https://doi.org/10.1007/s10462-009-9104-y

[11] Manzoni, L., Jakobovic, D., Mariot, L., Picek, S., & Castelli, M. (2020). Towards an evolutionary-based approach for natural language processing. arXiv preprint. arXiv:2004.13832. https://arxiv.org/abs/2004.13832

[12] Bungum, L., & Gambäck, B. (2016). Evolutionary Algorithms in Natural Language Processing. In Proceedings of the Workshop on Natural Language Processing and Computational Linguistics. https://www.semanticscholar.org/paper/Evolutionary-Algorithms-in-Natural-Language-Bungum-Gamb%C3%A4ck/d2424a7f83be06b3abd99485ee22f8fd90cb602c