METHODOLOGY FOR ASSESSING PATTERN COMPLEXITY

Alexey Myachin

National Research University Higher School of Economics
V.A. Trapeznikov Institute of Control Science of Russian Academy of Science

Introduction

Methods for pattern analysis, widely utilized in data analysis tasks, have several key areas of development. One critical area is the minimization of noise impact on final outcomes. Measurement errors, outliers, and other anomalies can distort analysis, making it essential to develop robust methods resistant to such factors. Another focus is dimensionality reduction, which helps decrease computational complexity and accelerate the processing of large datasets. A third area is ensuring interpretability and evaluating result quality. Employing metric indicators and visualization techniques aids in assessing the significance and reliability of conclusions, making the analysis process more transparent. Scalability of methods also remains a priority. The development of distributed algorithms and the use of cloud computing enable efficient processing of large-scale data while reducing analysis time. Optimizing the number of patterns plays a significant role in the interpretability of results. Too few patterns may distort the data structure, while an excessive number complicates interpretation. Finally, integration with traditional analytical methods broadens the scope for interdisciplinary research, enhancing model adaptability and practical relevance in areas such as medicine or financial analysis.

This study contributes to the third area by enhancing the interpretability of final results. To this end, the concepts of "empty pattern" and "pattern complexity" are introduced and explored.

Proposed Methods

The general formulation of the pattern analysis problem is presented in several works, e.g., [Aleskerov 2014; Myachin 2019]. The input data consist of n objects $x_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im})$, which need to be partitioned into non-overlapping subsets (patterns p) with similar structures based on a selected set of indicators.

Given the introduction of two new terms, the necessity of their use is justified as follows. Pattern analysis methods have been effectively applied to predictive models, such as forecasting COVID-19 incidence [Aleskerov et al. 2022] and analyzing tuition pricing in Russian universities [Dmitrienko, Myachin 2023]. A key objective of pattern analysis is the identification of dynamic groups of objects exhibiting similar developmental structures. Each pattern may represent a group of objects requiring specific improvements in accordance with predefined criteria. For such tasks, three scenarios are possible: increasing or decreasing the values of selected indicators does not change an object's membership in its original pattern; Changes in the indicators result in the object transitioning to another existing pattern; changes in indicators create a new structure, forming a previously non-existent pattern.

In the third scenario, the concept of an "empty pattern" is needed to predict changes in the developmental trajectory of the object under study. This approach accounts for the emergence of new patterns initially containing no objects. The use of probabilistic estimates allows for an accurate assessment of the likelihood of an object transitioning to a new state, considering changes in its developmental trajectory. The concepts of "empty pattern" and "pattern complexity" are defined as follows:

Definition 1. An empty pattern is a theoretically possible empty subset formed as a result of pairwise comparisons of objects, conducted within the framework of a given pattern analysis method.

Definition 2. Pattern complexity is defined as a value equal to one plus the minimum number of patterns required to transform one pattern into another, taking into account the existence of empty patterns. This methodology is particularly relevant for pattern analysis methods based on pairwise comparison of indicators [Myachin 2016a; Myachin 2016b].

Thus, the introduction of the "empty pattern" concept and its role in prediction, along with the definition of pattern complexity, contributes to a more precise description and modeling of processes occurring in complex dynamic systems. These concepts provide a methodological foundation for conducting in-depth analyses of the structure and developmental dynamics of objects belonging to different patterns.

Acknowledgements. This work has been supported by the grants the Russian Science Foundation, RSF No. 24-61-00030, https://rscf.ru/project/24-61-00030/

References

[Aleskerov 2014] Aleskerov F. et al. A Method of Static and Dynamic Pattern Analysis of Innovative Development of Russian Regions in the Long Run //Models, Algorithms and Technologies for Network Analysis: From the Third International Conference on Network Analysis. – Springer International Publishing, 2014. – C. 1-8.

[Aleskerov et al 2022] Aleskerov F. et al. Short-Term Covid-19 Incidence Prediction in Countries Using Clustering and Regression Analysis //International Conference on Computers Communications and Control. – Cham : Springer International Publishing, 2022. – C. 333-342.

[Myachin 2019] Myachin A. L. Pattern analysis in parallel coordinates based on pairwise comparison of parameters //Automation and Remote Control. – 2019. – T. 80. – C. 112-123.

[Dmitrienko, Myachin 2023] Dmitrienko A. S., Myachin A. L. Tuition Pricing in Russian Universities: How Competitor Universities Affect Tuition Costs? University Management: Practice and Analysis. 2023, Vol. 27, No. 2, pp. 75–88 (in Russian)

[Myachin 2016a] Myachin A. L. Pattern Analysis: Ordinal-Invariant Pattern Clustering// Management of Large Systems: Proceedings Collection. 2016, No. 61, pp. 41–59 (in Russian)

[Myachin 2016b] Myachin A. L. Pattern Analysis: Diffusion-Invariant Pattern Clustering. Control Science. 2016, No. 4, pp. 2–9 (in Russian)