Dr hab. inż. Karol Opara, prof. IBS PAN
Instytut Badań Systemowych PAN
ul. Newelska 6
01-447 Warszawa

Warsaw, 26th September 2025

Report for the PhD thesis “Multiple Representation-Based Ensembles for Time Series Classification” by Rauzan Sumara

Thesis scope and structure

Machine learning for time series often utilises representations such as global statistics, spectral features, shapelets, or embeddings. These methods capture different aspects of the input data, creating a potential for ensemble-based algorithms to outperform approaches based on a single representation.

The PhD thesis by Rauzan Sumara addresses time-series classification, which is a timely research topic of practical significance. The author introduces a method called Multiple Representation-Based Ensemble (MuRBE), which aggregates the outcomes of classifiers operating in four domains: features, dictionaries, intervals, and shapelets.

The thesis is well-written and follows the typical structure consisting of introduction, literature review, methods, results, and conclusions. The PhD candidate uses consistent and clear notations for mathematical formulae and pseudocodes and provides well-thought-out figures to illustrate the algorithmic ideas.

Detailed evaluation

The initial two chapters define and justify the research problem, while the literature review offers a comprehensive and current analysis of time-series classification. What I missed was a broader discussion of neural networks, which are becoming the dominant tool in many current applications. The thesis devotes just one paragraph to this group of methods, detailing only a single algorithm. The internal representations learned by other neural architectures, such as recurrent neural networks or transformers, could potentially focus on different aspects of information about the time-series properties and hence become a valuable addition to the MuRBE structure. Expanding the range of deep learning time series classifiers would also enhance the benchmarking comparison.

The PhD candidate has knowledge and skills related to convolutional neural networks, as he applies them in one of the classification frameworks introduced in the thesis, called

SAX-GCNN. That method preprocesses the temporal data by cutting it into intervals based on the percentiles of the marginal distribution. Each of these intervals is assigned a symbol. Consecutive symbols form words, and the resulting text is vectorised and processed through a stacked deep learning architecture presented in Figure 29. Such an approach is correct, but leaves me wondering if discretising the real-valued series and vectorising it back is better than directly using z-scores as input to the neural network.

Another classifier introduced in the thesis utilises a novel representation estimated through an autoregressive fractionally integrated moving average (ARFIMA). In contrast to classical ARMA processes, the model is capable of capturing long-range dependences. It also generates relatively small-dimensional representations of time series.

The two classifiers proposed by the PhD candidate, as well as two literature methods, provide the probabilities of each class as outcomes. The MuRBE framework first transforms them non-linearly, then computes the weighted average of the resulting scores and uses it to select the class. The non-linear transformation improves the predictive power against both the non-weighted case and other output aggregation methods, such as majority vote or proportional vote. Although I appreciate the quality of the final solution, I find its presentation confusing. First, despite the aggregation method operating on crisp values only and not utilising fuzzy arithmetic it was called "fuzzy rank-based". Second, Figure 25 suggests that probability estimates range from −1 to 1, even though the feasible values are between 0 to 1, which is also the case for all four base classifiers. Third, I do not see in what way "the fuzzy rank-based ensemble provides more flexible, dynamic and adaptive weights for individual models", given that the non-linear transformations proposed have no parameters and do not depend on any properties of the input. Finally, equation (5) should use the "arg min" operator rather than "min".

Chapter 5 describes plentiful, sound and appropriately statistically processed experiments. Benchmarking utilises 40 out of 112 datasets available in the UCR/UEA time series classification archive. However, a more comprehensive discussion of the inclusion criteria is necessary. We only learn about stratification by the groups of datasets in the repository and the omission of the least numerous groups. Still, the description suggests that the selection of individual classification problems might have been arbitrary rather than randomised, which could introduce bias. To bolster confidence in the reliability of the presented results, a more detailed discussion of the inclusion criteria would be beneficial, preferably illustrated with a flow diagram. An even more compelling approach would be to use as many datasets from the repository as feasible to analyse performance across the complete external benchmark rather than its subset.

The experiments run on the 40-dataset benchmark show very competitive performance of MuRBE, exceeding all of the analysed non-ensemble-based classifiers and being on par with the top ensemble-based methods. The variety of domains and characteristics of problems selected for benchmarking highlights the generalisation ability of MuRBE and its potential for practical applications.

Conclusion

The thesis offers a novel and effective solution to the time-series classification problem. The PhD candidate has exhibited a solid theoretical understanding of Information and Communication Technologies. Additionally, the thesis highlights the author's practical expertise in software engineering and his ability to conduct independent research. My overall assessment of the thesis is highly positive, and I recommend proceeding to further steps required to award the PhD degree to Mr Rauzan Sumara.

PODPIS ZAUFANY
KAROL ROMUALD
OPARA
26.09.2025 15:25:58 GMT+0200
Dokument podpisany elektronicznie podpisem zaufanym

Karol Opara