Time-series classification via ROCKET
The project

Anastasiia Prutianova [* 1] **Aziz Temirkhanov** [* 1 2] **Georgy Mkrcthyan** [* 1] **Nikita Bogdanov** [* 1]

Abstract

Modern time series classification approaches relies on various kinds of methods, including deep learning models. Vast majority of those algorithms are highly computationally complex. In this work, we apply random convolutions via ROCKET model and compare this approach, which claims to have a fraction of computational expense of other methods, with such acknowledged models as LSTM and XGBoost, and a moving averages as a baseline model. We study not only the differences in terms of accuracy but also a 'heaviness' of the models in terms of time it takes to fit them.

1. Introduction

Time-serious forecasting is a field of data science having lots of real-life applications like weather forecasting, financial marker analyses and trading, healthcare, climate change, and environmental studies, social studies, and many others.

In the 20th century, theoretical analyses of sequential data were done by Peter Whittle, introducing the ARMA model. ARMA and ARIMA are based on natural assumptions about the data like continuity and linearity to create plausible mathematical models. These simple yet effective models are still applied in many practical cases.

Natural language processing (NLP) was rapidly developing in recent years due to the emergence of neural networks. Different types of recurrent neural networks (RNN), such as Long short-term memory (LSTM) models and transformers that are applied in NLP, can be naturally generalized for the time-serious forecasting field due to the language "sequential" nature. Though RNNs became the state-of-the-art approach, they imply large volumes of training data and significant computational resources, making it difficult to deploy in real-life applications.

Recently ROCKET approach was proposed, using random convolutional kernels and a linear classifier to solve a serious classification problem. Its main advantage is training speed and simplicity. In our project, we aim to apply ROCKET in time series forecasting. We plan to compare it with state-of-the-art approaches: Moving Average model, boosting, and LSTM in terms of both quality and computational complexity.

2. Literature review

Generally, time-series classification problem is one of the most popular and usable problems in the field. Therefore, numerous different methods were developed as a solution to the issue.

To begin with, Gradient boosting is a very powerful machine learning ensemble method, which is used in classification and regression problems. XGBoost is specific implementation of gradient boosting, that uses advanced regularization which improves model generalization and reduces overfitting.

In recent work XGBoost was used for forecasting time series data: in (Wang & Guo, 2020) XGBoost was used for forecasting stock market volatility. XGBoost was used along with ARIMA model on set of datasets. And it was shown that results of XGBoost exceed results of ARIMA. On the other hand, research, that was using the other set of datasets showed, that ARIMA has a better predictive property, than XGBoost. Hence, we may say, that XGBoost may compete with ARIMA model depending on a provided dataset.

The long, short – term memory neural network is specific kind of recurrent neural network, designed to solve the vanishing gradient problem, by preserving the error that can be backpropagated through time and layers. Moreover, LSTMs overcome the weakness of learning long-range dependencies, a trait that is essential in time-series forecasting. The main difference of LSTM over others neural network is set of blocks, recurrently connected. Each block contains one or multiple recurrently connected memory cells and three multiplicative units – the input, output and forget gates.

[*]Equal contribution [1]Skoltech, Russia, Moscow [2]NRU HSE, Russia, Moscow. Correspondence to: Aziz Temirkhanov <atemirkhanov@hse.ru>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

Because of possibility of remembering information LSTM shows excellent predictive results. In (Paliari I., 2021), comparing LSTM to XGBoost and ARIMA it shows the best results overall on the set of datasets. The other work (Siami-Namini S., 2019) shows big improvement of LSTM over ARIMA model. Hence, we may say, that LSTM would have better performance on a provided dataset over XGBoost and ARIMA.

The best available at the current moment models are state-of-the-art approaches which are divided in several subgroups. First subgroup is focused on either frequency second is based on the variance and the last one takes into account presence of discriminative subseries. COTE or HIVE-COTE developed by the Bagnall A (Bagnall A, 2017), represents the first group of the models. BOSS, BOSS-VS and WEASEL are members of the second group. All of them are similar in time complexity having in general quadratic dependence both in signal length and number of observations, however, the worst performance is shown by the BOSS-VS but with the best scalability, on the other hand WEASEL is less scalable but has the best accuracy among the group. Shapelet Transform which represents the last group of models. Unlike first two groups Shapelet Transform has dependence of the 4th order on the length of the signal which makes it even less scalable than its competitors.

Despite the fact that state-of-the-art approaches shows very high performance their scalability on the bigger data is a major problem in terms of computational resources, which may restrict the availability of their usage.

As a supplement to above stated models Lucas B (Lucas B, 2019), and Shifaz A. (Shifaz A, 2019) suggested two more scalable methods: Proximity Forest and TS-CHIEF, which has lower quality of performance, however , has an opportunity to be used on bigger datasets as long as both of them has quasilinear in the number of training examples, but quadratic in time series length.

Another milestone in time-series classification problem is a group of convolutional networks which are well-known for high-quality performance on the image classification problem, and they turned out to be useful on the signal classification issue. Therefore, Ismail Fawaz (Ismail F, 2019) in his work showed how the ensemble of Inception like network having linear complexity in number of observations can compete with state-of-the-art models.

3. Plan

In order to proceed with those model and obtain an utter project reports, we come up with following plan.
First, let us state the list of methods we explore in this paper.

1. Moving averages - Prutianova Anastasiia
2. ROCKET - Mkrcthyan Georgy
3. XGBoost - Nikita Bogdanov
4. LSTM - Aziz Temirkhanov

Next, for every model, the next steps are applicable:

1. Model Implementation
    - Implement the method
    - Build a pipeline which processes the data in a right way for a certain approach
2. Experiments and evaluation
    - Design an experiments
    - Run the experiments, obtain graphs showing model's performance and evaluate the results
3. Report preparation
    (a) Technical details
        - There goes method mathematical and implementation details
        - In this section, we describe why and how the model works
    (b) Experiment details
        - In this section we discuss about the experiments design and its details
        - Also, we argue about necessity of such environment
    (c) Experiments results and conclusion
        - In this section we show the results of model evaluation along with graphs we obtained
        - Also, there are overall conclusion section where we compare all results and conclude our entire work

4. Moving average

4.1. Method details

4.2. Experiments and evaluation

4.3. Conclusion

5. ROCKET

5.1. Method details

5.2. Experiments and evaluation

5.3. Conclusion

6. XGBoost

6.1. Method details

6.2. Experiments and evaluation

6.3. Conclusion

7. LSTM

7.1. Method details

7.2. Experiments and evaluation

7.3. Conclusion

8. Conclusion

To sum up,...

References

Bagnall A, Lines J, B. A. L. J. K. E. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. In *Data Mining and Knowledge Discovery*, volume 31, pp. 606—-660. 2017.

Ismail F, Lucas B, F. G. P. C. S. D. W. J. W. G. I. L. M. P. P. F. Inceptiontime: Finding alexnet for time series classification. *arXiv preprint arXiv:190904939*, 2019.

Lucas B, Shifaz A, P. C. O. L. Z. N. G. B. P. F. W. G. Proximity forest: an effective and scalable distance-based classifier for time series. In *Data Mining and Knowledge Discovery*, volume 33, pp. 607—-635. 2019.

Paliari I., Karanikola A., K. S. A comparison of the optimized lstm, xgboost and arima in time series forecasting. In *2021 12th International Conference on Information, Intelligence, Systems Applications (IISA).*, pp. 1–7. IEEE, 2021.

Shifaz A, Pelletier C, P. F. W. G. Ts-chief: A scalable and accurate forest algorithm for time series classification. *arXiv preprint arXiv:190610329*, 2019.

Siami-Namini S., Tavakoli N., N. A. S. The performance of lstm and bilstm in forecasting time series. In *IEEE International Conference on Big Data (Big Data).*, pp. 3285–3292. IEEE, 2019.

Wang and Guo. Forecasting method of stock market volatility in time series data based on mixed model of arima and xgboost. In *China Communications*, volume 3, chapter 17, pp. 205–221. 2020.