UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Department of Informatics

University of Fribourg
Department of Informatics
Bd. de Pérolles 90
CH-1700 Fribourg
Phone. +41 26 300 84 65
Fax +41 26 300 97 26
http://exascale.info/

University of Fribourg, Dept. of Informatics, CH-1700 Fribourg

Jona Fontana

Fribourg
Switzeland

Dr. Mourad Khayati
Senior researcher
Phone +41 26 300 84 60
Fax +41 26 300 97 26
mkhayati@exascale.info

Fribourg, October 21, 2020

Comparison of Synthetic Time Series Data Generation Techniques

Work overview:

The rapid explosion of time series data in the recent years has brought a thrive of new time series analysis methods. Paradoxically, in the industry domains applications, large datasets are often highly homogeneous, restricting their use in fitting general time series analysis methods. This raises the need for a more careful evaluation of these methods. This can be done by simulating a more diverse set of time series data to enable reliable comparisons.

The basic idea behind data generation methods is to create synthetic examples in order to augment or create a dataset and help learn a model. While a lot of work has been done on data generation algorithms, less attention has been paid to finding better data generation methods, specifically for time series data. Some challenges arising from data generation methods include:

- **Data features.** While preserving the data features is crucial when generating data, the features of time series data are not fully utilized in current data generation methods.
- **Application task.** The data generation methods are task dependent. For example, the data generation methods applicable for time series classification may not be valid for time series anomaly detection.
- **valuation metric.** How to select the right metrics to evaluate the generation remains a challenge.

The objective of this thesis is to provide a better understanding of the different generation techniques approaches. This can be done by studying the association between each generation method and its intrinsic data properties (abrupt, seasonal, trendy, etc.), its task application (classification, anomaly detection, forecasting, etc.), and its evaluation metrics (RMSE, NMI, AutoCorr, etc.). Such a study could be significantly advantageous for it would help obtain generated data with optimal quality. The result will be a recommendation method that delegates the right generation method for a given dataset and application task, as well as the relevant set of metrics for its evaluation.

Tasks:

1. Get familiar with time series generation techniques [1, 2, 3, 4, 5, 6, 7] and create a taxonomy of existing techniques.
2. Analyze time series data and identify their main features.
3. Evaluate generation methods on different time series tasks such as classification, anomaly detection, and forecasting.
4. Perform an empirical evaluation that studies:
   - **Features.** Identify the features that support a given generation method.
   - **Application.** Identify the application tasks that a given generation method is fit for.
   - **Data.** Identify the types of data that a given generation method performs well on.
   - **Metrics.** Identify the evaluation metrics that are better fit for a given generation approach.
5. Propose a recommendation method of time series data generation for a given feature or use case.

References:

1. Kang, Yanfei, Rob J. Hyndman, and Feng Li. ”GRATIS: GeneRAting TIme Series with diverse and controllable characteristics.” Statistical Analysis and Data Mining: The ASA Data Science Journal (2020).
2. Esteban, Cristóbal, Stephanie L. Hyland, and Gunnar Rätsch. ”Real-valued (medical) time series generation with recurrent conditional gans.” arXiv preprint arXiv:1706.02633 (2017).
3. Yoon, Jinsung, Daniel Jarrett, and Mihaela van der Schaar. ”Time-series generative adversarial networks.” Advances in Neural Information Processing Systems. 2019.
4. Chen, Xi, et al. ”Infogan: Interpretable representation learning by information maximizing generative adversarial nets.” Advances in neural information processing systems. 2016.
5. Salem, Milad, Shayan Taheri, and Jiann Shiun Yuan. ”Anomaly Generation Using Generative Adversarial Networks in Host-Based Intrusion Detection.” 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, 2018.
6. Forestier, Germain, et al. ”Generating synthetic time series to augment sparse datasets.” 2017 IEEE international conference on data mining (ICDM). IEEE, 2017.
7. Laptev, Nikolay. ”AnoGen: Deep Anomaly Generator.” Outlier Detection De-constructed (ODD) Workshop. Available online: https://research. fb. com/publications/anogen-deep-anomaly-generator/(accessed on 20 August 2018). 2018.
8. Wen, Qingsong, et al. ”Time Series Data Augmentation for Deep Learning: A Survey.” arXiv preprint arXiv:2002.12478 (2020).

Starting date of thesis: 01.11.2020

Ending date of thesis: TBD

University of Fribourg
Department of Informatics
eXascale Infolab

Dr. Mourad Khayati, Abdelouahab Khelifati