Karlsruher Institut für Technologie

Institut für Programmstrukturen und Datenorganisation (IPD)
Lehrstuhl für Systeme der Informationsverwaltung, Prof. Böhm

Masterarbeit

Masterarbeit

Detecting Outlying Time-Series with Global Alignment Kernels / Ausreisererkennung von Zeitreihen mit Global Alignment Kerneln

Die Bearbeitung dieser Masterarbeit kann auf Deutsch oder Englisch erfolgen.

Detecting outliers in data is important in many applications, e.g., fraud detection, predictive maintenance, or finance. In such real-world settings, new observations are collected over time, i.e., the data are time-series. One is not only interested in finding single outlying observations ("point anomalies"), but rather outlying sets of consecutive observations, i.e., abnormal "segments", of the time-series. Detecting such outlying segments is challenging.

Support Vector Data Description (SVDD) [1, 2] is a state-of-the-art approach to detect outliers. SVDD fits a hypersphere around most observations, while minimizing its volume. Points that fall out of the sphere are declared as outliers [3]. However, SVDD does not work well with time-series, as the distances between segments are not easy to estimate with standard measures. Thus, one usually first maps the problem to a static one, by extracting domain-specific attributes for each segment [4]. SVDD then operates on this attribute space instead, where distances are easier to compute. This indirect way requires expert knowledge, which makes the application of SVDD impossible in many use cases and practical situations.

To cope with this issue, one could think about leveraging the so-called "Global Alignment Kernels" [5], to compute the distances between time-series segments via Dynamic Time Warping directly within SVDD. By doing so, one could bypass the attribute-extraction step and make the application of SVDD independent of the underlying domain.

The focus of this thesis is the design of a new outlier detection method combining SVDD and global alignment kernels. This results in the following tasks:

- Conduct an extensive literature review on global alignment kernels and SVDD.
- Implement a new approach to detect outliers, combining SVDD and global alignment kernels.
- Evaluate your approach against the current state-of-the-art methods and baselines.
- Explore, via a real-world use case, the differences between the outliers detected with your method and those obtained via the traditional application of SVDD as in previous work.

By working on these tasks the student will gain extensive knowledge about state-of-the-art outlier detection methods and the peculiarities of time-series data. The student will learn how to approach challenging research questions and learn how to perform systematic and reproducible experiments.

[1] D. M. Tax and R. P. Duin. "Support Vector Data Description". en. In: *Machine Learning* 54.1 (Jan. 2004), pp. 45–66.
[2] W.-C. Chang et al. "A revisit to support vector data description". In: *Dept. Comput. Sci., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep* (2013).
[3] B. Liu et al. "SVDD-based outlier detection on uncertain data". en. In: *Knowledge and Information Systems* 34.3 (Mar. 2013), pp. 597–618.
[4] M. Vollmer et al. "Energy Time-Series Features for Emerging Applications on the Basis of Human-Readable Machine Descriptions". en. In: *Proceedings of the Tenth ACM International Conference on Future Energy Systems*. Phoenix AZ USA: ACM, June 2019, pp. 474–481.
[5] M. Cuturi. "Fast global alignment kernels". In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. ICML'11. Madison, WI, USA: Omnipress, June 2011, pp. 929–936.

Ansprechpartner

Florian Kalinke, M.Sc.	florian.kalinke@kit.edu	Raum: 352
Am Fasanengarten 5	76131 Karlsruhe	Gebäude: 50.34