

Detection of Outliers in Time Series by Analyzing Cluster Transitions

Gas prices, market development, pandemic cases - time series data is everywhere. An important task in the analysis of this data is the detection of anomalies. Most methods do that by analyzing individual time series. A major issue with such solutions is that general factors, such as the development of the market, can mask anomalies related to individual events. In our work, we cluster time series at each point in time and analyze their transitions between different clusters. Since this solution requires a clustering with reliable and meaningful results, we make use of the Cluster Over-Time Stability Evaluation (CLOSE).

Method

The procedure for detecting anomalies can be described with the following steps illustrated in Figure 1:

1. Set a cluster transition threshold σ
2. Cluster time series based on CLOSE
3. Extract all cluster transitions
 - a. Weight them by the number of their occurrence
 - b. If a cluster transition has a weight smaller than threshold, add it to the set of outlier transitions
4. For each time series: if it contains a cluster transition from the outlier set, mark the corresponding subsequence as anomalous

Results

The method was applied to multiple real-world and generated data sets. One of the real-world datasets used for the evaluation is the EIKON data set, which contains data of publicly listed companies. The result of the detection of anomalous subsequences for it is shown in Figure 2. Most interesting about the identification of both, KR as well as UPS as outliers is the fact that they can be explained by related events. In case of UPS, it was the plane crash that prompted the company to reduce its expected return. Kroger (KR) is the largest grocery supermarket chain and the third largest retailer in the USA. Despite rising prices for food and energy, KR was also able to keep prices low in 2011-2013, which led to an additional increase in purchasing power and thus had a positive impact on net sales.

Current work in progress

Based on CLOSE, we are currently working on the following projects:

- Extraction of representative time series
- Imputation of missing data
- Data generation
- Maximum stable clustering of time series in real-time

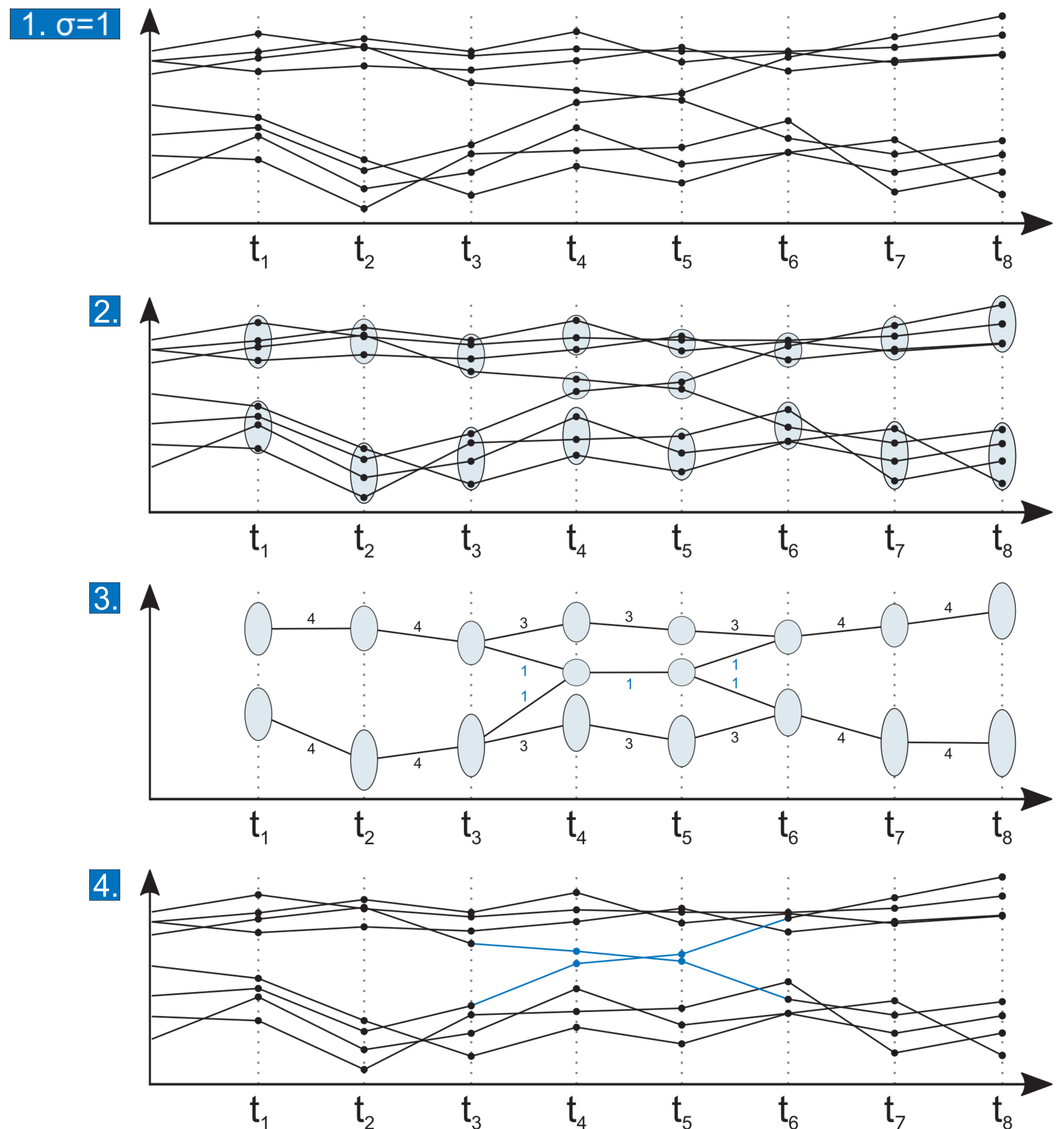


Figure 1: Stepwise procedure of the outlier detection method

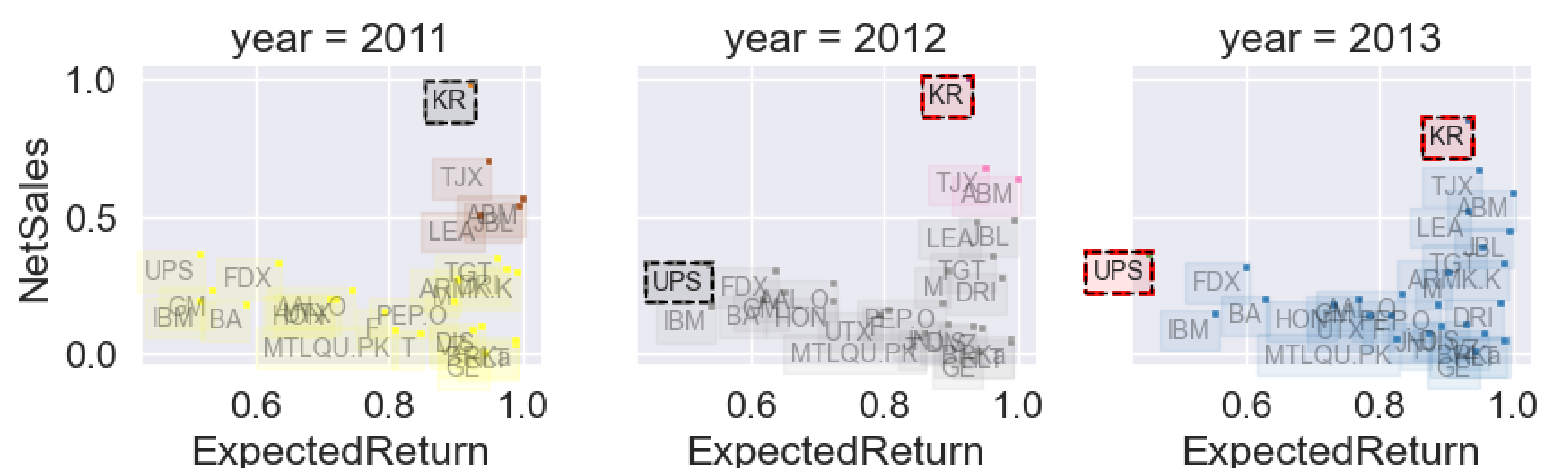


Figure 2: Outlier detection result for the EIKON data set