Visual Exploration of Local Interest Points in Sets of Time Series

Tobias Schreck[*]
University of Konstanz
Germany

Lyubka Sharalieva[†]
University of Konstanz
Germany

Franz Wanner[‡]
University of Konstanz
Germany

Jürgen Bernard[§]
Fraunhofer IGD
Darmstadt, Germany

Tobias Ruppert[¶]
Fraunhofer IGD
Darmstadt, Germany

Tatiana von Landesberger[‖]
TU Darmstadt
Germany

Benjamin Bustos[**]
Universidad de Chile
Santiago de Chile

Abstract

Visual analysis of time series data is an important, yet challenging task with many application examples in fields such as financial or news stream data analysis. Many visual time series analysis approaches consider a global perspective on the time series. Fewer approaches consider visual analysis of local patterns in time series, and often rely on interactive specification of the local area of interest. We present initial results of an approach that is based on automatic detection of local interest points. We follow an overview-first approach to find useful parameters for the interest point detection, and details-on-demand to relate the found patterns. We present initial results and detail possible extensions of the approach.

1 Introduction

Time series are a ubiquitous, important data type, and visual analysis of this data remains an important research topic [2]. Many approaches consider time series at a *global* level, where all time stamps of a given time series are visualized at once. However, one may also be interested in local properties of time series at different scales. An example are abnormal or frequently occurring outlying patterns (or motifs) in a time series. For large and many time series, it is not efficient to inspect all possible local patterns individually, but automatic support is required. Searching for time series which are locally [5] or globally [4] similar to user requirements is possible, however requires user specification of patterns.

Automatic methods can detect possibly interesting local patterns and guide users in identifying interesting data portions. A multitude of detectors based on generic [1] or domain-specific [6, 7] time series analysis exist. However, these often require setting of parameters which may not be easy to find in an explorative analysis scenario. We here propose an approach for detector-based time series analysis, which makes explicit the effect of up to two detector parameters in the analysis.

2 Approach

Our goal is to combine automatic time series interest point detection with interactive-visual exploration. Our analysis system should support (1) interactive identification of useful parameters and (2) appropriate visualization of the interest points for relation and comparison. We implemented a prototype based on the following components:

- The Control-chart (or Bollinger bands) method [1] is applied to detect local points of interest. It detects an interest point whenever a given time series value exceeds a band $\pm b\sigma$ of a multiple b of the time series standard deviation σ, calculated based on a moving average of length l (cf. Figure 1). Detected interest points are indicated by vertical bar markers. The Control-chart approach detects outlyingness, and parameters b and l influence the sensitivity of the detection.

- For a set of time series, we show a small multiple view of Control-chart graphs (cf. Figure 2(a)). This view allows to compare a set of time series for similarities regarding the interest point patterns at a given parameter pair (b,l). The parameters can be set interactively, updating the display in real time. The small multiple view can be sorted, based on selecting a given time series and arranging the remaining ones heuristically according to the spatial proximity of interest points occurrences.

- The distribution of detected interest points can be compared for different scales, that is, different values of the parameters (b,l). This is supported by an overview display. The parameters are sampled at regular intervals and a pixel-based view shows the resulting distributions in a 2D matrix view. Each occurrence of an interest point is shown by a set pixel. The overall display allows to easily compare different parameter settings and identify robust distribution patterns (cf. Figure 2(b)).

- Users can select single time series to sort the comparative views. In the pixel-based parameter space view, a mouse-over function allows inspection of contained time series in a focus-and-context fashion (cf. Figure 2(b) (bottom-right)). The time series segments responsible for the interest points with largest excess of the Control-chart can also be displayed for comparison and correlation (cf. Figure 3(b)).

3 Initial Results

We applied our approach to time series of daily return rates of German blue chip stocks as observed during 2005 and 2006 as an illustrative example. We sampled the Control-chart parameters ad-hoc at $b = \{0.5, 1.0, 1.5\}\sigma$ and $l = \{5, 10, 15, 20\}$ days. Figure 2(b) shows the resulting comparative view.

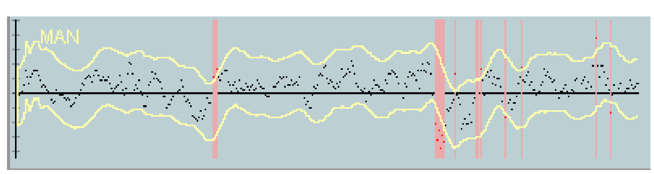
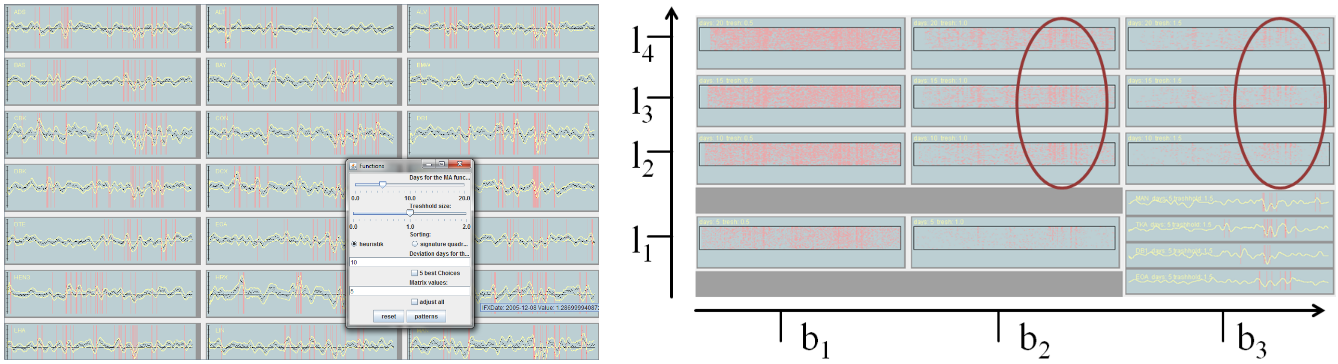


Figure 1: Original time series (black) with Control-chart (yellow) and detected interest points, based on Control-chart analysis (pink).

[*]e-mail:tobias.schreck@uni-konstanz.de
[†]e-mail: lyubka.sharalieva@uni-konstanz.de
[‡]e-mail:wanner@dbvis.inf.uni-konstanz.de
[§]e-mail:juergen.bernard@igd.fraunhofer.de
[¶]e-mail:tobias.ruppert@igd.fraunhofer.de
[‖]e-mail:tatiana.von_landesberger@gris.informatik.tu-darmstadt.de
[**]e-mail:bebustos@dcc.uchile.cl

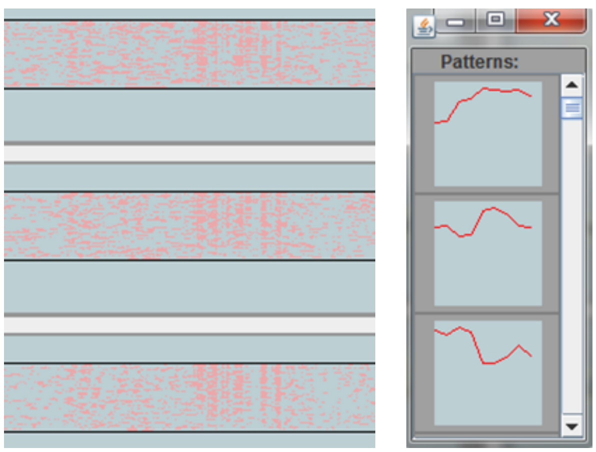
IEEE Conference on Visual Analytics Science and Technology 2012
October 14 - 19, Seattle, WA, USA
978-1-4673-4753-2/12/$31.00 ©2012 IEEE



(a) Small multiples of Control-chart views

(b) Pixel-based comparison of interest points at different scales

Figure 2: (a) Overview over time series with Control-chart lines and interest points (marked by pink bars). (b) Comparison of detection patterns (marked by pixels) across different parameter combinations (x-axis: detection threshold; y-axis: size of moving average).



(a) Interest point patterns

(b) Local patterns

Figure 3: (a) Zoom-in to a regular structure of detected interest points for many of the observed time series. (b) Detail-view of the most ... local time series patterns detected.

For the larger threshold values b and all considered moving average sizes, we observe regular structures of similar occurrences of many interest points (see circled areas). Figure 3(a) shows a zoom-in to a portion of the identified data area. This indicates that a significant amount of outliers exist in the set of time series which is robust over different parameter settings. A detailed inspection by mouse-over, illustrated in Figure 2(b) (bottom-right tile) indicates also, that the time series show similar curve shapes, thus indicating that a stock-market wide effect took place. Figure 3(b) furthermore shows a detail view of a number of time series patterns at this position, and can be used for comparing the actual curves that underly the detected set of interest points.

4 Next Steps

We identify the following steps to improve our approach:
(1) Additional and application-dependent interest point detectors should be considered. The Control-chart technique is focused on outlyingness. Detectors based e.g., on repetitive motifs, or similarity to user-selected or sketched local patterns could be useful.
(2) Our approach is based on local occurrences of defined patterns of similar size. In real applications, often more complex *event* patterns, possibly composed of many local patterns at different scales, lags, etc. are of interest. A framework for their formal description and detection has been proposed in [8] and could be a promising to support.
(3) We want to improve the analysis of local curve patterns at the detected interest points. Cluster analysis can reduce potentially many interest point curve shapes to a small number of prototypes. These can then be visually linked to the overview displays.
(4) Sorting time series according to similarity of occurring interest points is needed for correlation of many time series. Our current sorting relies on comparing spatial proximity of interest points. A more encompassing sorting criterion should also consider similarity of interest point curve shapes. The Signature-Quadratic Form Distance approach [3] can elegantly provide such a distance function and we want to test it.

5 Conclusions

We proposed to combine automatic interest point detection with visual representation and user interaction for identification and relation on local interest points in time series. We will build on this idea and explore novel methods for visual-interactive analysis of time series data based on local interest points in the future.

References

[1] I. Ada and M. Berthold. Unifying change – towards a framework for detecting the unexpected. In *IEEE Int. Conference on Data Mining Workshops*, pages 555–559. IEEE Computer Society, 2011.
[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer-Verlag New York Inc, 2011.
[3] C. Beecks, M. Uysal, and T. Seidl. Signature quadratic form distances for content-based similarity. In *ACM Multimedia*, pages 697–700, 2009.
[4] J. Bernard, J. Brase, D. Fellner, O. Koepler, J. Kohlhammer, T. Ruppert, T. Schreck, and I. Sens. A visual digital library approach for time-oriented scientific primary data. *Springer International Journal on Digital Libraries, ECDL 2010 Special Issue*, pages 1–13, 2011.
[5] H. Hochheiser and B. Shneiderman. Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.
[6] R. Kincaid. SignalLens: Focus+context applied to electronic time series. *IEEE Trans. on Visualization and Computer Graphics*, 16:900–907, 2010.
[7] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. Cleveland, S. Grannis, and D. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Trans. on Visualization and Computer Graphics*, 99:205–220, 2009.
[8] F. Moerchen and A. Ultsch. Efficient mining of understandable patterns from multivariate interval time series. In *Data Mining and Knowledge Discovery*, 2007.