

Survey: Visual Analysis Approaches to Time Series Prediction

Fabian Otto*

Technical University Darmstadt

ABSTRACT

1 INTRODUCTION

Predictive analytics is concerned with the prediction of future probabilities and trends based on observed events. It encompasses a multi-perspective approach that includes integrated reasoning, pattern recognition and predictive modeling associated with domain knowledge. However, the primary use of such systems tends to be reactive, meaning that analytic systems are typically In spatiotemporal data, analysts are searching for regions of space and time with unusually high incidences of events. In the cases that hotspots are found, analysts would like to predict how these regions may grow in order to plan decision support and preventative measures. Furthermore, analysts would also like to predict where future hotspots may occur.

2 TEMPORAL TIME SERIES

content

2.1 Simple Time Series

content

2.2 Time Series Prediction with Peak Preservation

content

2.3 Multivariate Time Series

Visual Analytics approaches for simple time series prediction (Sect. 2.1) offer a great variety for single time series. However, in a practical environment it is often necessary to evaluate multiple time series in parallel or deal with multidimensional input variables. In order to solve these problems, visual analytics approaches for simple time series are not sufficient anymore.

An older Visual Analytics approaches for multivariate time series prediction is from Ichikawa et al. [1]. Their goal was to predict multiple daytime stock prices and simultaneously visualize a set of prediction from different simulation systems within a single system. Therefore, their system supports multiple prediction for a single stock as well as predictions of multiple stocks. One major finding was, visualizing multivariate predictions in a 3-dimensional space creates large amounts of occlusion, thus it is not suitable. Instead the system utilizes line charts with cluttering control and color charts with level-of-detail control. The line charts support multivariate scaling, i.e. the time axis of the plot can be scaled differently to focus one certain areas, e.g. with less overlap. Further, the lines are colored differently, depending on the amount of overlap/uniform predictions. The color chart is composed of horizontal color-bands for each prediction, whereas a vertical color band can be seen as a period in time. In order to improve the visualization, the different color-bands are cluster based on the user-specified level-of detail. This results in a smaller amount of horizontal color-bands where the individual properties are diminished. In the workspace the

above plots can be displayed with different axis. This enables the user to compare a set of predictions for different parameter ranges (e.g. sales organizations) as well as different stocks simultaneously. Consequently, the user can get a overview or trend, but due to clustering and simply the amount of predictions displayed, he can hardly extract specific information.

2.4 Pattern and Anomaly Detection

Another system can be found from Steed et al. [5]. Their approach is focused on understanding patterns in log and imagery data collected by 3D printers.

3 SPATIOTEMPORAL TIME SERIES

[6]

The previous section (Sect. 2) presented an overview of systems for temporal time series prediction as well as pattern and anomaly detection. However, in other application areas such as crime prevention as well as emergency and epidemic intelligence, not only the temporal data is of value, but also locations. Typically, these locations fit into a hierarchical categorization structure, which can be filtered. Further, the data categories are processed either as aggregated time series over a spatial location (e.g. county, zip code, collection station) or represent a spatial snapshots of a small time aggregate (e.g., day, week).

One system which focuses on spatiotemporal prediction is from Maciejewski et al. [2]. It is based on their previous work [3, 4] and centers around on categorical geospatiotemporal event data, where events consist of locations in time and/or space, and each event fits into a hierarchical categorization structure. Specifically, they used a data set for detecting adverse health events using pre-diagnosis information from emergency departments. The system itself provides a colored geospatial window, which shows the percentage of events in a certain area, e.g. patients at an emergency department, which where classified with respiratory syndromes. Further, the user is able to apply filtering on a fine and coarse-grained level. The systems also differentiates between the time series and the geospatial prediction. The time series prediction is achieved by cumulative summation or a seasonal-trend decomposition based on locally weighted regression. For multivariate data, each event category is modeled as a separate time series signal. Equivalent to the time series prediction, the granularity/level of aggregation (e.g. state, county, etc.) for geospatial predictions can be adjusted by the user. Further, the system provides a Kernel Density Estimation approach which allows to display the spatiotemporal distribution on a more fine-grained level. In order to detect anomalies, e.g. outbreaks, (also see page 1) the system calculates the difference between the predicted and the actual values and highlights areas above a user specified threshold.

4 CONCLUSION

content

REFERENCES

- [1] Y. Ichikawa, T. Tsunawaki, I. Fujishiro, and H. Yoon. A visualization environment for multiple daytime stock price predictions. In *Proceedings of the 2nd VIIP International Conferences on Visualization, Imaging and Image Processing, Malaga, Spain, 2002*.

*e-mail: fabian.otto@stud.tu-darmstadt.de

- [2] R. Maciejewski, R. Hafen, S. Rudolph, S. G. Larew, M. A. Mitchell, W. S. Cleveland, and D. S. Ebert. Forecasting hotspots: A predictive analytics approach. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):440–453, April 2011. doi: 10.1109/TVCG.2010.82
- [3] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, M. Wade, and D. S. Ebert. Understanding syndromic hotspots - a visual analytics approach. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 35–42, Oct 2008. doi: 10.1109/VAST.2008.4677354
- [4] R. Maciejewski, B. Tyner, Y. Jang, C. Zheng, R. V. Nehme, D. S. Ebert, W. S. Cleveland, M. Ouzzani, S. J. Grannis, and L. T. Glickman. Lahva: Linked animal-human health visual analytics. In *2007 IEEE Symposium on Visual Analytics Science and Technology*, pp. 27–34, Oct 2007. doi: 10.1109/VAST.2007.4388993
- [5] C. A. Steed, W. Halsey, R. Dehoff, S. L. Yoder, V. Paquit, and S. Powers. Falcon: Visual analysis of large, irregularly sampled, and multivariate time series data in additive manufacturing. *Computers & Graphics*, 63:50–64, 2017.
- [6] J. Yue, A. Raja, D. Liu, X. Wang, and W. Ribarsky. A blackboard-based approach towards predictive analytics. In *AAAI Spring Symposium: Technosocial Predictive Analytics*, p. 154, 2009.