EuroVis Workshop on Visual Analytics (2016)
N. Andrienko and M. Sedlmair (Editors)

Visual-Interactive Exploration of Relations Between Time-Oriented Data and Multivariate Data

Jürgen Bernard[1,2], David Sessler [2], Martin Steiger[2], Martin Spott[3,4], and Jörn Kohlhammer[1,2]

[1]TU Darmstadt, Germany, [2]Fraunhofer IGD, Darmstadt, Germany, [3]HTW Berlin, Germany, [4]BT Research, Ipswich, United Kingdom

Abstract

The analysis of large, multivariate data sets is challenging, especially when some of these data objects are time-oriented. Exploring relationships between multivariate and temporal information, e.g., to identify patterns that support decision making is an important industrial analysis task. The target group of this design study are data analysts aiming at detecting fault patterns in a telecommunications network in order to spend maintenance budget more effectively. We present a visual analytics tool that provides overviews of multivariate data sets and associated time series. Users can select data subsets of interest in both attribute data and clustered time series data. Linked views consequently support the identification of relations between the two spaces. To ensure usefulness, the tool was designed in an iterative way, based on a careful characterization of the data, users, and tasks. A usage scenario demonstrates the applicability of the approach.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—User-centered design

1. Introduction

Large organizations continuously collect huge amounts of data about their operations. To spot opportunities and problems early, data needs to be explored for emerging patterns on an ongoing basis. As an example, telecommunication providers routinely record the state of their network through testing, fault reports by their customers and engineers as well as inventory information. Aiming for a continuous improvement of the telecommunications service, existing or emerging systemic problems need to be identified and eradicated as early as possible. Naturally, this becomes harder when all the obvious problems have been detected and solved.

The discovery of unexpected fault patterns is a complex monitoring task. In our application, the number of different *network and fault configurations* is at least in the order of 10^6. Combined with the problem of identifying *fault patterns that change over time* across such configurations, the task becomes insurmountable for human analysts. In practice, analysts can only explore a tiny fraction of this search space. Consequently they run the risk of missing latent or emerging problems. Similarly, they often do not succeed in finding reasonable explanations for existing problems. The high-level goal of the analysts is to reduce the number of faults in the network in a more effective and efficient way. To achieve this, they are interested in temporal fault patterns that explain particular problems described by configurations of network attributes. In turn, analysts try to identify configurations of network attributes that explain particular temporal fault patterns. They need visual support that eases the access to these complex types of data and seamlessly combines both types of relation-seeking problems in a visual-interactive approach. Currently, the most challenging problems hampering the analysts in their work environment are the extraction of features from the time-oriented data, the combination of heterogeneous data types, as well as establishing meaningful visualization and interaction designs.

In this work, we present a visual analytics solution for the exploration of relations between time series data and multivariate data. In our tool, coordinated views provide overviews of both multivariate data attributes and time-oriented data. The multivariate attributes are visualized with bar charts. Time-oriented data is displayed in a visual cluster analysis approach. Analysts can select interesting subsets in either the attribute space or the time series space. The selections are propagated to the other views, allowing the identification of interesting relations. The tool is the result of a design study conducted by visual analytics researchers and domain experts from the telecommunication provider. We report on a domain characterization phase, the results of an iterative design phase, and present a usage scenario to demonstrate the applicability of the tool. It is important to note that though the data has been derived from a real network, it has been anonymized into a manufacturing example.

DOI: 10.2312/eurova.20161124

Figure 1: *Our tool for seeking relations between multivariate attributes and time series data. We depict the attribute space on the left. Two visual cluster analysis results provide an overview of prominent time series patterns (upper right). For any given data subset, two line charts at the lower right show the development of two important parameters over time. In the example a user has selected a time series pattern (blue outline). The tool highlights interesting attribute values with blue bars.*

2. Related Work

Seeking Relations between Different Data Types To support the reduction of faults in their network, we need to enable analysts to identify error-prone network configurations. What makes a configuration interesting is the relationship between its defining multivariate attribute values and its time-dependent occurrence. With the identification of relationships, the primary task addressed is referred to as *relation seeking* [Shn96]. A variety of approaches for the analysis of heterogeneous and multi-modal data exist, some of them supporting visual comparison [GAW*11] and relation seeking [KH13]. However, visual-interactive interfaces revealing relationships between time series data and multivariate data are comparatively scarce. The particular challenge of this approach is to support the identification of relationships between these two data types in both directions, meaning that users are able to interactively define and adapt the 'target variable'. In the following, we review relation seeking approaches either supporting the mapping of time series data or of multivariate data.

Relating Time Series to Other Data Types We refer to the book of Aigner et al. for a general overview of visualization techniques for time-oriented data [AMST11]. Specifically, we focus on approaches at the synoptic level with an affinity to relation seeking [AA06]. Various visualization techniques support users in gaining an overview of thousands of temporal patterns, such as subsequence trees [LKL*04] or visual clustering approaches [SBVLK09]. Our aggregation approach is similar to the latter, but additionally provides brushing and linking interaction to relate interesting time series clusters with attribute visualizations. Various approaches combine time series aggregation with the visualization of relationships to non-temporal data. The cluster visualization approach by van Wijk and van Selow [VWVS99] relates time series clusters to a calendar view. Steiger et al. extended the approach by additionally relating time series clusters with geographical layouts [SBM*14]. The analysis of relations between time series and geographic information is a broad application field in general [AAD*10]. In several information visualization applications time series patterns are related to multivariate data, e.g., in geology [DKG15], finance [ZJGK10], or by example of industrial company analysis [GCML06]. In our previous works we have related time series clusters to additional metadata [BRS*12b] and metadata to time series clusters [BRS*12a], both strategies will be combined in this work. A special characteristic of our analysts' data is the existence of two equally important measures depending on time. One class of approaches for bivariate or multivariate time series data incorporates additional features [NAW13] [SBS11], while other approaches use projection techniques to represent multiple temporal dimensions in the visual space [BWK*13] [WVZ*15]. Our strategy differs in that we show two SOM-based data visualizations, one for each temporal dimension.

Identification of Relations in Multivariate Data A variety of statistical approaches exist that support the identification of interesting relations in multivariate data. We neglect approaches where relations are assessed at an *attribute-level*, because our analysts require a solution where relations are identified at the level of attribute values or *bins*, meaning that 'only' a subset of the data set shares particular relationships. A subset can be pre-defined, be the product of an interactive selection, or match an observation of some attribute. One class of approaches working at the bin-level makes use of contingency tables, where the frequency distribution of these observations is stored and displayed in a matrix metaphor (see, e.g., [Fri99] for an early multi-attribute approach). A visual-interactive technique using contingency tables is presented by Alsallakh et al. [AAMG12]. Similar to our approach, users can select individual data subsets of interest. Another class of approaches uses statistical dependency measures, to reveal 2D and multivariate relations [MBD*11] [BSW*14]. Yet other approaches working at the

bin-level use mixed data (see, e.g., Parallel Sets [KBH06]), or show relations between clusters and attributes (see, e.g., VisBricks [LSS*11]). However, all of these approaches fall short in revealing relations between multivariate data and time series data. Rare exceptions show layouts with relations between single attributes and time series [BRS*12a], or make use of data projection techniques to reveal relations with the visual variables position and color [SBSK15].

3. Domain, Data, and Task Characterization

The high-level goal of the analysts is to reduce the number of *faults* in the network. The challenge is to achieve a maximum of improvement at a minimum of cost. One quantification of success is the absolute number of faults (fault volume) that can be removed from the network in a certain period of time. More specifically, the analysts are interested in finding those parts of the network that produce a higher fault volume than others. A complementary information is the *fault rate*, i.e., the number of faults per network element. Together, a part of the network is interesting if it can be associated with many faults or more faults than expected (high fault rate), or a combination of the two. For example, parts with smaller fault volumes may be candidates for a maintenance exercise if their fault rate is very high.

A *fault* is described by attributes of the affected part of the network (location, product, technology, electrical measurements etc.), the fault itself (symptom, resolution if known), and its temporal information. The full data set contains more than 100 attributes with attribute value combinations (AVCs) at least in the order of 10^6. Since it is not feasible to look at these 10^6 fault time series of all AVCs by hand, we instead expect our visual analytics system to find common trend patterns across all the combinations and visualize the associated attribute values. This way, we ensure that the full search space is explored but broken down into a much smaller number of patterns that can be managed by analysts. The temporal information of faults and fault rates adds to the complexity of the challenge. The analysts expect spikes, sudden change, and upward trends to be interesting shapes in the temporal progressions that need to be identified by the system. However, it is important not to artificially restrict the set of shapes in advance. Thus, we choose a feature-based approach in combination with visual clustering.

Typical users of the system will be domain experts with basic statistical skills. They are familiar with the data, the network technology and business processes. Given such skills it is therefore important to hide away the analytical complexity and focus on representing the domain in intuitive and interactive visualizations.

4. The Visual-Interactive System

To address the aforementioned problems, we present a tool for the visual-interactive exploration of relations between time-oriented data and multivariate data. Three different views show the data from different perspectives. Combining these interactive views enables the analysts to explore relations between multivariate attributes and temporal data.

4.1. Data Abstraction and Functional Support

Together with the analysts, we agreed on a set of eight attributes most relevant for the analysis. The abstraction of the time-oriented data took several iterations. We calculated time series features and applied a visual clustering algorithm to provide overviews of both fault volume and fault rate patterns. Thus, we shifted the analysis task from the elementary to the synoptic level [AA06], allowing scalable analysis and significant insights. For every AVC, we applied the Piecewise Aggregate Approximation (PAA) descriptor [KCPM01] to compute aggregated features per time period of interest (in this case per week). Another user parameter was the time interval for time series patterns. Following our analysts' advice, we pre-calculated patterns of *four weeks* and *three months*. Since faults and fault rates have different value domains, we parallelized the analytical workflow, yielding two interesting *perspectives* on the temporal information. This is why we provide two instances of the Self-organizing Map (SOM) algorithm [Koh01] for visual cluster analysis. Our implementation allows analysts to visually observe the training of the iterative algorithm [SBVLK09]. The analysts welcomed this choice for providing an insight into the mechanics of the algorithm, which was known to them as 'some sort of sophisticated machine learning model' before.

4.2. Attribute View

The view in Figure 2 and Figure 1 (left) serves as an entry point for the analysis. It contains a list of bar charts, one for each attribute of the multivariate data, thus enabling the analyst to get a comprehensive overview of the distributions of the attributes. Every bar can be selected by the user to support brushing and linking in the other views. In the *fault distribution mode*, a classical bar chart visualization shows the distribution of observed values for each attribute (Figure 1). In the *relation seeking mode*, the height of bars in positive (blue) and negative (orange) y-direction indicates the deviation from the expected occurrence value given in the overall data set. As an example the currently selected data subset in Figure 1 has strong relations with a specific product ('Torsion Bar') and with a specific production step ('Milling'), to name just two.

4.3. Time Series Cluster View

The temporal information of the data set (see Figure 3) is presented in the upper right of the tool (see Figure 1). In two visualizations the results of the two SOMs with fault and fault rate patterns are shown, giving an overview of the temporal data as a whole. Together with the analysts, we decided to choose the hexagonal SOM variant, as proposed in earlier SOM-based approaches [Ves99] and in the original work by Kohonen [Koh01]. The results of two SOM trainings are shown, representing the *number of faults* and the *fault rates*. With the Time Series Cluster View, we depict information about the two most important variables of the data set depending on time. For each SOM a white-to-green colormap indicates the absolute number of faults represented in the cells. Looking at the grid as a whole, this is essentially a

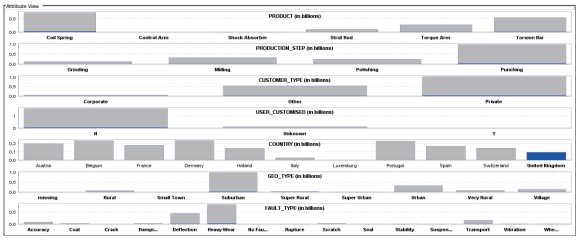

Figure 2: *The Attribute View in the fault distribution mode. The attribute value United Kingdom is selected (blue).*

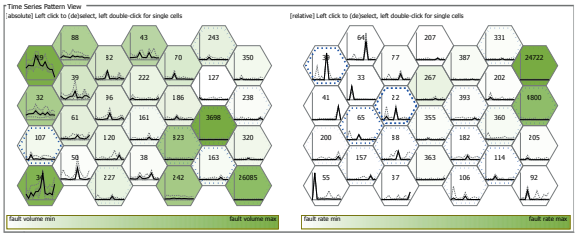

Figure 3: *Visualization of the clustering result (SOM). One Cell in the upper left is selected (blue outline).*

heat map. The saturation is a visual indicator of interestingness for any given cell. The visual representation of aggregated time series in each cell is designed as follows:

- A line chart shows the cell centroid as the representative of the data aggregate
- Dashed line charts show the the 25% and the 75% quantiles to represent the variation of the data over time
- A dashed outline of the hexagonal cell represents the ratio of currently selected temporal patterns

4.4. Time Series History

Following inquiries by the analysts, we provide an additional perspective on the data showing the temporal progression of selected subsets over years. The Time Series History is located at the lower right of the tool (see Figure 1). Analysts are now able to analyze the total *number of faults*, as well as the total *fault rate* over time. The areas highlighted with gray background color reflect the patterns selected in the Time Series Pattern View. The Time Series History is coordinated with the other views. Thus, temporal information of all configurations selected in the bar charts or in the SOM cells are summed up and displayed as single line charts. The number of such aggregated configurations can easily go into the hundreds or thousands, which is why we refrain from showing individual stacked line charts (cf. ThemeRiver [HHN00]). The rationale behind this combination of two well-known visualization techniques is to keep the analysis tool simple and intuitive for data experts while preserving the important information: the temporal change of the selected subset of the data.

5. Usage Scenario

As laid out in previous sections, analysts need to find network configurations with adverse fault developments. In Figure 1, we have selected a SOM cell with a particularly interesting fault pattern (blue outline): it shows (a) an overall upward trend, (b) a spike, and (c) a high fault volume indicated by a dark green background color. The Attribute View (in *relation seeking mode*) reveals the driving attribute values in blue: 'Torsion Bar', 'Milling', 'Other customers', and 'Deflection' to name just the most prominent. Without the tool, analysts would have to manually search through the entire AVC space and check all the associated time series for interesting changes. For the first time, a tool directly points analysts towards interesting relations. In addition the Time Series History (at the lower right of the tool) shows the global time axis for the selected fault pattern at a glance. Obviously, the highlighted area of the upper time series is equal to the selected fault pattern. Moreover, the highlighted area of the lower time series reveals two small spikes showing fault rate patterns.

The other way round, analysts can still explore AVCs by selecting values in the Attribute View and identify interesting temporal patterns related to the selected subset. We have selected the country 'United Kingdom' in Figure 2. The propagated selection can be seen in Figure 3. In the left SOM one of the cells (with 107 elements) has a considerably higher selection rate (dashed blue line), the time series pattern has a small early spike. In the SOM at the right the cells with 39, 22, and 65 elements have the highest selection rates (dashed blue lines). These three patterns show a similar progression with a small early spike followed by a clearly exposed spike roughly in the middle of the temporal pattern. The cell with 39 elements has the strongest spike, though. The analysts are now able to associate these patterns with the 'United Kingdom' as the country selected in the Attribute View.

In summary, our collaborating analysts feel they will be able to find many more interesting trend patterns and do so much quicker than before. Especially the possibility of switching the dependent variable between the attribute space and the two time series spaces is considered beneficial.

6. Conclusion

We presented a visual analytics tool for the exploration of relations between multivariate attributes and time series which we applied to complex telecommunication network fault data. Different views provide complementary perspectives of the data, brushing-and-linking enables analysts to focus on interesting attribute values or time series to reveal relations between the different data types. The analysts from the telecommunication provider involved in the design study expect a significant increase in both the efficiency and the effectiveness in detecting faults in their network by using the tool. Future work includes extensions for automatic highlighting of interesting patterns to further break down the huge search space for analysts.

References

[AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. doi:10.1007/3-540-31190-4. 2, 3

[AAD*10] ANDRIENKO G., ANDRIENKO N., DEMSAR U., DRANSCH D., DYKES J., FABRIKANT S. I., JERN M., KRAAK M.-J., SCHUMANN H., TOMINSKI C.: Space, time and visual analytics. *Int. J. Geogr. Inf. Sci. 24*, 10 (2010), 1577–1600. doi:10.1080/13658816.2010.508043. 2

[AAMG12] ALSALLAKH B., AIGNER W., MIKSCH S., GROLLER M.: Reinventing the contingency wheel: Scalable visual analytics of large categorical data. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 18*, 12 (2012), 2849–2858. 2

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data*, 1st ed. Human-Computer Interaction. Springer Verlag, 2011. doi:10.1007/978-0-85729-079-3. 2

[BRS*12a] BERNARD J., RUPPERT T., SCHERER M., KOHLHAMMER J., SCHRECK T.: Content-based layouts for exploratory metadata search in scientific research data. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (New York, NY, USA, 2012), JCDL '12, ACM, pp. 139–148. doi:10.1145/2232817.2232844. 2, 3

[BRS*12b] BERNARD J., RUPPERT T., SCHERER M., SCHRECK T., KOHLHAMMER J.: Guided discovery of interesting relationships between time series clusters and metadata properties. In *Knowledge Management and Knowledge Technologies (i-KNOW)* (New York, NY, USA, 2012), ACM, pp. 22:1–22:8. doi:10.1145/2362456.2362485. 2

[BSW*14] BERNARD J., STEIGER M., WIDMER S., LÜCKE-TIEKE H., MAY T., KOHLHAMMER J.: Visual-interactive Exploration of Interesting Multivariate Relations in Mixed Research Data Sets. *Computer Graphics Forum (CGF) 33*, 3 (2014), 291–300. doi:10.1111/cgf.12385. 2

[BWK*13] BERNARD J., WILHELM N., KRÜGER B., MAY T., SCHRECK T., KOHLHAMMER J.: MotionExplorer: Exploratory Search in Human Motion Capture Data Based on Hierarchical Aggregation. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 19*, 12 (2013), 2257–2266. doi:10.1109/TVCG.2013.178. 2

[DKG15] DASGUPTA A., KOSARA R., GOSINK L.: Vimtex: A visualization interface for multivariate, time-varying, geological data exploration. *Comput. Graph. Forum 34*, 3 (2015), 341–350. doi:10.1111/cgf.12646. 2

[Fri99] FRIENDLY M.: Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Computational and graphical Statistics 8*, 3 (1999), 373–395. 2

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (2011), 289–309. doi:10.1177/1473871611416549. 2

[GCML06] GUO D., CHEN J., MACEACHREN A. M., LIAO K.: A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Transactions on Visualization and Computer Graphics (TVCG) 12*, 6 (2006), 1461–1474. doi:10.1109/TVCG.2006.84. 2

[HHN00] HAVRE S., HETZLER B., NOWELL L.: Themeriver: Visualizing theme changes over time. In *Proceedings of the IEEE Symposium on Information Vizualization 2000* (Washington, DC, USA, 2000), INFOVIS '00, IEEE Computer Society, pp. 115–. 4

[KBH06] KOSARA R., BENDIX F., HAUSER H.: Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 12*, 4 (2006), 558–568. 3

[KCPM01] KEOGH E., CHAKRABARTI K., PAZZANI M., MEHROTRA S.: Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Inform. Syst. 3*, 3 (2001), 263–286. doi:10.1007/PL00011669. 3

[KH13] KEHRER J., HAUSER H.: Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 19*, 3 (2013), 495–513. doi:10.1109/TVCG.2012.110. 2

[Koh01] KOHONEN T.: *Self-Organizing Maps*, 3rd ed. Springer-Verlag Berlin Heidelberg, 2001. doi:10.1007/978-3-642-56927-2. 3

[LKL*04] LIN J., KEOGH E., LONARDI S., LANKFORD J. P., NYSTROM D. M.: Viztree: A tool for visually mining and monitoring massive time series databases. In *Very Large Data Bases* (2004), VLDB '04, VLDB Endowment, pp. 1269–1272. 2

[LSS*11] LEX A., SCHULZ H.-J., STREIT M., PARTL C., SCHMALSTIEG D.: Visbricks: Multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 17*, 12 (2011), 2291–2300. 3

[MBD*11] MAY T., BANNACH A., DAVEY J., RUPPERT T., KOHLHAMMER J.: Guiding feature subset selection with an interactive visualization. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (2011), pp. 111–120. doi:10.1109/VAST.2011.6102448. 2

[NAW13] NHON D. T., ANAND A., WILKINSON L.: Timeseer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 19*, 3 (2013), 470–483. 2

[SBM*14] STEIGER M., BERNARD J., MITTELSTÄDT S., LÜCKE-TIEKE H., KEIM D. A., MAY T., KOHLHAMMER J.: Visual Analysis of Time-Series Similarities for Anomaly Detection in Sensor Networks. *Computer Graphics Forum (CGF) 33*, 3 (2014), 401–410. doi:10.1111/cgf.12396. 2

[SBS11] SCHERER M., BERNARD J., SCHRECK T.: Retrieval and exploratory search in multivariate research data repositories using regressional features. In *ACM/IEEE Joint Conference on Digital Libraries* (New York, NY, USA, 2011), ACM, pp. 363–372. doi:10.1145/1998076.1998144. 2

[SBSK15] STEIGER M., BERNARD J., SCHADER P., KOHLHAMMER J.: Visual Analysis of Relations in Attributed Time-Series Data. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2015), Bertini E., Roberts J. C., (Eds.), The Eurographics Association. doi:10.2312/eurova.20151105. 3

[SBVLK09] SCHRECK T., BERNARD J., VON LANDESBERGER T., KOHLHAMMER J.: Visual cluster analysis of trajectory data with interactive kohonen maps. *Information Visualization 8*, 1 (2009), 14–29. doi:10.1057/ivs.2008.29. 2, 3

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages (VL)* (Washington, DC, USA, 1996), IEEE, pp. 336–. 2

[Ves99] VESANTO J.: Som-based data visualization methods. *Intelligent Data Analysis 3*, 2 (1999), 111 – 126. doi:http://dx.doi.org/10.1016/S1088-467X(99)00013-X. 3

[VWVS99] VAN WIJK J. J., VAN SELOW E. R.: Cluster and calendar based visualization of time series data. In *IEEE Symposium on Information Visualization (InfoVis* (Washington, DC, USA, 1999), IEEE Computer Society, pp. 4–. 2

[WVZ*15] WILHELM N., VÖGELE A., ZSOLDOS R., LICKA T., KRÜGER B., BERNARD J.: Furyexplorer: Visual-interactive exploration of horse motion capture data. SPIE Press. doi:doi:10.1117/12.2080001. 2

[ZJGK10] ZIEGLER H., JENNY M., GRUSE T., KEIM D.: Visual market sector analysis for financial time series data. In *Visual Analytics Science and Technology (VAST)* (2010), pp. 83–90. doi:10.1109/VAST.2010.5652530. 2