Eurographics Conference on Visualization (EuroVis), Posters Track (2016)
T. Isenberg and F. Sadlo (Guest Editors)

Towards Combining Attribute-based and Time Series-based Visual Querying

David Sessler[1], Martin Spott[2], Detlef Nauck[3], William Harmer[3], Jörn Kohlhammer[1] and Jürgen Bernard[4]

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

Abstract

We present a concept for the visual-interactive definition of meaningful subsets in data sets comprising multivariate attributes and time series data. Based on a generalization of requirements of a real-world user group, we propose a three-stage approach, combining visual-interactive querying, query filter analysis, and result exploration. The approach includes several design parameters that can easily be adapted in future design studies for alternative applications.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

Multivariate data and time series are two of the most common data types. Nearly every application domain gathers multi-modal information, consisting of multiple attributes, some of which also depend on time. The combined analysis of both data types in a unified approach can lead to new insight at the cost of additional design challenges. Related analysis goals can be broken down to low-level tasks such as filtering, visual comparison, or relation seeking. Depending on the analysts' prior knowledge about the data, visual-interactive systems need to address confirmatory or exploratory data analysis capability. To support analysts, both data spaces need to be accessed, characterized, and unified to facilitate the combined analysis. Our targeted user group, the analysts of British Telecommunications (BT), collect complex data of network service activities for a long time. This research effort is motivated by some of BT's analysts, who have an exploratory information need regarding performance attributes over time. The primary analysis goal is to retrieve meaningful data subsets, e.g., for downstream statistical testing, or for the communication of interesting patterns to facilitate decision making. For this purpose the challenge is to provide visual-interactive filtering capability to reveal relevant data subsets. In particular, this requires to combine visual-interactive filters for both the multivariate and the time series space in a unified approach. Provided that this challenge is addressed, we identify two downstream challenges. First, analysts are interested in how the different filter combinations relate to each other. Second, the presentation of retrieved subsets requires a visual interface that goes beyond traditional list-based result visualizations for two reasons. Arguments against list-based interfaces are because filtering does not enable a natural ranking of elements, and the number of relevant elements may still be significantly larger than 10, 20, or 30. Rather, the result space poses a new exploration space to be discovered.

We present a three-stage concept towards combining attribute-based and time series-based visual querying. The first stage provides visual interfaces for both the attribute space and the time series space, as well as faceted search and filter interaction. The second stage enables analysts to seek relations between intersecting filters. The third stage introduces a new visual interface for the exploration of retrieved subsets. The visualization technique combines data aggregation with outlier analysis, and presents search results in a similarity-preserving way.

2. Approach

We propose a three-stage approach towards the combined visual search and exploration of attribute facets and time series patterns.

2.1. Stage 1 - Faceted Search for Attributes and Time Series

Faceted search interfaces are a frequently applied technique in practice. We adopt the concept and combine it with standard bar-chart visualizations. In this way, analysts can gain an overview of the cardinality of the attributes, as well as the number of elements matching each of the observations (bins). The combination with faceted search enables the selection of subsets by different criteria, as applied in the illustrative example in Figure 1a and 1b. The visual interface automatically assigns colors for selected bins (here green for 'Milling', blue for 'UK' and orange for 'Suburban'). We preserve the color coding for the entire approach to allow linking different views. The query interface for time series patterns is inspired by visual cluster analysis approaches for time series data [BDF*15]. We present an overview of time series patterns, e.g., as a result of the Self-organizing Maps (SOM) algorithm. Based on the overview, analysts are able to filter the data set by selecting time series patterns of interest. In the illustrative example in Figure 1b two patterns are selected, visually represented with purple and red colors. The result of Stage 1 is a query to be executed in the database. Figure 1 proposes a concept to visually represent the query in a condensed way. In addition, analysts can be enabled to define the type of concatenations (AND, OR, etc.).

DOI: 10.2312/eurp.20161145

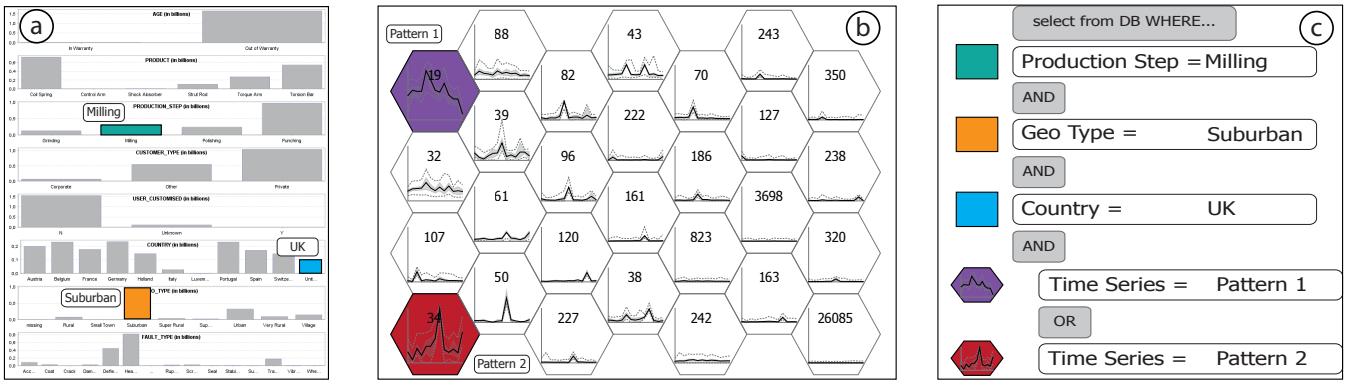

Figure 1: *Interfaces combining faceted search in multivariate attributes with query-by-example support in time series data. The visualizations a) and b) are the result of a recent design study [BSS*16], here applied to visual-interactive querying. Interface c) illustrates the visual representation of a query. Analysts may be interested in switching concatenations between AND and OR operations.*

2.1.1. Stage 2 - Visual Analysis of Intersecting Filters

In the second stage, analysts are able to interpret the set of applied filters. In many approaches, intersections between different subsets are of particular importance for gaining an understanding of the data set [AMA*14], or in this case the combination of filters defined by the analyst. We propose a matrix-based visualization since we assume that the number of filters is comparatively low and the structure provided by the position information enables an intuitive access to the desired information (see figure 2). For every intersection (cell of the matrix) the 'strength' of the relation between two filters is visualized. An obvious starting point is the visualization of pairwise occurrences with a contingency table. The visual encoding (here: growing point size) is a possible subject to be defined by analysts.

2.1.2. Stage 3 - Visual Exploration of Filtered Subsets

We address two challenges for the visualization of filtered subsets, i.e., the missing ranking of equivalent elements and the possibly large size of retrieved subsets. Figure 3 illustrates one design option to represent individual data elements in detail. However, the visual interface does not address the previously described challenges.

In Figure 4, we introduce a result visualization that resolves both challenges. Based on a similarity measure (default: time series similarity), we provide a similarity-preserving layout in 2D, e.g., in combination with a projection algorithm. We combine the approach with a data aggregation technique, i.e., a clustering algorithm. Finally, we propose the identification of outlier patterns to support both frequent pattern and outlier analysis [SBM*14]. As a result, Figure 4 combines both clusters and outliers in a single visualization. Color coding is used to link back to the applied filters. In the

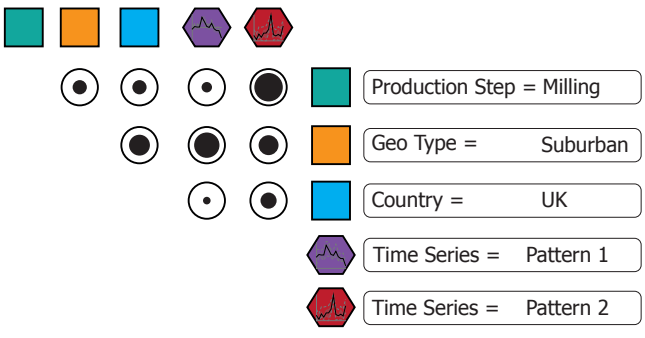

Figure 2: *Visual Interface revealing intersections between pairs of filters between different attributes.*

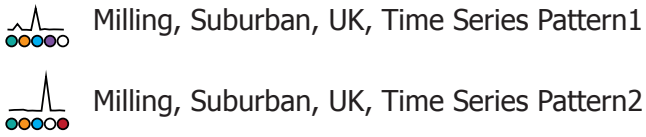

Figure 3: *Detailed result visualization for few elements. However, such standard list-based techniques are not applicable for the approach and need to be replaced by a more suitable visual interface.*

illustrative example, we assume that a user decided to concatenate attribute filters and time series patterns with an OR operation. Thus, for every entity the color codings show the match of the time series subset with the subsets defined by the attribute filters.

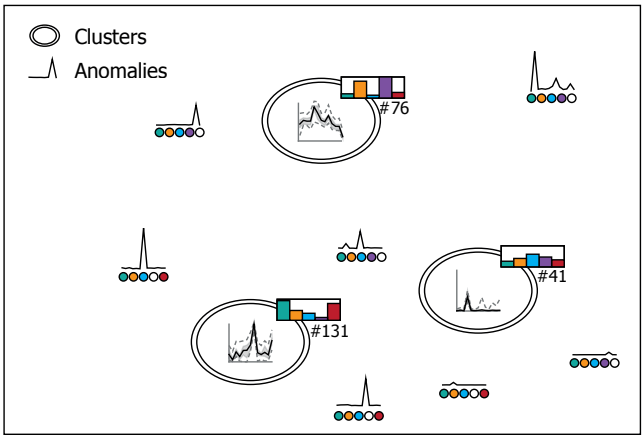

Figure 4: *Similarity-preserving layout of retrieved subsets. A clustering algorithm reveals frequently occurring time series patterns, outlier patterns are visualized separately.*

3. Future Work

We are still in an early phase within this design study. An obvious next step is the implementation of the concept on the basis of a careful data abstraction. We plan to implement an early prototype which will be refined in an iterative process, based on the analysts' feedback. Another point of interest is the reflection on the approach to generalize the visualization designs and the algorithmic support for other application domains. In particular, we identify similar information needs in collaborations in the medical, the Earth observation, and the energy domain.

References

[AMA*14] ALSALLAKH B., MICALLEF L., AIGNER W., HAUSER H., MIKSCH S., RODGERS P.: Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. In *EuroVis - STARs* (2014), Borgo R., Maciejewski R., Viola I., (Eds.), The Eurographics Association. doi:10.2312/eurovisstar.20141170. 2

[BDF*15] BERNARD J., DABERKOW D., FELLNER D., FISCHER K., KOEPLER O., KOHLHAMMER J., RUNNWERTH M., RUPPERT T., SCHRECK T., SENS I.: VisInfo: a digital library system for time series research data based on exploratory search - a user-centered design approach. *International Journal on Digital Libraries 16*, 1 (2015), 37–59. doi:10.1007/s00799-014-0134-y. 1

[BSS*16] BERNARD J., SPOTT M., STEIGER M., SESSLER D., KOHLHAMMER J.: Visual-Interactive Exploration of Relations Between Time-Oriented Data and Multivariate Data . In *EuroVis Workshop on Visual Analytics (EuroVA)* (2016), Andrienko N., Sedlmair M., (Eds.), The Eurographics Association. doi:10.2312/eurova.20151105. 2

[SBM*14] STEIGER M., BERNARD J., MITTELSTÄDT S., LÜCKE-TIEKE H., KEIM D., MAY T., KOHLHAMMER J.: Visual analysis of time-series similarities for anomaly detection in sensor networks. *Comput. Graph. Forum 33*, 3 (June 2014), 401–410. URL: http://dx.doi.org/10.1111/cgf.12396, doi:10.1111/cgf.12396. 2