Interactive Shape Specification for Pattern Search in Time Series

Paolo Buono
Dipartimento di Informatica
Università di Bari
Via Orabona, 4, Bari, Italy
+39-080-5442239
buono@di.uniba.it

Adalberto Lafcadio Simeone
Dipartimento di Informatica
Università di Bari
Via Orabona, 4, Bari, Italy
+39-080-5442299
simeone@di.uniba.it

ABSTRACT

Time series analysis is a process whose goal is to understand phenomena. The analysis often involves the search for a specific pattern. Finding patterns is one of the fundamental steps for time series observation or forecasting. The way in which users are able to specify a pattern to use for querying the time series database is still a challenge. We hereby propose an enhancement of the SearchBox, a widget used in TimeSearcher, a well known tool developed at the University of Maryland that allows users to find patterns similar to the one of interest.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI).

General Terms

Languages.

Keywords

Interactive visualization, interactive system, information visualization, visual querying.

1. INTRODUCTION

Time series analysis involves the use of algorithms and tools that allow users to better understand a phenomenon. Time series can be defined as an ordered sequence of measurement that describes a phenomenon. A user may perform analyses on time series in order to describe or to explain a phenomenon or to perform forecasting [1]. In both cases it is important to use algorithms that associate some behavior.

TimeSearcher is a time series visualization tool, recently updated to version 3.0 [1, 2]. This tool allows the user to interactively search recurring patterns in its data [3]. To accomplish this task, users can draw a box (the *SearchBox*) enclosing the pattern of interest, then through a contextual interface they can start the search. We made informal user testing in order to understand how much usable the interface was. Based on the observations made we improved the SearchBox widget.

2. TIME SERIES QUERYING

The search in TimeSearcher is composed by three steps. The user observes the time series and when (s)he finds an interesting pattern s(he) selects it and, by using the SearchBox, similar patterns are found. The users can tune the pattern search by interacting with the ToleranceHandle placed on the side of the box [3]. The ToleranceHandle sets the similarity degree. The process can be iterated. During the usability study we found that users would have liked to perform the pattern search starting from scratch, instead of searching first for an anomaly. In order to provide users with a widget that allows them to specify first a pattern shape and then start the search we improved the SearchBox with some interesting features.

3. SHAPE SPECIFICATION

Our goal was to allow users to specify the shape of a pattern in order to start the search in a time series. In literature there are some works that allow users to specify a query pattern. Don et al. [4] defined a set of typical patterns based on their shape; WizTree [5] allows users to specify the pattern behavior, after having transformed the series into an equivalent representation. Chortaras [6] permits to sketch the shape of the pattern, which is used as the input for finding similar patterns.
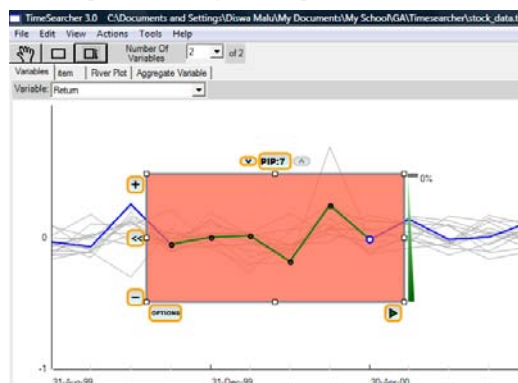

Figure 1 In the new SearchBox, the user can change the pattern in order to define the query pattern.

4. THE ENHANCED SEARCHBOX

As shown in Figure 1, as soon as the SearchBox is drawn on the interface, the time points belonging to the portion of the time series enclosed by the boundaries of the box are highlighted by colored disks. These disks will then be used as handles of the line segments to use for the purpose of pattern matching. The line formed by these handles is highlighted (in dark green). By

clicking and dragging the mouse over each disk, the user can then proceed to adjust the shape of the pattern. Users can also adjust the position of the active disk (the white one) by clicking on the *plus* and *minus* buttons located at the left corners of the box. Figure 2 shows the SearchBox after the user has dragged the white disc in order to decrease the third value of the pattern. The possibility to change a pattern was very appreciated by the users since they are not obliged to scan the dataset in order to find something to search for. Nevertheless, the users were still not completely satisfied because they could be interested to a specific pattern and they would like to depict it from scratch, eventually starting from a known shape. In order to satisfy this need we were inspired by Featurelens [4], a tool oriented to the analysis of text that uses a visualization technique similar to the one used in TimeSearcher.
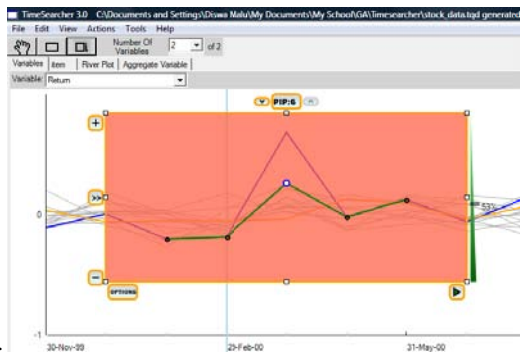

Figure 2 The user dragged the white disc into another position in order to change the original shape of the pattern.

We added the ShapesButton, a button that opens up a panel on the left side of the SearchBox containing several characteristic patterns (Figure 3). By clicking on it, the user will be presented with a side panel containing nine more buttons. Each one representing a common pattern that the user might be interested in adopting for the particular search that (s)he had in mind. These patterns are, from left to right: average, low peak, high peak (first row), increasing values, step up, step down (second row), decreasing values, valley, plateau (third row).
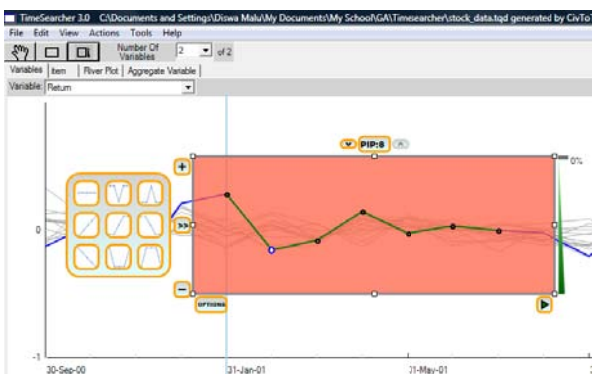

Figure 3 SearchBox with the ShapesButton exploded: the user can choose several pattern shapes in order to perform the query.

When the user moves its mouse cursor over one of these buttons, the pattern in the SearchBox is instantaneously changed to assume the shape shown in the button. This is, however, only a preview, the changes to the actual pattern shape will not be

applied until the user decides to click on the relevant button. Moving the mouse pointer outside the bounds of the ShapesButton will cause the pattern in the SearchBox to revert to its original shape.

5. DISCUSSION AND USER FEEDBACK

Our approach in pattern specification shows some similarities with QueryLines [7], a tool developed by Ryall et Al, which also allows user to perform time series querying. In QueryLines, the user can manually draw lines in the display area of the plot to select and sort the results. A query line can be of two types: "soft constraint" and "preference". QueryLines's shape specification features are more geared towards fine-tuning time series querying rather than explicit pattern analysis.

We performed informal tests on this new feature of TimeSearcher and we received positive feedbacks and suggestions. The possibility of manually selecting the pattern of interest was very appreciated by the users; now they need a way to specify null values or uninteresting parts of the shape. In addition, it would be useful to add the possibility to save patterns, whose shape may be of interest. Users could be able to do so by hitting a key combination, once they are satisfied with the shape of the pattern. "Custom" shapes will be accessible by clicking on an appropriate button in the related panel. It should be observed that drawing the SearchBox over the display area is required because the extents of the pattern area have to be known in order to perform the search.

6. REFERENCES

[1] Jank, W., Shmueli, G., and Wang, S., (2006). Dynamic, real-time forecasting of online auctions via functional models. In Proc. 12th ACM SIGKDD (Philadelphia, PA). ACM Press, New York, NY (2006), pp. 580-585.

[2] Hochheiser, H., and Shneiderman, B., (2004). Dynamic Query Tools for Time Series Data Sets, Timebox Widgets for Interactive Exploration. *Information Visualization*, vol. 3, No. 1. (March 2004), pp. 1-18.

[3] Buono, P., Aris, A., Plaisant, C., Khella, A., and Shneiderman, B., (2005). Interactive Pattern Search in Time Series, In Proc. of Visualization and Data Analysis, VDA 2005, SPIE, Washington DC (2005) pp. 175-186.

[4] Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B., and Plaisant, C., (2007). Discovering interesting usage patterns in text collections: integrating text mining with visualization. In Proc. Information and Knowledge Management, CIKM 2007, pp. 213-222.

[5] Keogh, E. and Kasetty, S., (2003). On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration, *Data Mining Knowledge Discovery*, vol. 7, No. 4. (October 2003), pp. 349-371.

[6] Chortaras, A., (2002). Efficient Storage, Retrieval and Indexing of Time Series Data, Department of Computing, Imperial College of London, UK, 2002.

[7] Ryall, K., Lesh, N., Lanning, T., Leigh, D., Miyashita, H., and Makino, S., (2005). QueryLines: approximate query for visual browsing. In Proc. of Human Factors in Computing Systems (CHI 2005) Short Paper, pp. 1765-1768.