

Bipeline: a Web-based Visualization Tool for Biclustering of Multivariate Time Series

Ricardo Cachucho, Kaihua Liu, Siegfried Nijssen, and Arno Knobbe

LIACS, Leiden University, the Netherlands,
`r.cachucho@liacs.leidenuniv.nl`, `k.liu.5@umail.leidenuniv.nl`,
`s.nijssen@liacs.leidenuniv.nl`, `a.j.knobbe@liacs.leidenuniv.nl`

Abstract. Large amounts of multivariate time series data are being generated every day. Understanding this data and finding patterns in it is a contemporary task. To find prominent patterns present in multivariate time series, one can use biclustering, that is looking for patterns both in subsets of variables that show coherent behavior and in a number of time periods. For this, an experimental tool is needed. Here, we present *Bipeline*, a web-based visualization tool that provides both experts and non-experts with a pipeline for experimenting with multivariate time series biclustering. With *Bipeline*, it is straightforward to save experiments and try different biclustering algorithms, enabling users to intuitively go from pre-processing to visual analysis of biclusters.

1 Introduction

The development of sensor networks has resulted in an explosion of time series data over the last years. These are large multivariate time series, where variables are collected synchronously over time. Thus, pattern mining of multivariate time series is becoming highly relevant, both in scientific research and industrial applications. Note that in the multivariate setting, not only patterns in one variable over time are relevant, but also relationships between multiple variables could provide useful insights. This task can be seen as clustering both time periods and variables, also known as biclustering [8–10].

Given a multivariate time series, it could be useful to try different biclustering algorithms. Also, one needs to optimize parameters across different steps, such as pre-processing, segmentation and biclustering itself. For each of these steps, there are many parameters to be optimized, leading to a large number of experiments. Furthermore, at each step, visual inspection is highly important for researchers to validate their findings. However, there is a lack of tools for this process.

We propose *Bipeline*, a web-based visualization tool that provides a pipeline for applying biclustering to multivariate time series. This tool is readily accessible to anyone via a web-based interface, allowing them to navigate through multiple experimental settings. Parameters can be interactively tuned, with web components such as checkboxes, sliders and drop-down menus. At each step of the biclustering process, feedback is provided by means of visualizations, with plots such as pre-processed time series, segmentation boundaries and biclusters. One or more biclusters can be plotted with a simple selection procedure.

2 Related Work

Until now, biclustering software tools with a graphical user interface have been developed to deal with biological gene expression data. BicOverlapper [1] is a tool for visual inspection of gene expression biclusters, introducing a novel visualization algorithm *Overlapper* to represent biclusters. Similarly, BiCluster Viewer [2] is a visualization tool for efficient and interactive analysis of large gene expression datasets. BicAT [3] implements multiple biclustering algorithms, for visualization and analysis of biclusters for expression data. BiGGEsTS [?] provides an environment for biclustering time series gene expression data.

All tools mentioned above integrate techniques for pre-processing and biclustering analysis, specifically for gene expression data. Their main purpose is to support biologists with the analysis and exploration of the gene expression data. However, these tools do not support biclustering analysis for multivariate time series. Also, most of them do not provide a pipeline experiment environment. Pipeline provides such a pipeline, where intermediate results can be inspected and saved. Using a friendly and interactive plotting environment, both non-experts and experts can pre-process, segment and analyze biclusters for multivariate time series.

3 Tool Overview

Pipeline is a web-based application that provides a pipeline to pre-process, segment and bicluster multivariate time series. An online version is available [12], which is compatible with all modern web browsers and across different client platforms. Both the user interface in the web browser and the server are implemented using R Shiny package [5]. In Figure 1, the system architecture illustrates the experimental pipeline and how each individual step relates to the other steps:

Importing: Users can upload datasets and have a first view of the data table and descriptive statistics (*minimum, maximum, mean, ...*). This first inspection, although useful, is not enough to assess the quality of the data.

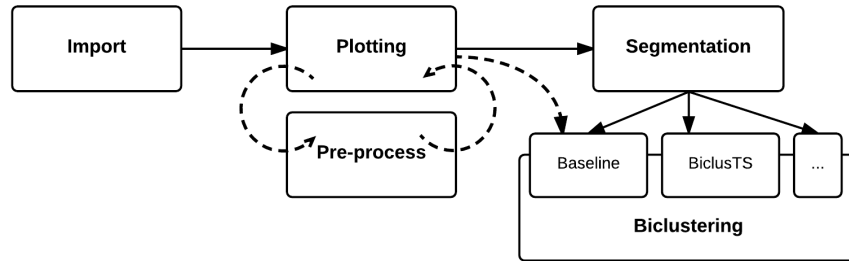


Fig. 1: A overview of *Pipeline* architecture.

Plotting: To gain further insight into the time series, it is crucial to have a visual inspection of the time series. The plotting panel includes multiple interactive plotting views, using a plotting R package *dygraphs* [6]. An example of these plots is illustrated in Figure 2(a). These interactive plots allow zoom in and out functionality, which is a highly desirable functionality for visual inspection of large time series.

Pre-processing: This panel allows preliminary handling of data such as: excluding variables, normalization, conditional removal and replacement of data, and outlier removal. Users can alternate between *plotting* (Figure 2(a)), and *pre-processing* (Figure 2(b)) until satisfied, then export the pre-processed data by clicking the *Save* button.

Segmentation: This allows segmentation of the data, one of the steps necessary for the biclustering as suggested by [10]. By default, all variables share the same parameter settings: *window size*, *overlap* and *threshold* can be easily tuned. For greater flexibility, the user can dynamically create new tabs to set the parameters for individual variables. Additionally, a *minimum segment size* is customizable, and the tool will merge short segments to its most similar contiguous segment. Segmentation results can be visualized (Figure 2(c)), saved and (re-)loaded, allowing the results to be used during the next step, biclustering.

Biclustering: In *Bipeline*, we implement a number of biclustering algorithms, group in three categories. The *baseline algorithms* allow users to try well-known biclustering algorithms (e.g., Cheng & Church) [8, 9], that have been implemented using R package *biclust* [7]. *Segmentation + Baseline* biclusters the time series using an average representation of each segment, instead of using individual rows. *Segmentation + BiclustTS* is a novel algorithm [10] introduced to recognize similarities between segments, using probability density-difference estimation [11]. All biclusters are plotted in colored blocks, as shown in Fig.2(d). Users can select the biclusters they want to see, and the plot will respond with a real-time update.

Multiple features are shared by both *Segmentation* and *Biclustering*. Plots and parameter tables from different experiments are kept in history, allowing users to navigate back and forth to compare results and optimize parameters. During computationally expensive tasks, the front-end displays a progress bar, while the back-end server is busy carrying out the calculations. Furthermore, interactive web components can be saved into images with a single click.

4 Conclusion

We propose *Bipeline*, a web-based visualization tool, which provides a pipeline for applying biclustering to multivariate time series. Its main features include: visual inspection at multiple stages, interactive zoom in and out plotting, easy navigation, storage of results, and saving plots and experimental settings using a single click. *Bipeline*'s intuitive web-based design, makes it accessible both to experts and non-experts, and compatible across platforms.

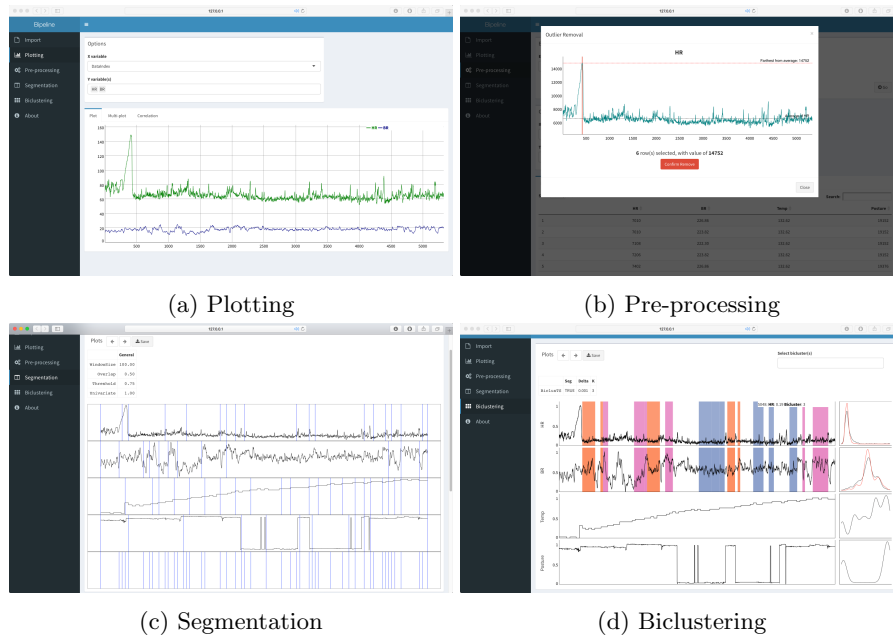


Fig. 2: *Bipeline* user interface.

References

1. R. Santamara, R. Thern, L. Quintales, BicOverlapper: A tool for bicluster visualization. *Journal Bioinformatics*, pp. 1212–1213, 2008.
2. J. Heinrich, R. Seifert, M. Burch, D. Weiskopf, Bicluster viewer: a visualization tool for analyzing gene expression data, *Proc. ISVC*, pp. 641–652, 2011.
3. S. Barkow, S. Bleuler, A. Prelic, P. Zimmermann, E. Zitzler. BicAT: a biclustering analysis toolbox, *Journal Bioinformatics*, pp. 1282–1283, 2006.
4. J. Goncalves, S. Madeira, A. Oliveira, BiGGEsTS: integrated environment for biclustering analysis of time series gene expression data, *Journal BMC*, pp. 1–11 2009.
5. W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, shiny: Web Application Framework for R. R package version 0.13.1, 2016.
6. D. Vanderkam, J. Allaire, J. Owen, D. Gromer, P. Shevtsov, B. Thieurmél, dygraphs: Interface to 'Dygraphs' Interactive Time Series Charting Library. R package, 2016.
7. S. Kaiser, R. Santamaria, T. Khamiakova, M. Sill, R. Theron, L. Quintales, F. Leisch, E. DeTroyer, biclust: BiCluster Algorithms. R package version 1.2.0, 2015.
8. Y. Cheng, G. Church, *Biclustering of Expression Data*, *Proc. Eighth International Conference on Intelligent Systems for Molecular Biology*, pp. 93–103, 2000.
9. S. Madeira, A. Oliveira, *Biclustering Algorithms for Biological Data Analysis: A Survey*, *Journal IEEE/ACM Trans. Comput. Biol. Bioinformatics*, pp. 24–45, 2004.
10. R. Cachucho, S. Nijssen, K. Liu, A. Knobbe, *Biclustering Multivariate Time Series*, Under review for *Proc. ECML-PKDD*, 2016.
11. M. Sugiyama, T. Kanamori, T. Suzuki, M. Plessis, S. Liu, I. Takeuchi, *Density-Difference Estimation*, *Proc. NIPS*, pp. 683–691, 2012.
12. <http://fr.liacs.nl:7000>