Analyzing Parameter Influence on Time-Series Segmentation and Labeling

Martin Röhlig, Martin Luboschik, Heidrun Schumann * Markus Bögl, Bilal Alsallakh, Silvia Miksch †

ABSTRACT

Reconstructing processes from measurements of multiple sensors over time is an important task in many application domains. For the reconstruction, these multivariate time-series can be automatically processed. However, the outcomes of automated algorithms often vary in quality and show strong parameter dependencies, making manual inspections and adjustments of the results necessary. We propose a visual analysis approach to support the user in understanding parameters' influences on these results. With our approach the user can identify and select parameter settings that meet certain quality criteria. The proposed visual and interactive design helps to identify relationships and temporal patterns, supports subsequent decision making, and promotes higher accuracy as well as confidence in the results.

1 INTRODUCTION

In many application domains dealing with time-series analysis, it is required to decompose multivariate time series into a sequence of segments and assign labels to these segments. This task is described as a segmenting and labeling problem. The goal is to reconstruct time evolving activities, processes, or physical phenomenons generating these time-series in a more conceptional way. The data stems, for example, from sensor system measurements. Possible application scenarios are process reconstruction in air traffic control, oil process plants, or in navigation of autonomous robots [4].

Usually automatic algorithms, such as Hidden Markov Model based techniques [5], compute the segments and assign labels to these segments. Although these algorithms are very fast, they are limited in accuracy, for example, if exceptional patterns occur, or if the input data contains missing values or outliers. Recently an approach to support domain experts in overcoming some of the limitations in the automated segmentation algorithms was proposed [1]. In this approach, users, like domain experts, can visually inspect the automatic segmentation results and identify exceptions or special patterns in the segmentation. These results are shown along with the raw input data. Hence, the user is enabled to examine the results in relation to the raw data, to investigate reasons for segmentation problems, and to adjust the results.

An open problem in this approach is the lacking support to choose a parameter setting for the segmentation algorithm that results in adequately labeled segments. Due to the large range of parameters, an iterative individual visual inspection of multiple parameter settings and the multiple segmentation and labeling results is cumbersome. Furthermore, it does not allow for a comparison of parameter settings to explore the parameters' influence and in this way to guide the user to fine-tune the used algorithms. In this regard, the aforementioned problem is related to the field of parameter visualization, aiming for a linked analysis of several hundred parameter settings and the resulting labeled segments. Recent approaches from this domain [3] facilitate such analysis using compact color-driven encodings of parameters and data. Nevertheless, labeled segments over time – which are categorical data – are not covered yet.

*University of Rostock, E-mail: {martin.luboschik, martin.roehlig, heidrun.schumann}@uni-rostock.de

†Vienna University of Technology, E-mail: {boegl, alsallakh, miksch}@cvast.tuwien.ac.at

For this reason, we aim to close this gap and propose a new approach which enables the interlinked analysis of parameters and labeled segments over time. We support the user in the analysis of how the parameters influence the segmentation and labeling results by visual and interactive means. Providing a compact overview visualization of the different parameter settings and associated segment output in combination with interactively applicable ordering strategies allows the user to explore parameter dependencies, to adjust the segmentation and to find labeled segments that are stable across different parameterizations. Thus, our approach supports the analysis and selection of appropriate parameter settings which results in a higher accuracy and confidence in the segmentation and labeling results.

2 VISUAL DESIGN

The basic idea of our approach is to show the parameter combinations used for the automated algorithms aligned with the labeled segments over time in one compact overview visualization. Our approach follows the concept presented in [3]. Beside dealing with categorical data, we extend this idea further by integrating a compound view of the labeled segments that provides user guidance.

Visualizing the labeled segments The amount of labeled segments and different parameter combinations requires a compact encoding of the information. Therefore, we resort to a row-wise pixel-based visualization, depicting the segmentation and labeling results of each parameter combination as a sequence of colored stripes over time (Fig. 1b). To distinguish the individual labels from each other, we apply a qualitative color scale, and, if applicable, match it with existing color conventions in the problem domain. The compact row-wise encoding of labeled segments provides an overview and facilitates the comparison of different segmentation and labeling results. As we point out in Section 3, dedicated rendering algorithms or visual cues are necessary to ensure a minimum visibility of individual segments or to hint at overplotting due to screen space limitations.

Visualizing the parameters The parameter combinations are initially a subset of equally sampled parameterizations from the parameter space. These combinations are displayed row-wise next to the respective segmentation results (Fig. 1a). Each parameter is assigned to a separate column with the parameter values being color coded. The color coding of the columns reflects the type and value range of the represented parameters. For example, numerical parameter values are encoded with a sequential color scale.

Ordering of rows The row-wise alignment of parameter combinations and labeled segments enables analyzing their dependencies. The order of these rows has substantial impact on the possible correlations and regularities revealed by our visualization. For this reason, we investigated different ordering strategies. We seek an order that places similar rows close together either according to the labeled segments or according to the parameters themselves. The first option focuses on the detection of labeled segments with similar value distributions and temporal patterns. In contrast, an ordering with regard to the parameters fosters the interpretation of the parameter influence on the segmentation and labeling outcomes. For both options, we apply specific metrics and ordering procedures. As an example the Edit-Distance [2] could be utilized as a similarity measure for labeled segments over time. To support alternative setups, interests and tasks, the user can interactively choose between ordering methods and select certain time intervals and parameter subsets to adjust the arrangements accordingly.

Compound view of labeled segments The above design decisions allow for an exploration of individual segmentation and labeling results as well as parameter dependencies. To gain a summarized overview of multiple segmentation results at once, we introduce the compound view (Fig. 1c-e), that summarizes the different labels at each time step. Special considerations are needed to compute the compound segment boundaries based on the labeled segments they aggregate. A simple solution would be to equally divide the time span and aggregate the segments for each interval. The compound view encompasses three separate sequence views and is displayed at the bottom of the labeled segments. The first view consists of a single sequence of colored stripes (Fig. 1c). These stripes encode the most probable segment across all parameter combinations for a specific time interval. This could be done by calculating the majority, but could also employ more complex aggregations, for example, involving the confidence values. To communicate the confidence in this labeling, we make use of a second sequence view right below the previous one (Fig. 1d). The overall confidence is depicted by a gray scale, in which high values are indicated by dark gray. To clearly identify high confidence, we enlarge the stripes which are above a certain confidence level (Fig. 1d). A third sequence view provides further details and encodes the ratio of the different labels contributing to the compound label at each interval using stacked bars (Fig. 1e). Uncertain segments, missing data, and segments with low ratio are summarized into a bar with specific color (gray bars in Fig. 1e). Details about these summarized segments are provided on demand. Interactively hovering the parts of the stacked bars highlights the respective segments in the labeled segments view.

In this way, the three components of the compound view allow to easily trace predominant labels over time while simultaneously giving the user a sense for the stability of the label selection. Moreover, it facilitates an interactive data folding: Labels which are of high confidence and stable across different parameterizations for a long time period, can be automatically collapsed – triggered by the user – to shorten the exploration of long term time series without loss of information.

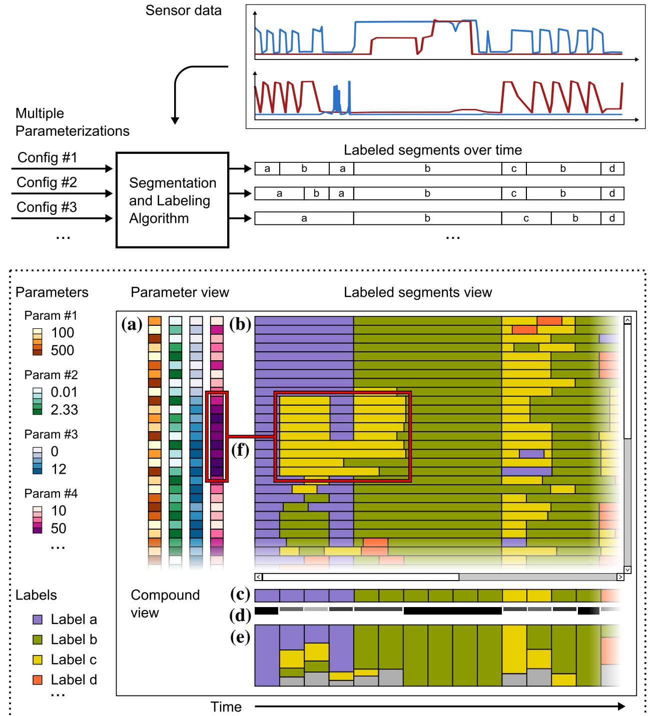

Figure 1: Our approach for analyzing parameter influence on segmentation results. Different parameterizations (a) are shown next to the ordered labeled segments (b) to allow exploration of the parameter settings and the respective outcome. The compound view (c-e) shows the most probable decision of the labeled segments (c), along with the corresponding confidence values (d). More detail on what the decision relies on is shown in the stacked bar chart (e). Specific patterns and correlations can be revealed interactively (f).

3 Discussion and Future Work

The presented approach helps to build an understanding of automatically labeled segments over time and in general aims to support the analysis of time-evolving categorical data. Involving the user in the analysis loop enables selecting different parameter combinations. The interactive sorting allows for the detection of parameter influence on labeled segments and the compound view shortens the time needed to explore long term time-series without information loss. However, even with our compact visual design the amount of parameter combinations and time-series segments can quickly exceed the available pixels in the dedicated display space. This can lead to overplotting and therefore impairs the interpretation of temporal patterns and parameter dependencies. As a countermeasure we plan to integrate graphical indicators hinting at overplotting occurrence as well as interactive means, like zoom and pan, to support the user in the data exploration process.

Another topic for future work is the extension of our approach to also take into account the provenance of the data, the different automated methods, and their associated parameter spaces. Our goal is not to limit the user to identify acceptable parameter settings for one algorithm, but to enhance the search to find the best possible combinations of parameters as well as automated segmentation and labeling methods. In doing so, we aim at gaining further insights and supporting the user in the task of process reconstruction with the desired accuracy.

Proposing and discussing our approach encourages further research in the visual analysis of parameter dependencies with regard to time-series segmentation and labeling. We aim at applying our approach to the actual scenarios presented in [1] for determining how to improve the visual design and evaluate its effectiveness. We also aim to tackle the open issues mentioned above such as determining compound segment boundaries, sorting strategies of the rows, and supporting larger numbers of parameters.

Acknowledgements This work was supported by the German Research Foundation (DFG) as part of VASSIB (SPP 1335), and by the Austrian Federal Ministry of Science, Research, and Economy via CVAST, a Laura Bassi Centre of Excellence (No. 822746).

References

[1] B. Alsallakh, M. Bögl, T. Gschwandtner, S. Miksch, B. Esmael, A. Arnaout, G. Thonhauser, and P. Zöllner. A visual analytics approach to segmenting and labeling multivariate time series data. In *EuroVis Workshop on Visual Analytics (EuroVA)*. Eurographics, 2014.

[2] L. Chen and R. Ng. On the marriage of lp-norms and edit distance. In *International conference on Very large data bases (VLDB), Volume 30*, pages 792–803. VLDB Endowment, 2004.

[3] M. Luboschik, S. Rybacki, F. Haack, and H.-J. Schulz. Supporting the integrated visual analysis of input parameters and simulation trajectories. *Computers & Graphics*, 39:37–47, 2014.

[4] J. M. Molina, J. Garcia, A. C. B. Garcia, R. Melo, and L. Correia. Segmentation and classification of time-series: Real case studies. In *IDEAL 2009*, pages 743–750. Springer, 2009.

[5] P. Wang, H. Wang, and W. Wang. Finding semantics in time series. In *ACM SIGMOD Conf. on Management of Data*, pages 385–396, 2011.