Supporting Visual Parameter Analysis of Time Series Segmentation with Correlation Calculations

C. Eichner, H. Schumann and C. Tominski

University of Rostock, Institute of Computer Science, Germany

Abstract

Parameter analysis can be used to find out how individual parameters influence the output of an algorithm. We aim to support the visual parameter analysis of algorithms for the segmentation of time series. To this end, we automatically search for correlations between parameters and the segmentation outputs. Correlations are not only determined globally, but also locally within parameter subspaces. Calculated correlations are used to visually emphasize parameter and value ranges with high influence on the segmentation. By interactive exploration, the analyst can study the multidimensional parameter space in depth.

CCS Concepts

•***Human-centered computing*** → ***Visual analytics;*** *Visualization application domains; Information visualization;* •***Mathematics of computing*** → *Time series analysis;*

1. Introduction

Algorithms play a crucial role in the context of visual analytics. They process data D in a variety of ways and produce different kinds of output results O depending on the data and a set of parameters $D \times P_1 \times \cdots \times P_n \rightarrow O$. However, it is not always clear if and how the involved parameters influence the computed results. This question can be addressed with parameter space analysis [SHB*14].

We are interested in the parameter space analysis of algorithms for the segmentation of time series. Segmentation algorithms partition the time axis into segments and assign to each segment a certain segment type. A segmentation result can then be described as a sequence of segments $\mathbf{s} = (s_1, s_2, \ldots, s_k)$, where each segment $s_i = (l_i, t_i)$ is characterized by its length l_i and type t_i. In order to arrive at appropriate analysis results, users need to know how the choice of different parameter values affect the segmentation. In other words, how does the segmentation change when certain parameters are adjusted? To help users answer this question, we integrate correlation calculations and highlighting mechanisms into an existing parameter space visualization.

2. Visualization

The visualization depicts parameterizations and segmentation sequences as color-coded rows in a tabular form [RLK*15]. Each row represents a parameterization $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and its corresponding sequence of segments $\mathbf{s} = (s_1, s_2, \ldots, s_k)$. Figure 1a shows an example with five parameters and the obtained segments for 768 different parameterizations.

Sorting the rows according to the values of a parameter enables the user to study the impact that a particular parameter may have on the segmentation. Figure 1a has been sorted according to the third parameter, which has three different values represented as white, gray, and black. Sorting with respect to this parameter does not reveal any interesting pattern among the segmentations. The user would need to perform many different sorting operations and inspect many different visual representations in order to derive insight about parameter dependencies. To make this procedure less time-consuming, we compute correlations between parameters and segmentation sequences and highlight strong correlations in the visualization to make dependencies easier to find.

3. Calculation of Correlations

The calculation of correlations is done in two steps. First, each sequence of segments $\mathbf{s}$ has to be characterized by a feature vector $\mathbf{f}$. In our case, it makes sense to consider the *number of segments* c and the *average segment length* l of $\mathbf{s}$ per segment type $t \in T$. As a result, we get feature vectors of the form $\mathbf{f} = (c_1, l_1, c_2, l_2, \ldots, c_m, l_m)$ where $m = |T|$ is the number of different segment types.

The second step is to calculate the correlation between each of the parameters and each component of the feature vectors using the Pearson correlation coefficient. Correlation coefficients close to zero indicate that a parameter has no influence, whereas values close to 1 or -1 suggest a high influence. For example, if the calculation yields a correlation coefficient of 0.9 between parameter P_2 and feature component c_7, we can say that the second parameter has a strong influence on how many segments get assigned the seventh segment type.

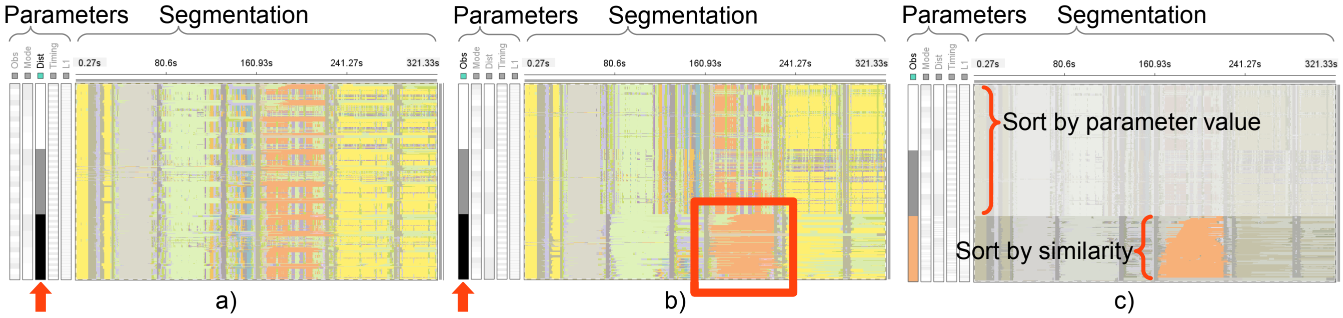

Figure 1: *Three views of the same segmentation results. Each view depicts the parameter space with five parameters on the left side in gray and the segmentation results on the right side. Each row represents a segmentation result as color-coded segments. Rows are sorted by the values of the marked parameter. From left to right: (a) Parameter with minor influence: There is no recognizable correlation between segment properties and parameter values. (b) Parameter with influence on the length of segments. Low parameter values are depicted at the bottom and produce long orange segments. No additional rearrangement of rows is used. (c) Highlighting to emphasize the correlation between long orange segments and low parameter values. The segmentation results are sorted to place similar orange segments next to each other.*

The standard way of computing correlation coefficients is to consider all parameterizations and segmentation sequences, or in other words, all rows. However, such a global search does not account for correlations between subsets. Figure 1b shows an example. The average length of orange segments seems to correlate only weakly with the first parameter when looking globally at all segmentation results. On the other hand, when we focus on the parameter range depicted in black (lowest parameter value), we can clearly see that longer orange segments occur only in this range. In order to detect such patterns automatically, we perform a subspace-search for correlations.

The procedure starts with reasonably small subsets. A good starting point is to search for correlations in subsets that share the same parameter value or have parameter values in the same range. Then, each of these subsets is expanded to create increasingly bigger subsets. Dynamic programming is applied to calculate the correlation coefficients for all of these subsets. To increase the calculation efficiency, partial sums of the Pearson correlation formula that have already been calculated for smaller subsets are reused for the larger subsets.

As a result, we now know which parameters correlate with which components of the feature vectors and hence have a strong impact on the outcome of the segmentation algorithm. We know this not only globally, but also for subsets of the data. The next step is to highlight the correlations visually.

4. Highlighting and Interactive Exploration

The highlighting has to communicate three aspects: (i) the influencing parameter, (ii) the subrange in which the influence takes effect, and (iii) the segment property being influenced. As illustrated in Figure 1c, this is accomplished by the following steps. First, the uninvolved parameters and segments with unaffected segment type are dimmed. Second, all rows are sorted according to the values of the influencing parameter and the subrange of influence is marked. Then, only the rows within the subrange are sorted according to their similarity. To this end, we compute the edit distances between rows and reorder them such that the sum of distances between all neighboring rows is minimal. As can be seen in Figure 1c, visual clutter is reduced, the important information stands out, and the comparison of similar segments becomes easier.

Based on this visual highlighting, analysts can interactively explore all aspects of the calculated correlations. We support two strategies to conduct such an interactive exploration:

- Parameter first: The parameters are automatically divided into parameters that have strong correlations and parameters with only weak or no correlations. The analyst successively examines interesting parameters, starting with those that have the strongest correlations. For each parameter one gets to see all correlations that involve the selected parameter. This helps users to estimate potential parameter influences.
- Segmentation first: The user searches for possibly interesting properties of segments in the visualization and manually selects regions to be inspected in more detail. Then, the user gets to see all those correlations that have an effect on segments in the selected region. This way, the user can evaluate which parameter might have caused the observation. Correlations that match the region best and those with high correlation coefficients are shown first, because they are likely the best explanations.

During exploration analysts can freely switch between both strategies to incrementally get more and more insight into the influence of each parameter of the segmentation algorithm.

5. Summary

This work proposed to support the analysis of parameter dependencies of segmentation algorithms by a novel integration of correlation calculations, interactive highlighting and exploration, and a suitable visual representation. The approach is ongoing work and has been applied in the context of activity recognition [RLK*15, BBB*18].

References

[BBB*18] Bernard J., Bors C., Bögl M., Eichner C., Gschwandtner T., Miksch S., Schumann H., Kohlhammer J.: Combining the automated segmentation and visual analysis of multivariate time series. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2018). conditionally accepted. 2

[RLK*15] Röhlig M., Luboschik M., Krüger F., Kirste T., Schumann H., Bögl M., Alsallakh B., Miksch S.: Supporting activity recognition by visual analytics. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2015), pp. 41–48. doi:10.1109/VAST.2015.7347629. 1, 2

[SHB*14] Sedlmair M., Heinzl C., Bruckner S., Piringer H., Möller T.: Visual parameter space analysis: A conceptual framework. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2161–2170. doi:10.1109/TVCG.2014.2346321. 1