Categorizing Uncertainties in the Process of Segmenting and Labeling Time Series Data

M. Bögl[1] and C. Bors[1] and T. Gschwandtner[1] and S. Miksch[1]

[1]TU Wien, Austria

Abstract

The segmenting and labeling of multivariate time series data is applied in different domains, e.g. activity recognition or sensor states. This involves several steps of (pre-) processing, segmenting, and labeling of time intervals, and visually exploring the results as well as iteratively refining the parameters for all the processing steps. Within these processes different uncertainties are involved and relevant. In this poster we identify and categorize important uncertainties in this problem domain. We discuss challenges for visually communicating these uncertainties throughout the segmenting and labeling process.

CCS Concepts

•***Human-centered computing*** → ***Visualization theory, concepts and paradigms;*** •***Mathematics of computing*** → *Time series analysis;*

1. Introduction

Segmenting multivariate time series data into time intervals and associating labels to these intervals from a label set is done in different application domains (e.g., the recognition and labeling of activities based on sensor measurements, different states in smart buildings, speaker recognition in audio tracks). The field of segmenting and labeling time series comprises various challenges that are currently being worked on. Throughout the process of segmenting and labeling one main challenge is the contingency of uncertainties on segmenting and labeling results, how different uncertainties are relevant during the process, and how these are introduced, altered, and how they influence the decision making. In this abstract we first briefly describe the process of segmenting and labeling, the influence of uncertainty, and we provide a categorization of different types of uncertainties. We follow up with some considerations that are important for integrating uncertainty representations in such an exploration environment supporting the process of segmenting and labeling.

2. Segmenting and Labeling Process

In principle, segmenting and labeling of time series can be described as a three stage process (see Figure 1). Given multivariate time series data, mostly from sensor measurements, multiple preprocessing steps need to be applied as a first step to make the raw data useful for subsequent processing. The second step is the application of appropriate segmentation and labeling algorithms to provide decisions on the segments and the labels. Given the output of the segmenting and labeling step, it is possible to derive the segmenting and labeling results and visually communicate these results in an interactive exploration environment to the user. Different parameter configurations of the algorithms result in multiple segmenting and labeling outcomes. The user explores them to find the most adequate result sets. Enabling adjustments in the parameters and the algorithms through the exploration environment constitutes an iterative process that results in findings and insights by the user as well as in refined results, together with segmented and labeled time series. In Figure 1 we illustrate the process of segmenting and labeling of multivariate time series data, which we adapted from [BBB*18] with more details on involved uncertainties.

3. Uncertainty in the Segmenting and Labeling Process

In previous work, we described how the segmenting and labeling process can be enhanced by using the knowledge generation model by Sacha et al. [SSS*14] and discussed aspects of uncertainties in the segmenting and labeling results [GSB*15]. In the following, we further elaborate a categorization of uncertainty types that are relevant in segmenting and labeling time series data and describe their relations to the particular processing steps.

3.1. Categorization of Uncertainties

Investigating uncertainties involved in the above mentioned process, we identify several aspects of uncertainty and categorize them to four types of uncertainty which may foster the informed execution of the segmentation pipeline. They appear across different stages in the process and also affect the results and decisions in different aspects (see Figure 1). Because of how and where they

DOI: 10.2312/eurp.20181126

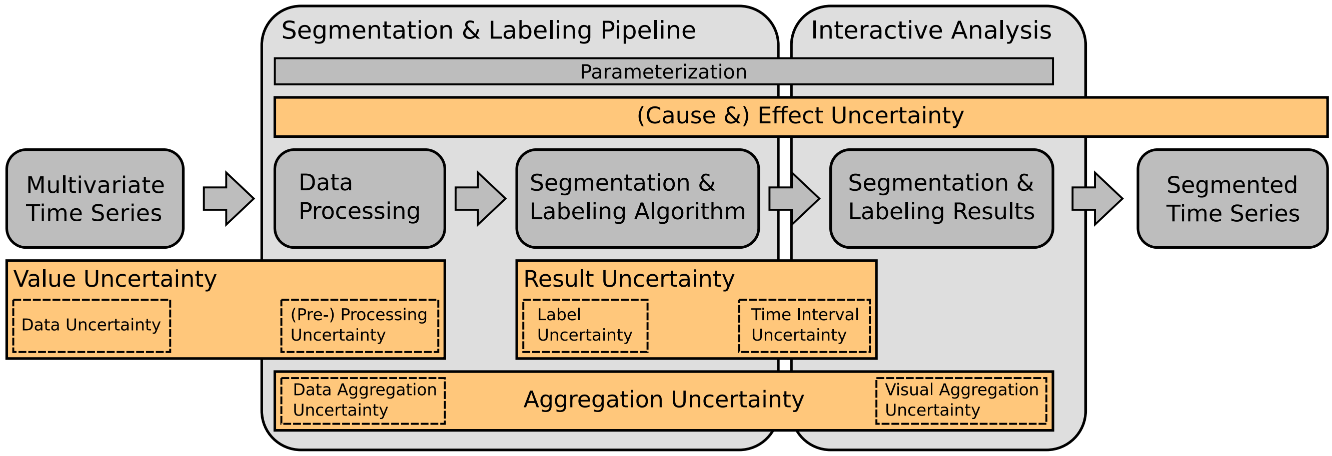

Figure 1: *Uncertainty types involved in the process of segmenting and labeling multivariate time series data. In this figure you can observe the relation between uncertainties and process steps. The vicinity of the uncertainties to the corresponding steps imply where they are generated in and which they influence, e.g. value uncertainties are coming from the input time series or generated by data processing.*

appear as well as what effect they have on the data or the visual representation, we group them into the following categories:

- The *Value uncertainty* includes (a) *data uncertainty* contained in the input data, and (b) *(pre-) processing uncertainty* that is introduced by changes to the value domain (e.g., noise reduction) in the original time series.
- The *Result uncertainty* contains (a) *label uncertainty* where information about the winning label and the respective uncertainty information is available in the form of probabilities for each potential label. These are provided by the segmentation models. The result uncertainty also contains (b) *time interval uncertainty*, which indicates the uncertainty about the definite start/end of a label interval. It is constructed through the implicit changes of label probabilities and winning label transitions inherent in label uncertainty.
- *Aggregation uncertainty* can be comprised of (a) *data aggregation uncertainty*, propagated from value domain uncertainty of (pre-)processing steps and can be related to *value uncertainty*. (b) Visual aggregation uncertainty is introduced by visually representing segmentation results, if aggregation is required due to insufficient screen/pixel-space.
- *(Cause &) effect uncertainty* is not explicitly recorded in the segmentation pipeline, but stems from consecutively performing interactive exploration, re-adjustment, and comparison of different algorithms and/or parametrizations.

Each type of uncertainty obligates special considerations regarding the visual representation within an interactive exploration environment and how the different uncertainty properties can be indicated in tandem. In the following section we briefly discuss these challenges and give a short outlook on future work and next steps.

4. Discussion and Conclusion

To generalize, when visualizing uncertainties, different aspects require attention depending on the stage of analysis within the three step process. To support visual interactive analysis a comprehensive overview of the segmentation results is required, which also includes adequately communicating the respective types of uncertainty. Our categorization of different types of uncertainty is a first step towards formulating metrics that help to communicate these uncertainties. For instance, a value uncertainty metric would be a numeric measure that represents the aggregated uncertainty inherent in the input data as well as the uncertainty introduced by different pre-processing steps. Given such a composite measure, a simple uncertainty visualization technique can be employed for comprehensive overview (e.g., color gradients, saturation). In future work we will investigate different visual representations for exploring details on the composition of these multi-faceted (pre-)processing uncertainties to comprise value uncertainty.

Different processing algorithms will contribute different dimensions of value uncertainty and must be visualized accordingly. But also, user's tasks need to be appropriately reflected in the visualization, hence it is necessary to make the uncertainty representation configurable to allow focusing on particular types. Consequently the interactive explorations need to provide an interface that allows such customization.

In summary the categorization of uncertainty types involved in the segmenting and labeling process allows to identify and describe the uncertainties relevant in each of the stages of the process. For future work this will help to use the categories and specify different designs for representing the different types of uncertainty. In the end this should provide a guideline for choosing appropriate visual representations in an interactive exploration environment for segmenting and labeling time series data.

Acknowledgements This work was supported by the Austrian Science Fund (FWF), Project No. I 2850-N31, Lead Agency Procedure (DACH) "Visual Segmentation and Labeling of Multivariate Time Series (VISSECT)".

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), Eurographics. in-press. 1

[GSB*15] GSCHWANDTNER T., SCHUMAN H., BERNARD J., MAY T., BÖGL M., MIKSCH S., KOHLHAMMER J., RÖHLIG M., ALSALLAKH B.: Enhancing time series segmentation and labeling through the knowledge generation model. In *Poster Proceedings of EuroVis Conference* (2015), Eurographics. Poster. 1

[SSS*14] SACHA D., STOFFEL A., STOFFEL F., KWON B. C., ELLIS G., KEIM D.: Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1604–1613. doi:10.1109/TVCG.2014.2346481. 1