Visual Support for Rastering of Unequally Spaced Time Series

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

[a] *Vienna University of Technology, Austria*

Cleansing and wrangling [2, 1] – preprocessing data and transforming it into a usable form – constitutes an important step for subsequent analysis. In many application domains, e.g., environmental sensor measurements, datasets are created with varying interval lengths. Specifically with time series data, established analysis methods require the data to be structured, e.g., being equally spaced. By rastering time series, unevenly distributed time points and their corresponding values are aggregated and binned into evenly spaced time intervals, while still retaining the original data's structure. Rastering the original data alters it to (1) trade consistent value distribution for accuracy of the original values, (2) achieve more accurate value representation by smoothing measurement inaccuracies, and (3) reduce data size by lowering the time series resolution. Users require knowledge about the data domain and temporal aspects of the data to generate an adequately transformed time series usable for subsequent analysis. Rastering introduces uncertainty, which users are predominantly not made aware of in further analysis.
We propose a Visual Analytics (VA) approach to effectively support users during the analysis and validation of time series raster parametrizations. VA intertwines interactive visualization, analytical methods, perception and cognition to ease the information discovery process. Our conceptualized VA framework allows users to transform unequally spaced time series data into equally spaced rasters. It facilitates finding appropriate parametrizations and analyzing the rastering outcome. By providing quality measures and uncertainty information, users receive contextual knowledge on quality issues occurring during rastering to assess the outcome of the time series rastering. For different time series characteristics it is necessary to adapt the rastering algorithm and feedback information accordingly. We provide considerations for handling special use cases and domain specific properties and suggest well-fitting measures to deal with intricacies in the data.
Keywords: Time Series Preprocessing, Visualization, Visual Analytics.

References

[1] T. Gschwandtner, W. Aigner, S. Miksch, J. Gärtner, S. Kriglstein, M. Pohl, and N. Suchy. TimeCleanser: A Visual Analytics Approach for Data Cleansing of Time-oriented Data. In *Proc. of the 14th International Conf. on Knowledge Technologies and Data-driven Business*, i-KNOW '14, pages 18:1–18:8, New York, NY, USA, 2014. ACM.

[2] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. v. Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011.