Haßler, Marc; Kohlschein, Christian; Meisen, Tobias
Similarity analysis of time interval data sets – a graph theory approach. (English)


Summary: Comparison of entities, i.e., the measurement of their similarity, is a frequent, but challenging task in computer science. It requires a precise and quantifiable definition of similarity itself. Are two texts equal, if they overlap in a majority of their composing words? Does a pair of pictures resemble the same content? What defines the sameness of two songs? While certain distance-based approaches, e.g., Minkowski, make for a good starting point in defining similarity, there is no one-size-fits-all approach. In this work, we tackle a particularly interesting problem, namely, the definition of a similarity measure for comparing time interval data sets. Our approach regards the data sets as disjoint parts of a bigraph, thereby allowing for an application of methods from graph theory. We present both a formal definition of the similarity of two time intervals and our methods as well as concrete use-case from the medical domain, thus demonstrating the applicability for real-world scenarios.

For the entire collection see [Zbl 1411.62018].

MSC:
62M10 Time series, auto-correlation, regression, etc. in statistics (GARCH)
62P10 Applications of statistics to biology and medical sciences; meta analysis
05C90 Applications of graph theory

Keywords:
graph theory; time interval data set; similarity analysis; medical data analysis; distance measures

Full Text: DOI