外 国 語 要 旨

Time-series Visualization adopting Overview and Summarization

Advanced Science, Graduate School of Humanities and Sciences, Aki Hayashi

This thesis presents time-series visualization adopting overview and summarization. Recent information systems accumulate variety of time-series data, including system logs such as Web access logs and transaction logs, and life logs such as location logs obtained from social networking services. Music data including MIDI (Musical Instrument Digital Interface) can be also treated as time-series data. Analysis focusing on periodic features of time-series data has great potential for improvements on prediction and contents or situation understanding. They can derive high economic effects such as efficient recommendation and marketing. Information visualization has evolved since it supports human's visual understanding of abstract information by presenting concrete images. There have been also recent emerging studies on “Visual Analytics” which aims at analysis of natural or social phenomenon applying conventional visualization techniques. Overview and summarization generally have important roles on visual analytics, where users firstly overview the tendencies of whole data, then observe the important portions in detail by summarizing the information. However, time series data are so massive that there are few effective techniques realizing overview and summarization. The goal of this thesis is realization of sophisticated visualization techniques adopting overview and summarization for massive time-series data. In particular, the thesis focuses on the installation of automatically extracted indices aiming at improved comprehensibility and selection of fruitful portions for detailed observation while applying summarization mechanisms. The thesis presents three methods for specific applications, including location log visualization, music visualization, and visualization of general multi-variate time-series data.

The first visualization method features general system logs including transactions and access logs. The system log analysis is important for decision making in the various fields such as Web cite operation and marketing. However, these logs are often massive containing a lot of variables (attributes) and huge number of items. Therefore, the method switches highly comprehensive heatmap-based overview that draws the time-varying tendencies of the selected variable. In addition, the method features variable recommendation for selecting better variables such as a variable that includes a lot of abnormal behaviors. Users can adjust time-slices by assigning appropriate variables to the X-axis, to observe important portion of the data in detail. They can also reduce items to be drawn applying sorting and clustering algorithms. The thesis introduces use cases applying real transaction and access logs, and the evaluation with feedbacks of both non-expert users and expert analysts.

The second visualization method features life logs including location logs. Life log analysis is helpful for various applications such as recommendation services. In the case of recommendation systems, it is not always easy to determine the suitable timing and contents for the recommendation to end users. Analysis methods to understand detailed dynamic status of users would be useful to determine the timing and contents. The method focuses on the periodicity of logs based on multiple time-scales and implements “regular behavior measure” to extract the regularity of users: how the action reflects their habits. The method supposes that the service can recommend places irregularly visited by other users, such as sightseeing places during irregular activities such as trips of the target user. The method realizes geometric-map-based overview to observe common locational regularity among multiple users to assist the determination of the recommendation contents. The thesis also proposes a user summarization method which selects the regularity transition of a target user to determine the recommendation timing, using lists of regularity transitions of multiple users. The thesis introduces use cases for real check-in logs. Also, the thesis presents quantitative and qualitative evaluations of the method.

The third visualization method “Colorscore” features more specific MIDI data. Scores of orchestral music are too complicated for ordinary people to read, because they contain too many staves of instrument. Colorscore supports two requirements for composers, arrangers, and players: overview and arrangement based on the role of the phrases. Colorscore realizes the overview summarizing role changes and repetitive constitution considering all parts of entire music. The horizontal summarization that shrinks unimportant bars is proposed for summarization towards bars, and the vertical summarization that reduces the parts is proposed for rearrangement remaining essential phrases. The thesis discusses the effectiveness of Colorscore by use cases applying multiple orchestral music and feedbacks of both novice users and music experts.

The most important advantage of the proposed methods is the implementation of the following mechanisms: automatically extracted indices that improves comprehensibility of visualization results, and selection of the fruitful portions which are worth to observe in detail applying summarization mechanisms. To establish visualization systems for various time-series data, the implementation of such indices considering analysis goals, data characteristics as well as periodic feature in time axis would be effective to design overview and time- and item-based summarization.