Interactive Interpretation of Serial Episodes: Experiments in musical analysis

Béatrice Fuchs, Amélie Cordier

Interactive Interpretation of Serial Episodes:
Experiments in musical analysis

Béatrice Fuchs[1], Amélie Cordier[2]

[1] Université de Lyon, UJML3, IAE, LIRIS, F-69 008, Lyon, France. beatrice.fuchs@liris.cnrs.fr

[2] Université de Lyon, Hoomano, LIRIS, F-69 100, Villeurbanne, France. amelie.cordier@hoomano.com

This is a summary of a paper presented at EKAW-2018 (Fuchs & Cordier, 2018).

The context of this work is the study of sequential data that can be represented with sequences of timestamped events. The aim is to explore these sequences with sequence mining to discover *serial episodes* which are frequent event subsequences that occur frequently in data (Mannila *et al.*, 1997). The domain of melodic analysis is studied in this work : the aim is to highlight the structure of a musical piece by discovering its main melodic patterns. The episodes produced by the miner are examined by a user generally an expert of the domain who have to identify relevant episodes and interpret them. Meanwhile in the interpretation step, the user has to face to a recurrent overabundance of mining's results which makes difficult the identification of interesting ones. There is a real need to adopt a rigorous approach to methodically manage this step and assist the user's work. For this, we propose a visual and interactive approach to assist the interpretation of serial episodes.

An Interactive approach to the interpretation of serial episodes

We propose to assist the interpretation task by managing combinatorial redundancy in order to focus on relevant episodes. The assistance combines iteratively ranking and filtering useless episodes to help focusing on relevant ones. It has been exemplified in the Transmute prototype, a web-based application enabling user's interaction with events sequences and serial episodes that are represented graphically on a timeline with customisable icons.

The interpretation process consists in the main iterative steps : ranking, selection and filtering. The user can choose measures to rank episodes and then select among them to display their occurrences in the sequence. When a choice is made, a filtering process is triggered to clean up other episodes that can no longer be selected following the previous selections of the user. Finally, the user can interpret the episodes by attaching them annotations and record the model resulting from the interpretation into a knowledge base.

The ranking of episodes is performed thanks to several objective interestingness measures which estimate the relative importance and compactness of the episodes in the sequence. The first measure is the event coverage indicator which is the number of distinct events of the occurrences of an episode. The second measure is the spreading indicator which is the number of events of the sequence in the time intervals of the episode occurrences. The noise indicator is the difference between these two previous indicators and corresponds to the number of events of the sequence in the time intervals of the episode occurrences. Temporal measures may also be used when event duration are known.

The selection of an episode by the user triggers the filtering process which is based on the event coverage of the selected episode. The remaining episodes are examined and occurrences having at least an event in common with the event coverage are discarded. The support is consequently updated and episodes whose support becomes less than the given frequency threshold are discarded. This results in removing combinatorial redundancy around the chosen episode and leads to a gradual diminution of the remaining episodes, allowing to the user a better focus on other episodes.

Experiments

The experiment aims to verify the ability of the approach to improve the ranking of episodes and as a consequence, to a lower effort from the expert. For this, three musical pieces have been chosen and for each of them, an expert gave the relevant episodes to find, which we name the expert episodes. The miner was launched with parameters to ensure the presence of the expert episodes and the mined episodes were ranked using interestingness measures. The ranks of the expert episodes in the mining results are used to assess the hypothetical effort of the expert. The smaller the rank, the less important is the effort of the user to find the expert episodes. Two situations are compared in two tables : without filtering and with filtering. Without filtering, the effort is the highest rank of expert episodes. With filtering, the effort is the sum of the lowest ranks of expert episodes, since the examination is resumed at the beginning after each filtering and ranking, each time an expert episode in found. Experiments show an important diminution (> 80%) of the effort for the three pieces with the ranking-filtering strategy. A counterpart is that some expert episodes may be discarded, and the recall measure may be quite bad in some cases.

Discussion, limits

The combination of both filtering and ranking is conclusive but may sometimes lead to a lower recall. A first experiment has been conducted to test the usability of the Transmute prototype by users, but a full scale experiment with more complex pieces and domain experts remains to be conducted. This approach is suitable only if data lend to compression. Moreover, the Transmute prototype can not handle more than several thousands of results.

Related works

The sur-abundance of results in data mining is a major issue for a long time. Among related works we can mention the *compactness* that measures gaps in episode occurrences (Tatti, 2014), pattern selection based on their ability to compress data (MDL) : (Rissanen, 1978; Vreeken *et al.*, 2011) for itemsets and (Lam *et al.*, 2014) for sequences and finally *human in the loop* approaches (Bertini & Lalanne, 2009). More recent approaches claim that interestingness is subjective in essence and take into account the goals of the user and its knowledge (van Leeuwen, 2014) but out approach is quite different in the sense that the user has a more active role in the process.

Références

BERTINI E. & LALANNE D. (2009). Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery : Integrating Automated Analysis with Interactive Exploration*, p. 12–20 : ACM.

FUCHS B. & CORDIER A. (2018). Interactive interpretation of serial episodes : experiments in musical analysis. In C. FARON-ZUCKER & C. GHIDINI, Eds., *Knowledge Engineering and Knowledge Management, 21st International Conference - EKAW-2018*, LNAI 11 313, p. 131–146, Nancy, France : Springer.

LAM H. T., MÖRCHEN F., FRADKIN D. & CALDERS T. (2014). Mining compressing sequential patterns. *Statistical Analysis and Data Mining*, **7**(1), 34–52.

MANNILA H., TOIVONEN H. & VERKAMO A. I. (1997). Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, **1**, 259–289.

RISSANEN J. (1978). Modeling by shortest data description. *Automatica*, **14**(5), 465 – 471.

TATTI N. (2014). Discovering episodes with compact minimal windows. *Data Min. Knowl. Discov.*, **28**(4), 1046–1077.

VAN LEEUWEN M. (2014). Interactive data exploration using pattern mining. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, p. 169–182. Springer.

VREEKEN J., LEEUWEN M. & SIEBES A. (2011). Krimp : mining itemsets that compress. *Data Mining and Knowledge Discovery*, **23**(1), 169–214.