Temporal Pattern Mining in Dynamic Environments

Andreas D. Lattner

Dynamic scenes with many different objects and interrelations changing over time demand complex representations. The identification of frequent patterns and prediction rules in such scenes would be very valuable as associations in the data could be discovered or a system's performance could even be improved by utilizing the new information in the behavior decision process.

1 Introduction

Many domains feature a dynamic characteristic and it would be useful to learn temporal patterns, e.g., in logistics, sports, and medicine. In the logistics domain, for instance, there might be many different kinds of objects like different transport vehicles (e.g., trucks, ships, or planes), different actors or organizations (e.g., storages, transport companies, manufacturers), highways or tracks, reloading points, etc. Different events can occur like traffic jams, weather events, break down of a transport vehicle, or delay of some goods. It would be valuable to identify repeating patterns that lead to certain situations in order to predict a traffic jam or a delay and initiate some counter-actions in order to avoid financial loss or penalty payments. Such a pattern could be, for instance: If the traffic density is medium and increasing on highway X on a Friday afternoon and the weather is rainy, it is likely that there will follow a traffic jam on highway Y.

In the requirements analysis, different demands on representation formalisms for dynamic scenes as well as for patterns to be mined from dynamic scenes have been identified from a soccer scenario of the RoboCup simulation league. Among other demands the representation of objects and relations and the temporal validity of these relations have been detected to be crucial. As various actions and events in dynamic scenes can occur concurrently, it is important that the representation also supports this concurrency. A comprehensive investigation of the state of the art has led to a number of relevant approaches covering parts of the requirements. Particularly of interest were the approaches dealing with association rule mining, especially extensions for sequential, temporal, or relational data. *WARMR* [3] can be used to mine relational association rules but has no direct means to represent the temporal dimension. The work of Höppner [5] which addresses rule learning from interval-based temporal data uses interval relations in order to represent temporal interrelations among states in the input data. A third relevant approach by Lee [6] allows for mining first-order sequential rules but does not support concurrently occurring intervals of relations.

2 Mining Temporal Patterns

Based on the defined requirements and on the analyzed approaches, the temporal pattern mining approach *MiTemP* has been developed. It can mine temporal patterns from time interval-based relational data and additional conceptual information about objects and their interrelations in dynamic scenes. The relevant concepts of the approach have been defined formally. Besides formal definitions of dynamic scenes and their schemata as well as definitions for patterns and prediction rules, a partial order has been defined for the generalization relation between patterns. Following ideas of Lee [6], an optimal refinement operator has been defined that guarantees a complete and non-redundant generation of frequent patterns from the given representation. In the conceptual chapter of this thesis, it has also been shown how *WARMR* can be used in order to mine temporal patterns albeit it generates many redundant ones.

The pattern mining approach is based on the association rule mining algorithm *Apriori* [1]. It starts with the most general (empty) pattern and successively performs refinement operations to those patterns that still exceed the minimal frequency threshold. Five refinement operations have been set up: lengthening, temporal refinement, unification, concept refinement, and instantiation. Concept restrictions can represent the information that specific variables can only be bound to instances of certain concepts in the concept hierarchy.

In order to represent temporal relations in patterns, a new concise set of mutual exclusive and jointly exhaustive interval relations has been set up by combining ideas from Allen's and Freksa's interval relations [2, 4]. For the interval relations (before/after, older/younger & contemporary, head-to-head), a composition table and an temporal reasoning algorithm has been set up. The reason for the new set of interval relations was to reduce complexity and to focus on important relations for prediction rules. However, the set of interval relations can be easily replaced without changing the mining algorithm, e.g., by using Allen's interval relations [2].

It can easily happen that a huge number of patterns is generated. Therefore, different means to restrict the relevant pattern space have been introduced. It is possible to disable single refinement types (e.g., no instantiation) or to restrict the maximal refinement level. If it is known before mining that only certain patterns with some predicates are of interest, a bias can be set up consisting of partial conjunctive patterns. If such a bias is defined, the patterns that are inconsistent with this bias are filtered out during pattern mining. Another way to reduce the number of patterns is a selection of patterns to be refined at each refinement level. In the