Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data
by I. Batal, D. Fradkin, J. Harrison, F. Moerchen, M. Hauskrecht

1 Summary
This paper introduces a framework that mines temporal patterns in complex multivariate time series data. Since multivariate temporal time series could be noisy and inaccurately reported, first they are converted into time-interval sequences. Such sequences are then used to build temporal patterns backwards in time using temporal operators, which describe relation between two sequences (before and co-occur). Authors are focused on recent temporal patterns (RTPs), i.e., patterns that can be identified using recent data. RTPs are useful for monitoring and solving event detection problems.

Experimental evaluations have been performed on electronic health record data (EHRs), and confirmed efficiency of the newly introduced approach. Precision and recall of generated patterns is computed by using medical diagnosis guidelines to verify patterns.

2 Strong Points
- All definitions in the paper, are explained with simple and easy to understand examples.
- Efficiency (number of generated patterns) of the proposed framework is proved formally, and compared with a naive approach.
- Authors use a few different pruning techniques to further limit the number of generated candidate patterns. Thanks to this the framework generates an order of magnitude less patterns with high precision and recall, comparing to the naive approach.
- A dataset for experiments have been carefully prepared. All experiments and their results are explained in details.

3 Weak Points
- The first step of processing time series points is to abstract them into time intervals. However, authors defined only the way of computing the value of such intervals, but neither their length, nor start time, nor end time. It is not clear how end points of time-intervals have been chosen. Have they been taken in the middle of time distance between reported time points? Based on interpolation?
- The goal of the problem defined in this paper is to find patterns. However, authors did not specify explicitly ways of evaluating patterns. The intuition is to get patterns with as many time series variables as possible. But is it correct intuition? Does it work?
- Although the paper is well written, reading it with understanding requires some effort.

4 Questions and Discussion Points
More Temporal Relations. One of the newly defined temporal relations between time intervals in the paper is co-occur, which covers many relations defined by Allen, including overlaps and contains (F. Allen. Towards a general theory of action and time. Artificial Intelligence, 23:123-154, 1984.). Why authors merged overlaps and contains? These two relations are substantially different. Even using g, which model max. noise of reported timestamps, these two relations can be easily identified in the time interval series. Authors should consider both relations overlaps and contains independently or at least provide some explanations, why is it not important.

Patterns Evaluation. All generated patterns are treated independently, and they are not compared. It would be interesting to order them, based on different criteria, e.g., number of used time series variables. Another idea is to find similar patterns, i.e., define a similarity measure for patterns. Having such measure one can cluster patterns into superpatterns, which could be used to generate more patterns or to reduce the current ones. What would be a correlation measure to compare two temporal patterns?