Mining Temporal Relationships with Multiple Granularities in Time Sequences

Claudio Bettini, X. Sean Wang and Sushil Jajodia

Presented by

Qasim Iqbal

EE380L Data Mining

- Introduction
- Classical techniques in time series
- Definition of the problem
- Definitions
- Approach used by the authors
- Results
- Conclusions

Introduction

- Huge amount of data is collected everyday in the form of event time sequences
  - Stock shares
  - computer access via a terminal
  - bank transactions
  - events related to malfunctions in an industrial plant

Classical approaches to “knowledge extraction” in Time Sequences

- Prediction
- Forecasting
- Standard prediction approach involves constructing an underlying model which gives rise to the observed sequences
- Takens and Packard have provided basis for a diffeomorphism (one-to-one differential mapping) between a finite window of time series; a non-linear autoregression of the form

$$y(k) = g[y(k-1), y(k-2),..., Y(k-T)]$$

- Irie, Miyake, Honnik, Stinchcombe, White and others have shown that a feedforward neural network with an arbitrary number of neurons is capable of approximating any uniformly continuous function

Problem: Symbolic Vs. Numeric

- At the lowest level a time series can be interpreted purely as numbers
  - After all, normally data is collected as discrete samples of an underlying continuous waveform
- However, at the higher level the same time series can interpreted differently, as a symbolic tag, or perhaps a semantic model
  - A stock value going down from, say, $30 to $20 can be assigned a tag: Stock_fall.
  - It is possible to find interesting (temporal) relationships dealing only with symbolic tags
  - Neural networks (and other statistical pattern recognition techniques) work purely with numbers

Framework

- FIR Network Model
  - for stationary models
- Kalman Filter
  - Non-stationary models

Problem definition

- A lot of research followed that examined *sequences of events* for knowledge extraction leading to find similarity or interesting patterns in the event sequences
- However, an important aspect for such a discovery process was largely missing from the literature: discovering temporal patterns or relationships that involve multiple granularities
- Events occurring in the same day, or week or months etc. are important aspects of a data mining paradigm

Some definitions

- An event is a pair $e = (E, t)$ where E is an event type and t is a positive integer called the *timestamp* of e
- An event sequence, σ, is a finite set of events
- A granularity is a mapping μ from the set of positive integers (used as index) to subsets of the time domain such that for all positive integers i and j with i < j
  - $\mu(i) \neq \Phi \wedge \mu(j) \neq \Phi$ implies that each number in $\mu(i)$ is less than all numbers in $\mu(j)$
  - $\mu(i) = \Phi$ implies that $\mu(j) = \Phi$

Example

Temporal constraints with Granularities

Even structure with multiple granularities

The Discovery Problem

An Event structure

Finding frequent complex event types

Experimental event structure

- E_1 occurs after E_0, but within the same or the next two business days
- E_2 occurred the next business day of E_1, or the business day after
- E_3 occurred after E_2, but in the same business week of E_2

Results

Problems and overcoming

- The algorithm executes in polynomial time
- To overcome this the authors
  - eliminate inconsistent event structures
  - reduce the event sequence
  - reduce the occurrences of the reference event type to be considered,
  - scan the event sequence for each candidate complex event type, to find out if the frequency is greater than the minimum confidence value

Conclusions

The authors

- introduced and studied the notion of temporal constraints with granularities and event structures
- presented a time automaton with granularities for finding event sequences that match event structures
- defined event-discovery problems and provided a practical procedure that exploits the properties of granularities and event structures