Mining Segment Wise Periodic Patterns in Time-Related Databases

Jiawei Han, Wan Gong, Yiwen Yin
SFU
Presented by Andrew Foss
UofA

Only some things may be periodic, and then not every time

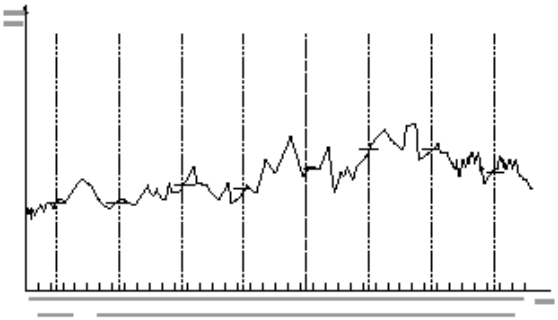

Figure 1: A segment-wise periodic pattern: This stock goes up every Wednesday.

Look for events that occur periodically above a certain confidence level

Example of a periodic sequence:

132113412341

Starting at offset 1, 3 repeats every 4th digit - *3**

If it occurs 3 times out of 4 it satisfies a 75% confidence

Use Data Cubes:

a reference cube and a working cube

Collect a set of objects based on the mining query

Each is associated with a time series

Digitise the Attribute Dimension

Every slice except 'All' becomes a bit array.
The time dimension becomes a time-index and a period-index.

A working cube consists in general of a time-index, a period-index, a time-related attribute and one or more non-time-related attributes.

A T-slice consists of the complete time plane and the entire domain of the time-related attribute dimension.

These bit-arrays are mined.

The process

1. Find one-cycle patterns with confidence >= conf_min
2. Generate candidate I-cycle patterns
3. Check pattern existence

Generating Candidate Sets

Apriori

If a k=cycle is frequent all j-cycles in it with j < k must be frequent

Join to create a candidate set and ***Prune*** those i-cycle patterns that contain (i-1)-cycle sub-patterns that are not in the i-1 candidate set.

Testing candidate Sets

- Do a bit-and
- Count the pairs
- Check the threshold

Viz. The 3-candidate set is {3, 3, {{0,1}, {1,1}, {2,2}}}

agrees @ ?%

Performance and Benefits

More efficient than

- Apriori w/o exploring the bit-array cube
- Using a cube but not exploring apriori

Can be used on any dimension that could contain cycles

Drilling can seek cycles at different levels of granularity - *if a pattern is cyclic at a finer scale it must be cyclic with the same period at a rougher scale*

Future Work

Extend to arbitrary time periods

(unexpected - like 6 weeks)

Reduce the search space for confidence < 100% where the known rules (such as multiples of the periodicity) don't apply