**Mining Precise-positioning Episode Rules from Event Sequences**

**Xiang Ao¹, Ping Luo¹, Jin Wang², Fuzhen Zhuang¹, Qing He¹**

Institute of Computing Technology, CAS, China¹ University of California at Los Angeles, USA²

---

**MOTIVATION**

**Traditional Episode Rule**

Given a frequent episode \( \alpha \), a traditional episode rule in the form of \( lhs \rightarrow rhs \) is generated straightforwardly: The antecedent \( lhs \) is the prefix of \( \alpha \) and the consequent \( rhs \) is the last event in \( \alpha \), if its confidence is larger than a user-specified threshold.

![Image of traditional episode rule example](image)

From Fig.1, \( <D,A> \rightarrow <B> \) is a traditional episode rule which indicates it is within 2 time intervals after the occurrence of \( <D,A> \) that \( B \) will occur (with 100% confidence).

---

**Limitation of Traditional Episode Rule**

**Example:** In stock investment application, we can map price change ratios to events and use candlestick charts to represent events. Red bars denote price increase of a stock, and green bars denote prices decrease.

- The episode rule \( <D,A> \rightarrow <B> \) predicts correct in the following two cases, however we will lose money in Case 2 if we long the stock after we observed the antecedent of the rule.

![Image of traditional episode rule limitation](image)

**Precise-positioning Episode Rule (PER)**

We define precise-positioning episode rule in the form of:

\[
\Gamma = \alpha \xrightarrow{\Delta t} \beta
\]

- \( \alpha \): a traditional episode, as the antecedent;
- \( \beta \): a fixed-gap episode, as the consequent;
- \( \Delta t \): the time constraint between the antecedent and the consequent.

![Image of precise-positioning episode rule](image)

---

**Mining ALGORITHM & EFFICIENCY**

1. **MIP-ENUM Algorithm**

The basic idea of MIP-ENUM is to enumerate PER candidates by concatenating discovered traditional episode with fixed-gap episode and subsequently filter the invalid ones according their confidence values.

![Image of MIP-ENUM algorithm](image)

2. **MIP-TRIE Algorithm**

Data structure: PER-trie stores valid PER compactly.

Algorithm: MIP-TRIE(DFS) and MIP-TRIE(PRU).

- We use PER-trie to store all valid PER given an antecedent \( \alpha \) and propose two algorithms to build complete PER-trie.
  - MIP-TRIE(DFS) expands the PER-trie by a recursively depth first search manner.
  - MIP-TRIE(PRU) adopts an improved traverse strategy with pruning technique.

![Image of MIP-TRIE algorithm](image)

**Efficiency Comparisons**

Dataset: Retail -- http://fimi.cs.helsinki.fi/data/

Observations: 1. MIP-TRIE(PRU) outperforms MIP-TRIE(DFS) and MIP-ENUM algorithm; 2. MIP-TRIE algorithms significantly outperform MIP-ENUM.

---

**EFFECTIVENESS of PER**

**DATASET:** 150 related industry sector pairs of China stock market from Jan. 1, 2010 to Aug. 29, 2014.

**EVT SEQ. CONSTRUCTION:** UP (if the price increases) and DN (otherwise) for each industry sector.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-UP</td>
<td>B-DN</td>
<td>B-DN</td>
<td>B-DN</td>
<td>B-UP</td>
<td>B-UP</td>
<td>B-UP</td>
<td>B-DN</td>
<td>B-UP</td>
<td>B-UP</td>
</tr>
</tbody>
</table>

Fig.3 The example stock industry sector event sequence.

- A and B denote stock industry sectors.

**SETTINGS:** We use first 4-year sequence as the training set to mine PER on each sequence and degrade PER whose \( \Delta t = 5 \) to traditional episode rule (denoted as TDR), then test prediction ability of them on the rest.

**COMPARISON:** For PER, we trade strictly according to the rule; for TDR, we trade after antecedent occurs and close out either consequent appears or the maximal occurrence window for consequent reaches.

**MEASURE:** We close out when the float loss exceeds a stop-loss threshold during the holdings by TDR. We compute the return of holdings and visualize the winning rate of PER under different stop-loss thresholds.

**VENUE & CONTACT INFORMATION**

The 33rd IEEE International Conference on Data Engineering, San Diego, California, USA, April 19-22, 2017.

Email: [ao.xiang, luop]@ict.ac.cn, jinwang@cs.ucla.edu, [zhuangfz, heq]@ics.ict.ac.cn

Homepage of MLDM Group, ICT, CAS: http://mldm.ict.ac.cn

Xiang’s personal homepage: http://mldm.ict.ac.cn/MLDM/~aox

Public account on Wechat: