Stock Pattern Search & Data Mining

i) Development of a Time Series Database
ii) Time Series Streaming in Mobile Environment
iii) Time Series Data Mining API

Supervisor: Dr. Korris Fu-lai Chung

Department of Computing
Hong Kong Polytechnic University
Focus of this Project

- **Stock Information Retrieval (Time Series Search Engine)**
  - Given a technical pattern, find its occurrences in a selected stock time series.

- **Stock Data Mining (Investment Knowledge Discovery)**
  - Given a collection of stock time series, mine the association of patterns
A website has been set up: http://www.tsdm.org
Problems being addressed

- Time series segmentation
- Time series clustering
- Time series pattern matching
- Time series pattern representation
- Intra-stock data mining
- Inter-stock data mining
- Time series visualization in mobile devices
- Time series data stream mining
Why Time Series Segmentation?

- Time Series is difficult to process, analyze, query and mine
- Better to transform it into meaningful symbols
- Based on the time series segments or patterns, one may conduct queries like:
  - When did Microsoft’s stock price exhibit a head-and-shoulder pattern?
    ⇒ a stock time series query problem
  - Are there any other stocks also exhibiting a head-and-shoulder pattern nearby temporally?
    ⇒ a stock data mining problem
Mining Time-Series and Sequence Data

- Time-series database
  - Consists of sequences of values or events changing with time
  - Data is recorded at regular intervals
  - Characteristic time-series components
    - Trend, cycle, seasonal, irregular

- Applications
  - Financial: stock price, inflation
  - Biomedical: blood pressure, EEG, etc.
  - Meteorological: weather
Mining Time-Series and Sequence Data

Time-series plot

Price History - International Business Machines Corp... (3/17/97 - 3/20/98)

- IBM
- MSFT
- INTC

A customer

An item

A customer

An item

A customer
Stock Data Mining

- There exist at least two types of stock data mining
  - Intra-stock data mining
  - Inter-stock data mining

- Intra-stock data mining (as elaborated)
  - Each stock is treated as a customer (sequential association analysis) and
  - Each price movement is treated as a transaction/item
  - i.e. MSFT: Go-up, Go-down, Go-up, Go-up, ..., etc.
  - Hence, frequent sequence like “Go-up, Go-up, and then Go-down” can be mined.
Stock Data Mining (cont.)

- Inter-stock data mining
  - Each time window is treated as a transaction
  - Using non-sequential association rules
  - The behavior of different stocks within a time window will form the list of items
  - One can use the candlestick method to characterize the behavior, e.g. open-high-close-low (OHCL), open-low-close-low (OLCL), etc.
  - We will have:
    
    MSFT  IBM  INTC
    TID#1: OHCL OLCL OLCH
    TID#2: OLCL OHCH OHCL
Inter-stock data mining
Stock 0008 with pattern Head and Shoulder appears from 1997 to 1999 -- Best Match 93.38%