Sequential Pattern Mining in ebanking from Contiguous Code Subsequence

Name: Yip Chi Kin
Date: 3-12-2005

• Motivation

Response from eBanking

- Identifying internal and external consumer base and their needs
- Diversify into innovative consumer credit products
- Market domination through innovative schemes and promotional activities
- Striving for an early market delivery and capture
- Brand recognition, product awareness and image building
- Building alternate and cost effective distribution channels
- Expand consumer credit footprint

Benefits of eBanking

- Data Mining for Marketing
- years of data available to customers and banks
- Retain and attract profitable customers
- Customized design for customer base and bank image

• Review of Literacy

Vision and Values

To ensure delivery of desired banking services to customers in cost effective way, with convenience, security and protection from any unauthorized access. The "value triad" proposes a "value proposition" that considers the ***customer*** (relationship), ***product*** (information), and ***application*** (automation). Changing any one of the triad's components changes the value proposition.

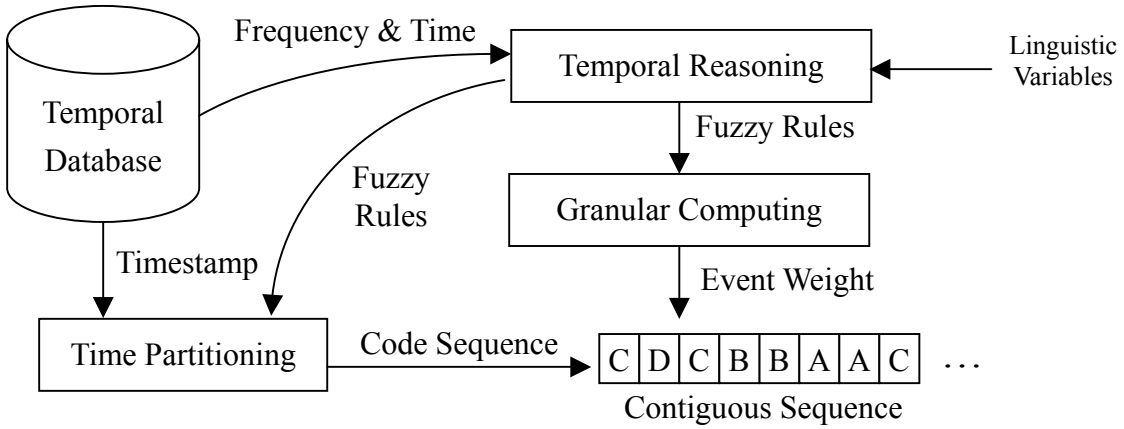
System Features

- FAT Server Technology
- Powerful SQL Database
- Secure Web Server
- User-friendly expert website design

• Methodology

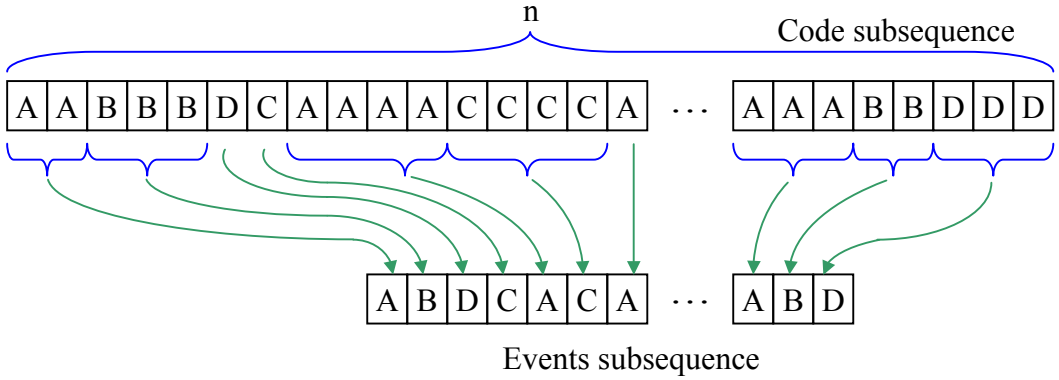
The information Granulation theory is applied to discrete web-log session, and a **Subsequence Miner** uses different algorithms in order to find out optimal *substring*, *superstring*, *subsquence* and *supersequence* of web-based temporal data. The dataset included customers' profiles.

Strategy of discretization temporal dataset



e.g. Time interval = 1 second , weblog duration = one year ,
Code length of Personalization Profile = n = 365×24×60×60 = 31536000

Contiguous Sequence = Artificial DNA



Temporal Data Mining Aspects

- Prediction
- Correlation
- Regression
- Benchmarking
- Periodic Pattern Mining
- Temporal Association Finding
- Causality Analysis
- Sequential Event Patterns
- Threshold selection
- Frequency Analysis
- Anomaly Detection
- Clustering and Classification

• Subsequence Matching

The steps of mining techniques as below:

Temporal data → Fuzzy rules → Codes with weight → Sequence matching → Event patterns of eCRM (or browsing behavior)

Some matching algorithms could be applied same as bioinformatics:

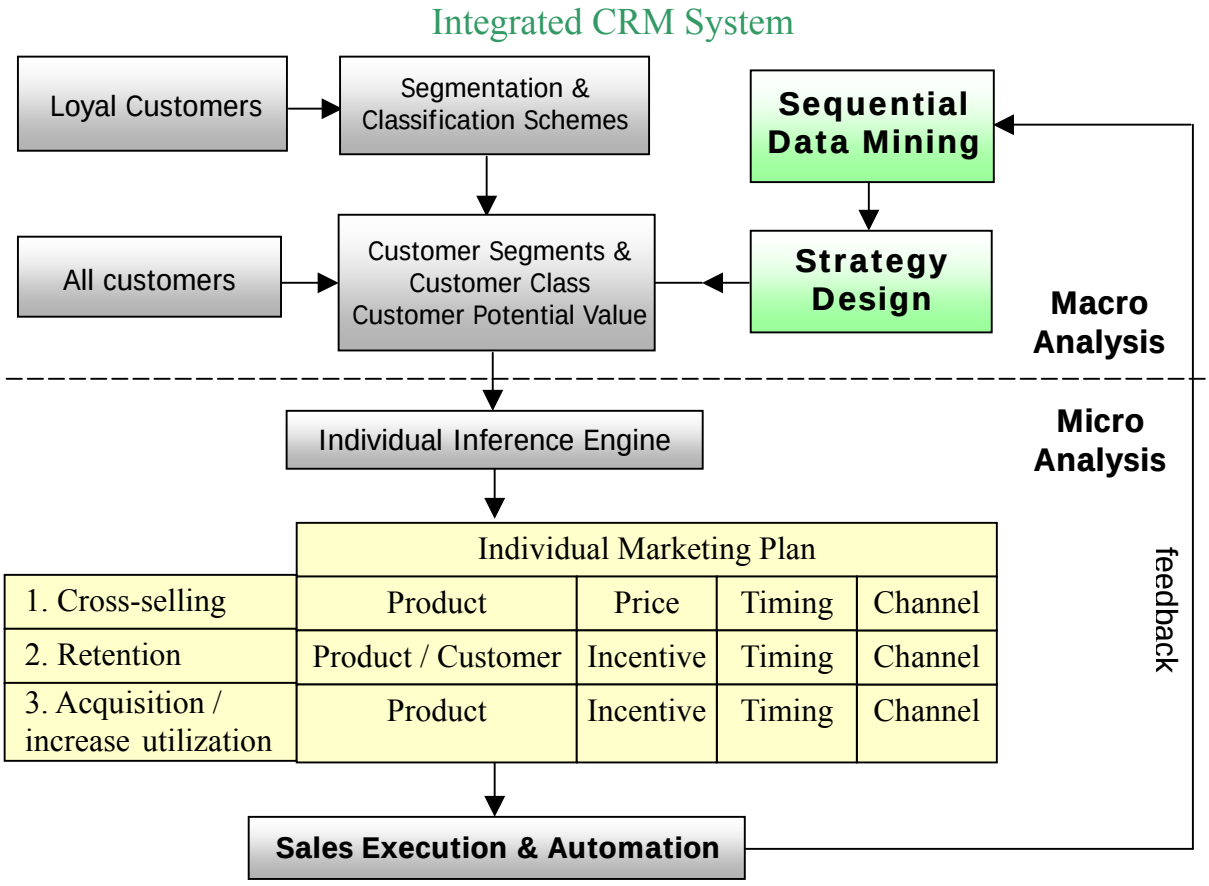
- Mapping
- Similarity search
- Subsequence comparison
- Multiple alignment

According to above techniques, subsequence codes could be ***classification***, ***clustering***, ***prediction*** and ***association rules*** of data mining tasks in order to obtain **marketing strategies**.

• Electronic CRM

How do you integrate marketing, sales, and support activities?

By having good electronic customer relationship management (eCRM). CRM's goal is to create a synergy among sales, marketing, and customer-service activities within an organization in order to obtain and retain customers. The followings are shown the main applications of eBanking.



According to the above road map of CRM, there are some mining areas as following:

- Item Selection for Marketing --- What are the most profitable items?

  Item ranking with cross-selling effect by Size-constrained selection and Cost-constrained selection
- Cross-selling or Up-selling with Collaborative Filtering

  --- Finds users whose taste is similar to you and uses them to make recommendations.

  Statistical Collaborative Filters, Probabilistic Collaborative Filters, Bayesian Filters, Association Rules
- Maximal Profit Item Selection with Cross-Selling Considerations (Association-Rule Based)
- Technology fundamental are Selling by Information, Selling by Relationship and Selling by Automation. (eCRM of eBanking solutions)

• Macro Analysis

Classification

- by Loyal Customer *e.g.*

  Have Payroll account
  Have Credit Card
  Have either Loan/Deposit account
  At least 10 banking transactions in the last 2 months
  At least one year tenure
- by Potential Customers (Job, Address, Lift-stage, Credit Card Merchant, … etc)
- by Interesting Products (Insurance, Loan, Mortgage, UT, … etc)
- by Synchronous Actions (Cross-selling, Up-selling, … etc)
- by Navigation Patterns (Page sequence, Browse Duration, Revisit page, … etc)

Grouping of classifications could be applied in *banner advertising*, *hypertext* and *hyperlink implementation*, *eBanking performance*, … etc.

Segmentations

- By occupation
- Privy scheme

e.g.

	Single	Married	Married with kids	Retired
Public				
Private				
Luxury housing				

- By shareholders variables

e.g.

		Profitability		
		Low	Medium	High
Tenure	New			
	1 – 3			
	3 +			

Actionable marketing strategies

- *Strategic focus on values*

		Future Value: Low	Future Value: High
Pass Value	High	Let go/stay	Retention / Loyalty Program
	Low	De-marketing	Aggressive selling

- *Strategic focus on class and relationship*

Customer Class	Good Relationship	Bad Relationship
Good Potential	Retention	Cross-selling
Average Potential	Cross-selling	Increase utilization
Poor Potential	Minimum Marketing effort	Demarketing

Classifications and Segmentations could be identified by weight values, in order to give scores or benchmarks for marketing strategy design.

• Micro Analysis

Cross-selling strategies

- Product upgrade rules *e.g.* Deposit ⇒ Card ⇒ Loan ⇒ Insurance ⇒ UT (Class)
  Classic ⇒ Gold ⇒ Platinum (Card)
- Product bundle rules *e.g.* Deposit ⇒ UT
  Mortgage ⇒ Home Insurance
- Event driven rules *e.g.* Newly wedded ⇒ Gold card ∧ marriage loan
  Change in address ⇒ Selling a property ⇒ Lump sum UT
- Exclusion rules *e.g.* If "Loans" then NO "Deposit"
- Product prioritization rules *e.g.* If "good relationship" then "prioritize product by margin"

Retention strategies

1. Qualification for retention

e.g. A customer is qualified if relationship score ≥ 2

2. Product retention rules

Product	**Possible retention actions**		**Possible events to trigger retention**	**Minimum attribution probability for retention action**
	Non-Bargain hunter	**Bargain hunter**		
Card	Waive annual fee	20% discounts on air ticket	Drop in spending complaint on services	0.7
Mortgage	Cash rebate + (p–1)	p–2 for 10 years	Inquiry	0.6

3. Customer retention rules

e.g. **IF** potential customer ∧ no mortgage **THEN** up-selling UT products

Acquisition and Utilization

1. Qualification

e.g.

Product	Minimum utilization to trigger utilization program
Card	Spending/Limit ≤ 0.05
OD	Lending/Limit ≤ 0.1

2. Family Bundle Rules

e.g. **IF** total deposits from all family members > 0.5 million **THEN** Deposit rate = S + 1

IF Total OD facilities from all family members > 1 million **THEN** Lending rate = P

3. Product Bundle Rules

e.g. **IF** card spending > 10K per month **THEN** High rate on deposit

IF card spending > 10K per month **THEN** Monthly discount on insurance saving plan

The above examples could be obtained by data mining (sequential pattern Mining)

• Result Applications

Customer class clustering

- Banner advertising
- Special offers or discount
- Banking services (personal or business)

Cross-selling or Up-selling

- Electronic customer relationship management (eCRM)
- Marketing Strategy Design
- Item Selection for Marketing
- Maximal Profit Item Selection
- Collaborative Filtering

Personalization

- Classify individuals profiles or demographics (Rules-based personalization)
- Recommending information for similar customers (collaborative filtering)
- Determine interestingness (usage-based personalization)
- Predict future needs (Inference personalization)

Sequential Pattern Mining

- Frequent Navigation Subsequence
- Potential Navigation Patterns
- Sequential Event Patterns (Product-oriented events)

• Conclusion

The Finance Industry in this age faces a large number of challenges. These challenges include need for improved customer service, reduction of call center costs, increased productivity, accelerated access to information and decreased operational costs, to name a few. After overcoming these obstacles, the industry is faced with an even bigger dilemma of attaining these goals without compromising the security. Not only must financial services firms protect information; they must document their security practices to demonstrate compliance with regulatory demands.

In this demanding scenario, technology plays a vital role in giving you an edge over competitors. Use of agile information systems that are flexible, highly secure, integrated and fast in implementation is the key to success. Such systems allow financial services firms to perform risk assessments and actively monitor, validate and secure Web application requests before they reach back-end infrastructure and data.

Most Finance related firms are transitioning from legacy systems to web based applications in order to provide the fastest and most secure services to their clients. Internet based technologies enhance their business and increase revenues for e.g. **online banking**. Emulating their success, smaller firms are also venturing into the web domain. Many companies are using customized web applications to proactively **secure customer and corporate information** and also to expedite transactions and calculations. **Several Banking software, internet Banking applications, financial software, mortgage software, data mining application and data warehousing applications** are available today to cater to the needs of the industry.

Customized web based applications are a tailor made solution for information flow in a company. Critical areas in the company like **customer relationship management, data storage/mining, data security, knowledge management, sales and marketing, employee management, compliance etc**. are some of the key beneficiaries of the customized web based solutions by data mining of eBanking.

eCRM is deliver the right product, ***faster than the competitors***, to the right customer with the right price at the right time and the right channel (eBanking).	
Before mining is **Product Based Selling**	After mining is **Customer Based Selling**
Selling a good product by • Advertising • Personal Selling	Selling a good product by • Information • Relationship • Automation

• Appendix

eBanking links

HSBC eBanking http://www.hsbc.com.hk
Sequential pattern mining http://www.cs.hku.hk/pakdd01/page-tutorial.htm#tutorial_3
Association rules model http://library.witpress.com/pdfs/abstracts/DATA04/DATA04016AU.pdf
Usage Behavior Analysis by Using K-Means Algorithm in eBanking
http://www.ait.ac.th/digital_gms/Proceedings/C55_SURACHAI_WIWATTANACHARO.pdf
A better return on data http://www.informationweek.com/596/96bank.htm
eBanking delivery channels http://www.itcnasia.com/icit/presentations/day_2/4th_session/Javed_Y_Edhi.pdf
eBanking Platform Upgrade Cycle http://www.crm2day.com/news/crm/114753.php
Online banking services http://www.bitpipe.com/rlist/orgtypegrp/ALL/Online-Banking-Services.html
Online banking grow http://www.crm2day.com/news/crm/EpVuVFZlpVmXjNSUka.php
Internet banking http://www.datacenterinc.com/internet.htm
CRM http://www.bitpipe.com/rlist/term/CRM.html?src=googlecrm_content
Financial services http://72.14.203.104/search?q=cache:GYSOD3VGxXoJ:www.digitalinsight.com/pdf/lessons.pdf+&hl=zh-TW
eBanking Platform
http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/scc/&toc=comp/proceedings/scc/2004/2225/00/2225toc.xml&DOI=10.1109/SCC.2004.1358011

HSBC e-banking Services:

Banking	Investments	Market Info
Account Overview	Overview	Rates
Transfer – Hong Kong	Stocks	Share Watch
Transfer – Overseas	Unit Trusts	Real-time Quote Summary
View & Pay Bills	Bonds / CDs	Delayed Quote Summary
Autopay	ELI	Warrant Info
Time Deposits	Others	Top 10 Movers
	Transaction History	Indices
	Order Status	SEHK News
	Buy	Stock Code Search
	Sell	Daily Market Commentary
	Stop Loss	Investment Watch
	Two-way	Futures

Cards	Loans	Insurance
Apply Online	Mortgages	Travel
Instructions	Personal Loan Plan	Home
	Personal Tax Loan	Helper
		Medical
		Accident
		Life
		View Your Policy
		Make a Claim
		Amend Your Policy

MFP	**ORSO**	eAlerts	My Details	Email

Temporal Data types

- ***Irregularity***: Many types of numerical temporal data are not equally paced.
- ***Asynchronousness***: In distributed computing environments like sensor networks, data from different sources tend to be not aligned and hence cannot apply synchronous methods.
- ***Distributed analysis***: A trend in temporal data analysis is to perform data filtering, transformation and analysis as close as possible to the data sources to avoid the prohibitive amount of data being transmitted and analyzed. This new computing paradigm calls for a new theoretical foundation.
- ***Streaming Data***: Some temporal data is stored only temporally and requires near real-time analysis.
- ***Heterogeneous data types***: It is very common that temporal data is partly categorical events and partly numerical time series. It remains to be an interesting challenging to best analyze all possible data in a uniform way.
- ***Huge Volume***: The stream of data can be huge for a long, continuous observation period. Many types of measurements can be obtained from a large number of data sources. This requires designing scalable solutions in analyzing a large volume of temporal data, in terms of both the large number of data points and the large number of types of measurements.

Project Conceptual Modeling

Conceptual information modeling and data mining, is now an emerging *granular computing* paradigm of information processing. This Coding characteristics is a conceptual methodology for context and time. This type of modeling is context-oriented by using codifying linguistic variables and temporal partitioning. The context associated with client individually, especially in natural language information contexts. Discretization coding method for linguistic variables and time applied fuzzy quantification of information granules. The granulate time of events, millions of codes manipulated individually by computers, this fundamental computing paradigm shift as a meta-information processing. In fact, human behaviors in daily life, including most web-based conceptual information modeling in practice, could easily take the assumption of the objective reality. Actually, this mathematical modeling tried to demonstrate in ebanking contexts (another word is situation).