Expert Systems with Applications 37 (2010) 8042–8057

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An analytic approach to select data mining for business decision

Jia-Lang Seng [a,*], T.C. Chen [b]

[a] Department & Graduate School of Accounting, College of Commerce, National Chengchi University, Taipei 115, Taiwan
[b] MICRONIX Technologies Inc., Taipei, Taiwan

ARTICLE INFO

Keywords:
Data mining
Business decision
Selection model

ABSTRACT

Due to the information technology improvement and the growth of internet, enterprises are able to collect and to store huge amount of data. Using data mining technology to aid the data processing, information retrieval and knowledge generation process has become one of the critical missions to enterprise, so how to use data mining tools properly is user concern. Since not every user completely understand the theory of data mining, choosing the best solution from the functions which data mining tools provides is not easy. If user is not satisfied with the outcome of mining, communication with IT employees to adjust the software costs lots of time. To solve this problem, a selection model of data mining algorithms is proposed. By analyzing the content of business decision and application, user requirements will map to certain data mining category and algorithm. This method makes algorithm selection faster and reasonable to improve the efficiency of applying data mining tools to solve business problems.

1. Introduction

1.1. Research background

Due to the information technology improvement and the growth of Internet, enterprises are able to collect and to store huge amount of data. People gradually realize that data is not equal to information that data should be further analyzed and extracted. Professionals are trained to analyze and interpret data, but the increases in data amount, data type, and analytical dimensions. Information technology has gone beyond storage, transmission, and processing. Data needs to be converted into information and knowledge in order to support decision making.

1.2. Research issue

Enterprises use data mining tools to support knowledge discovery and decision making. In this research, we develop a selection model to solve the research issue. This model recommends the most suitable algorithm after marketing professionals and analysts describe the business problems using a standard procedure and format. This model provides an algorithm standard as the foundation of dynamic data mining modeling.

This selection study of data mining mainly has two parts: the commercial problems analysis and the data mining algorithms analysis. Commercial problems analysis contains a general set of 22 problems. These problems relate to the banking, finance, insurance, telecommunication, retail, and manufacturing applications. They are classified into 12 business application categories according to their characteristics. Data mining algorithms analysis discuss five main types of data mining: association, classification, prediction, clustering, and profiling. In each category, several typical algorithms are listed and each of them is used to discover similar knowledge in different situations. The algorithm concepts, parameters, processes, and characters are compiled and compared to other algorithms in the same category.

1.3. Research limitation

Data mining has been applied to business area for more than two decades. There are countless application cases for different industries, different information requirement, and lots more circumstances. We adopt the literature survey to summarize approaches and concepts from our literature review to formulate the selection framework and produce the processing details into broader concepts and terms. This selection model focuses on matching the applicable data mining method with the characteristics of business decision and application.

2. Literature review

2.1. Data mining

The amount of data continues to grow at an enormous rate even though the data stores are already vast. The primary challenge is

* Corresponding author.
E-mail addresses: seng@nccu.edu.tw, jia.lang.seng@gmail.com (J.-L. Seng), jljan2@gmail.com (T.C. Chen).

0957-4174/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2010.05.083

how to make the database a competitive intelligence by converting seemingly meaningless data into useful information. How this challenge is met is critical because enterprises are increasingly relying on effective analysis of the information simply to remain competitive. By knowledge discovery in databases, interesting knowledge, regularities, and high-level information can be extracted from the relevant sets of data in databases and be investigated from different angles. From a data warehouse perspective, data mining can be viewed as an advanced stage of on-line analytical processing (OLAP). However, data mining goes far beyond the narrow scope of summarization-style analytical processing of data warehouse systems by incorporating more advanced techniques for data understanding (Han & Kamber, 2001). Many people treat data mining as a synonym for another popularly used term, Knowledge Discovery in Databases, or KDD. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery in databases. Choosing the data mining algorithm includes selecting method(s) to be used for searching for patterns in the data such as deciding which models and parameters may be appropriate and matching a particular data mining method with the overall criteria of the KDD process. Data mining searches for patterns interest in a particular representational form set of such representations including classification rules or trees, regression, clustering, sequence modeling, dependency, and line analysis. The mining results which match the requirements will be interpreted and marshaled, to be taken into action or be presented to interested parties in the final step. The concept of data mining contains all activities and techniques utilizing the collected data to get implicit information and analyzing historical records to gain valuable knowledge.

2.1.1. Data mining method

Data mining methods refer to the function types that data mining tools provide. The conceptual definition of each data mining method and the assortment basis always differ for the ease of explanation, the consideration of present situation, or researcher background. Classification, association, prediction, clustering are usually the common methods in different works, while the term description, summarization, sequential rule may not always be used and listed in the first place.

Data mining methods include techniques which evolve from artificial intelligence, statistics, machine learning, OLAP and so on. These most often mentioned methods are classified into five categories according to their function types in business applications are shown in Table 1.

2.1.2. Data mining modeling

Data mining modeling is the critical part in developing business applications. Business applications, such as "cross-selling", will be turn into one or more of business problem and the goal of modeling is to formulate these business problems as data mining tasks.

Table 1
Data mining literature categories.

Author	Data mining categorization
Fayyad, Piatetsky-Shapiro, and Smyth (1996), Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy (1996)	Classification, regression, clustering, summarization, dependency modeling, link analysis, and sequence analysis
Berry and Linoff (1997)	Classification, estimation, prediction, affinity grouping, clustering, and description

Fayyad, Piatetsky-Shapiro, and Smyth (1996) and Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy (1996) state that data mining algorithms consists largely of some specific mix of three components:

- *The model:* There are two relevant factors. They are the function of the model and the representational form of the model. A model contains parameters that are to be determined from the data.
- *The preference criterion:* A basis for preference of one model or set of parameters over another, depending on the given data. The criterion is usually some form of goodness-of-fit function of the model to the data, perhaps tempered by a smoothing term to avoid over fitting, or generating a model with too many degrees of freedom to be constrained by the given data.
- *The search algorithm:* The specification of an algorithm for finding particular models and parameters, given data, a model, and a preference criterion.

In Berry and Linoff (1997), a mapping between business tasks and data mining techniques had been partially done. The choice of what data mining techniques to apply at a given point in the knowledge discovery processes depends on the particular data mining task to be accomplished and on the data available for analysis. The requirements of tasks dedicate to the functions of mining and the detail characteristics of tasks influence the feasibility between mining methods and business problems. The so called detail characteristics include data types, parameter varieties, hybrid approaches and so on. Slightly difference in the model will cause enormous performance changes. So the modeling stage affects the quality of data mining tools heavily (Table 2).

2.2. Business decision

Data mining assists many kinds of analysis works and the most popular business applications collected from the literature are listed as follows.

Finance and insurance

- Financial product cross-selling identifies additional types of products that customers could purchase which they currently are not purchasing. For example, a saving account customer would like to buy life insurance products of the original company.
- Telemarketing and direct mail marketing need marketing name list so data mining is in charge of figuring out the most responsive and valuable customers for certain products.
- Market segment analysis is one of the basic marketing researches where similar customers will be divides into same segmentations.
- Product mix analysis considers customer needs, marketing strategy and product line completeness to find out the best product mix.
- Credit card fraud detection and bad debt collection monitors and detects unusual activities of fraud and the occurrence of bad debt.
- Customer churn analysis is used to analysis the transaction record and behavior pattern to know the possibility of customer attrition, and then, enterprise will apply different recovery actions to customer base on their churn tendency and contribution to the company.
- Credit card application and personal loan rating classifies the customer into different level for loan or credit card applications.

Table 2
Correspondence between business applications and data mining methods.

Technique	Classification	Estimation	Prediction	Affinity grouping	Clustering	Description
Standard statistics	✓	✓	✓	✓	✓	✓
Market basket analysis			✓	✓	✓	✓
Memory-based reasoning	✓		✓	✓	✓	
Genetic algorithms	✓		✓			
Cluster detection					✓	
Link analysis	✓		✓	✓		
Decision trees	✓		✓		✓	✓
Neural networks	✓	✓	✓		✓	

Retail

- Member customer cross-selling analyzes customer transaction records to further introduce them other products.
- Promotional product mix discovers the items that often be purchased together so these products can be placed closely or be sold in pairs.
- Advertisement name list analysis uses analytical tools to generate the customer list from the customer database for marketing activity to get a higher responsive rate and reduce marketing costs.
- Market segmentation divides the market into several parts by certain characters.
- Direct marketing needs analytical software to figure out the most responsive and valuable customers for marketing.
- Inventory analysis analyzes inventory and sales data to calculate the trend of future requirement. This information can be the accordance of stock replenishing and marketing decision making.

Telecommunication

- Customer segmentation see Retail (4).
- Direct marketing see Retail (5).
- Charge making is important to telecommunication enterprises because the charge policy is the key point which customers concern. The basic objective of charge making application is to design charge policies according to customer calling pattern.
- Link analysis associates customers with their communicators. The activities radiate from single customer will link several persons. The relationship among these people can be roughly described and the calling pattern can be used to design more customize charge policy.
- Customer retention tracks customer retention rate and analyses customer behavior change to support customer relationship management.

Manufacturing

- Production yield analysis uses prediction methods to detect defect clusters on product quickly to improve the yield and to reduce the production failure lost.
- Production process control finds out the factors in the processes that cause failure and classify factors which cause defect into groups to analyze the quality issues.

Security investment

- Stock and foreign exchange market prediction builds a simulation model to Predict the fluctuation of stock price and foreign exchange.
- Direct marketing see Finance and insurance.

2.3. Business applications related works

The primary goal of data mining is discovery of new patterns and deeper insights within the data. New pattern discovery is used in marketing to make predictions about buyer behavior, to understand customer preference, and to manage customer relationship. Deeper insights of service and product designing improve the quality and efficiency of process. Business application can be decomposed into several business problems. Business problems can be divided into problem processing steps and problem characteristics which are derived from problem descriptions involves processes of a business decision and application are shown in Table 3.

2.3.1. Cross-selling

Cross-selling applications include financial product cross-selling and retail member customer cross-selling. The advantages of cross-selling strategy are threefold. First, targeting customers with products they are more likely to buy should increase sales and therefore increase profits. Second, reducing the amount of people targeted through more selective targeting should reduce costs. Finally, it is an established fact in the financial sector that loyal customers normally possess more than two products on average; therefore, persuading customers to buy more than one product should increase customer loyalty. To achieve cross-selling effects,

Table 3
Business application categories.

Application categorization	Industry sector	Business applications
Cross-selling	Financial	Product cross-selling
	Retail	Member customer cross-selling
Direct marketing	Financial	Telemarketing and direct mail marketing
	Retail	Advertisement name list analysis
	Retail	Direct marketing
	Telecom	Direct marketing
	Security	Direct marketing
Segmentation analysis	Financial	Market segment analysis
	Retail	Market segmentation
	Telecom	Customer segmentation
Product mix analysis	Financial	Product mix analysis
	Retail	Promotional product mix
Fraud detection and bad debt collection	Financial	Credit card fraud detection and bad debt collection
Churn analysis	Financial	Customer churn analysis
	Telecom	Customer retention
Credit rating	Financial	Credit card application and personal loan rating
Stock and exchange prediction	Security	Stock and foreign exchange market prediction
Inventory analysis	Retail	Inventory analysis
Charge making	Telecom	Charge making
Telecom link analysis	Telecom	Link analysis
Production Yield analysis	Manufacturing	Production yield analysis
	Manufacturing	Production process control

knowing who would be interested in what product is the key. The overall objective was to discover characteristics of current customers which can then be used to target all other customer segments in order to classify them into potential promotion targets and unlikely purchasers.

2.3.2. Direct marketing

Direct marketing can be found in many industries. Direct marketing is returning as a valuable tool in modern business as mass marketing loses effectiveness (Forcht & Cochran, 1999). The purpose of direct marketing is to reduce cost and increase efficiency by selling products or services to customers directly instead of the utilization of merchandisers or wholesalers. Direct marketing asks for immediate customer response or purchase rather than long-term product image build-up. Sales information is sent to certain customer via media such as mail, television, email, Internet, and customers can response to the enterprise from indicated channels. Direct marketing is a matured business function. Sales representatives can be aided by statistic tools to interpret the customer characters of previous transactions in the first place, and then significant characters are chosen to be the accordance to make marketing decisions. Customers conform to those characteristics predefined by sales persons will be classified into the same group, and suitable marketing campaign can be made and assign to this group.

2.3.3. Segmentation analysis

Segmentation is essentially aggregating customers into groups with similar characteristics such as demographic, geographic, or behavioral traits, and marketing to them as a group (Bose, 2002). Facing the market with diverse demands, applying market segmentation strategy can increase the expected returns. Much of marketing research focuses on examining how variables such as demographics and socioeconomic status can be used to predict differences in consumption and brand loyalty. Segmentation problem should be considered as two different situations known character parameters and unknown character parameters. Character parameters are known means segmentation analysis deals with customers who have transactional or behavioral records stored in the enterprise database and the analytic parameters are predefined and are derived from analyzer interests.

2.3.4. Product mix analysis

Product mix analysis is one of the main processes in product management and sales strategy planning. Enterprises usually own many products to meet different customer demands and each product has its positional strategy and promotional activities so product mix analysis is introduced to coordinate these synchronous marketing strategies. The purpose of creating several products in same product area is to exhaust the market size, strengthen company position, and keep potential competitors out.

2.3.5. Fraud detection

The purpose of fraud detection is to reduce the cost caused by criminal activities. The idea is to mine system audit data for consistent and useful patterns of user behavior and then keep these normal behaviors in profiles. While a transaction occurs, system will compare current transaction with existed profile to analyze where this transaction is dubious. A dubious transaction means it has high probability to be a fraud so this transaction should not be accepted and an alarm will be sent to the customer and the business manager.

2.3.6. Churn analysis

Churn management consists of developing techniques that enable firms to keep their profitable customers and it aims at increasing customer loyalty. The adherents of customer retention argue that retaining customers improves profitability, mainly by reducing the costs incurred in acquiring new customers.

2.3.7. Credit card application and personal loan rating

In financial industry, credit risk is the oldest form of risk in the financial markets. It is crucial for corporate financing and our lives. Banks need to investigate customer credit histories by loan experts before issuing cash cards and approving credit limits, but the expertise is scarce and difficult to duplicate. Credit rating is even more important task today as financial institutions have been experiencing serious competition during the past few years. It is impossible both in economic and manpower terms to conduct all works with the explosive size growths (Lee, Chiu, Chou, & Lu, 2006). In order for banks to make more precise judgments and better decisions, an automatic and scientific evaluation model about customers credit rating should be integrated into a simplified rating process so that applications for cash cards can be processed quickly and accurately to reduce loss from bad debt.

2.3.8. Stock movement prediction

Stock movement prediction research can also utilize data mining techniques to build analysis model from historical data (Gavrilov, Anguelov, Indyk, & Motwani, 2000; Lu, Han, & Feng, 1998; Povinelli & Feng, 1999). The performance of stock market is the result of all participator collaboration. Investor behavior has particular principles such as time sequence and is affected by certain factors. Juridical person and investors will actively invest certain industry and periodically turn to another based on the company strategy. The knowledge of sequential phenomenon mentioned above can be mined so sequential pattern has been performed broadly in this area. Traditionally, the prediction models are developed with statistical and regression formulas but the combination of machine learning and statistical methods can provide better performance and accuracy. Statistical and regression approaches have long been used to forecast numerical information in financial area and stock market. Machine learning is relatively new technique. Machine learning enables more complex prediction approach and calculation which approximate reality better, hence, it usually produces more accurate prediction results.

2.3.9. Inventory analysis

How to develop an effective inventory management method to properly allocate resources and stocks and how to deliver products into customer hands on time is enterprise key competence. Retailer can use association analysis to know the causal relationship between sales condition and required products. The causal relationship can be used to build requirement prediction model to forecast the requirements in the future. The complexity and difficulty of related requirements which makes inventory management efficient going down usually upraises with the amount of items. The "support" concept in association approach replaces the manmade probability forecasting of pair requirements to get a more objective assessment. The itemsets derived from previous analysis step can be further clustered to know the product sales features and requirement in different periods. Replenishment of stock with this information will become precise, reducing inventory cost and avoiding insufficient inventory at the same time.

2.3.10. Charge making

Pricing is now a dramatically different area than it has been historically. New software tools give enterprises a new way of setting, optimizing, and enforcing pricing changes within the organization. With the best tools, an integrated view of customers, their past purchases, benchmarked pricing by segment and size of purchase, relationship data, and comparison of trends over time – all are available to provide decision support to field sales representatives, sales managers, marketers and general managers (Davidson & Simonetto, 2005).

2.3.11. Link analysis

Link analysis is based on a branch of mathematics called graph theory and it has yielded promising results in solving some problems such as analyzing telephone call patterns. Each telephone call is a relationship between two end points so call patterns are naturally presented as graphs with thousands or millions of nodes and edges. Link analysis is playing two roles in analyzing the cellular phone data (Berry & Linoff, 1997). The first is the power of visualization. Second, link analysis can apply the concept generated by visualization to larger sets of customers. The power of visualization provides the ability to see some of the graphs representing call patterns make patterns for things like inertia or influence much more obvious. In the second stage of link analysis, the previous retrieved pattern is applied to the whole transactional database to find other correspondence ones. These chosen customers match certain features and some marketing opportunities may fit to them.

2.3.12. Production yield analysis

A high priority goal for wafer manufacturing is finding the most probable causative factors that discriminate between low yield and high yielding product by quickly examining the historical data already being collected. A combination of self-organizing map (SOM) neural networks and rule induction is used to identify the critical poor yield factors from normally collected wafer manufacturing data. The SOM neural network algorithm performs a type of "multivariate, non-linear regression". Rule Induction is an additional unsupervised data mining algorithm that works synergistically with the cluster map. It generates logical expressions that identify the data attributes that most discriminate between clusters, thus explaining the clusters.

2.4. Summary

Data mining has a large family composed of various algorithms and the scope is still expanding because researchers devote to improve the efficiency and accuracy of the existed algorithms. Most researches in data mining area focus on improving efficiency and accuracy of single business decision and application. Fewer efforts are devoted into the discussion of applicability and fitness. The more complex the application is, the larger the gap comes into existence between applications and users. We propose a selection model to match these business requirements to data mining categories and connect complex data mining concepts with business problems.

3. Research approach

3.1. Business decision and application analysis

The basic concepts and contents of business decision and application analysis come from the literatures. In our literature review, 12 business decision and application areas are surveyed. Most of these literatures focus on only one or two application issues and describe specific approaches. The development of selection model requires both conceptual analysis and operational definition of business decision and applications. Applications are usually composed of several problems to be solved and the systematically descriptions of these problems are the basis of modeling. In this research, we decompose each application into four parts to establish a standard description. Those four concepts are defined as follows.

1. Business decision and application (BA) activities in certain business sector (e.g. cross-selling).
2. Processing steps (PP) steps of solving business problem. Each step gets certain information which matches certain pattern from data by considering problem characteristics (e.g. customer segmentation analysis of cross-selling).
3. Processing characteristics (CH) information which needs to be assigned or predefined in processing steps (e.g. customer background data of cross-selling).
4. Processing outcome (PO) required analytical result of problem processing step (e.g. customer profile of segmentation analysis).

These four concepts have hierarchical relationships. Business problem is one of the processes in business decision and application and processing steps construct business problem. Processing characteristic indicates the information required in the mining processes. To start our analysis procedure, first, each business decision and application is decomposed by several processing steps as shown in table. Usually, different approaches are applied to the same business decision and application when available data or analysis postulation differs are shown in Tables 4 and 5.

Business decision and applications are described by normal comprehension of the problems to make the concepts approachable. These comprehensible descriptions are not precise enough and contain too many synonyms for similar notions in different problems. For example, some problem solving steps belonging to different business decision and applications are actually the same process applying to different data areas. To make the meaning of each processing step clear, we define the function of each step as shown in Table 6.

Each business decision and application can be solved by several analytical activities which are called "processing steps". One processing step may involve several characteristics to accomplish the analysis. Those business decision and application analysis concepts and their relationships are listed in table. Processing outcome is the result of processing step and the answer which analyst wonders. Each processing step leads to an outcome are shown in Tables 7 and 8.

Processing characteristics and processing outcomes defined above involve processing steps so we match these characteristics and outcomes for each processing step is shown in Table 9.

3.2. Data mining algorithms analysis

The purpose of data mining algorithm analysis is to discern each algorithm by concretely describing its features. The main components of a business decision and application are algorithm and data structure. An algorithm analysis is essential to develop a selection model. Four concepts are defined to describe each algorithm. Four concepts are defined to describe each algorithm, their definitions are as follows.

1. Data mining method: an analytical tool used to discover implicit information from large database (e.g. Apriori algorithm).
2. Algorithmic steps: the functional type which is achieved by mining steps (e.g. association).
3. Input data unit: the input content and format of an algorithm (e.g. item in association).
4. Output data unit: the output content and type of an algorithm (e.g. itemset in association).

Four types of data mining categories are considered and described in table. Data mining category is summarized from many algorithms which have similar use are shown in Tables 10 and 11.

Table 4
Literature review of business applications.

	Item	Literature number													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Business application	Cross-selling	˘	˘						˘						
	Direct marketing	˘					˘			˘					
	Segmentation analysis	˘								˘	˘				
	Product mix analysis	˘									˘				
	Fraud detection	˘											˘		
	Churn analysis					˘		˘							
	Credit rating														˘
	Stock movement prediction											˘			
	Inventory analysis														
	Charge making													˘	
	Telecom link analysis	˘		˘											
	Product yield analysis				˘										
Processing step	Classify customers	˘		˘		˘	˘			˘	˘		˘	˘	˘
	Classify product defects				˘										
	Detect unusual activities	˘											˘		
	Match campaigns to potential customers		˘	˘			˘			˘	˘			˘	
	Find relationships of products in transactions	˘							˘						
	Find relationships in customers' characteristics	˘	˘												
	Winback customers							˘							
	Make replenishment decision	˘													
	Predict stock price trend											˘			
	Estimate the revenue of pricing model	˘												˘	
	Estimate the revenue of product mix	˘									˘				
	Predict the tendency of churn					˘		˘							
	Credit rating customers														˘
Processing characteristic	Customer background data	˘	˘	˘		˘	˘	˘		˘	˘			˘	˘
	Customer transaction data	˘	˘	˘		˘	˘	˘	˘		˘		˘	˘	˘
	Product basic information	˘							˘					˘	
	Product sales data	˘													
	Manufacturing data				˘										
	Causes and effects			˘											
	Number of classes to be classified	˘			˘										
	Stock historical data											˘			
Processing outcome	Customer profile	˘	˘	˘		˘	˘	˘		˘	˘		˘	˘	
	Exception report	˘											˘		
	Credit rating report														˘
	Product profile	˘							˘						
	Product mix	˘							˘		˘				
	Pricing model	˘												˘	
	Prospect list	˘	˘	˘			˘			˘	˘			˘	
	Defects distribution				˘										
	Loyalty report						˘	˘							
	Stock price trend prediction											˘			
	Revenue estimate report	˘									˘			˘	
	Replenishment decision	˘													
	Calling pattern	˘		˘											

1. Berry and Linoff (1997); 2. Anand, Patrick, Hughes, and Bell (1998); 3. Ahn and Ezawa (1997); 4. Gardner and Bieker (2000); 5. Min, Min, and Emam (2002); 6. Pitta (1998); 7. Lejeune (2001); 8. Agrawal, Imielinski, and Swami (1993); 9. Forcht and Cochran (1999); 10. Pritscher and Feyen (2001); 11. Lu et al. (1998); 12. Bose (2006); 13. Netessine, Savin, and Xiao (2004); 14. Lee et al. (2006).

Table 5
Business applications decomposition.

Business application	Processing steps	Sources
Cross-selling	Find relationship of customers' characteristics Match campaigns to potential customers	Agrawal et al. (1993), Anand et al. (1998), Berry and Linoff (1997)
Direct marketing	Classify customers Match campaigns to potential customers	Berry and Linoff (1997), Forcht and Cochran (1999), Pitta (1998)
Segmentation analysis	Classify customers Match campaigns to potential customers	Berry and Linoff (1997), Forcht and Cochran (1999)
Product mix analysis	Classify customers Match campaigns to potential customers Estimate the revenue of product mix	Pritscher and Feyen (2001)
Fraud detection	Classify customers Detect unusual activities	Bose (2006), Brause et al. (1999), Chan et al. (1999)
Churn analysis	Classify customers Predict the tendency of churn Winback customers	Min et al. (2002), Lejeune (2001)
Credit rating	Classify customers Credit rating customers	Lee et al. (2006), West (2000)
Stock movement prediction	Predict stock price trend	Lu et al. (1998),
Inventory analysis	Find relationship of products in transactions Make replenishment decision	Bansal, Vadhavkar, and Gupta (1998)
Charge making	Classify customers Match campaigns to potential customers Estimate the revenue of pricing model	Davidson and Simonetto (2005), Netessine, Savin, and Xiao (2004)
Telecom link analysis	Classify customers Match campaigns to potential customers	Ahn and Ezawa (1997), Berry and Linoff (1997)
Product Yield analysis	Classify product defects	Gardner and Bieker (2000)

Table 6
Processing step definition.

Processing steps	Explanation
Classify customers	Divide all customers into several groups
Classify product defects	Divide all product defects into several groups according to some characteristics
Detect unusual activities	Judge whether a payment or a transaction is made by it's owner
Match campaigns to potential customers	Assign customers to suitable marketing campaign according to customers' characteristics
Find relationship of products in transactions	Discover common products or sequential purchasing patterns from transactions
Find relationship of customers' characteristics	Discover related and common characteristics of customers of certain products
Winback customers	Provide attractive marketing program which meets customers' needs to winback them
Make replenishment decision	Predict future sales to replenish stock
Predict stock price trend	Predict stock price fluctuation over time
Estimate the revenue of pricing model	Calculate total revenue of pricing model to decide which model is the best
Estimate the revenue of product mix	Predict total revenue of each product mix to decide which product mix plan is the best
Predict the tendency of churn	Predict the possibility that a customer leave
Credit rating customers	Evaluate a customer's condition to know the risk of transaction

3.2.1. Association rule algorithms analysis

The problem of discovering all association rules can be theoretically decomposed into two steps.

1. Find all sets of items (itemsets) that have transaction support above minimum support which defined by analyzer. The support for an itemset is the number of transactions that contain the itemset. Itemsets with minimum support are called large itemsets, and all others small itemsets.
2. Use the large itemsets to generate the desired rules. For every large itemset I, find all non-empty subsets of I. For every such subset a, output a rule of the form $a \rightarrow (I - a)$ if the ratio of support(I) to support(a) is at least minimum confidence. We need to consider all subsets of I to generate rules with multiple consequents.

Table 7
Processing characteristic definition.

Processing characteristic	Explanation
Customer background data	Including demographic characteristics like gender, age, habitation, job, appointment, family and so on
Customer transaction data	Including interaction and transaction records between customer and organization. Time, channel, amount, payment often been recorded in this kind of data
Product basic information	Including intrinsic attributes of all kinds of products. For example, the weight, price, and nutritional constituent of food, the price and capacity of a hotel room, and content of service
Product sales data	Sales condition observed which base on product unit
Manufacturing data	Including all settings, conditions, and interactions of variables in manufacturing processes
Causes and effects	Including relationship or reasoning rule which is predefined to guide analysis process
Number of classes to be classified	Predefined or known possible types of analytical result
Stock historical data	The rising and falling history of stock price

The algorithms which provide association analysis are based on Apriori algorithm and the most famous application using association analysis is market basket analysis. Many advanced algorithms are proposed to do more complicated analysis such as single dimension, multi-level, and multi-dimensional association rules.

3.2.2. Classification rule algorithms analysis

The function of classification is to distinguish data into different group, sounds like the job clustering do. Classification is often referred to as supervised learning. The analysts give the system certain rules to use in searching the database. In contrast, clustering is frequently referred to as unsupervised learning because the system is left to freely examine the data and discover patterns on its own (Forcht & Cochran, 1999).

Table 8
Processing outcome definition.

Processing outcome	Explanation
Customer profile	Customer's profile which describes his behavioral or transactional habits
Exception report	Giving judgment whether a transaction has a high probability to be an unusual activity or fraud
Credit rating report	Giving judgment to the degree of a customer's credit
Product profile	Generating a list of products which match certain characteristics
Product mix	Giving the best product mix decision
Pricing model	Giving the best pricing model decision
Prospect list	Generating a list of customers who match certain characteristics
Defects distribution	Grouping defects according to their processing condition
Loyalty report	Giving judgment of the degree of a customer's tendency of leaving
Stock price trend prediction	Forecasting daily or long-term situation of stock price fluctuation
Revenue estimate report	Prediction result of product sales considering customer segment, promotion, or other situation
Replenishment decision	Predicted amount and type of inventory replenishment
Calling pattern	Description of customer's calling behavior including time, frequency, and target

Approaches which can provide classification function include decision tree, Bayesian statistics, case-based reasoning, generic algorithm, *k*-nearest method, and neural network. Each algorithm involves lots of issues and different practices behind its back. For the use of this research, we simply describe these algorithms by looking at key factors of them as shown in Table 12.

3.2.3. Prediction algorithms analysis

Several software packages such as SAS and SPSS have existed to solve regression problems because statistic techniques matured quite early in business area. There are two types of prediction techniques. The major technique is numerical regression and the other can be applied to the modeling of categorical response variables. Traditional prediction methods come from statistic area, for instance, linear regression and non-linear regression. Although most researches do not take statistical regression as mining method, they still participate in many mining processes to help checking the significance of decision variables.

Regression model include two kinds of variables, response variable and predictor variable. Target variable, Y, called a response variable, is modeled as a function of another random variable, X, called a predictor variable. Take linear regression as example, the function is, $Y = \alpha + \beta X$, where α and β are regression coefficients specifying the T-intercept and slope of the line, respectively. The coefficients of the model can be trained and adjusted to produce more precise predictive result.

Linear regression is widely used, owing largely to its simplicity. Linear regression model is limited to continuous data and normal distribution, so generalized linear models are developed to cover the analytical needs of processing categorical data. One widely applied method of generalized linear models is logistic regression. We choose logistic regression to be the representative method of prediction methods.

3.2.4. Clustering algorithms analysis

The cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. As a branch of statistics cluster analysis has been studied extensively for many years focusing mainly on distance-based cluster analysis. Clustering is frequently referred to as unsupervised learning. It does not rely on predefined classes and class-labeled training examples.

Table 9
Mapping processing step, processing characteristic, and processing outcome.

Problem processing steps (PP)	Processing characteristic (CH)	Processing outcome (PO)
Classify customers	Customer background data	Customer profile
	Customer transaction data	Prospect list
	Causes and effects	Calling pattern
Classify product defects	Manufacturing data	Defects distribution
	Number of classes to be classified	
Detect unusual activities	Customer transaction data	Customer profile Exception report
Match campaigns to potential customers	Customer background data	Prospect list
	Customer transaction data	
Find relationship of products in transactions	Customer transaction data	Product profile
	Product basic information	Product mix
	Product sales data	
Find relationship of customers' characteristics	Customer background data	Customer profile
	Customer transaction data	
Winback customers	Customer background data	Customer profile
	Customer transaction data	
Make replenishment decision	Product basic information	Replenishment decision
	Product sales data	
Predict stock price trend	Stock historical data	Stock price trend prediction
Estimate the revenue of pricing model	Customer background data	Pricing model
	Product basic information	Revenue estimate report
Estimate the revenue of product mix	Customer background data	Product mix
Predict the tendency of churn	Customer background data	Customer profile
	Customer transaction data	Loyalty report
Credit rating customers	Customer background data	Credit rating report
	Customer transaction data	

Table 10
Data mining categories.

Category	Input data unit	Crucial parameter	Output data unit
Association	Object	Support, confidence	Association between objects
Classification	Compound object, case	Class	Class
Prediction	Condition	Coefficient	Experimental conclusion
Clustering	Object	Radius	Clusters

3.3. Mapping business characteristic to data mining method

After the analysis of both business side and mining side, the characteristics of business side must map to concepts of mining

Table 11
Literature review of data mining methods.

	Item	Literature number														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Data mining method	Apriori	ˇ	ˇ						ˇ				ˇ			
	Decision tree	ˇ				ˇ	ˇ	ˇ				ˇ				ˇ
	Bayesian belief network			ˇ												
	K-nearest neighbor	ˇ														
	Case-based reasoning	ˇ														
	Generic algorithm	ˇ														
	Neural network	ˇ			ˇ					ˇ	ˇ			ˇ		
	Logistic regression														ˇ	
	K-means	ˇ										ˇ				
Algorithm step	Association	ˇ	ˇ						ˇ							
	Classification	ˇ		ˇ		ˇ		ˇ		ˇ	ˇ	ˇ		ˇ	ˇ	ˇ
	Prediction	ˇ											ˇ			
	Clustering	ˇ			ˇ							ˇ				
Input data unit	Continuous data	ˇ											ˇ			
	Discrete data	ˇ	ˇ	ˇ	ˇ	ˇ		ˇ	ˇ	ˇ	ˇ	ˇ		ˇ	ˇ	ˇ
	Inference rule			ˇ												
	Time sequence	ˇ						ˇ			ˇ		ˇ			
	Predefined parameter	ˇ										ˇ				
Output data unit	Set	ˇ	ˇ						ˇ				ˇ			
	Predefined class	ˇ		ˇ				ˇ		ˇ		ˇ		ˇ	ˇ	ˇ
	Probability distribution			ˇ											ˇ	
	Weighted sum	ˇ									ˇ					
	Cluster	ˇ			ˇ							ˇ				
	Rule	ˇ				ˇ						ˇ				ˇ

1. Berry and Linoff (1997); 2. Anand et al. (1998); 3. Ahn and Ezawa (1997); 4. Gardner and Bieker (2000); 5. Min et al. (2002); 6. Pitta (1998); 7. Lejeune (2001); 8. Agrawal et al. (1993); 9. Forcht and Cochran (1999); 10. Pritscher and Feyen (2001); 11. Lu et al. (1998); 12. Bose (2006); 13. Netessine et al. (2004); 14. Lee et al. (2006).

Table 12
Classification algorithms analysis.

Data mining method	Algorithm steps	Input data unit	Output data unit
Decision tree	Judgment contain Explicit rule, link, and parameter value	Single object	Class
Bayesian belief network	The decision depends on occurrence probability and some reasoning rules	Single object	Probability
K-nearest neighbor	Find the most similar historical solution	Single object with numerical data	Class
Case-based reasoning	Find the most similar historical solution	Single object with mixed type data	An approximate solution
Generic algorithm	Find the most similar historical solution or create a new solution	Single object with mixed type data	Fittest rule
Neural network	The synthetic judgment using weighted summary	Single object with mixed type data	Weighted summary

Table 13
Mapping processing steps to algorithm steps.

Processing steps	Algorithm steps			
	Association	Classification	Prediction	Clustering
Classify customers		ˇ		ˇ
Classify product defects				ˇ
Credit rating customers		ˇ		
Detect unusual activities		ˇ		
Estimate the revenue of pricing model		ˇ		
Estimate the revenue of product mix		ˇ		
Find relationship of customers' characteristics	ˇ			
Find relationship of products in transactions	ˇ			
Make replenishment decision		ˇ		
Match campaigns to potential customers		ˇ		
Predict stock price trend			ˇ	
Predict the tendency of churn		ˇ		
Winback customers		ˇ		

side to find applicable data mining method. Processing steps, processing characteristics, and processing outcomes have been dedicated in business decision and application analysis to decide the information needs so looking for their corresponsive technical solution would be the next step.

The analysis in precious section is prepared for selection model development. The concept of processing step in business side is matched to the algorithm step in mining side. The purpose of each processing step is to look for the potential relation between analytical units such as customers or products sales. Algorithm steps in each data mining method are designed to search, compare, or calculate input data to generate filtered or condensed information. Because processing steps can be actualized as algorithm steps, mapping of these two concepts can bridge business decision and

Table 14
Mapping processing characteristic to input data unit.

Processing characteristic	Input data unit				
	Continuous data	Discrete data	Inference rule	Time sequence	Predefined parameter
Causes and effects			✓		
Customer background data		✓			
Customer transaction data		✓		✓	
Manufacturing data		✓			
Number of classes to be classified					✓
Product basic information		✓			
Product sales data		✓		✓	
Stock historical data	✓			✓	

Table 15
Mapping processing outcome to output data unit.

Processing outcome	Output data unit					
	Set	Predefined class	Probability distribution	Weighted sum	Cluster	Rule
Calling pattern		✓				
Credit rating report		✓				
Customer profile	✓					✓
Defects distribution					✓	✓
Exception report		✓				
Loyalty report		✓				
Pricing model		✓				
Product mix	✓					
Product profile	✓					
Prospect list		✓	✓		✓	
Replenishment decision				✓		
Revenue estimate report		✓				
Stock price trend prediction	✓					

application to data mining method to some extent. The corresponding concepts in the literatures are listed in Table 13.

Processing characteristics related to processing step are another key point of mapping. Characteristics in business perspective contain interested data area or additional consideration of processing. Input data units include data type or necessarily predefined parameter of a data mining method, algorithm step can not proceed if input data exceeds algorithm capability. Processing characteristics in business view usually mix the concepts of data content and data type together so we have to divide these concepts apart to be the consideration of selection decision. The essential natures of processing characteristic are expressed by mapping characteristics to technical definitions which are input data units of data mining method. The relations between processing characteristics and input data unit are presented in Table 14.

Processing outcome is the last key which links business decision and application to data mining method. In different business decision and applications, required information differs due to different domain data definition and analytical objectives. In fact, some outcomes from different processing steps use similar measures and forms, so we combine various kinds of processing outcomes into 13 typical ones. Output data units of data mining method gathered from the literatures are also consolidated into six types. The relation between processing outcome and output data unit is shown in Table 15.

3.4. A selection model of data mining methods

3.4.1. Research structure

Business decision and application analysis has been decomposed into four main parts. Data mining algorithm is described by four major concepts. After the analysis process, the business requirements and algorithm functions link together to address the algorithm selection problem. Three tasks including business decision and application analysis, data mining algorithm analysis, and data analysis are performed to develop the decision model. The business decision and application is divided into conceptual analysis and operational analysis. Conceptual analysis contains the steps of business decision and application, business problem, and problem processing. The result of conceptual analysis is transformed to operational analysis where problem characteristics are captured. The algorithm conceptual analysis has properties of data mining category, input data units, and output data units. Business decision and application operational analysis and algorithm conceptual analysis will match in the data mining selection process. Thus, the algorithm operational analysis checks the algorithm steps. Data used in the algorithmic process is collected from data sources such as data warehouse, data mart, and operational database are shown in Fig. 1.

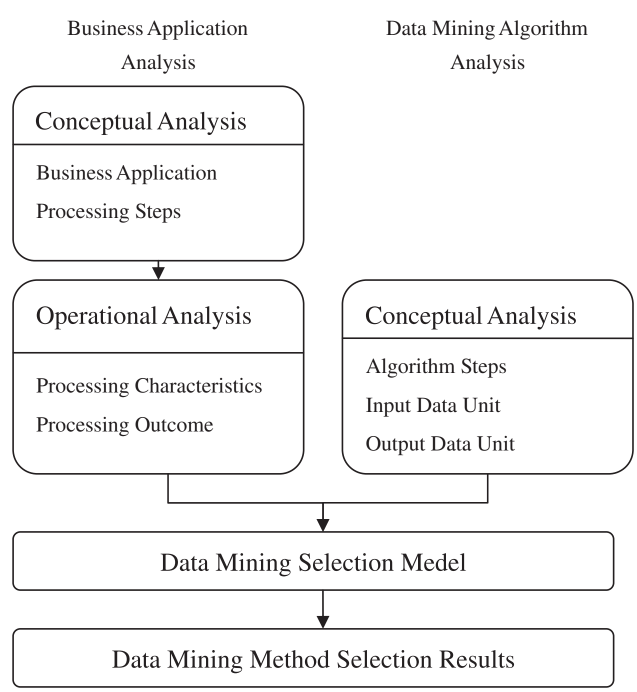

Fig. 1. Research structure.

Table 16
A selection model example step one.

Business application	Processing steps
Direct marketing	1. Classify customers 2. Match campaigns to potential customers

To show as a straight through example, we select the direct marketing application. The direct marketing is to upraise the response rate of marketing activities by recognizing customer features and bringing out customized marketing campaigns. The selection model starts from the business decision and application conceptual analysis then the processing of characteristics, then the mapping to features to determine the applicable data mining method are shown in Tables 16 and 17.

1. Direct marketing is solved by two business processing steps.
2. Each processing step needs to produce an outcome to match the required information type and the processing characteristics.
3. Now, characteristics of business decision and application are obtained in order to select a proper algorithm. Business decision and application analysis maps to the features of data mining by the following three steps.
   (Step 1) There are four algorithms matching the processing step as shown in Table 18.
   (Step 2) Bayesian network is ruled out because this application does not provide inference rule as shown in Table 19.
   (Step 3) Hence, only decision tree provides segmentation and rule to distinguish customers as shown in Table 20.

3.4.2. Business side

Business side functions are described by the order in the selection process. Each main component is abbreviated to save space.

Business application (BA) is an activity in a certain business sector. BA can be partitioned into several problem processing steps (PP). The relation is presented in Function (Function 1).

$$BP \rightarrow nPP \tag{Function 1}$$

Table 17
A selection model example step two.

Processing step	Processing outcome	Processing characteristics
1. Classify customers	Customer background data Customer transaction data	Customer profile
2. Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list

Table 18
A selection model example step three.

Business side			Mining side			Selection result
Processing steps	Classify customers	→	Algorithm steps	Classification	→	Decision Tree Bayesian belief network Neural network *K*-nearest neighbor

Table 19
Selection model example step four.

Business side			Mining side			Selection result
Processing characteristics	Customer background data, customer transaction data	→	Input data unit	Discrete data, time sequence	→	Decision Tree Neural network *K*-nearest neighbor

Table 20
Selection model example step five.

Business side			Mining side		
Processing outcome	Customer profile	→	Output data unit	Rule	Decision Tree

The arrow represents "decide" relation, and "*n*" denotes the number of PP steps. Function (Function 1) illustrates one business application can decide multiple processing steps. The n value depends on the application requirements.

During processing, there is much data involved which is called processing characteristic (CH). Problem characteristics are the condition of business problem. They are prerequisite of problem solving are decided by analyst or expert. Each processing step usually takes multiple characteristics. The relation is presented in Function (Function 2).

$$PP \rightarrow nCH \tag{Function 2}$$

Every processing step must generate an output according to analytic input. Each processing step has a processing outcome (PO) which is presented in Function (Function 3).

$$PP \rightarrow PO \tag{Function 3}$$

3.4.3. Mining side

Mining side algorithms are defined by four components. The data mining method (DM) is identified by three characters Algorithm Steps (AL), Input Data Unit (IN), and Output Data Unit (OT). The expression is presented in Function (Function 4).

$$\mathrm{DM} = \{\mathrm{dm}(x,y,z) | x \in \mathrm{AL}, y \in \mathrm{IN}, z \in \mathrm{OT}\} \tag{Function 4}$$

The "dm" is an instance of the set of data mining methods. It has three characters including x, y, and z. The character x belongs to the set AL, y belongs to the set IN, and z belongs to the set OT. By combining these characters together, we can choose the algorithm.

The algorithm step is the problem type that a data mining method can solve. In Function (Function 5), the object "al" is an instance of the set AL and "al" can be found in AL only if the Algorithm Steps (AL) which cites IN exists. AL is presented in Function (Function 5).

$$\mathrm{AL} = \{\mathrm{al}(\mathrm{IN}, \mathrm{AL}) | \mathrm{AL}(\mathrm{IN}) \text{ is true}\} \tag{Function 5}$$

Output Data Unit is the output and format of an algorithm. OT comes from the operation of AL referencing IN. It is presented in Function (Function 7).

$$\mathrm{OT} = \mathrm{AL}(\mathrm{IN}) \tag{Function 7}$$

3.4.4. Selection model

Drawing out from the business side and mining side, we are able to develop Function (Function 15).

let $\quad BA = (PP, CH, PO)$
then
$$DM = \{dm(AL, IN, OT) | PP \in AL, CH \in IN, PO \in OT\} \tag{Function 15}$$

The element "algorithm" in Function (Function 15) is an instance of the set of Algorithms. CH, PO, PT come from the business application analysis. If $CH \in IN$, it means CH can be found in the set of algorithm of IN. By combing the business side and the mining side analysis, a matched algorithm is selected.

4. Prototype implementation

In this section, we describe a prototype that realizes the theoretical design and illustrates the feasibility of selection model. We describe the design and development of the prototype in the section.

4.1. Prototype system design

4.1.1. Database design

Database design is critical because we use data to characterize data mining methods and business applications. Every data mining method can be described by AL, IN, and OT. Every business application can be described by PP, CH, and PO. Business application and data mining method are the two main entities in Fig. 2. Data mining method is defined by three attributes of algorithm step, input data unit, and output data unit. There are three weak entities of processing step, processing characteristic, and processing outcome. Processing step depends on business application. Processing characteristic and processing outcome depend on processing step. There are also three "map to" relationship between processing characteristic and input data unit, processing step and algorithm steps, processing outcome and output data unit. The number besides these "map to" relationship means, for example, a processing

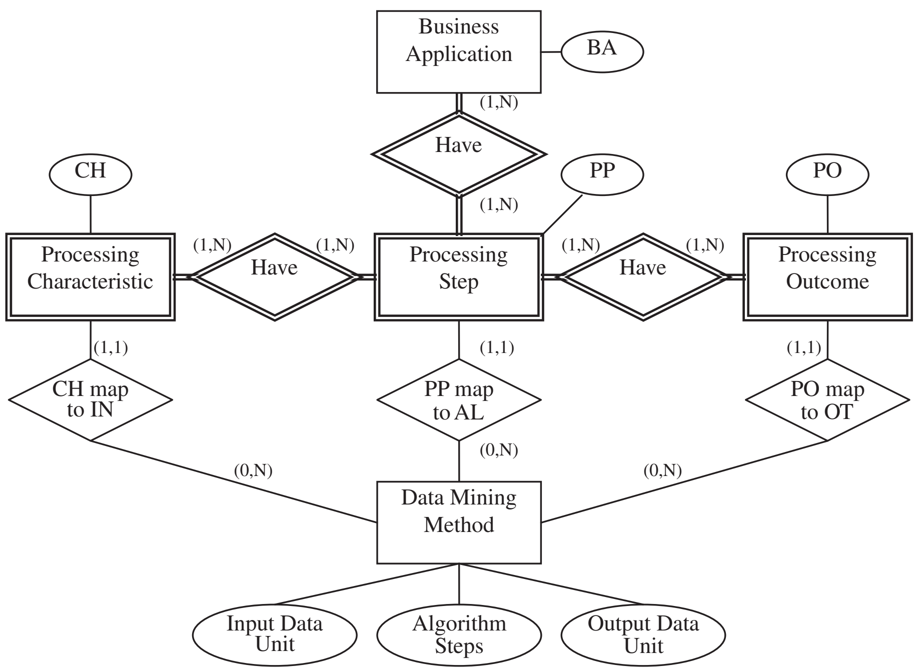

Fig. 2. ER diagram for selection model database.

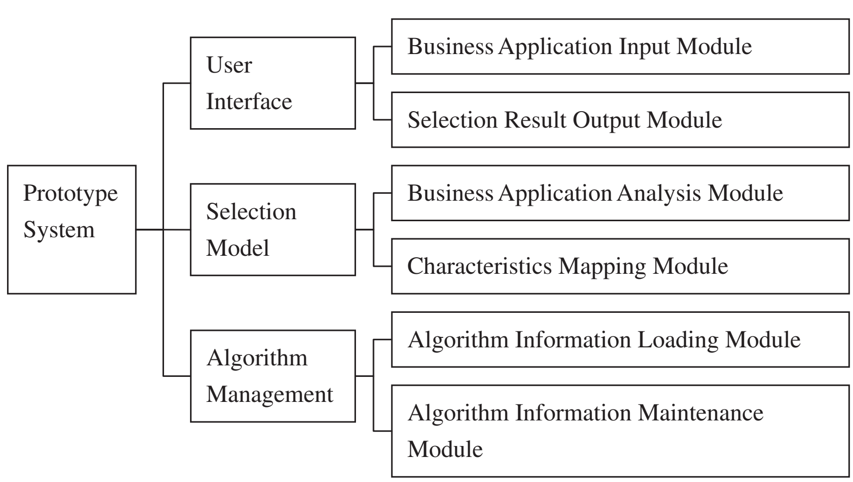

Fig. 3. Prototype system modular design.

characteristic can find zero or multiple corresponding input data unit on the other side is shown in Fig. 2.

4.1.2. Modular design

Modular design includes user interface, selection system, and algorithm management system. User interface handles business decision and application information input and selection result output. Selection model deals with the analysis tasks. Algorithm management provides database access and data manipulation capabilities. Algorithm definition and mapping rules are important to selection model are shown in Fig. 3.

Table 21
Test cases analysis.

Business application	Processing steps	Processing characteristics	Processing outcome
Cross-selling	Find relationship of customers' characteristics	Customer background data Customer transaction data	Customer profile
	Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list
Telecom link analysis	Classify customers	Customer background data Customer transaction data Causes and effects	Calling pattern
	Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list
Product yield analysis	Classify product defects	Manufacturing data	Defects distribution
Churn analysis	Classify customers	Customer background data Customer transaction data	Customer profile
	Predict the tendency of churn	Customer background data Customer transaction data	Customer profile
Direct marketing	Classify customers	Customer background data Customer transaction data	Customer profile
	Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list
Churn analysis	Predict the tendency of churn	Customer background data Customer transaction data	Customer profile
	Winback customers	Customer background data Customer transaction data	Prospect list
Cross-selling	Find relationship of products in transactions	Customer background data Customer transaction data	Product profile Product mix
Segmentation analysis	Classify customers	Customer background data	Prospect list
	Match campaigns to potential customers	Customer background data	Prospect list
Inventory analysis	Find relationship of products in transactions	Product basic information Product sales data	Product profile
	Make replenishment decision	Product basic information Product sales data	Replenishment decision
Product mix analysis	Classify customers	Customer background data Customer transaction data Number of classes to be classified	Product profile
	Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list
	Estimate the revenue of product mix	Customer background data Customer transaction data	Revenue estimate report
Stock movement prediction	Predict stock price trend	Stock historical data	Stock price trend prediction
Fraud detection	Classify customers	Customer transaction data	Customer profile
	Detect unusual activities	Customer transaction data	Exception report
Charge making	Classify customers	Customer background data Customer transaction data	Customer profile
	Match campaigns to potential customers	Customer background data Customer transaction data Product basic information	Pricing model
	Estimate the revenue of pricing model	Product basic information	Revenue estimate report
Credit rating	Classify customers	Customer background data Customer transaction data	Customer profile
	Credit rating customers	Customer background data Customer transaction data	Credit rating report
Cross-selling	Classify customers	Customer background data Customer transaction data	Customer profile
	Match campaigns to potential customers	Customer background data Customer transaction data	Prospect list
Fraud detection	Classify customers	Customer transaction data	Prospect list
	Detect unusual activities	Customer transaction data	Exception report
Churn analysis	Find relationship of customers' characteristics	Customer transaction data	Customer profile
	Classify customers	Customer transaction data	Prospect list
	Winback customers	Customer transaction data	Prospect list

Table 22
Experiment result comparison.

Business application	Processing steps	Original decision	Selection result	Comparison result
Cross-selling	Find relationship of customers' characteristics	Apriori	Apriori	Same
	Match campaigns to potential customers	Decision tree	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Telecom link analysis	Classify customers	Bayesian belief network	Bayesian belief network	Same
	Match campaigns to potential customers	Bayesian belief network	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Product yield analysis	Classify product defects	Neural network	Neural network, *k*-means	Superset
Churn analysis	Classify customers	Decision tree	Decision tree	Same
	Predict the tendency of churn	Decision tree	Decision tree	Same
Direct marketing	Classify customers	Decision tree	Decision tree	Same
	Match campaigns to potential customers	Decision tree	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Churn analysis	Predict the tendency of churn	Decision tree	Decision tree	Same
	Winback customers	Decision tree	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Cross-selling	Find relationship of products in transactions	Apriori	Apriori	Same
Segmentation analysis	Classify customers	Neural network	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-means, *k*-nearest neighbor, logistic regression, neural network	Superset
	Match campaigns to potential customers	Neural network	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Inventory analysis	Find relationship of products in transactions	No	Apriori	Superset
	Make replenishment decision	Neural network	Neural network	Same
Product mix analysis	Classify customers	*k*-means	*k*-means	Same
	Match campaigns to potential customers	Decision tree	Decision tree	Same
	Estimate the revenue of product mix	No	No	Same
Stock movement prediction	Predict stock price trend	Apriori	Apriori	Same
Fraud detection	Classify customers	Neural network	Decision tree	Different
	Detect unusual activities	Neural network	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Charge making	Classify customers	Logistic regression	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-means, *k*-nearest neighbor, logistic regression, neural network	Superset
	Match campaigns to potential customers	Logistic regression	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
	Estimate the revenue of pricing model	Logistic regression	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Credit rating	Classify customers	Decision tree	Decision tree	Same
	Credit rating customers	Decision tree	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Cross-selling	Classify customers	Decision Tree	Decision tree	Same
	Match campaigns to potential customers	No	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-means, *k*-nearest neighbor, logistic regression, neural network	Superset
Fraud detection	Classify customers	*K*-nearest neighbor	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-means, *k*-nearest neighbor, logistic regression, neural network	Superset
	Detect unusual activities	No	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset
Churn analysis	Find relationship of customers' characteristics	Association rule algorithm	Apriori	Same
	Classify customers	Classification algorithm	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-means, *k*-nearest neighbor, logistic regression, neural network	Superset
	Winback customers	No	Bayesian belief network, case-based reasoning, decision tree, generic algorithm, *k*-nearest neighbor, logistic regression, neural network	Superset

Table 23
Comparison result.

Comparison result	Percent
Same	44.12
Superset	52.94
Subset	0
Different	2.94

5. Research experiment

5.1. Experimental design

Experiment is designed to prove the validity of the selection model. Since the research model is developed using literature analysis. The validity appraisal is to enter the training and new literature to examine if the selection result matches the answer given in the literature. A correctness measurement is provided at the end of experiments.

5.2. Test cases

Fourteen literatures were reviewed in the literature review to be used as the test cases. To be able to enter data into selection model, we first analyze the test cases in table. After entering information in Table 21 to selection model, the selection results are collected in Tables 22 and 23.

We aggregate the result of the test cases and prepare table. The correctness appraisal is described as follows. Based on the observation, we achieved 97% coverage of business decision and application can be matched with applicable data mining method.

1. *Same*: Business applications can find the most suitable data mining method.
2. *Superset*: Business applications without specific requirements can find several applicable data mining methods.
3. *Subset*: There is no such kind of literatures in test cases.
4. *Different*: The percentage 2.94% means only one out of 14 test cases is correct.

6. Research implication and conclusion

6.1. Research implication

The objective of this research is to enable decision makers to find most suitable data mining method. From the literature review, we discover that most algorithm researches skip conceptual analysis of business problems. It is hard for users to understand algorithms. The selection for business decision and application needs a practical framework which provides straightforward analysis to find the right method. To solve this problem, we develop an approach to select data mining method by mapping business decision and application characteristics and data mining features. In the approach, the business decision and application analysis and mining analysis are conducted. In the model, business decision and application is characterized by processing step, processing characteristic, and processing outcome. Each step in the selection model has its operational definition. In this model, data mining capability is described by the algorithm steps, input data unit, and output data unit. The relation type which an algorithm can find is defined as the algorithm steps. The data type which an algorithm can process is defined as the input data unit. The analysis result type which an algorithm can generate is defined as the output data unit. These three types map to business decision and application characteristics. The model decomposes the algorithms into four parts. They are data mining method, algorithm steps, input data unit, and output data unit.

6.2. Conclusion

Each business decision and application is described by four parts. They are business decision and application, business problem, problem processing steps, problem characteristics. Each data mining algorithm is depicted by four parts. They are data mining method, input data unit, output data unit, and algorithmic step. After systematically mapping the features of business decision and application and the characteristic of data mining method, the selection model generates the matched set of choice. In terms of future research work, the following study will be continuously carried out. More practical selection can be achieved by considering the features of data side. Extension of data side enables precise description of mining procedure. Data side means analysis in data type, data format, data structure, and data definition.

Acknowledgement

This research is partially sponsored by National Science Council research Grant No. NSC95-2416-H-004-006.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD*, Washington.

Ahn, J. H., & Ezawa, K. J. (1997). Decision support for real-time telemarketing operations through Bayesian network learning. *Decision Support Systems, 21*, 17–27.

Anand, S. S., Patrick, A. R., Hughes, J. G., & Bell, D. A. (1998). A data mining methodology for cross-sales. *Knowledge-Based Systems, 10*, 449–461.

Bansal, K., Vadhavkar, S., & Gupta, A. (1998). Neural networks based data mining applications for medical inventory problems. *International Journal of Agile Manufacturing, 1*(2), 187–200.

Berry, M. J. A., & Linoff, G. (1997). *Data mining techniques for marketing, sales, and customer support*. John Wiley & Sons Press.

Bose, R. (2002). Customer relationship management key components for IT success. *Industrial Management & Data Systems, 102*(2), 89–97.

Bose, R. (2006). Intelligent technologies for managing fraud and identify Theft. In *Proceedings of the third international conference on information technology new generation* (pp. 446–451).

Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In *Proceedings of the 11th IEEE international conference on tools with artificial intelligence, Chicago* (pp. 103–106).

Chan, P. K., Fan, W., Prodromidis, A. L., & Stolfo, S. J. (1999). Distributed data mining in credit card fraud detection. *Intelligent Systems and Their Applications, 14*(6), 67–74.

Davidson, A., & Simonetto, M. (2005). Pricing strategy and execution an overlooked way to increase revenues and profits. *Strategy & Leadership, 33*(6), 25–33.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39*(11), 27–34.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data dining*. Cambridge AAAI/MIT Press.

Forcht, K. A., & Cochran, K. (1999). Using data mining and data warehousing techniques. *Industrial Management & Data Systems, 99*(5), 18–19.

Gardner, M., & Bieker, J. (2000). Data mining solves tough semiconductor manufacturing problems. In *Conference on knowledge discovery in data* (pp. 376–383), Boston: ACM Press.

Gavrilov, M., Anguelov, D., Indyk, P. & Motwani, R. (2000). Mining the stock market which measure is best? In *Conference on knowledge discovery in data*, Boston, MA, USA (pp. 487–496).

Han, J., & Kamber, M. (2001). *Data mining concepts and techniques*. Academic Press.

Lee, T. S., Chiu, C. C., Chou, Y. C., & Lu, C. J. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis, 50*, 1113–1130.

Lejeune, M. A. P. M. (2001). Measuring the impact of data mining on churn management. *Internet Research Electronic Networking Applications and Policy, 11*(5), 375–387.

Lu, H. J., Han, J., & Feng, L. (1998). Stock movement prediction and *N*-dimensional inter-transaction association rules. In *ACM SIGMOD workshop on research issues on data mining and knowledge discovery*, Seattle (pp. 12.1–12.7).

Min, Hokey, Min, Hyesung, & Emam, Agmed (2002). A data mining approach to developing the profiles of hotel customers. *International Journal of contemporary Hospitality Management, 14*(6), 274–285.

Netessine, S., Savin, S., & Xiao, W. (2004). Revenue management through dynamic cross-selling in e-commerce retailing. Working paper.

Pitta, D. A. (1998). Marketing one-to-one and its dependence on knowledge discovery in databases. *Journal of Consumer Marketing, 15*(5), 468–480.

Povinelli, R. J., & Feng, X. (1999). Data mining of multiple non-stationary time series. In *Proceedings of artificial neural networks in engineering*. St. Louis, Missouri (pp. 511–516).

Pritscher, L., & Feyen, L. (2001). Data mining and strategic marketing in the airline industry. *Data Mining for Marketing Applications*.

West, D. (2000). Neural network credit scoring models. *Computers & Operations Research, 27*, 1131–1152.