Mining Customer Behavior Knowledge to Develop Analytical Expert System for Beverage Marketing

Chun Fu Lin #1, Yu Hsin Hung #2, and Ray I Chang #3

*1#3 Department of Engineering Science and Ocean Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan (R.O.C.)

#2 Information Technology Office, National Taiwan University Hospital, No.7, Chung Shan S. Rd., Taipei 10002, Taiwan (R.O.C.)

Abstract—**Consumer relationship management (CRM) requires detailed information and business knowledge for successful adoption. Data mining techniques are widely used in business administration, the financial industry, and marketing. Mining techniques provide decision administration reference for enterprises by integrating useful information and discovering new information from different perspectives. In this study, we applied data mining technique and statistics and utilized questionnaires in CRM to analyze customer behavior. The Chinese tea market is famous worldwide, customizing the tea service is a special trend in chain stores, and customer behavior analysis is essential for the tea market. This study aims to develop a customer behavior analysis expert system (CBAES) in which a decision tree is used to identify relevant knowledge and personalize merchandise based on association rule framework of consumer behavior analysis in chain store beverage marketing. Identifying consumers' preferences and providing optimal purchase strategy using this approach is a helpful characteristic of customers and facilitates marketing strategy development.**

Keywords—**Consumer relationship management, expert system, decision tree algorithm, marketing**

I. Introduction

Customer relationship management (CRM) has attracted a large amount of attention from researchers because it can improve customer satisfaction and loyalty while increasing the profitability of the enterprise. For example, customer behaviour analysis and marketing these business strategies are required to optimize customers' satisfaction. Technical support for customer behaviour analysis can not only reduce the risk of marketing but also the cost of the client service. Accordingly, facilitating personal service via analysing customer behaviour to optimize customers' satisfaction has become a popular trend in CRM. Therefore, the application of data mining techniques in personalizing service aims at providing the customer with customized purchase recommendations through interactions and, further, to meet the requirements of different customers.

Data mining techniques which utilize methods at the intersection of machine learning, database administration, and statistics have been applied widely to marketing [1] and behavioural analysis [2]. The mining mechanism can extract specific patterns from large datasets and transform the mined data into a comprehensible structure that has further utility. Data mining has been recently applied to develop business intelligence. The dramatically increasing amount of customer information has also increased the complexity and difficulty of marker-based decision making for enhancing customization services. Using mining techniques, purchase information or customer behaviour information can be integrated as useful market knowledge which can assist in predicting customer preferences or needs.

Chinese tea has been exported in great quantity and it is famous around the world. To date, this industry extends into developing various tea products, fruits, syrup, milk, or local specialty foods such as tapioca balls mixed with the traditional Chinese tea beverage, the typical example of which is pearl milk tea. The taste can also be adapted according to customer preferences. A customized tea service that provides personalized tea merchandise to the customers of all ages has therefore become an innovation in tea enterprises in Taiwan. A large tea enterprise may establish hundreds of tea beverage chain stores to provide this customized tea service. For these enterprises, customer preference analysis is essential for making effective marketing strategies, as well as to maintain a stable customer relationship. Accordingly, we take tea beverages as an example of developing a customer behaviour analytical expert system (CBAES) in which data mining techniques are integrated to facilitate knowledge of the customization process. Web techniques are used to empower an expert system to interact with customers to successfully attract their interest and efficiently maintain the customer relationship.

II. Relate work

Marketing is the management process through which products and services move from concept to customer. McCarthy proposed the import theory-marketing mix 4P's model, Marketing includes the coordination of "Place", "Product", "Promotion", and "Price" [3] Customer relationship management (CRM) manages the enterprise's interactions with current and future customers, and it also applies the most important rules in the 4P model [4]. Efficient CRM could improve the quality of products or service, enhance additional value, and inspire purchasing interest for the enterprise to secure more benefit. This stream of papers focuses on enhancing customer satisfaction by developing marketing service strategies based on CRM. Recently, customer behavior has been used as a basis for pricing and inventory management [5] [6] and customer preference analysis has been claimed as a necessity for effective CRM [7]

. However, customers' purchase behavior can be influenced by both external environmental factors [8] [9] and internal mental status [10] causing difficulty in evaluation. Therefore, the psychological theory of needs has been employed in CRM based on the realization that customers' preferences and requirements are critical for business revenue and product promotion. According to Abraham Maslow's (1943) hierarchy of needs, there are five needs in human beings: physiological, safety, love and belonging, esteem, and self-actualization. These needs can be extended to manage customer behavior and, further, to enhance customer satisfaction. In this study, we try to incorporate this need theory into our CBAES. Specifically, the proposed CBAES is aimed to assistant customers achieve self-actualization in choosing a favorite beverage based on their preferences and personal decisions. During the process, physiological needs are satisfied via online shopping, safety needs are achieved through artificial intelligence (AI) computing, love and belonging needs are satisfied through the consideration of emotions, and esteem needs are achieved through personalized services which take personal preferences into account.

In this study, we take Chinese tea beverage market as an example to illustrate how customers can automatically get inferences for purchasing suggestions using data mining techniques. Data mining is an effective tool for data preparation and pattern discovery in various applications [11]. The four main pattern discovery methods employed in data mining technologies are association rules, sequential patterns, clustering, and classifications. Discovering new patterns from different perspectives through integrating information can provide decision-makers with administrative reference points. With advances in commerce, the importance of business information mining is becoming greater. There is an increasing interest in employing data mining in marketing strategy decisions, which makes data mining the focus of a new and growing research community [12]-[14]. In particular, the decision tree technique has been suggested as an effective vehicle for meeting individual requirements. For example, clustering and association rule algorithms have been applied in promotion strategies and feature selection has been employed to increase individual purchase satisfaction and to minimize unnecessary information effects on decision making. This study therefore aims at developing CBAES, in which the decision tree technique is employed to provide adaptive purchase recommendations for customers with varied backgrounds and personal traits. This study aims at exploring customers' preferences as well as providing them with online suggestions based on their behavior and experience when they are purchasing a product. To enhance such a convenient service and purchasing self-efficacy, we develop CBAES based on the web technique and the agent technique.

III. METHODOLOGY

In CBAES, theories of need hierarchy, customer behavior theory, agent technology and data mining were employed to infer customers' preferences when purchasing a tea beverage. The architecture of CBAES is illustrated in Fig. 1. In CBAES, the decision tree algorithm and knowledge management (KM) are employed to develop business intelligence for analyzing customer preferences and optimizing purchase suggestions. More specifically, the recommendation learning module saves recommended rules into the knowledge database via XML technique, and marketing information is analyzed by the decision tree algorithm. In addition, the agents in CBAES act as a medium to feedback information to the mobile device, allowing the customer to obtain real-time information. Accordingly, CBAES can personalize purchase service procedures as well as provide suggestions based on input information.

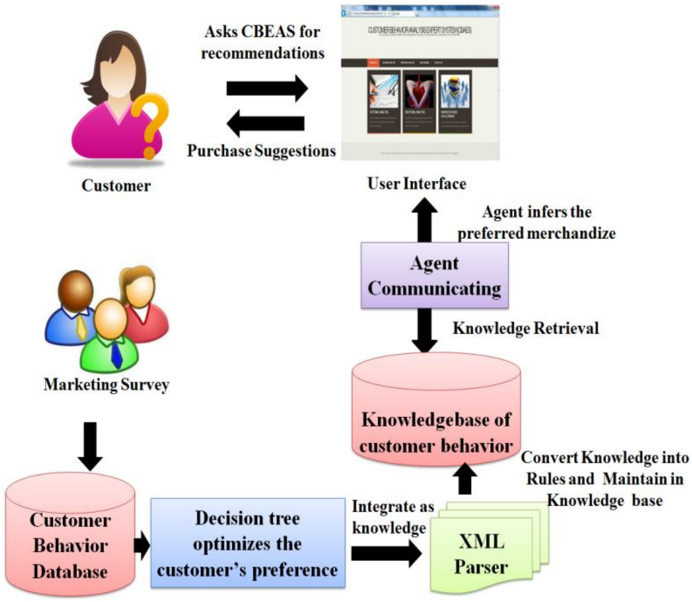

Fig. 1 The architecture of CBAES.

A. *Customer behavior analysis expert system (CBAES)*

First, the external customer characteristics, emotional characteristics, and purchase information are collected in the customer behavior database using a questionnaire survey. The survey results serve as input information for problem evaluation. Next, customer preference problems are resolved using a decision tree algorithm (See Fig. 2). Node selection is essential for the tree split process when a node divides rows into two child nodes. When all possible outcomes and actions find a place in the decision tree, the gain ratio is employed to determine which decision path leads to the most desirable outcome. Feature selection empowers each tree to choose the adaptive node using a mathematical evaluation method to improve the performance of the algorithm.

Pseudo-code of the preference analysis application is described as follows

Input: Customer behavior Variables v_i (i = the variable attribute)
Output: The optimal path of customer preference rules *CP*
Check for base cases
For each variable v
Find the normalized information gain(GR) from splitting on v
Let v_{best} be the attribute with the highest normalized information gain
Create a decision node that splits on v_{best}
Subsists CP obtained by splitting on v_{best}, and add those nodes as children of node
While maximum iterations or achieved the objective
return *CP*
End

Fig. 2 Pseudo-code for customer preference decision tree algorithm.

Feature selection improves the performance of the preference analysis optimization by choosing the most significant features, and enhances accuracy by analyzing correlations between significant features and the forecast variable. In the feature selection module, parameters include the tea merchandise type, the combinational beverage preference, the gender, and the emotional status. The forecast variable was tea merchandise and enterprise. To normalize features, the following formula for gain ratio (Equation 1) was used to calculate the ratio between each feature and the forecast variable. Here *GR* stands for Gain ratio, *p* presents probability and *E* means entropy.

$$GR = \frac{(E_{before} - E_{after})}{Split\ Gain}$$

$$St. \quad E = \sum_{i=1}^{e} -p_i \log p_i \qquad \text{Equation 1}$$

In this step, the best combinational decision path is filtered based on the expected values. Moreover, path evaluation results are maintained in the purchase knowledge database. The algorithm is repeated or the present optimal result is output depending on whether the process has achieved the main goal or the maximum amount of data. Recursive training improves intelligence of the computing operation to enhance optimization and the optimal result is integrated as knowledge. Then XML parser converts this knowledge into a rule using XML format and maintains it in the knowledge base. When the customer asks for a recommendation, the agent retrieves the knowledge based on the customer's characteristics to infer the customer's preference in real-time. In CBEAS, the user can obtain real-time purchase recommendations on a mobile device, and the dynamic automatic inference preference can enhance convenience.

B. System design of CBAES

Fig. 3 depicts the user interface of CBAES which consists of three parts: knowledge development and external and emotional analysis. The customer's behaviour is determined by the individual emotion variable and the external variable. Development of the knowledgebase is as demonstrated in Fig. 4. The CBAES establishes a multi-environmental analysis including external and emotional customer characteristics. Data from 784 participants is employed in the decision tree algorithm to mine useful information (See Fig.5).

Fig. 3 User Interface of CBAES.

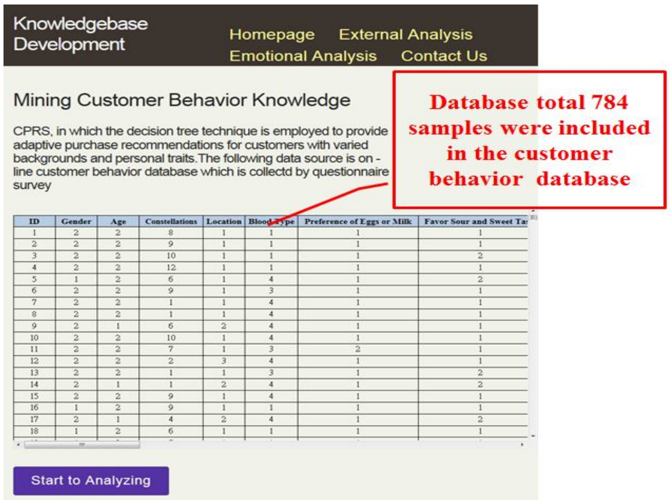

Fig.4 Knowledgebase conduction.

In CBAES, the agent interacts with the customer and the knowledge base to infer the result, based on individual characteristics such as the external variable. The customer wants analysis with external analysis, and therefore they filled the questionnaire (See Fig. 6) and provide agent feedback to the knowledge base to optimize the solution (See Fig. 7). Fig. 8 demonstrates the emotional state and agent response for the appropriate recommendation.

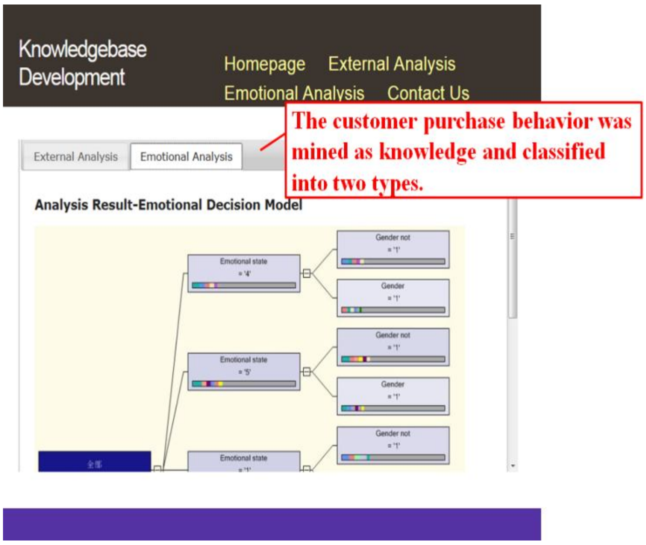

Fig. 5 Decision tree algorithm and mining the information.

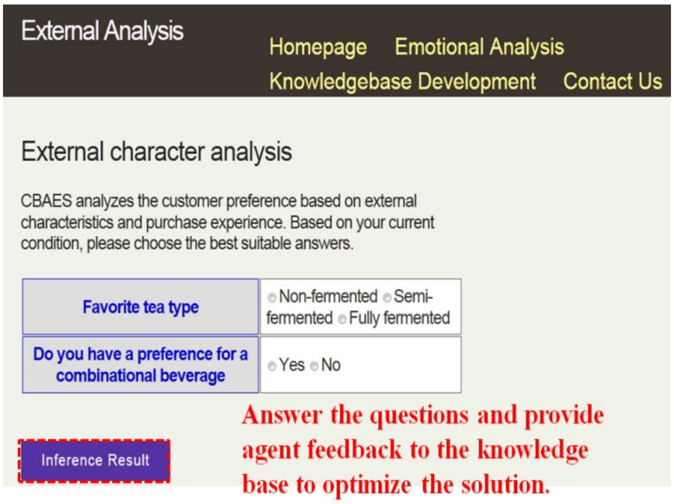

Fig.6 User Interface for providing external analysis.

Fig. 7 The inference result of external analysis.

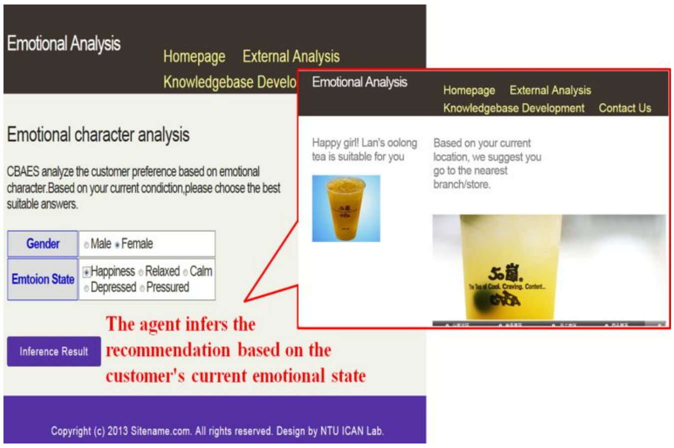

Fig. 8 The emotional analysis and the outcome.

IV. EXPERIMENT

A. *Experimental design*

This study aims at mining knowledge to develop the knowledgebase for the agent retrieval process. The data source of customer purchase behavior is therefore essential for evaluating preferences. We collected data from 758 participants via a marketing survey. The data was saved in the customer behavior database as training guidelines. The collected information is displayed in Table 1. For the purpose of adapting the inference result in real cases, the psychological variables, demographic variables, and purchases experiences were included in the research questions (see Table 1) .

TABLE I
THE VARIABLE OF CUSTOMER PREFERENCE SURVEY.

Variables	**Description**
Personal traits-Age	18 – 26 years old (mean = 21)
Gender	Male (type = 0), Female (type = 1)
Emotional state	Relaxed (type = 2) Calm (type = 3) Depressed (type = 4) Pressured (type = 5)
Has a preference for a combinational beverage	Yes (type = 1) No (type = 2)
Favorite tea type	Non-fermented (type=1) Semi-fermented (type=2). Fully fermented (type=3)
Favorite beverage product	Beverage (type 1 –type 12)
Favorite beverage enterprise	Enterprise (type 1 –type 8)
Frequency	Numerical

B. *Feature selection of purchase behaviour*

Findings from the training data reveal that the four features employed in this study have a gain ratio greater than 0.05. The order of the features is as follows: tea type, the preference for a combinational beverage, emotional state, and gender (See Table 2).

TABLE II
THE GAIN RATIO RESULTS

Variables		Gain Ratio	Priority
External	Favorite tea type	0.117	1
	Has a preference for a combinational beverage	0.017	2
Emotional	Emotional state	0.015	3
	Gender	0.008	4

C. *The result of decision tree analysis*

Feature selection results were defined as inputs and the purchase suggestion was defined as the output in the decision tree algorithm. The dataset of 758 participants was employed to identify purchase preferences in this study. Tea type and the preference for a combinational beverage are important external variables, whereas gender and emotional state are emotional variables for customer behaviour. Therefore, we chose these variables as features and included them in the decision tree algorithm separately to examine their relationship with customer preferences. The computation result was integrated in the form of a decisions model and a rule. The rule consisted of the optimal enterprise and product based on the customer preference which adapts to the emotional or external environment.

1) *External character decision model:* Gain ratio was used to determine decision priority. Therefore, tea type (T_{type}) has priority in entering the decision making module. Fig. 9 illustrates decision results of the purchase suggestion and external features. Next, the second highest ratio variable–preference feature (R_{type}) was applied to the first processing result as a branch guideline to create nodes. The decision model was obtained after completing the decision algorithm (See Fig. 9), and we then integrated the optimal enterprise (E_{type}) and products (P_{type}) as a rule for CPBAES to infer customer preference using their external environment (See Table 3). For example, if a customer prefers non-fermented tea and combinational beverages, then the agent infers the optimal suggestion to be “oolong tea” from enterprise A.

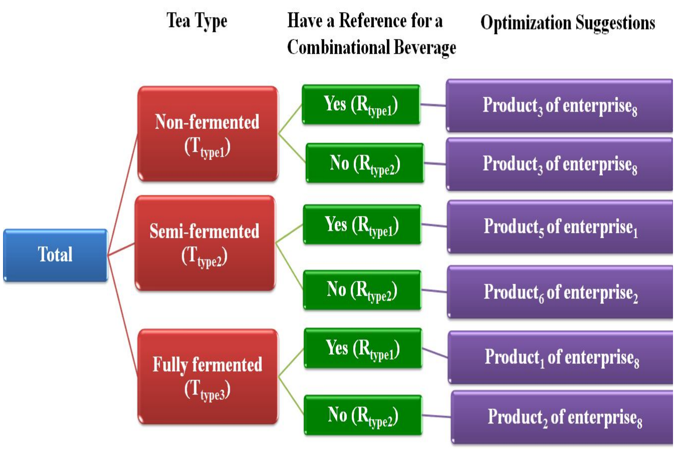

Fig. 9 External character decision model.

TABLE III
THE GENERATED RULES BASED ON EXTERNAL CHARACTER.

rule 1	If (T_{type1} ∩ P_{type1}) → (E_{type3} ∩ P_{type8})
rule 2	If (T_{type1} ∩ P_{type2}) → (E_{type3} ∩ P_{type8})
rule 3	If (T_{type2} ∩ P_{type1}) →(E_{type5} ∩ P_{type1})
rule 4	If (T_{type2} ∩ P_{type2}) → (E_{type6} ∩ P_{type2})
rule 5	If (T_{type3} ∩ P_{type1}) → (E_{type1} ∩ P_{type8})
rule 6	If (T_{type3} ∩ P_{type2}) → (E_{type2} ∩ P_{type8})

2) *Emotional character decision model:* In this study, we examine customer preference in the physical environment using emotional state and gender as emotional character decision model. Fig. 10 illustrates mining results of the decision tree algorithm. The results were integrated into ten rules for CBAES to analyse the customer’s current emotional states and to infer a purchase suggestion (see Table 4). This analysis result emphasizes emotional features instead of physical items compared to an analysis of the external environment. For example, if a male customer’s emotional state is happy, the CPRS then suggests that he drinks “pearl milk tea” of enterprise B. Table 4 demonstrates the rules in the potential environment.

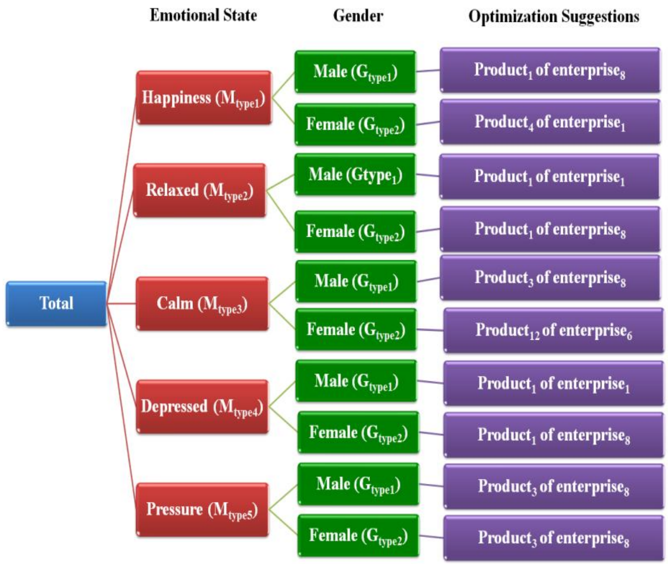

Fig. 10 Emotional character decision model.

TABLE IV
THE GENERATED RULES BASED ON EMOTIONAL CHARACTER.

rule 1	If (M_{type1} ∩ G_{type1}) → (E_{type1} ∩ P_{type8})
rule 2	If (M_{type1} ∩ G_{type2}) → (E_{type4} ∩ P_{type1})
rule 3	If (Mt_{type2} ∩ G_{type1}) → (E_{type1} ∩ P_{type1})
rule 4	If (M_{type2} ∩ G_{type2}) → (E_{type1} ∩ P_{type8})
rule 5	If (M_{type3} ∩ G_{type1}) → (E_{type3} ∩ P_{type8})
rule 6	If (M_{type3} ∩ G_{type2}) → (E_{type12} ∩ P_{type6})
rule 7	If (M_{type4} ∩ G_{type1}) → (E_{type1} ∩ P_{type1})

rule 8	If ($M_{type4} \cap G_{type2}$) → ($E_{type1} \cap P_{type8}$)
rule 9	If ($M_{type5} \cap G_{type1}$) → ($E_{type3} \cap P_{type8}$)
rule 10	If ($M_{type5} \cap G_{type2}$) → ($E_{type3} \cap P_{type8}$)

CBAES includes external and emotional environmental analysis of participants as predictors for product suggestion. The gain ratio analysis shows that the tea types, the preference for a combinational beverage, gender, and emotional state are important predictors. In this study, we used real data to illustrate how the inferred result can be employed to personalize a customer's purchase suggestions. The data was obtained from 38 participants aged 18 to 47 years. Notably, the participants had a 95% chance of obtaining above-average satisfaction for the inference result (See Table 5). Strong evidence suggests that the proposed decision tree algorithm is highly accurate in predicting the optimal purchase suggestion for customers. This supports the claim that customer behavioural analysis, based on the tree mechanism, can meet individual requirements and enhance relationship management efficiency. Successful personalized learning should sell and manufacture a product that customers need and wish to buy and take. In the CBAES, customer behaviour and information are recorded in the knowledge database, and this information can be retrieved or maintained through a knowledge agent approach. The customer can also administer the environment using knowledge management. When the customer receives feedback, the system can also receive training knowledge, which will help it adjust to determine the learner's cognitive loading status in the optimization.

TABLE V
THE SATISFACTION[1] TEST OF INFERENCE RESULTS

	Mean	**S.D**[2]	***t-value***	***p-value***[3]
Emotional infer result is reasonable	2.37	0.79	4.96	00000
Emotional infer result is interesting	2.58	0.79	3.27	00005
External infer result is reasonable	2.34	0.75	5.44	0.0000
External infer result is interesting	2.47	0.73	4.47	0.0000

[1]. Satisfactions is measured with 5-point Likert type scale with response options of strongly agree to strongly disagree (type=1~5)
[2]. S.D =Standardize Deviation
[3]. p_value <0.05

V. CONCLUSION

In this study, we selected an essential feature to discuss customer relationship management and add to the expert system. CBAES implements Taiwan's special-tea market as an example. A large amount of real cases were added in the database and mined as knowledge rules. Our experiment shows that CBAES can successfully develop the knowledge inference model and it can also be applied as an expert system for customer operation. The main contributions of the CBAES are evaluating the features related to customer purchase via decision tree algorithm and integrating the rule-based knowledgebase with the expert system. A multi-agent system uses applied knowledge to interact with the customer and personalize product suggestions to improve the efficiency of knowledge sharing and retrieval, and also enhance the convenience of human interaction with the computer. Moreover, based on the web technique, CBAES includes the integration of cross-domain knowledge and the use of real-time information sharing. The finding in this study is that the mining algorithm and rule inference aspect of the system could help customers make optimal purchase decisions online. CBAES provides immediate feedback to the customer, which could assist the related enterprise to tailor marketing strategy and optimize purchase effect. In future studies, we would expect this system to be widely used in other industries, and develop a customer preference recommendation application on a mobile device.

REFERENCES

[1] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Journal of Decision Support Systems*, vol. 50, no. 3, pp. 559-569, 2011.

[2] M. Thelwall, D. Wilkinson, and S. Uppal, "Data mining emotion in social network communication: Gender differences in MySpace," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 1, pp. 190-199, 2010.

[3] E. J. McCarthy, "Basic marketing: a managerial approach," RD Irwin, 1978.

[4] H. Tohidi, "CRM technology as a strengthening factor to business outcomes," *AWERProcedia Information Technology and Computer Science*, 1, 2012.

[5] X.Chen, and D. Simchi-Levi, "Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case," *Operations Research*, vol. 52, no. 6, pp. 887-896, 2004.

[6] Taudes, and C. Rudloff, "Integrating inventory control and a price change in the presence of reference price effects: a two-period model," *Mathematical Methods of Operations Research*, vol. 75, no. 1, pp. 29-65, 2012.

[7] T. Allard, B. Babin, J. C. Chebat, and M. Crispo, "Reinventing the branch: An empirical assessment of banking strategies to environmental differentiation," *Journal of Retailing and Consumer Services*, vol. 16, no. 6, pp. 442-450, 2009.

[8] De Nisco, and G. Warnaby, "Urban design and tenant variety influences on consumers' emotions and approach behavior," *Journal of Business Research*, In Press, 2012.

[9] Brunner-Sperdin, M. Peters, and Strobl, A , "It is all about the emotional state: Managing tourists' experiences," *International Journal of Hospitality Management*, vol. 31, no. 1, pp. 23-30, 2012.

[10] R. Kittler, and W. Wang, "Data mining for yield improvements. *Proceedings from MASM*, 2000.

[11] M. Suman, T. Anuradha, and K. M. Veena, "Direct marketing with the application of data mining," *Journal of Information Engineering and Applications*, vol. 1, no. 6, pp. 1-4, 2012.

[12] U. Marjanović, D. Graĉanin, and B. Lalić, "Web stores in Serbia: recommendation for e-business strategy implementation," *Information Technology in Management*, 968, 2012.

[13] Warf, "Global E-Commerce," In *Global Geographies of the Internet*, pp. 77-113, Springer Netherlands, 2013.

[14] J. R. Méndez, F. Fdez-Riverola, E. L. Iglesias, F. Díaz, and J. M. Corchado, "Tracking concept drift at feature selection stage in SpamHunting: An anti-spam instance-based reasoning system," *Lecture Notes in Computer Science*, 4106, pp.504-518, 2006.