DISCOVERING RULES FOR PREDICTING CUSTOMERS' ATTITUDE TOWARD INTERNET RETAILERS

Lina Zhou
Department of Information Systems
University of Maryland, Baltimore County
zhoul@umbc.edu

Wei-yu Kevin Chiang
Department of Information Systems
University of Maryland, Baltimore County
wchiang@umbc.edu

Dongsong Zhang
Department of Information Systems
University of Maryland, Baltimore County
zhangd@umbc.edu

ABSTRACT

Customers' attitude toward online shopping is the key to the survival and profitability of Internet retailers in the intensely competitive market. Aiming to automatically discover knowledge for predicting customers' attitude toward Internet retailers in relative to traditional retailers, we examine two classification approaches, decision trees and neuro-fuzzy systems, which are capable of generating such knowledge in form of rules. The neuro-fuzzy approach has rarely been investigated in the Internet retailer context. We compare the two knowledge discovery approaches using data sets for two types of products collected from an empirical survey. The results show that while the performance of the two approaches is comparable, the neuro-fuzzy model is superior to the decision tree in handling uncertainty and imprecision of the data sets. Motivated by the potential value that knowledge discovery and Web services can add to existing services, we also propose an architecture that enables agile Web services in Internet retailing. The insights on methodology and the Internet retailing application gained from this study suggest a number of interesting issues for future research.

Keywords: Internet Retailer, Decision Tree, Neuro-fuzzy Approach, Web Services, Agility

1. Introduction

With the advance of Internet technology, e-businesses are striving to reach an unprecedented large population and start to take on new forms. An e-business can be built on top of a wide range of e-business models [Singh 2002]. This paper selects business-to-consumer (B2C) Internet retailers and examines customers' attitude and patronage behavior toward them. Understanding what motivates customers to adopt and patronage Internet retailers is important because it is the key to Internet retailers' survival in the intensely competitive market. The competition comes not only from the e-commerce market, but also from alternative channels such as traditional retailers [Chen et al. 2002]. Compared with a traditional retailing environment, enticing customers to an Internet retailing environment is far more challenging. It is because that the online environment requires customers to make substantive behavioral changes in adopting and trusting e-commerce technologies and making informed decisions using technologies [Bhattacherjee 2000]. Therefore, obtaining knowledge about customers' attitude toward Internet retailers can help businesses develop effective marketing strategies for attracting and retaining customers and gain competitive advantage.

There is a growing body of literature on examining customers' acceptance of Internet shopping [Meyer and Johnson 1995; Liang and Huang 1998; Devaraj et al. 2002; Kwak et al. 2002; Shim et al. 2002; Pavlou 2003]. Several factors, including customers' perception of convenience, product offerings, production information, site design, financial security of Internet stores, and trust, are found influential to customers' satisfaction with Internet shopping [Liang and Huang 1998; Szymanski and Hise 2000; Gefen et al. 2003; Vatanasombut et al. 2004]. Undoubtedly, improving the performance of those factors can potentially attract more customers and increase the effectiveness of Internet retailing services. In reality, however, due to resource constraints such as technical feasibility, cost feasibility, and organizational feasibility, it is almost impossible for Internet retailers to improve all of the factors affecting customers' perception simultaneously. This reveals a need to focus on a subset of key attributes that have been identified in the relevant literature.

To this end, this paper takes a step toward automatically discovering knowledge for predicting customers' attitude toward B2C Internet retailers by using classification approaches. We select classification approaches based on their knowledge discovery capability in general and knowledge interpretability in particular. Rules usually consist of several *if* patterns and one or more *then* patterns [Winston 1992]. They are known as straightforward and easy to understand and verify, and thereby can be easily applied by retailers. They can represent associations and causal-effect relationships between different variables or factors. Knowledge in form of rules can reveal patterns in customers' perception of Internet retailing services. Therefore, we target at rules in this study.

Rules can be discovered with heuristic approach or empirical approach. In general, a heuristics-based method highly depends on experts' knowledge. Due to continuous evolution of business models and diversified Internet retailing services, it becomes infeasible to acquire heuristic knowledge from domain experts. In addition, such knowledge also tends to be subjective and inflexible. An empirical method aims to discover knowledge from relevant data automatically. It excels in efficiency, adaptability and cost-effectiveness. Without proper guidance, however, an empirical approach may produce a lot of 'superfluous' knowledge. The quality of data also affects the quality of discovered knowledge. Therefore, we leverage the strength of heuristics-based and empirical approaches for knowledge discovery in our solution.

Classification techniques have been widely applied to facilitating decision making in an e-business environment [Mena 1999]. The decision tree is one of the most widely used approaches to learning deterministic rules. Considering imprecise and uncertain nature of the data about Internet retailing services, a neuro-fuzzy approach is able to generate fuzzy rules. Therefore, in this study, we select two classification approaches, decision tree and neuro-fuzzy models, for predicting customers' attitude toward Internet retailers. In particular, we are interested in the following research questions:

- Which knowledge discovery approach, decision tree or neuro-fuzzy model, can predict customers' attitude toward Internet retailers more effectively?
- How to implement knowledge discovery models to make them interoperable with heterogeneous service applications?

A neuro-fuzzy model is a hybrid approach that harvests the parallel learning ability of neural networks and exploits the tolerance for imprecision, uncertainty, approximate reasoning, and partial truth of fuzzy systems [Mitra 2002]. As a result, it achieves tractable, robust, and low-cost solutions. To our best knowledge, the neuro-fuzzy approach has rarely been used to examine customers' behavior. The comparison between the performance of decision tree and neuro-fuzzy approaches will help us understand the strength and weakness of both approaches. In the light that key attributes influencing customers' attitude may vary with product types, it is highly desired that the selected approaches be evaluated on different products.

Once an effective approach to acquiring knowledge about customers' motivation to accept (or reject) Internet retailers is identified, the knowledge discovery model itself can be turned into a new type of e-service. Such a type of service extends the traditional notion of e-service [Rust and Lemon 2001] by taking on a value-adding and supporting role. In particular, it can help increase the agility of existing Internet retailing services by discovering knowledge tailored to specific product services. Moreover, the new type of e-service should be implemented in a way that makes it easy to interact with heterogeneous retailing applications. Web services have become a *de facto* standard for integrating and reusing disparate applications via open, XML-based standards [Peltz 2003]. They offer a loosely coupled and service-oriented architecture, which bridges service providers and consumers and increases the interoperability, flexibility, and accessibility of business services. Web services allow providers to focus more on function and quality instead of implementation details of e-businesses [Sutor 2004]. They are playing an increasingly important role in interacting with outside partners in B2B and B2C applications as well as cross-functional services integration within an enterprise. Therefore, we propose using Web services to deploy knowledge discovery models for Internet retailing. The benefit of Web services to service deployment, integration, and agility can lead to significant cost reduction and customer relationship improvement.

The remainder of the paper is organized as follows. In Section 2, we introduce two types of rule knowledge and corresponding knowledge discovery approaches. In addition, we select important service attributes of Internet retailers by drawing on findings from previous literature. In Section 3, we describe data collection process and analyze results of the selected approaches. Then, we present a knowledge discovery enabled architecture for agile Web services in Section 4. Finally, we conclude the findings and highlight future research directions.

2. Methodology

2.1. Deterministic vs. fuzzy rules

The antecedent as well as the consequent parts of a rule can be represented with crisp values or qualitative symbols indicating fuzziness. According to the level of involved uncertainty, rules can be classified into two categories: deterministic rules and fuzzy rules. In the former case, all variables hold crisp values. For example, a deterministic rule can be "*If a product can be delivered within* ***three*** *days then the customer will buy it"*. In the latter case, however, variables in the antecedent(s) or consequent(s) of rules are expressed with fuzziness. The above example will be restated as a fuzzy rule like *"If a product can be delivered* ***quickly*** *then the customer will buy it"*.

Compared with fuzzy rules, deterministic rules are more precise but less flexible. For example, three delivery days may be short enough for Internet customers who purchase books, but too long for Internet food customers. Even for the same type of products, the normal delivery days may vary across regions and over time. It is difficult and unrealistic to generate deterministic rules for all possible scenarios. On the other hand, fuzzy rules allow imprecision and uncertainty while representing a causal-effect relationship. Given a fuzzy rule, the interpretation of terms in a rule can be adjusted according to the context of individual retailers. Therefore, both fuzzy rules and deterministic rules are important to business decision making.

In this study, decision tree and neuro-fuzzy approaches are used to discover the above two types of rules for predicting customers' attitude toward e-services respectively. In particular, decision trees are developed to produce deterministic rules, while neuro-fuzzy systems are employed to generate fuzzy rules. The latter belongs to soft computing paradigm, which provides a flexible information processing capability for handling real-life situations [Zadeh 1994].

2.2. Neuro-fuzzy systems

Neuro-fuzzy computing combines the merits of neural networks such as massive parallelism, robustness, and the capability of learning with those of fuzzy logic approaches in modeling imprecise and qualitative knowledge and transmission of uncertainty [Mitra and Hayashi 2000]. It enables people to build more intelligent decision-making systems. Such a combination results in self-tuning neuro-fuzzy systems that are interpretable, into which prior knowledge can be inserted. It models the behavior of experts by compiling their knowledge in linguistic rules [Mitra and Hayashi 2000].

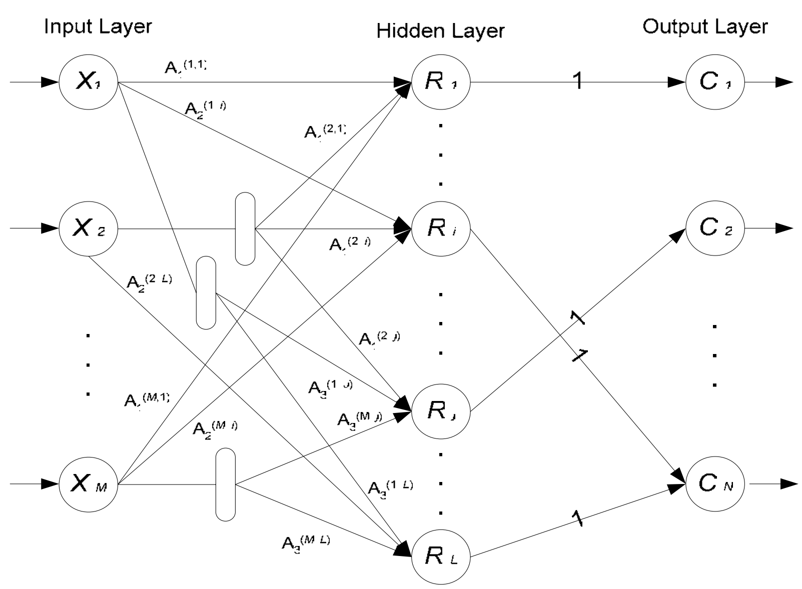

Figure 1: An Architecture of the Neuro-fuzzy System (adapted from [Nauck et al. 1997])

The integration of neural networks and fuzzy systems can take on various forms by changing the hybridization approach. In this paper, we choose a neuro-fuzzy system (See Figure 1), which is a fuzzy system augmented by a neural network to enhance some of its characteristics such as flexibility, speed, and adaptability [Mitra and Hayashi 2000]. In particular, the fuzzy system is trained by a (heuristic) learning algorithm derived from the neural network. Therefore, the underlying process is still fuzzy reasoning, where the connection weights of the network correspond to the parameters of fuzzy reasoning [Keller et al. 1992; Nauck et al. 1997]. Using the back-propagation (BP) learning algorithm, the neuro-fuzzy system can identify fuzzy rules and learn membership functions through fuzzy reasoning or optimizing existing ones. The BP algorithm computes output errors and propagates them backwards from the output layer towards the input layer. The errors are used to change parameters of the fuzzy system locally [Nauck et al. 1997].

Fuzzy sets (e.g., low, middle, and high) represent vague knowledge, for they must overlap to a certain degree without passing each other. They are building blocks of the antecedent part of a fuzzy classification rule, while the consequent part consists of a crisp value assigned to a class. The rules are selected based on values of their performance indicators.

There are many issues involved in developing a neuro-fuzzy classifier, such as choosing the membership function, partitioning the domain, and learning classification rules. Moreover, the derived classifier should be interpretable and testable. The common procedure of creating a neuro-fuzzy classifier is that the structure of the classifier is first created to fit the data as well as possible, and then the training of the classifier is conducted through an iterative process in order to determine its parameters, aiming to improve the accuracy without loosing the interpretability.

It should be pointed out that the multi-layer perception layout in Figure 1 is for the convenience of visualizing the structure and flow of data. In particular, the first layer consists of input variables (X_i, $i = 1 \ldots M$), the hidden layer (R_j, $j = 1 \ldots L$) and associated network connections represent fuzzy rules, and the output layer nodes denote output variables (C_k, $k = 1 \ldots N$), with one node for each class. A connection weight between an input node X_i and a rule node R_j is denoted by $\mathrm{A}_s^{(i,j)}$, where index s represents the selected fuzzy sets of the partition [Nauck 1999]. The output is computed by a maximum operation instead of a weighted sum.

A satisfactory interpretation of learning results of neuro-fuzzy computing is not always guaranteed, especially in multi-dimensional problems. Therefore, there is a need to enhance the learning algorithm of a neuro-fuzzy system with techniques for simplifying the obtained results [Nauck and Kruse 1997]. One solution is to use rule pruning strategies, which fall into one of the two categories [Nauck 1999]: antecedent-based and consequent-based. In addition to simplifying results, pruning in fuzzy rule learning also increases the generality of the learning results.

2.3. Decision trees

Decision trees start with the entire population as a single group. Then, based on criteria, such as entropy minimization from information theory, attributes are gradually selected and their values are used to split the group into sub-groups. This process continues until a termination condition is reached [Quinlan 1993]. Decision trees are well known for their ability to reasonably interpret results, from which decision rules can easily be induced. Nonetheless, tree pruning is usually necessary in order to avoid extraneous results [Quinlan 1987].

The C4.5 algorithm [Quinlan 1993] for inducing classification models of decision trees is an extension of the basic ID3 algorithm. It uses a greedy, top-down algorithm to produce a decision tree. If all data items belong to the same class, C4.5 keeps a decision tree as a leaf. Otherwise, it will recursively try to find the best attribute to split data items into sub leaves. Information gain from splitting is a common criterion for selecting the best attribute. If no gain is obtained from splitting, the set of mixed data items forms into a leaf node, which is labeled according to the most frequent class of data items in that set. From a decision tree, a rule can be derived by following attributes and decision criteria on each path from the root to a leaf node. C4.5 can determine how deep a decision tree grows and how errors will be reduced through pruning [Winston 1992]. Pruning decision trees not only help reduce errors, but also avoid the over-fitting problem [Mingers 1989].

2.4. Input and output features of classification models

All classification methods require training samples with both input and targeted output to build models. Since the ultimate goal of this study was to predict customers' attitude toward Internet retailers, the output feature of each classification model was customer type, which could be one of the two values — Internet customer and traditional customer.

The input includes retailing service attributes that are potentially useful for predicting customers' attitude. The attributes were selected based on heuristic knowledge derived from transaction cost theory and the related work on examining consumers' behavior in Internet retailers [Liang and Huang 1998; Girard et al. 2002; Kwak et al. 2002]. They were pre-tested in a pilot study and modified slightly according to the feedback. The final list of input attributes is shown in Figure 2.

Prior studies on customers' behavior have been primarily focused on identifying influential factors on consumers' perception and acceptance of Internet shopping in general (e.g., [Szymanski and Hise 2000; Kwak et al. 2002; Shim et al. 2002; Vatanasombut et al. 2004]) rather than factors specific for individual products. It has been suggested that consumers' perception of retailing service attributes tend to vary across product types [Liang and Huang 1998]. Therefore, in order to examine effectiveness and generality of the selected classification methods, books and food items, which are representatives of two distinct product categories with regard to customers' patronage tendencies, were selected in this study. Note that books belong to "low touch" items [Economist 2000], which typically require little examination before purchase, thus are popular products in Internet retailing. Food items are the opposite.

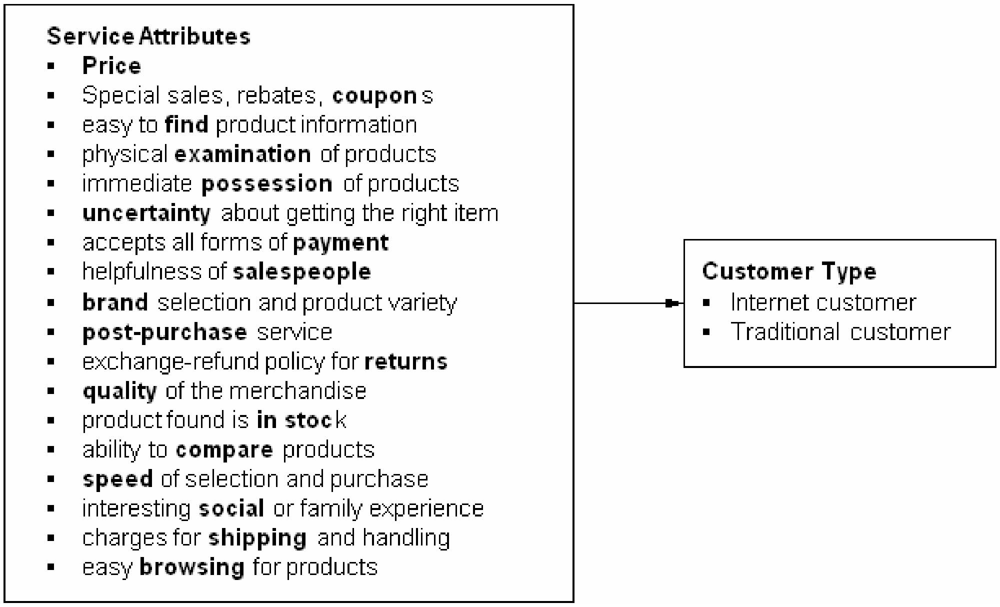

Figure 2: Input and Output Features of Classification Models

3. Knowledge Discovery and Results Analyses

3.1. Data sets and evaluation metrics

Data sets used in this study consisted of responses collected from an online research survey, in which 140 college students from a large mid-Western university and other adults participated. For simplicity, the survey questions are summarized in simple terms in Figure 2 (the individual highlighted words will be used in the rules learnt hereafter). The respondents were asked to rate Internet retailers in relative to traditional retailers of each product in terms of retailing service attributes on a scale ranging from (1) absolutely low, through (4) about the same, to (7) absolutely high. The patronage behavior of customers was rated on a 6-point scale, and was split in the middle and transformed into either Internet customers (4-6) or traditional customers (1-3).

The data sets were used to build classification models for predicting customers' perception of Internet retailers. Performance of the models was assessed in terms of accuracy using 10-fold cross-validation. Accuracy is calculated as the percentage of customers that were accurately predicted by the models. In order to fully reflect the performance of classification models, accuracy was computed for all customers, Internet customers, and traditional customers separately.

3.2. Results of decision trees

We employed J48 [Witten and Frank 1999] as the decision tree system, which is a revision of the popular C4.5 [Quinlan 1993]. The structures and evaluation results of decision trees on books and food items services are listed in Table 1.

Table 1: Performance of the Decision Trees

Products	Error-reduced pruning	Accuracy (%)			Number of leaves	Size of trees
		All customers	Internet customers	Traditional customers		
Food items	Before	74.8	16.7	90.8	12	23
	After	74.1	3.3	93.6	1	1
Books	Before	77.1	85.3	60	14	27
	After	70	93.7	20	2	3

The final decision trees for predicting customers' attitude toward food and book retailers are displayed in Figure 3 and Figure 4, respectively. Each leaf node in a tree represents one of the target categories: I (Internet customers) and T (traditional customers). A decision rule can be constructed by following a path starting from the root to any leaf node (See Figure 2 for meaning of each node in context). The two leaf nodes that are highlighted in each of the trees can be interpreted as follows:

R1: If customers' perception of attractiveness to special sales and rebates in Internet food retailers is not higher and that of product comparison is lower than those in traditional food retailers, they are traditional customers.

R2: If customers' perception of attractiveness to special sales and rebates in food retailers is higher and that of uncertainty in getting the right item is lower than those in traditional food retailers, they are Internet customers.

R3: If, compared with the perception of traditional book retailers, customers' perception of Internet book retailers is at most high in physical examination of products and convenience of comparing products, and is higher in the chance of having products in stock, they are Internet customers.

R4: If customers' perception of the problem of physical examination of products in Internet book retailers is much higher than that in traditional book retailers, they are traditional customers.

```
coupon <= 4
|  comparison <= 3: T (48.0/2.0: R1)
|  comparison > 3
|  |  price <= 3: T (11.19)
|  |  price > 3
|  |  |  possession <= 5
|  |  |  |  service <= 3
|  |  |  |  |  brand <= 4: T (2.0)
|  |  |  |  |  brand > 4: I (2.0)
|  |  |  |  service > 3: T (25.81/2.0)
|  |  |  possession > 5
|  |  |  |  uncertainty <= 6
|  |  |  |  |  in-stock <= 3: T (2.0)
|  |  |  |  |  in-stock > 3
|  |  |  |  |  |  refund <= 4: T (2.0)
|  |  |  |  |  |  refund > 4: I (10.0/1.0)
|  |  |  |  uncertainty > 6: T (5.0)
coupon > 4
|  uncertainty <= 3: I (8.0: R2)
|  uncertainty > 3
|  |  comparison <= 6: T (20.0/4.0)
```

Figure 3: The Decision Tree for Customers of Food Item Retailers

```
examination <= 5
|  comparison <= 5
|  |  in-stock <= 4
|  |  |  coupon <= 4
|  |  |  |  uncertainty <= 1: T (6.84)
|  |  |  |  uncertainty > 1
|  |  |  |  |  speed <= 2: I (6.84)
|  |  |  |  |  speed > 2
|  |  |  |  |  |  salespeople <= 2: T (11.0/1.0)
|  |  |  |  |  |  salespeople > 2
|  |  |  |  |  |  |  payment <= 3: T (3.0)
|  |  |  |  |  |  |  payment > 3
|  |  |  |  |  |  |  |  comparison <= 4: I (13.0/1.0)
|  |  |  |  |  |  |  |  comparison > 4: T (3.0/1.0)
|  |  |  coupon > 4
|  |  |  |  payment <= 2: I (20.88)
|  |  |  |  payment > 2
|  |  |  |  |  service <= 2: T (2.0)
|  |  |  |  |  service > 2
|  |  |  |  |  |  search <= 1: T (2.0)
|  |  |  |  |  |  search > 1
|  |  |  |  |  |  |  quality <= 5: I (21.92/3.92)
|  |  |  |  |  |  |  quality > 5: T (4.84/1.84)
|  |  in-stock > 4: I (18.41/1.16: R3)
|  comparison > 5: I (15.12)
examination > 5: T (11.16/1.08: R4)
```

Figure 4: The Decision Tree for Customers of Book Retailers

The predictive performance of a rule is indicated by the number in parentheses following the corresponding leaf node. For example, the leaf node R1 is attached with (48.0/2.0), which indicates that there was a mix of 48 traditional retailer customers and 2 Internet retailer customers in this node. The decimal number was generated by the way the algorithm used fractional instances to handle missing values [Witten and Frank 1999].

Since the bottom part of a decision tree is often based on relatively few samples and could be very inaccurate, it is desirable to have the tree pruned [Quinlan 1987]. Pruning decision trees is a process of replacing a sub-tree with a branch or a leaf node, which may lead to the reduction of expected errors. After further pruning the trees based on error reduction criteria, the decision tree for food-item customers was reduced to a single node: T (93.0/20.0). It indicates that all of the customers were classified as traditional customers. The decision tree for book customers was also pruned to including two leaf nodes: exam <= 5: I (85.91/23.0); exam > 5: T (8.09/1.09). In this case, only "physical examination of products" attribute was incorporated into the decision tree.

3.3. Results of neuro-fuzzy models

We utilized NEFCLASS-J [Nauck 1999] as the neuro-fuzzy system in this study, which is a Java implementation of fuzzy systems augmented by neural networks. It is platform independent and user friendly.

Triangular membership functions were selected for all input variables. The number of fuzzy set was empirically set to 3 (i.e., large, medium, and small) for food customers and 5 (i.e., very large, large, medium, small, very small) for book customers, and the maximum number of rules was heuristically set to 18 for both types of products. Since the sample size was unbalanced between the two target classes, we selected the best rules for each class in the result to enable rules to be generated for both majority and minority classes. The performance of neuro-fuzzy models is reported in Table 2.

Table 2: Performance of Neuro-fuzzy Models

Products	Max # of rules	Accuracy (%)			Number of rules	Size of fuzzy sets
		All customers	Internet customers	Traditional customers		
Food	Auto*	99.3	96.7	100	131	3
items	18	77.7	46.7	86.2	3	3
Books	Auto	99.2	98.9	100	126	5
	18	73.57	97.9	22.2	5	5

*"Auto" means that the number of rules is automatically determined by the system.

As shown in Table 2, the number of rules automatically generated by the systems was very large. It was also noted that many rules were very complex and only applicable to a few data samples. Consequently, both generality and interpretability of the learning results were compromised. Therefore, we further simplified the rules by pruning the results based on several criteria, such as correlation of retailer attributes, classification frequency, minimum frequency, and support of fuzzy sets [Nauck 1999]. The effect of pruning on neuro-fuzzy learning output was similar to that on decision trees output. The pruned neuro-fuzzy classifiers consisted of the fuzzy rules, as shown in Table 3.

Table 3: Rules in pruned neuro-fuzzy models

Food items:

1. if *comparison* is large then *Internet*;
2. if *comparison* is small then *Traditional*;
3. if *comparison* is medium then *Traditional*.

Books:

1. if *examination* is very small then *Internet*;
2. if *examination* is medium then *Internet*;
3. if *examination* is small then *Internet*;
4. if *examination* is large then *Internet*;
5. if *examination* is very large then *Traditional*;

It is fairly straightforward to interpret fuzzy rules listed in the above table. For example, Rule 1 for food items indicates that customers, who perceive the convenience of comparing products in Internet retailers higher than in traditional retailers, are Internet customers. Conversely, Rule 3 indicates that customers, who perceive the convenience of comparing products in Internet food retailers as smaller in relative to in traditional retailers, are traditional customers. The meanings of the linguistic symbols, such as large, medium, and small, were learnt with rules by the system at the same time. The mapping between fuzzy sets and real values revealed that respondents who rated the convenience of comparing products 4.8 (out of 7) or higher had the largest membership values in the large fuzzy set; respondents who rated 3.1 or below had the largest membership values in the small fuzzy set, and the values of middle fuzzy set fell in between. The rules for book customers can be interpreted in a similar manner. For example, Rule 5 for books indicates that customers who perceive the problem of physically examining products in Internet retailers to be very large in relative to that in traditional retailers are traditional customers. A further examination of the large fuzzy set of examination attribute revealed that it was mapped to the value range of 5.3 and above.

3.4. A comparison of decision trees and neuro-fuzzy models

The large number of rules generated by neuro-fuzzy models and decision trees indicated that there was great variability in the input data, and the relationships between input and output were quite complex. Too many rules with respect to the sample size implied that many of them may be superficial, and interpretation of the learning results would be difficult. Therefore, we pruned the rule bases to simplify the results. The considerable reduction in the number of rules after pruning confirmed that some of the variables in the original rules were not important and those rules rarely provided maximum degree of fulfillment for the class.

The comparison of Table 1 and Table 2 shows that the performance of the decision tree and neuro-fuzzy approaches is comparable. The neoro-fuzzy approach is designed to simulate the imprecise and uncertain nature of human's cognitive process for solving a problem by allowing overlaps in dividing the range of an input variable. It provides an alternative solution to uncertain problems at the cost of precision. The close call between decision tree and neuro-fuzzy models implies that the neuro-fuzzy approach is able to not only handle uncertainty and imprecision but also achieve equivalent or even better performance.

Missing values in the data sets were replaced with group means before being fed into decision trees. However, such a pre-processing step is not necessary for the neuro-fuzzy approach, for it has the capability of handling missing values inherently, which demonstrates the advantage of neuro-fuzzy approach: flexibility and uncertainty handling.

It is notable from Figure 4 that variable *comparison* is used more than once with different thresholds along the same path from the root to a leaf node. This may increase classification accuracy but reduce interpretability. This is not an issue with generating rules using the neuro-fuzzy approach, for a fuzzy rule is involved with at most one fuzzy set for each input variable. Nonetheless, the expression of fuzzy rules may seem verbose sometimes. For example, if we incorporate decision-rule type of representations, the first four rules for books, as shown in Figure 4, can be alternatively denoted as: if examination <= large, then Internet. This motivates us to explore fuzzy decision trees [Mitra et al. 2002] in future research.

The rules discovered by decision trees and neuro-fuzzy models were similar to each other, especially for book customers. Both models included *examination* as the only variable in the antecedents of their rules. However, compared with decision rules, fuzzy rules were intuitively more interpretable, partially due to their conformance to human's qualitative and imprecise knowledge. For example, a rule with the antecedent of "*the problem of physical examination of products is greater than 5"* is less informative than one with "*the problem of physical examination of products is very large*". However, interpretable solutions usually have to be obtained with the user's cooperation. The user must decide whether a solution's readability is sufficient or not, and be ready to influence the learning process when necessary [Nauck 1999]. There are a number of parameters that the user can control in a neuro-fuzzy system. Better and faster solutions may be obtained, as users get more familiar with the data and the problem domain.

4. A Knowledge Discovery Enabled Architecture for Agile Web services

The rules discovered in this study have broad managerial implications. First, they provide knowledge about retailer attributes that are crucial to customers' choice of Internet retailers or traditional mortar retailers. Second, the rules reveal specific levels of service attributes that account for Internet or traditional customers. For example, for Internet food customers, greater improvement of convenience in comparing food may significantly increase their patronage frequency. Likewise, for traditional book retailers, it is important to allow customers to physically examine products to a great extent in order to attract and retain customers. This kind of knowledge can be utilized directly by Internet retailers, traditional retailers, and multi-channel retailers in their decision-making process. Third, the knowledge discovery process may be turned into a new type of e-services to add values by enhancing the agility of existing Internet retailing services.

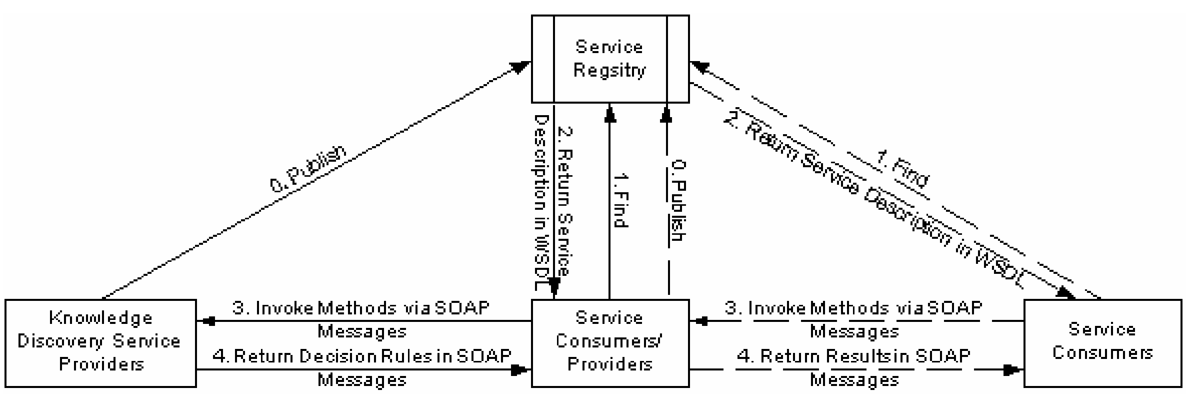

Figure 5: A knowledge discovery enabled architecture for agile Internet retailing Web services

More and more Internet retailers (e.g., Amazon) are encapsulating their services with Web services. Web services are loosely coupled software components with the programmatic interfaces made available for application-to-application communication over the Web [Booth et al. 2004]. They are reshaping the e-business industry. IDC predicts that by 2007, Web services will be a US$21 billion market and 80 percent of enterprises will have implemented Web services. Web services provide a new opportunity for e-businesses to improve their operation efficiency and customer retention [Sutor 2004]. They bring new opportunities to the e-business because today's business processes need agility to quickly adapt to the changing customer needs and market conditions [Peltz 2003]. Implementing the knowledge discovery e-service with Web services technologies help fully realize its potential in diverse applications. As a result, business agility can be achieved not only by the dynamic discovery and composition of Web services, but also by the update of Web services. The latter can be achieved with knowledge discovery models presented in this paper.

Figure 5 shows a knowledge discovery enabled architecture for agile Web services. The focus of service registry is the definition of a set of Web services (e.g., in Web Services Description Language (WSDL) [Chinnici et al. 2004]) supporting the description and discovery of: 1) businesses, organizations, and other Web services providers, 2) the Web services that are available, and 3) the technical interfaces which are used to access those services [Bellwood et al. 2002]. The communication between service providers and consumers is enabled via SOAP ((Simple Object Access Protocol) messages [Karmarkar et al. 2004]. Figure 5 extends the general architecture of Web services [Booth et al. 2004] by inserting a component of secondary service providers on the left. Both of the left and right triangles include general processes engaging Web services. The difference lies in the nature of service consumers and service providers. The service providers at the left end provide knowledge discovery services to the service customers in the middle. The providers in the middle in turn provide retailing services to the customers at the right end.

The left-hand side triangle in Figure 5, where components are connected with solid lines, represents an extension of Web services implementation enabled by the knowledge discovery models. The leftmost service provider offers services to other service providers (e.g. Internet book retailers). When invoking such a service (step 3 on a solid line), a consumer should encapsulate data as well as other information in a SOAP message, which is then sent to the service provider. Once the provider discovers knowledge from the data, the knowledge will be forwarded to the consumer. For example, if a service provider implements a neuro-fuzzy model as a Web service, the SOAP message returned from the provider will include generated fuzzy rules. In B2C e-business, an Internet retailer usually partners with several other business providers (e.g., credit card services and shipping services) to serve customers. If the discovered knowledge reveals that book customers are more concerned about charges for shipping and handling, an Internet book retailer will be motivated to team up with other providers offering cheaper logistic services. For Internet retailers that have implemented Web services, the discovered rules can directly help them adjust their business processes and strategies attract more customers. It can be achieved, for example, by applying the rules in Web services discovery and composition. As a result, the agility of Internet retailers is increased via the use of Web services. All of the above processes are transparent to the service consumers. Thus, Internet retailers can continuously improve their services without affecting applications of service consumers.

The architecture shown in Figure 5 may be implemented in a variety of forms. For example, the two types of services may be merged and published as a single service in the registry. They may be called simultaneously by the same consumer during dynamic composition of Web services in Internet retailing.

5. Conclusions and Future Research Directions

The objective of this study is to discover knowledge for predicting customers' attitude toward Internet retailers. We employed neuro-fuzzy and decision tree approaches in learning rule knowledge for book and food customers. The discovered knowledge can be directly used to support business strategic planning and decision making of Internet retailers and multi-channel retailers. The results showed that classification performance of the two selected methods was comparable. Our investigation also revealed strengths and weaknesses of each approach. On one hand, the neuro-fuzzy approach has advantages over the decision tree in terms of uncertainty handling, interpretability, and flexibility. On the other hand, the expression of fuzzy rules may be verbose and the performance of neuro-fuzzy approach is partially dependent upon human's familiarity with the problem and the data.

The knowledge discovery models can be transformed into a new type of e-services. They aim to help Internet retailing services to increase their agility by adjusting service provisions according to the discovered knowledge from their customers. Web services become an ideal platform for implementing such e-services to interoperate with other existing e-business services. The proposed knowledge discovery Web services are not only essential to increasing the agility of e-business, but also enabling efficient re-adjustment of e-business in a cost-effective manner.

The current investigation suggests several directions for further study:

- *Improvement on the choice of retailer attributes*. Current attribute selection was based on the extant literature and our pilot study. Some of the attributes were proved ineffective in our empirical study. In addition, other features (e.g., demographic information and experience of customers or seasonal factors) may play a role in determining the preferences of customers while choosing between Internet retailers and traditional retailers.
- *Integration of decision tree and neuro-fuzzy approaches*. In view of the efficiency of decision trees and the ability of incorporating prior knowledge into neuro-fuzzy systems, we can use decision trees to generate prior knowledge for neuro-fuzzy systems, which may give the latter a jump-start. We may also integrate the two approaches by introducing fuzziness into the decision trees [Mitra et al. 2002].
- *Integration of fuzzy inputs with fuzzy outputs*. In the context of e-business, customers' choice of purchasing channel is likely to be fuzzy rather than clear-cut. In particular, a customer may patronage traditional retailers most of the time, but occasionally purchase from Internet retailers. Therefore, it would be interesting to combine fuzzy inputs with fuzzy outputs in predicting customers' patronage tendency.
- *Assessing knowledge discovery models with data sets with different class distribution*. In view of the nature of products, data sets collected for some types of products may be balanced but others not. Neuro-fuzzy systems may mitigate the effect of unbalanced data on resulting rules by adjusting learning parameters, but as a type of nonparametric tools, decision trees are highly sensitive to the distribution of data. For example, in this study, all the customers of food items were simply treated as Internet customers in the pruned decision tree.
- *Applying the knowledge discovery techniques to support decision making in new e-business contexts*. We selected Internet retailing as the application domain to empirically compare rule generation approaches and to develop the architecture for agile Web services. Many other types of e-businesses such as B2B e-Commerce and B2C e-services may benefit from the technology as well.

The main business goals of Internet retailers include attracting more customers and reducing the operation cost. The practical implications of predicting customers' attitude towards Internet retailers are far-reaching.

RERERENCES

Bellwood, T., L. Clément, C. von Riegen, et al, "UDDI Version 3.0.1," http://uddi.org/pubs/uddi_v3.htm, 2003.

Bhattacherjee, A., "Acceptance of E-Commerce Services: The Case of Electronic Brokerages". *IEEE Transactions on Systems, Man & Cybernetics: Part A*, Vol. 30, No. 4: 411-421, 2000.

Booth, D., H. Haas, F. McCabe, E. Newcomer, M. Champion, C. Ferris and D. Orchard, "Web services architecture," http://www.w3.org/TR/ws-arch/, 2004.

Chen, L.-d., M. L. Gillenson and D. L. Sherrell, "Enticing online consumers: an extended technology acceptance perspective," *Information and Management*, Vol. 39, No. 8:705-719, 2002.

Chinnici, R., M. Gudgin, J.-J. Moreau, J. Schlimmer, S. Weerawarana et al. Web Service Description Language (WSDL), http://www.w3.org/2002/ws/desc/, 2004.

Devaraj, S., M. Fan and R. Kohli, "Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics," *Information Systems Research*, Vol. 13, No. 3: 316-334, 2002.

Economist (2000). E-Commerce: Shopping around the world: 5-54.

Gefen, D., E. Karahanna and D. W. Straub, "Trust and TAM in online shopping: An integrated model," *MIS Quarterly*, Vol. 27: 51-90, 2003.

Girard, T., R. Silverblatt and P. Korgaonkar, "Influence of Product Class on Preference for Shopping on the Internet," *Journal of Computer Mediated Communication*, Vol. 8, No. 1, 2002.

Karmarkar, A., M. Gudgin and Y. Lafon, "SOAP Resource Representation Header," http://www.w3.org/TR/2004/WD-soap12-rep-20040428/, 2004.

Keller, J. M., R. R. Yager and H. Tahani, "Neural network implementation of fuzzy logic," *Fuzzy Sets System*, Vol. 45: 1-12, 1992.

Kwak, H., R. J. Fox and G. M. Zinkhan, "What products can be successfully promoted and sold via the Internet?" *Journal of Advertising Research*, Vol. 42, No. 1: 23-38, 2002.

Liang, T. and J. Huang, "An empirical study on consumer acceptance of products in electronic markets: A transaction cost model," *Decision Support Systems*, Vol. 24: 29-43, 1998.

Mena, J., Data mining your Web site. Digital Press, USA, 1999.

Meyer, R. and E. J. Johnson, "Empirical Generalizations in the Modeling of Consumer Choice," *Marketing Science*, Vol. 14, No. 3: 180-189, 1995.

Mingers, J., "An empirical comparison of pruning methods for decision-tree induction," *Machine Learning*, Vol. 4: 227-243, 1989.

Mitra, S., "Data mining in soft computing framework: A survey," *IEEE Transactions on Neural Networks* Vol. 13, No. 1: 3-14, 2002.

Mitra, S. and Y. Hayashi, "Neuro-fuzzy rule generation: Survey in soft computing framework," *IEEE Transactions on Neural Networks*, Vol. 11, No. 3: 748-768, 2000.

Mitra, S., K. M. Konwar and S. K. Pal, "Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: Generation and evaluation," *IEEE Transactions on System, Man, and Cybernetics*, Vol. 32, No. 4:328-339, 2002.

Nauck, D., F. Klawonn and R. Kruse. Foundations of neuro-fuzzy systems. Chichester, U.K., Wiley, 1997.

Nauck, D. and R. Kruse, "New learning strategies for NEFCLASS," Proceedings of the Seventh International Fuzzy Systems Association World Congress (IFSA), Prague, June 25-29, 1997.

Nauck, U., Design and implementation of a Neuro-fuzzy data analysis tool in Java, Braunschweig, Institut Fur Betriebssysteme and Rechnerverbund, 1999.

Pavlou, P. A., "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model," *International Journal of Electronic Commerce*, Vol. 7, No. 3: 101-145, 2003.

Peltz, C. (2003). "Web services orchestration - A review of emerging technologies, tools, and standards," http://devresource.hp.com/drc/technical_white_papers/WSOrch/WSOrchestration.pdf, Hewlett Packard, Co. 2003.

Quinlan, J. R., Simplifying decision trees, *International Journal of Man-Machine Studies*, Vol. 27: 221-234, 1987.

Quinlan, J. R., C4.5: Programs for machine learning. San Mateo, CA, Morgan Kaufmann Publishers, 1993.

Rust, R. T. and K. N. Lemon, E-Service and the Consumer, *International Journal of Electronic Commerce* Vol. 5, No. 3: 85-102, 2001.

Shim, J. P., Y. B. Shin and L. Nottingham, "Retailer Web site influence on customer shopping: An exploratory study on key factors of customer satisfaction," *Journal of the Association for Information Systems*, Vol. 3: 53-76, 2002.

Singh, M., "Evolving E-business models: A process of technology," E-business in Australia: Concepts and cases, D. Waddell (ed), Australia, Pearson Education: 8-22, 2002.

Sutor, B. (2004). Web services wish list 2004: Better living through Web services. TechTarget.

Szymanski, D. M. and R. T. Hise, "e-Satisfaction: An initial examination," *Journal of Retailing*, Vol. 76, No. 3: 309-322, 2000.

Vatanasombut, B., A. C. Stylianou and M. Igbaria, "How to retain online customers?", *Communications of the ACM*, Vol. 47, No. 6: 65-69, 2004.

Winston, P. H., Artificial Intelligence, Reading, MA, Addison Wesley, 1992.

Witten, I. H. and E. Frank, Data Mining: Practical machine learning tools and techniques with Java implementations, Morgan Kaufmann, 1999.

Zadeh, L. A., "Fuzzy logic, neural networks, and soft computing," *Communications of the ACM*, Vol. 37: 77-84, 1994.