User-Centric Personalization to Predict User Purchases Based on the Discovery of Important Association Rules Using Rough Set Data Analysis *

Jiye Li (jiye@cse.yorku.ca)
Department of Computer Science and Engineering, York University
4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

Nick Cercone (ncercone@yorku.ca)
Faculty of Science and Engineering, York University
4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

Robin Cohen (rcohen@ai.uwaterloo.ca)
David R. Cheriton School of Computer Science, University of Waterloo
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

Rajan Lukose (rajan_lukose@hp.com)
HP Labs, 1501 Page Mill Road, Palo Alto, California, USA 94304

Jing Zhou (jzhou7@email.uncc.edu)
Belk College of Business, University of North Carolina Charlotte
9201 University City Blvd, Charlotte, NC, 28223

Abstract. In this paper, we present a model to extract important rules from user browsing history in an online purchasing database that makes use of user-centric data. Users' behaviours across all web sites visited is gathered into a database. This database is then mined for important association rules in order to predict the potential online buyers for certain products. Our research includes a method for constructing features to reflect online purchases based on the user-centric data collected from across multiple websites. It also introduces a new Rule Importance Measure based on the rough sets theory that provides an objective determination of the most appropriate rules to employ for the prediction task. Through experiments using a user-centric clickstream dataset from an online audience measurement company (showing customer online search experiences on search engines and shopping sites), we demonstrate how the Rule Importance Measure can be well adapted to predict online product purchases. In particular, we are able to isolate those user-centric features that are most important for predicting online purchases.

Keywords: Association Rules, Data Mining For Modeling User Behaviours, User-Centric Personalization, User Purchasing Prediction, Web-based Personalization, Rough Set Data Analysis, Rule Importance Measure

---

* This paper or a similar version is not currently under review by a journal or conference, nor will it be submitted to such within the next three months. This paper is void of plagiarism or self-plagiarism as defined in Section 1 of ACM's Policy and Procedures on Plagiarism.

1. Introduction

Personalization towards individuals has recently become an important focus for business applications, such as personalized home pages and a personalized shopping cart. In an online shopping application, individuals' online purchasing options and online browsing experiences may be personalized as well. Such personalization is helpful to predict customers' interests and to recommend relevant advertisements of interested products to facilitate customers' online shopping experiences.

A user-centric approach to personalization is one that models the behaviour of individual users in order to make predictions about their preferences for future purchasing. Figure 1 below illustrates a prototype of a user-centric personalization system, combining data mining and machine learning algorithms on predicting online product purchases. User-centric data is collected and stored in a database. Features related to user-centric clickstream data are selected and the data is preprocessed for the prediction engine. The search terms users input into the search engines, and the search terms they use on the leading online shopping stores are considered as strong indications of their purchasing interests, and the terms are categorized first to classify potential users into different product purchasing categories. Classification algorithms such as decision trees (Quinlan, 1993), logistic regression (Agresti, 1996) and Naïve Bayes (Maron, 1961), association rules algorithms such as apriori (Agrawal and Srikant, 1994), and other prediction algorithms are applied in the following steps to further predict whether a user is an online buyer or non-buyer according to the observed browsing behaviours across multiple websites.

In this paper, we present research that models a user's browsing histories across multiple websites in order to make predictions about that user's interests, of use in personalizing products and advertisements to this user. We discuss the value of user-centric personalization, in contrast with the currently popular approach of site-centric personalization. This leads to a framework that retains the core ideas of site-centric personalization but incorporates a rich model of a user's behaviors and interests, including web sites visited, search terms used and products purchased.

We then introduce a specific approach for employing an association rules algorithm in order to perform the prediction task for personalization. Features to represent the user data are constructed. We then apply a method for reducing the number of association rules to consider. This latter task is achieved by employing a novel technique for determining the importance of association rules, based on rough sets theory (Pawlak, 1992), referred to as the Rule Importance Measure.

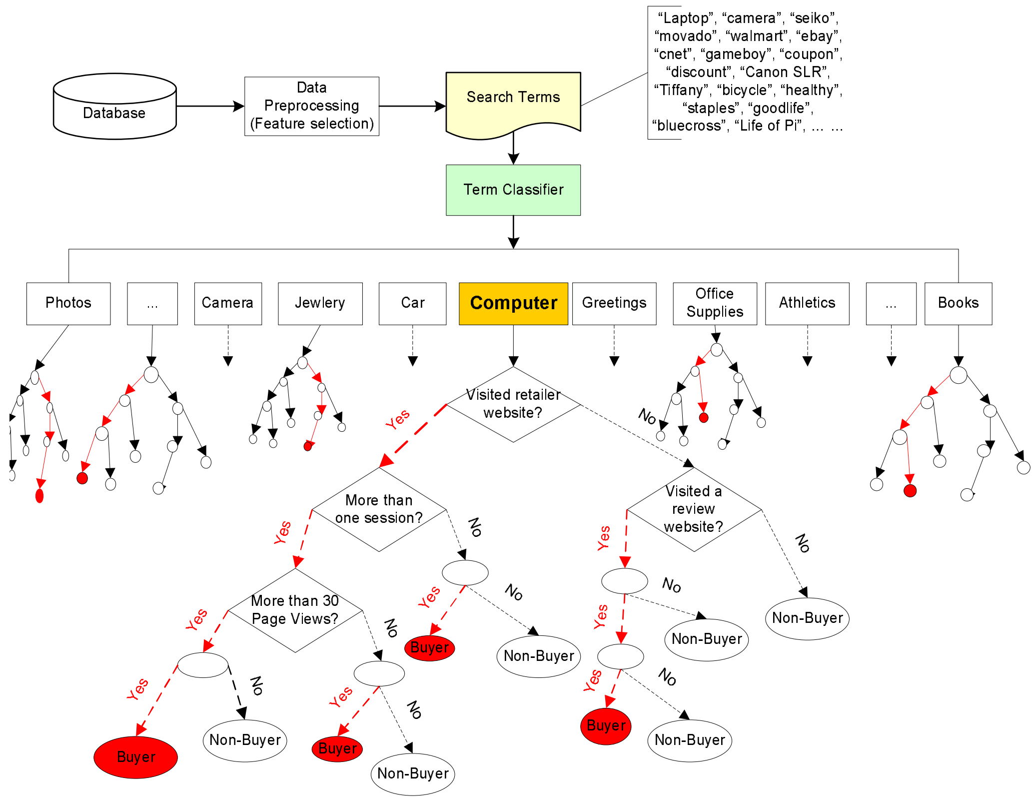

Figure 1. Prototype for Online Product Purchasing System

We demonstrate the value of our proposed approach through experiments conducted on a clickstream dataset from an online audience measurement company (showing customer online search experiences on search engines and shopping sites) for a project at HP Labs in Palo Alto. Using this dataset, we are able to show that a valuable and reasonably-sized set of rules can be determined using our Rule Importance Measure, of use in predicting user behavior, towards personalization.

We argue for the benefit of data mining for modeling user behaviour, in order to deliver personalization to users. We discuss how to develop a system for user-centric personalization through effective data mining by determining: how to model user-centric data, what kinds of features to extract from such data to model the users' intentions and how to predict product purchasing using user-centric personalization.

2. An Approach for User-Centric Personalization

Clickstream data collected across all the different websites a user visits reflect the user's behaviours, interests, and preferences more completely than data collected from the perspective of a single website. For example, we would expect that we could better model and predict the intentions of users who we know not only searched on Google but also visited the HP shopping website and the Dell website, than if we know only one of those pieces of information. The complete data set is termed user-centric data, which contains site-centric data as a subset.

2.1. Site-Centric Personalization

Current research on clickstream data analysis is centered around what is called site-centric data (Padmanabhan et al., 2001). The site-centric personalization systems collect customers' browsing histories based on the clickstream data from the individual web site perspective, and personalizations are generated according to these clickstream data to recommend items to the Internet users who browse this specific web site. Predictions can be generated for a new customer based on profile matching of existing customers (using such attributes as name, location, gender, occupation, IP address, operating system and browser information), browsing histories (using information such as the web pages the customers visited during a certain period of time, application tasks and their sequences the customers performed during a certain period of time), and the preferences of the browsed items (using for instance the fact that some customers have expressed great interest on specific items or tasks, whereas some customers show no interest on the same items or tasks).

Each customer thus has his/her individual profile collected. The more customer profiles a personalization system collects, the more data becomes available for precise recommendations. These user profiles are then saved either in flat files or are loaded into a database. After the data is collected, preprocessing of the original data is conducted, which includes tasks such as missing values processing, discretization, normalization and so on.

Then, personalizations are generated by performing data mining on the database, using certain rule generation algorithms, such as association rule algorithms, clustering algorithms, classifications and so on. The amount of personalization can be huge when first generated; therefore, post-processing for the generated results is performed in this stage.

The rules generated based on customers' profiles therefore serve as the available knowledge base for personalization systems. In real world situations, when the personalization systems observe a new customer whose profile is an exact match or similar match to the profiles in the database, the recommendations from the personalization database are generated and provided to this new customer.

Figure 2 shows a model for a site-centric data personalization system. Data collected from different users, including the browsing histories, personal preferences and demographic data, are sent for creating the personalization engine. When a new customer comes, based on the browsing histories and the demographic information, the engine recommends personalized interesting items (such as web pages) to this new user.

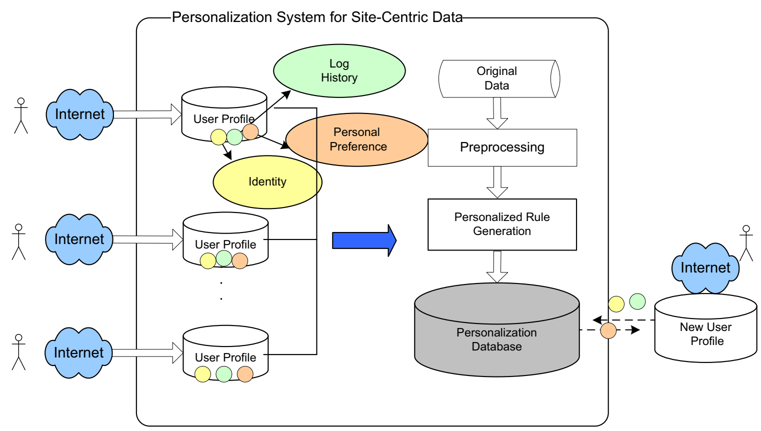

Figure 2. Personalization for Site-Centric Data

2.2. User-Centric Personalization

Much current research on clickstream data personalization focuses on site-centric personalization (Padmanabhan et al., 2001). With site-centric data, however, it is sometimes difficult to fully capture customers' online shopping behaviours for precise personalization modeling and predictions. We explain one of the difficulties in the following example.

Using an online shopping and retail website such as shopping.com or Amazon.com as an example, we consider the online clickstream data collected by this site as site-centric data. Knowledge such as customers' demographic information, the web pages the customers visited, the time the customers spent on each of the particular web pages, the incoming and outgoing URLs for each of the customers and so on is collected on the server side. To assist a new customer, information on the previous web pages this customer has visited is collected, and recommendations are suggested based mostly on the behaviour of similar buyers who are observed to make a purchase at this site. For those customers who visited but do not make a purchase at such sites, although later they may make a purchase elsewhere (e.g., HP shopping websites), this site captures the browsing histories for people who visited, but such browsing information is not considered to be important for making online product purchase predictions. The available information is thus not fully captured and utilized. Demographic information for customers' background are also taken into consideration for making recommendations. In user-centric personalization, the limitations of not effectively capturing complete information collected from only some sites no long exists. Users' web histories across multiple websites are all used towards the construction of the engine.

The personalization techniques for site-centric data are quite mature, which are techniques originating from traditional web log mining, machine learning and data mining. Given the differences between site-centric data and user-centric data, it is important to study whether these site-centric personalization techniques can be applied to user-centric data, and whether new issues (in terms of data collection, data preprocessing, user behaviour modeling and so on) and new challenges should be taken into consideration for user-centric data personalization.

User-centric data is collected for each of the individual users. The data contains users' browsing histories on all the web sites they visit and their own preferences of interested web sites. Then, in order to make a prediction for a specific customer, his/her profile is compared against a database of other customers' data that has been preprocessed and mined, for example using association rules. Figure 3 depicts a sample user-centric personalization system.

3. The Rule Importance Measure for Evaluating Association Rules

Our approach for user-centric personalization includes the generation of association rules from the dataset, in order to predict the user's future behaviors. As discussed in the Introduction section, we make sure of a novel method for producing a smaller set of valuable rules to be used, a rough set based rule evaluation technique referred to as the Rule Importance Measure. In this section, we describe this measure in detail. In the following section 4, we discuss how to use this measure in the context of user-centric personalization.

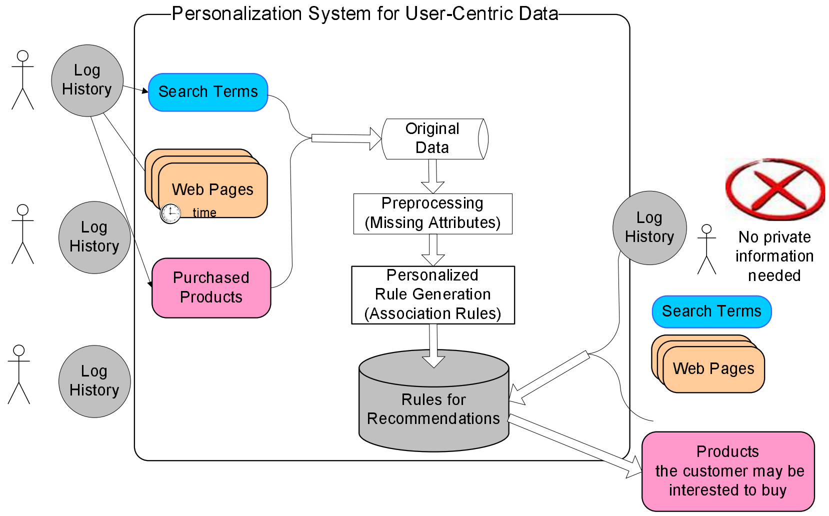

Figure 3. Personalization for User-Centric Data

3.1. Background

Knowledge discovery in databases is a process of discovering previously unknown, valid, novel, potentially useful and understandable patterns in large data sets (Fayyad et al., 1996). Data mining is one of the activities in this interactive process. Data mining encompasses many different techniques and algorithms, including clustering, classification and association rule algorithms.

In a knowledge discovery system, the original data set is first being preprocessed. This step includes the processing of data instances with missing attribute values, inconsistent data instances, data discretization, feature selection and so on. Several approaches on processing data with missing attribute values, i.e., semantic methods and objective methods such as ignoring data instances containing missing attributes, can be used optionally. Inconsistency may exist when two or more data instances contain the same condition attribute values but different decision attribute values. These data instances may be removed if the data mining algorithms cannot process inconsistent data. After the data is preprocessed, various data mining algorithms, such as association rules, classification rules, sequential patterns, can be applied for rule generations. For example, a rule such as "Japanese cars with manual transmission and light weight usually have high mileage", can be generated by a classification algorithm from a data set of cars which contain the mileage of the cars and features such as the manufacture, the model, the transmission, the weight and so on (Hu, 1995). Such rules are used for making predictions. Rule evaluations are performed based on the generated rules, and then useful rules are presented as the output of the system.

A challenging problem in rule generation is that an extensive number of rules are extracted by data mining algorithms over large data sets, and it is infeasible for human beings to select important, useful, and interesting rules manually. How to develop measures to automatically extract and evaluate interesting, relevant, and novel rules becomes an urgent and practical topic in this area. Many existing methods such as rule interestingness measures and rule quality measures from statistics and information theory areas were reported in (Bruha, 1996; Hilderman and Hamilton, 1999; Tan and Kumar, 2000).

In order to improve the utility of the rules that emerge during knowledge discovery, we introduce a novel technique referred to as the Rule Importance Measure which is applied in the context of generating association rules, and which draws from a theory for knowledge discovery known as rough sets theory. We clarify these two terms below.

Rough sets theory was first introduced by Pawlak in the 1980's (Pawlak, 1992). An early application of rough sets theory to knowledge discovery systems was introduced to identify and remove redundant variables, and to classify imprecise and incomplete information. Reduct and core are the two important concepts in rough set theory.

A data set can be represented as a decision table, which is used to specify what conditions lead to decisions. A decision table is defined as $T = (U, C, D)$, where U is the set of objects in the table and $U \neq \phi$, C is the set of the condition attributes and D is the set of the decision attributes. A reduct of a decision table is a set of condition attributes that are sufficient to define the decision attributes. A reduct does not contain redundant attributes towards a classification task. It is often used in the attribute selection process to reduce the redundant attributes, and to reduce the computation cost for rule generations. There may exist more than one reduct for each decision table. Various approximation algorithms are used to obtain reduct sets (e.g. (Bazan et al., 2000)). The intersection of all the possible reducts is called the *core*. The core is contained in all the reduct sets, and it is the essential part of the whole data. Any reduct generated from the original data set cannot exclude the core attributes. Various algorithms exist to determine the core (e.g. (Hu et al., 2004)).

Table I shows a car data set displayed as a decision table where the decision attribute is the category of mileage and the condition attributes are various features of the car such as the transmission type and the number of cylinders. The reducts for this data set are displayed in Table II, and the core attributes for this data are *make_model* and *trans*.

Table I. Artificial Car Data Set

make_model	cyl	door	displace	compress	power	trans	weight	Mileage
USA	6	2	Medium	High	High	Auto	Medium	Medium
USA	6	4	Medium	Medium	Medium	Manual	Medium	Medium
USA	4	2	Small	High	Medium	Auto	Medium	Medium
USA	4	2	Medium	Medium	Medium	Manual	Medium	Medium
USA	4	2	Medium	Medium	High	Manual	Medium	Medium
USA	6	4	Medium	Medium	High	Auto	Medium	Medium
USA	4	2	Medium	Medium	High	Auto	Medium	Medium
USA	4	2	Medium	High	High	Manual	Light	High
Japan	4	2	Small	High	Low	Manual	Light	High
Japan	4	2	Medium	Medium	Medium	Manual	Medium	High
Japan	4	2	Small	High	High	Manual	Medium	High
Japan	4	2	Small	Medium	Low	Manual	Medium	High
Japan	4	2	Small	High	Medium	Manual	Medium	High
USA	4	2	Small	High	Medium	Manual	Medium	High

Table II. Reducts Generated by Genetic Algorithm for the Artificial Car Data Set

No.	Reduct Sets
1	{make_model, compress, power, trans}
2	{make_model, cyl, compress, trans}
3	{make_model, displace, compress, trans}
4	{make_model, cyl, door, displace, trans, weight}

The association rule algorithm was first introduced in (Agrawal and Srikant, 1994), and is commonly referred to as the apriori association rule algorithm. It can be used to discover rules from transaction datasets. The algorithm first generates frequent itemsets, which are sets of items that have transaction support more than the minimum support; then based on these itemsets, the association rules are generated which satisfy the minimum confidence.

Association rule algorithms can be used to find associations among items from transactions. For example, in *market basket analysis*, by analyzing transaction records from the market, we could use association rule algorithms to discover different shopping behaviours such as, when customers buy bread, they will probably buy milk. This type of behaviour can be used in the market analysis to increase the amount of milk sold in the market.

An association rule is a rule of the form $\alpha \rightarrow \beta$, where α and β represent itemsets which do not share common items. The association rule $\alpha \rightarrow \beta$ holds in the transaction set D with *confidence* c if $c\%$ of transactions in D that contain α also contain β. The rule $\alpha \rightarrow \beta$ has *support* s in the transaction set D if $s\%$ of transactions in D contain $\alpha \cup \beta$. Using Table I as the data set, we can use association rule algorithms to generate rules that are used for making decisions. For example, the rule Light-weight $\rightarrow$ High-mileage can be extracted with a confidence of 100% (2 transactions that have Light-weight also have High-mileage), and a support of 14.3% (2 transactions out of 14 have both features).

A problem with using the association rules algorithm in large data sets is that there are usually *too many rules generated* and it is difficult to analyze these rules. The Rule Importance Measure is described in detail in the following section. It applies rough sets theory to association rules generation in order to evaluate association rules and thus improve their utility.

3.2. Defining the Rule Importance Measure

The general model on which we compute the Rule Importance Measure is shown in Figure 4.

We begin with original data in the form of a decision table with possibly several condition attributes and a single decision attribute. First during the data preprocessing step, the inconsistent data instances and the data instances containing missing attribute values are processed. Inconsistency exists in a decision table when two or more data instances contain the same condition attribute values but different decision attribute values. These data instances must be removed. To remove them, we first sort the whole data set according to the condition attributes, excluding the decision attributes. Then we select data instances that contain the same condition attribute values, but different decision attribute values. They are removed during this stage. Discretization algorithms, such as equal frequency binning or entropy

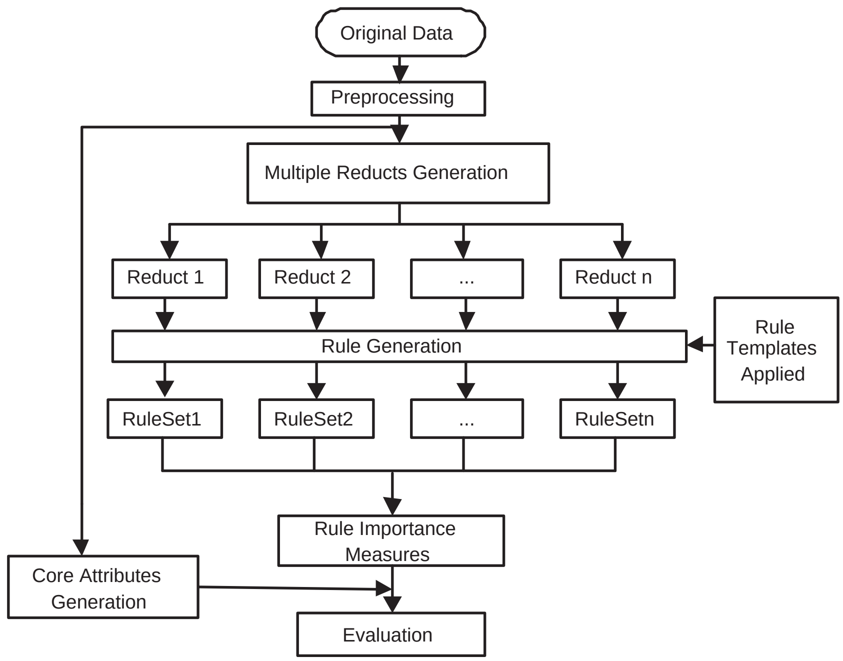

Figure 4. How to Compute the Rule Importance

algorithm (Øhrn, 1999), are also applied during this stage if necessary. For example, an equal frequency discretization algorithm can be used to divide each of the condition attributes from the data into intervals with an equal number of data values in the interval. In so doing, the continuous data values are processed into discrete values to be used by association rules algorithm. Core attributes are generated at the end of the data preprocessing stage. It is worthwhile to mention that core generation requires no inconsistencies in the data set.

After the data is preprocessed, multiple reducts are generated. Various rough set software offer approximation algorithms for multiple reduct generation, using a genetic algorithm approach. For example, ROSETTA's genetic algorithm (Øhrn, 1999) generates multiple reducts; RSES (Bazan et al., 2000) provides a genetic algorithm for a user defined number of reducts generation, which is appropriate in cases of larger data sets for generating representative reducts. In our experiments in Section 4.2 we use the RSES genetic algorithm to generate multiple reducts with the option of full discernibility.

After multiple reducts are generated, a decision table using the condition attributes contained in the reduct together with the decision attribute is used as the input data for rule generation. Let us take Table I as an example to explain how the rules are generated from a given reduct. This artificial car data set has 4 reducts. Reduct No.1 contains condition attribute “make_model”, “compress”, “power” and “trans”. We construct a new decision table with only the condition attributes from reduct No.1, and the original decision attribute “Mileage”, as shown by Table III.

Table III. Decision Table based on Reduct No.1

make_model	compress	power	trans	Mileage
USA	High	High	Auto	Medium
USA	Medium	Medium	Manual	Medium
USA	High	Medium	Auto	Medium
USA	Medium	Medium	Manual	Medium
USA	Medium	High	Manual	Medium
USA	Medium	High	Auto	Medium
USA	Medium	High	Auto	Medium
USA	High	High	Manual	High
Japan	High	Low	Manual	High
Japan	Medium	Medium	Manual	High
Japan	High	High	Manual	High
Japan	Medium	Low	Manual	High
Japan	High	Medium	Manual	High
USA	High	Medium	Manual	High

After obtaining the decision table III for Reduct No.1, rule templates, such as

$$\alpha_1, \alpha_2, \ldots, \alpha_n \Rightarrow \beta$$

are applied in the rule generation step in order to isolate the decision attributes, where α_1, α_2, ..., α_n and β are different (condition and decision) attributes in the data set. Depending on different applications and the expected results, rule templates for desired types of rules and for subsumed rules are defined prior to the rule generation and are applied during the rule generation process. Rules such as

$$make\_model = USA, compress = medium \rightarrow Mileage = medium$$

are generated based on Table III. Since there are multiple reducts obtained as shown in Figure 4, for each reduct, an individual decision table containing condition attributes from this reduct with original decision attributes can be generated. Rules based on such decision tables can then be generated. Therefore, multiple rule sets are obtained after the rule generations for multiple reducts.

Note that rules generated from different reduct sets can contain different representative information. If only one reduct set is being considered to generate rules, other important information might be omitted. Using multiple reducts, some rules will be generated more frequently than other rules. A list of rules with their importance has obtained after the rule generation process. We consider the rules that are generated more frequently more important. The Rule Importance Measure is used to evaluate the importance of association rules.

The definition of the Rule Importance Measure is presented in Eq. 1. Let n be the number of reducts generated from the decision table $T(U, C, D)$. Let $RuleSets$ be the n rule sets generated based on the n reducts. $ruleset_j \in RuleSets$ $(1 \leq j \leq n)$ denotes individual rule sets containing rules generated based on reducts. $rule_i$ $(1 \leq i \leq m)$ denotes the individual rule from $RuleSets$. RIM_i represents the Rule Importance Measure for the individual rule. Thus the Rule Importance Measures can be computed by the following

$$RIM_i = \frac{|\{ruleset_j \in RuleSets | rule_i \in ruleset_j\}|}{n}. \tag{1}$$

The following example shows how to compute the Rule Importance Measure. We use the UCI Iris (Newman et al., 1998) data set as an example, which is a data set containing three classes of Iris plants, which are Iris setosa, versicolour and virginica. For the four attributes, we use "*sl*" to stand for attribute "sepal length", "*sw*" for "sepal width", "*pl*" for "petal length" and "*pw*" for "petal width". There are $n = 4$ reducts available for rule generations. For each of the reducts, the rule sets generated based on the reduct are shown below.

Table IV. Reducts and Rule Sets for Iris Data

Reducts	Rule Sets
$\{sl, sw, pl\}$	$\{sl = 4.4 \rightarrow setosa, sw = 2.9 \rightarrow versicolor, pl = 1.9 \rightarrow setosa, \ldots\}$
$\{sw, pl, pw\}$	$\{sw = 2.9 \rightarrow versicolor, pl = 1.9 \rightarrow setosa, pw = 1.1 \rightarrow versicolor, \ldots\}$
$\{sl, pl, pw\}$	$\{sl = 4.4 \rightarrow setosa, pl = 1.9 \rightarrow setosa, pw = 1.1 \rightarrow versicolor, \ldots\}$
$\{sl, sw, pw\}$	$\{sl = 4.4 \rightarrow setosa, sw = 2.9 \rightarrow versicolor, pw = 1.1 \rightarrow versicolor, \ldots\}$

Rule $sl = 4.4 \rightarrow setosa$ is generated across 3 rule sets, therefore the rule importance is $RIM = \frac{3}{4} = 75\%$. For rules $sw = 2.9 \rightarrow versicolor$,

$pl = 1.9 \rightarrow setosa$, $pw = 1.1 \rightarrow versicolor$, they are all generated from 3 of the 4 rule sets, therefore their rule importance is also 75%.

The Rule Importance Measure is calculated independently of the core attributes.

3.3. Complexity Analysis

We present the time complexity for our approach of generating important rules. Suppose there are N data instances in the data set, and M attributes for each data instance, N' is the number of distinct values in the discernibility matrix (Pawlak, 1992) which is a matrix composed of attributes for computing the core and the reduct, r is the number of multiple reducts for the data set, the time complexity in the worst case is analyzed as follows. The time complexity for multiple reducts generation is $O(N'^2)$ (Vinterbo and Øhrn, 2000). The core generation takes $O(NM)$ (Hu et al., 2004). The apriori association rules generation takes $O(NM!)$ (Agrawal and Srikant, 1994); therefore, it takes $O(rNM!)$ to generate multiple rule sets for multiple reducts. The calculation of the rule importance for the total rules k generated by the multiple rule sets takes $O(k \log k)$. In general, r is much smaller than N, therefore the time complexity of our approach is bounded by $O(N'^2 + NM + NM! + k \log k) \approx O(NM!)$ in the worst case.

The Rule Importance measure originated in (Li and Cercone, 2006) and the above discussion provides a measure of the worst case analysis of the calculation of Rule Importance from an entire data set.

3.4. The Rule Importance Measure in Use

As shown in Figure 4, the Rule Importance Measure is used to rank the rules generated by counting the rule frequencies appearing across all the rule sets. Rules with their individual importance measures are ranked according to Eq. 1 and returned from the model.

In the evaluation stage of the model, core attributes play an important role for evaluating these ranked rules. Rules with 100% importance contain only the core attributes. Rules that contain more core attributes are more important than rules that contain fewer or no core attributes. Since core attributes are the most representative among all the condition attributes, more important rules contain more representative attributes, which are the core attributes. Therefore by checking for the presence of the core attributes in the rules, we can evaluate the ranked rules with their rule importance.

The Rule Importance Measure is simple, quick, easy to compute; it provides a direct and objective view of how important a rule is. The following more detailed example illustrates how the rule importance measure ranks rules according to the importance of a rule.

The data set used in the following example is an artificial data set about cars (Hu, 1995), as shown in Table I. The condition attributes are *make_mode, cyl, door, displace, compress, power, trans, weight. Mileage* is the decision attribute. There are 14 instances. The data set does not contain missing attribute values.

For the Car data set, ROSETTA software generates 4 reducts as shown in Table II. The core attributes are, *make_model* and *trans.*

Since we are interested in predicting the mileage of a car based on the model of a car, the number of doors, the compression, the weight as well as other factors related to a car, we would like to extract rules which have the decision attribute "*mileage*" on the consequent part of the rules. Therefore we specify the template for desired rules as shown by Eq. 2.

$$\langle make\_model, cyl, \ldots, weight \rangle \rightarrow \langle mileage \rangle. \tag{2}$$

And if a rule

$$\langle make\_model = Japan, weight = medium \rangle \rightarrow \langle mileage = High \rangle \tag{3}$$

is generated, rules such as Eq. 4

$$\langle make\_model = Japan, trans = manual, weight = medium \rangle \rightarrow \langle mileage = High \rangle \tag{4}$$

are removed, because this rule can be subsumed by the previous rule.

We generate the rule sets based on these 4 reduct sets with $support = 1\%, confidence = 100\%$, and we also rank their rule importance, as shown in Table V.

From Table V, the first 2 rules have an importance of 100%. This observation matches our experiences on cars. The auto transmission cars usually have a lower mileage than the manual cars. Japanese cars are well known for using less gas and providing higher mileage. The rule "Door=4 → Mileage=Medium" has a lower importance because the number of doors belonging to a car does not really affect car mileage. We noticed that the two rules with importance of 100% contain core attributes and only core attributes to make a decision of mileage. For the rest of the rules with importance less than 100%, the attributes on the left hand side of a rule contain non-core attributes. This observation suggests that core attributes are important when evaluating the importance of the rules. Our method of generating rules with reduct sets is efficient. There are 6,327 rules generated from the original data without using reducts or rule templates. 13 rules are generated using reducts and rule templates.

Table V. The Rule Importance for the Artificial Car Data Set

No.	Selected Rules	Rule Importance
1	Trans=Auto → Mileage=Medium	100%
2	make_model=Japan → Mileage=High	100%
3	make_model=USA, Compress=Medium → Mileage=Medium	75%
4	Compress=High, Trans=Manual → Mileage=High	75%
5	Displace=Small, Trans=Manual → Mileage=High	50%
6	Cyl=6 → Mileage=Medium	50%
7	make_model=USA, Displace=Medium, Weight=Medium → Mileage=Medium	25%
8	Power=Low → Mileage=High	25%
9	make_model=USA, Power=High → Mileage=Medium	25%
10	Compress=Medium, Power=High → Mileage=Medium	25%
11	Displace=Small, Compress=Medium → Mileage=High	25%
12	Door=4 → Mileage=Medium	25%
13	Weight=Light → Mileage=High	25%

3.5. The Benefits of the Rule Importance Measure

We have conducted several experiments to discover the benefits of using the Rule Importance Measure during the knowledge discovery process. This includes an experiment on a sanitized real-world geriatric care data set from the Dalhousie University Faculty of Medicine to determine the survival status of a patient, giving all the symptoms he or she shows (Li and Cercone, 2006). We used survival status as the decision attribute and the 44 symptoms of the patient as condition attributes. The experimental results show that we can make dramatic reductions in the number of rules that can be used for knowledge discovery (218 rules ranked by the Rule Importance Measure compared to $2,626,392$ rules generated from the original data without considering this measure) and we can generally provide some rules with a high measure.

We also note some benefits of the Rule Importance Measure in comparison with other rule evaluation methods. Various rule interestingness measures exist (Hilderman and Hamilton, 1999), including considering as interesting rules with a certain level of support and confidence. But these measures are applied after rule generation is performed on the entire data set, therefore requiring more computational resources, whereas the Rule Importance Measure reduces the amount of data required for rule generation by selecting only important attributes from the original data. The Rule Importance Measure differs from the method of rule quality, introduced in (Bruha, 1996), which is often applied in the post-processing step during the rule extraction procedure, to evaluate whether rules overfit the data. The rule quality measure is used to remove low quality rules from the set of all rules generated. Since the rule importance measure considers the representative attributes contained in the reducts, it therefore operates with much fewer rules.

In our experiments on the geriatric care data set, we also compared the Rule Importance Measure with the interestingness measure of *confidence* (Agrawal and Srikant, 1994). Given the antecedent of a rule existing in the data set, confidence measures the probabilities of both the antecedent and consequent of the rule appearing together in the data set. The higher the probability, the more interesting the rule is considered to be. We examined all the rules with confidence of 80% or higher and discovered that many of them had low measures of Rule Importance. This pointed out several rules that might have been considered interesting, but were in fact less important for providing the most effective geriatric care (e.g. not being able to walk long distances turned out to be as interesting as having a severe heart problem, but in fact the latter is more important). These experiments serve to provide some first steps in comparing the Rule Importance Measure to other methods for ranking rules, in knowledge discovery.

As the notion of Rule Importance Measure evolves and matures, a meaningful empirical comparative evaluation can be undertaken. At this time, the closest measures to the Rule Importance Measure are the merit of the rule interestingness and rule quality measures. Detailed comparisons and discussions for these measures are available in (Li, 2007). Given a data set, let I represent the set of association rules generated by any given algorithm from this data set, II represent the set of rules containing important attributes from the multiple reducts, and III be the set of rules containing attributes from the core of this data. We use the following Figure 5 to illustrate why rules generated using the Rule Importance Measure are more important compared to other rules. Reducts of a data set contain attributes that are sufficient to define the decision attributes. The Rule Importance Measure considers such attributes as more important, therefore association rules containing attributes from the reducts, as in set II, are considered to be more important than other rules, as in set $(I \setminus II)$. Core is the essential information of the data set; therefore, rules containing the attributes from the core are considered to be the most important rules, as in set III. Using the Rule Importance Measure, one can obtain a list of more important rules and at the same time reduce the computation effort in generating redundant rules.

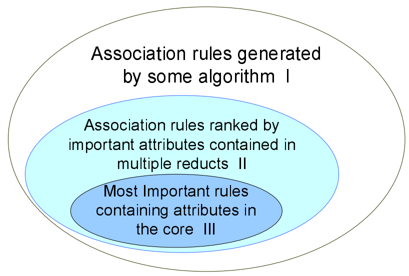

Figure 5. Comparison of Rule Measures

4. Online Purchasing Prediction

The essential purpose of this adaptive web personalization project is to predict users' online purchasing behaviours based on all the websites a user visited. The motivation of the experiment is to demonstrate the usage of the Rule Importance Measure which is applied to rank the important rules extracted from the experimental data, to predict the potential online buyers for certain products.

4.1. Experimental Data

Nielsen//NetRatings MegaPanel data [1] is used as our testbed for this adaptive web personalization project. Nielsen is an online audience measurement company, which is the premier provider of the media-quality internet data. The MegaPanel data offers the overall, in-depth profiles of customer behaviours. The data is collected over the complete customers' online search experiences on both leading search engines (such as Google, Yahoo) and shopping websites (such as Amazon, BestBuy). The data collection processes are designed in such a way that the average customers' online behaviours and their retention rate are consistent with the goal of representative sampling of internet users.

The data collected over 8 months (from November 2005 to June 2006) amounted to approximately 1 terabyte from more than $100,000$ households. For each URL there are time stamps for each internet user. Retailer transaction data contains more than 100 online leading shopping destinations and retailer sites. The data also contains travel

[1] http://www.nielsen-netratings.com/

transaction data, such as air plane, car and hotel reservation histories. There are also users' search terms collected in the URL data. The search terms are collected from top search engines and comparison shopping sites. In addition, additional search terms are extracted and customized by HP (e.g., from Craigslist.org, which is a website for online classifiers and forums).

4.1.1. *Feature Construction*

Features are important elements representing the experimental data. Feature construction is usually conducted in the data collection process. In our experimental data, features reflecting online purchases are not directly available from the original source of the data. Since our aim is to predict online product purchases using user-centric data, we focus on constructing features that can reflect the users' browsing and searching behaviours across multiple websites. There are 26 online product categories available in our experimental data. In this experiment, we limit the online purchasing product category to be personal computers, including desktops and laptops.

The intuition for extracting such features towards our user-centric personalization task is that, during an online purchasing event, in general, people would search for the product category using the leading search engines (such as Google or Yahoo); they would also visit the websites of retailers (such as Dell) who sell this product for detailed product information; they would check how other customers consider this product at some review websites (such as CNET); they would also commonly visit websites offering coupons or discounts for certain products.

Since site-centric data are collected as a subset of user-centric data, traditional features for site-centric clickstream analysis are considered as part of our feature sets. Features such as "the number of sessions the user spent on certain website", "the sub URLs visited" and "the total time spent per session of visit" are extracted.

User-centric features related to searches across multiple websites such as "search terms used across multiple search engines and websites", "whether visited retailer websites", "whether visited review websites" and "whether made an online purchase" are extracted.

According to the above mentioned site-centric and user-centric related features, we construct 28 features that are used in the following experiments for predicting purchase of personal computers, as shown in Table VI. December 2005 data is used for this experiment. In the feature descriptions, "NNR" stands for Nielsen//NetRating; HP customized sites stand for additional searches or websites extracted and customized by HP (such as Craigslist). The HP customized sites includes all the sites pre-classified by NNR.

4.1.2. *Decision Table*

December 2005 data is used for this experiment. We consider the 28 features as shown in Table VI as condition attributes; we consider whether a person is a buyer or non-buyer for personal computers in December 2005 as the decision attribute. For a decision table $T = (C, D)$, C = {feature sets containing 28 features}, D = {Yes, No} indicating whether a person is a buyer or non-buyer. With $83,635$ users and 28 features, we create a decision table as shown in Table VII.

After the data is processed in the format of a decision table, we then apply the equal frequency (Chiu et al., 1991) approach to discretize the data.

4.2. Rule Importance Measures

Recall the generation of the Rule Importance Measure in Figure 4. After the input data is preprocessed, the multiple reducts are generated. We use the genetic algorithm provided by RSES (Bazan et al., 2000) for multiple reducts generation. The reducts are shown in Table VIII.

We use apriori association rule generation to obtain prediction rules. Since the goal of this experiment is to predict whether an internet user is a potential online buyer of personal computers, our interest is to generate rules which lead to the predictions of buyers or non-buyers of computers. We specify the following two rule templates as shown in Eq. 5 and Eq. 6 that are applied during rule generations, in order to constrain the association rule algorithm.

First, we specify that only decision attributes (buyer or non-buyer) can be on the consequent part of a rule, and there may exist more than one feature on the antecedent part of the rule. The antecedent leads to a decision (buyer or non-buyer) which is represented by the consequent part.

$$\langle Feature_1, Feature_2, \ldots, Feature_n \rangle \rightarrow \langle Decision \rangle \tag{5}$$

Secondly, we specify the subsumed rules using the following constraint. Given a rule represented by Eq. 6.

$$\langle Feature_1, Feature_2, ..., Feature_m \rangle \rightarrow \langle Decision \rangle \tag{6}$$

the following rules

$$\langle Feature_1, Feature_2, ..., Feature_m, Feature_s \rangle \rightarrow \langle Decision \rangle \tag{7}$$

$$\langle Feature_1, Feature_2, ..., Feature_m, Feature_p \rangle \rightarrow \langle Decision \rangle \tag{8}$$

Table VI. 28 User-Centric Features for Computer Purchases in December 2005 Data

No.	Feature ID	Feature Description	No. of Users who satisfy the feature	Value range
1	G1a	Whether searched “laptop” on Google before purchasing	279	{Yes, No}
2	G1b	# of sessions this user searched “laptop” on Google before purchasing	279	{0, ..., N}
3	G1c	# of sessions this user searched “laptop” before purchasing on all NNR	647	{0, ..., N}
4	G1d	# of sessions this user searched “laptop” before purchasing on all NNR & HP customized search	1,012	{0, ..., N}
5	G2a	# of page views on Google before purchasing	41, 778	{0, ..., N}
6	G2b	# of page views on all NNR before purchasing	69, 219	{0, ..., N}
7	G2c	# of page views on all HP customized search before purchasing	70, 192	{0, ..., N}
8	G3a	# of sessions on Google before purchasing	41, 778	{0, ..., N}
9	G3b	# of sessions on all NNR before purchasing	69, 219	{0, ..., N}
10	G3c	# of sessions on all HP customized sites before purchasing	70, 192	{0, ..., N}
11	G5a	# of page views per user who searched “laptop” on Google before purchasing	279	{0, ..., N}
12	G5b	# of page views per user who searched “laptop” on all NNR websites before purchasing	647	{0, ..., N}
13	G5c	# of page views per user who searched “laptop” on HP customized websites before purchasing	1,012	{0, ..., N}
14	G6c1	# of sessions a user visited a hardware manufacturers or multi-category computers/consumer electronics sites and NNR sites before purchasing	48, 130	{0, ..., N}
15	G6c2	# of sessions a user visited a hardware manufacturers or multi-category computers/consumer electronics sites and HP customized sites before purchasing	48, 627	{0, ..., N}
16	G6d1	# of page views a user visited a hardware manufacturers or multi-category computers/consumer electronics sites and NNR sites before purchasing	48, 130	{0, ..., N}
17	G6d2	# of page views a user visited a hardware manufacturers or multi-category computers/consumer electronics sites and HP customized sites before purchasing	48, 627	{0, ..., N}
18	G15	# of sessions the user searched “coupon” or “review” before purchasing	3,208	{0, ..., N}
19	G6a	Whether this user visited the hardware manufacturers or multi-category computers/consumer electronics sites and HP customized sites before purchasing	48, 627	{Yes, No}
20	G6b	Whether this user visited the hardware manufacturers or multi-category computers/consumer electronics sites and NNR sites before purchasing	48, 130	{Yes, No}
21	G20a	Whether this user visited the hardware manufacturers or multi-category computers/consumer electronics sites before purchasing	50, 041	{Yes, No}
22	G20c	# of sessions this user visited the hardware manufacturers or multi-category computers/consumer electronics sites before purchasing	50, 041	{0, ..., N}
23	G20d	# of page views this user visited the hardware manufacturers or multi-category computers/consumer electronics sites before purchasing	50, 041	{0, ..., N}
24	G14a	Whether this user made a purchase (of any product category) in the past month (November)	25, 029	{0, ..., N}
25	G14b	Whether this user made a purchase of computer hardware, or computer software, or consumer electronics categories in the past month (November)	5, 400	{Yes, No}
26	G14c	# of purchases of computer hardware, computer software, or consumer electronics category the user made in the past month (November)	5, 400	{0, ..., N}
27	G11	# of time (seconds) this user spent to visit the hardware manufacturers or multi-category computers/consumer electronics sites before purchasing	50, 041	{0, ..., N}
28	G16	Whether this user visited a review site before purchasing (In the URL table, pag_addres contain %cnet%)	12, 323	{Yes, No}

Table VII. Decision Table for Classifications

User ID	Condition Attributes 28 Features								Decision Attribute {Yes, No}
ID	G1a	G1b	G1c	G1d	...	G14c	G11	G16	{Yes, No}
1	Yes	2	0	2	...	7	5200	No	Yes
2	Yes	5	1	7	...	2	413	Yes	No
3	No	0	0	1	...	0	622	No	Yes
...	...	...	...	...	...	...	...	...	...
83,635	Yes	1	0	3	...	0	342	No	Yes

Table VIII. Reducts Generated by Genetic Algorithm for Decision Table VII

No.	Reduct Sets
1	{G2c, G3a, G3b, G14a, G11, G6b, G16}
2	{G2a, G2c, G3b, G6d1, G14a, G11, G16}
3	{G2a, G2c, G3b, G6c2, G14a, G11, G16}
4	{G2a, G2b, G2c, G6d1, G14a, G11, G16}
5	{G2a, G2c, G3b, G14a, G11, G6a, G16}
6	{G2a, G2c, G3b, G20a, G14a, G11, G16}
7	{G2c, G3a, G3b, G6c2, G14a, G11, G16}
8	{G2b, G2c, G3a, G20c, G14a, G11, G16}
9	{G2c, G3a, G3b, G20d, G14a, G11, G16}
10	{G2c, G3a, G3b, G20a, G14a, G11, G16}

can be removed because they are subsumed by Eq. 6.

The classes of online buyers and non-buyers are very imbalanced in this data set. Among the $83,635$ number of users, only 449 are buyers, which take 0.53% of the total number of users. It is trivial to obtain higher confidence rules by simply generating rules to predict the non-buyers based on the features. However, this will not satisfy the purpose of doing research to predict online buyers. We therefore set the values of support and confidence to be lower in order to generate rules that can be used to predict both buyers and non-buyers. We generate the rule sets based on these 10 reduct sets with $support = 0.01\%, confidence = 5\%$.

There are 75 rules generated by using the Rule Importance Measures and rule templates. We rank their rule importance, as shown in Table IX[2] In comparison, without the using the Rule Importance Measure, $16,178,963$ rules are generated.

Table IX. The Rule Importance for Decision Table VII

No.	Selected Rules	Rule Importance
1	G2c=0, G14a=1, G11 $\geq$ 622 $\rightarrow$ buyer	100%
2	G16=0 $\rightarrow$ non-buyer	100%
3	G11=0 $\rightarrow$ non-buyer	100%
4	G3b=0, G11 $\geq$ 622 $\rightarrow$ buyer	80%
5	G2c<13, G3a=0, G14a=0, G11 $\geq$ 622 $\rightarrow$ buyer	50%
6	G2a=0, G2c<13, G14a=0, G11 $\geq$ 622 $\rightarrow$ buyer	50%
7	G2b $<$ 10 $\rightarrow$ non-buyer	20%
8	G2c $<$ 13, G20a = 1, G14a = 1, G16=0 $\rightarrow$ buyer	20%
9	G2b $<$ 10, 1 $\leq$ G20c $<$ 4, G14a=1 $\rightarrow$ buyer	10%
10	G2a=0, G2c $<$ 13, G11 $\geq$ 622, G6a=1 $\rightarrow$ buyer	10%
...	...	...

4.2.1. *How to Interpret the Rules*

To clarify what the rules represent in English, we present two rules from Table IX as examples.

Rule No.1: *If an online user has not searched on any of the HP customized search sites, but this user made an online purchase (of any product category) in the previous month, and this user spent more than 622 seconds visiting a hardware manufacturer or multi-category computers/consumer electronics sites, then this user may be a potential online buyer of personal computers.*

[2] The results shown in this table have been adjusted from the discretization result for interpretations. Let us take the attribute G11 as an example. With the equal frequency interval = 3, the attribute values for $G11$ are discretized into G11 = 0, which represents the values of G11 falling into interval (0, 261]; G11 = 1, which represents the values of G11 falling into interval (261, 622]; and G11 = 2, which represents the values of G11 falling into $(622, \infty)$. Therefore for attribute value G11 $\geq$ 622, it is originally discretized as G11 = 2.

Rule No.3: *If an online user did not visit any hardware manufacturer or multi-category computers/consumer electronics sites, then this user may not be a potential online buyer of personal computers.*

4.2.2. *Discussions*

The Rule Importance Measure provides an efficient view for important and representative knowledge contained in this user-centric clickstream data. Such extracted rules are useful to predict whether an online purchase will happen for certain users according to the observed online searching and browsing behaviours. From the list of generated rules in Table IX, we also obtain the degree of how important these rules are. These results are useful to help provide a diverse view of online purchase predictions.

In order to generate rules for possible online buyer prediction, the value of the support is quite low. The following study provides one explanation for this situation. According to a study published by comScore [3] in December 2004 about the results for internet users' potential on purchasing electronics and computer products, the results indicate that 92% of the internet users purchase the products offline after searching on the internet. Only a small percentage of internet users would make a purchase eventually, although 85% of such purchases happen after 5 or 12 weeks of the initial search. The current experimental data contains users' online browsing behaviours occurring within one month. Therefore the occurrence of online purchasing in our data is low. It is also observed that many consumers search online but later purchase the products in the store. Such a situation also leads to a smaller number of online buyers.

Through our case study, we also found certain user-centric features are more important than others on predicting online purchases, namely the features that arise in the Rule Importance Measure. For example, the number of page views an internet user spent on search websites is an indication of this person's interests. The fact that some user made a purchase online previously indicates such a user is more likely to conduct online purchases.

For this user-centric web personalization application, in addition to the study of using the Rule Importance Measure for evaluating important rules for online purchases, we also conducted related experiments on using classification algorithms including decision trees, logistic regression and Naïve Bayes for online product purchasing prediction based on this user-centric experiment data. This serves as preliminary work on applying classification algorithms for user-centric web person-

[3] http://www.comscore.com/press/release.asp?press=526

alization purchasing predictions. From our experiments, we observed that logistic regression provides a lower precision than the C4.5 decision tree, although it provides a flexible option to adjust the precision and recall for the classifiers. Naïve Bayes assumes the independence between each of the features. It is a simple classification model, although the precision is lower than logistic regression. The classification experimental results we have obtained on this user-centric clickstream data demonstrate effective product level prediction (Li, 2007).

5. Related Work

Zhu (Zhu, 2006) recently developed a user-side web personalization system "Web-IC" to predict information content (IC) pages that a web user will be interested to visit. The motivation of this system is to help web users locate these IC pages everywhere on the web for the users themselves based on their own behaviours. The words contained in the web pages a user visits, as well as the actions (such as back pages browsing or follow-up pages) the user makes on such pages are taken into consideration as users' interests for behaviour modeling. It is shown that classifiers built from such features as extracted from user-side browsing properties can effectively predict the interested webpages for the users (Zhu et al., 2003).

Although our work is similar in the fact that we both are interested in personalization towards the user-side, the purposes of our experiments and the background of the adaptive web personalization project are different from this work. We do not consider the content information (words) inside the webpages; we do not collect the user's web actions on the web pages (such as back page browsing, follow-up pages). We are interested in predicting online product purchases instead of predicting interested web pages. We are also interested in studying how site-centric algorithms can be adapted for user-centric personalization.

Other researchers studying user-centric personalization include Lieberman, who developed the Letizia web search agent for web page recommendation (Lieberman, 1995), Billsus and Pazzani, who query users to get feedback for recommending news web pages (Billsus and Pazzani, 1999) and Ardissono *et al.* who customize the presentation of a website advertising a product to a user, based on a monitoring of the user's interests (Ardissono et al., 2005). Our research is different because we are focused on how to develop effective data mining techniques for personalization.

Various user modeling researchers have also emphasized the importance of personalization in e-commerce settings. (Schafer et al.,

1999) discusses how various commercial companies have made use of methods for discovering relationships between items a customer has already purchased and other items which may be promoted to these customers, in a recommender system approach. The aim of this research is similar to ours, namely towards the ultimate goal of promoting new products to buyers. Our approach, however, involves the processing of web logs with massive amounts of data, which leads us to focus on the development of effective data mining techniques and to suggest the use of rule importance measures to assist in drawing out key relationships of use to marketers.

6. Concluding Remarks

Through a case study of a user-centric web personalization, we show how the Rule Importance Measure can be utilized and adapted to an actual system. It is used throughout the personalization system to extract important rules for purchasing predictions. The end results indicate the extracted important rules are useful to predict whether an online purchase will happen for certain users according to the observed online searching and browsing behaviours. We also have interesting experimental results on discovering prominent features for user-centric personalization applications.

The Rule Importance Measure provides a rank of how important the rules are. It would be interesting to study the cutoff threshold for extracting important rules given certain applications. The Rule Importance Measure demonstrated is based on association rule generation. We believe such measures can be widely applied towards other rule generations such as classification rules and sequential patterns (Agrawal and Srikant, 1995). In addition to extending the proposed evaluations to more application domains, we are also interested in exploring their values in a general rule evaluation framework, such as a three-level framework for the theoretical foundations of measuring and quantifying discovered knowledge based on utility theory (Y.Y. Yao, 2003; Yao et al., 2006).

In our case study of user-centric web personalization using the Rule Importance Measure, we also determined several valuable areas for future research towards improved personalization for users. First, classifying imbalanced data is a challenging process. Since most people are not online buyers, in our data set, the majority class belongs to non-buyers, and a very small percentage are buyers; out of $83,635$ number of users, the two classes of buyers and non-buyers are divided as 449 vs. $83,186$ users. Without a controlling method (such as forcing the decision tree to branch), the C4.5 decision tree classifies all the users as non-buyers. We would like to use the techniques such as (Sheng and Ling, 2006; Sun et al., 2006) from recent research on classifying imbalanced data to help solve the classification difficulties. Secondly, feature constructions require a mix of domain knowledge and a data miner's expertise. Features that can better describe an online buyer or non-buyer's intentions still need to be studied and brought into the experiment. As we perform additional experiments to determine user purchasing behaviour for different types of goods, we may also learn more about how best to design the features needed for any prediction task.

Another problem is that standard evaluation for user-centric personalization systems has yet to be formulated. For the future, we hope to work on developing appropriate benchmarks for evaluating user-centric personalization systems.

In our current work, we discussed online product purchasing predictions. The online users may display various intentions in addition to online shopping, such as blog searching, news reading and so on. It would be interesting to study the user intentions, to develop models to capture such intentions for more precise, task-oriented personalization. Finally, we are especially interested in exploring methods for developing richer user models, and investigating techniques for predicting approximate purchasing time for user online purchases.

For future research, it would also be interesting to explore the integration of our approach for user-centric personalization based on user browsing history with other systems to model customers within the organization, towards a more comprehensive marketing campaign for customers. This approach would be in tune with what (Fink and Kobsa, 2000) refers to as "user modeling servers", where an enterprise-wide model of the user can be built up from various sources, towards a more comprehensive representation that is employed by various applications.

References

Quinlan, J. C. C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., 1993.

Agresti, A. An introduction to Categorical Data Analysis. John Wiley & Sons, 1996.

Maron, M. E. Automatic Indexing: An Experimental Inquiry. *J. ACM*, **8**(3):404–417, 1961.

Agrawal, R. and Srikant, R. Fast Algorithms for Mining Association Rules. In Jorge B. Bocca and Matthias Jarke and Carlo Zaniolo, editors, *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, pp. 487–499, Morgan Kaufmann, 1994.

Padmanabhan, B., Zheng, Z., and Kimbrough, S. O. Personalization from incomplete data: what you don't know can hurt. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 154–163, 2001.

Fayyad, U., Piatesky-Shapiro, G. and Smyth, P. From Data Mining to Knowledge Discovery: an Overview. *Advances in Knowledge discovery and data mining (AAAI/MIT Press)*, pp. 1–34, 1996.

Hu, X. *Knowledge Discovery in Databases: an Attribute-Oriented Rough Set Approach.* PhD thesis, University of Regina, Regina, Canada, 1995.

Pawlak, Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Norwell, MA, USA, 1992.

Bazan, J. G., Nguyen, H. S., Nguyen, S. H., Synak, P. and Wróblewski, J. Rough set algorithms in classification problem. *Rough set methods and applications: new developments in knowledge discovery in information systems*, pp. 49–88, 2000.

Hu, X., Lin, T. Y. and Han, J. A new rough sets model based on database systems. *Fundam. Inf.* **59**(2-3), pp. 135–152, 2004.

Hilderman, R. and Hamilton, H. Knowledge discovery and interestingness measures: A survey. Technical Report CS-99-04, Department of Computer Science, University of Regina, Regina, 1999.

Pang-Ning Tan and Vipin Kumar. Interestingness Measures for Association Patterns: A Perspective. KDD 2000 Workshop on Postprocessing in Machine Learning and Data Mining, Boston, MA.

Newman, D. J., Hettich, S., Blake, C. L. and Merz, C. J. UCI Repository of machine learning databases. http://www.ics.uci.edu/∼mlearn/MLRepository.html University of California, Irvine, Dept. of Information and Computer Sciences, 1998.

Li, J. and Cercone, N. Empirical Analysis on the Geriatric Care Data Set Using Rough Sets Theory. Technical Report CS-2005-05, School of Computer Science, University of Waterloo, Waterloo, 2005.

Bruha, I. Quality of Decision Rules: Definitions and Classification Schemes for Multiple Rules. In Machine Learning and Statistics, The Interface, Edited by G. Nakhaeizadeh and C. C. Taylor, John Wiley & Sons, Inc., pp. 107–131, 1997.

Li, J. and Cercone, N. Introducing a Rule Importance Measure. *Transactions on Rough Sets* **4100**, pp. 167–189, 2006.

Øhrn, A. *Discernibility and Rough Sets in Medicine: Tools and Applications.* PhD thesis, Norwegian University of Science and Technology, Trondheim Norway, 1999.

Bazan, J. G., Szczuka, M. S. and Wróblewski, J. A New Version of Rough Set Exploration System. *RSCTC '02: Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing*, pp. 397–404, 2002.

Chiu, D. K. Y., Wong, A. K. C. and Cheung, B. Information Discovery through Hierarchical Maximum Entropy Discretization and Synthesis. *Knowledge Discovery in Databases*, pp. 125–140, 1991.

Zhu, T. *Goal-Directed Complete-Web Recommendation.* PhD Thesis, University of Alberta, Alberta, Canada, 2006.

Li, J. *Rough Set Based Rule Evaluations and Their Applications.* PhD Thesis, University of Waterloo, Waterloo, Canada, 2007.

Zhu, T., Greiner, R. and Häubl, G. Learning a Model of a Web User's Interests. *User Modeling*, pp. 65–75, 2003.

Lieberman, H. Letizia: An Agent That Assists Web Browsings. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pp. 924–929, 1995.

Billsus, D. and Pazzani, M. J. A hybrid user model for news story classification. *UM '99: Proceedings of the seventh international conference on User modeling*, pp. 99–108, 1999.

Ardissono, L., Goy, A., Petrone, G. and Segnan, M. A multi-agent infrastructure for developing personalized web-based systems. *ACM Trans. Inter. Tech.*, **5**(1): 47–69, 2005.

Sheng, V. S. and Ling, C. X. Thresholding for Making Classifiers Cost-sensitive. *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, Boston, Massachusetts, USA, 2006.

Sun, Y., Kamel, M. S. and Wang, Y. Boosting for Learning Multiple Classes with Imbalanced Class Distribution. *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006)*, Hong Kong, China, 2006.

Staal A. Vinterbo and Aleksander Øhrn. Minimal approximate hitting sets and rule templates. Int. J. Approx. Reasoning, **25**(2):123–143, 2000.

Rakesh Agrawal and Ramakrishnan Srikant. Mining Sequential Patterns. Proceedings of the Eleventh International Conference on Data Engineering, pp. 3–14, Washington, DC, USA, 1995.

Y.Y. Yao and Y.H. Chen and X.D. Yang. A Measurement-Theoretic Foundations of Rule Interestingness Evaluation. Foundations and Novel Approaches in Data Mining Series, pp. 41–59, Springer-Verlag, Berlin, 2006.

Y.Y. Yao. A Step Towards The Foundations of Data Mining. Data Mining and Knowledge Discovery: Theory, Tools, and Technology V, The International Society for Optical Engineering, pp. 254–263, 2003

Josef Fink and Alfred Kobsa. A Review and Analysis of Commerical User Modeling Servers for Personalization on the World Wide Web. User Modeling and User-Adapted Interaction., **10**(2-3): 209–249, 2000.

Schafer, J.B., Konstan, J.A., and Riedl, J. Recommender Systems in Electronic Commerce. Proceedings of the ACM Conference on Electronic Commerce (EC-99), Denver (CO), pp. 158–166, 1999.