INFORMS Journal on Computing
Vol. 20, No. 1, Winter 2008, pp. 34–45
ISSN 1091-9856 | EISSN 1526-5528 | 08 | 2001 | 0034

DOI 10.1287/ijoc.1060.0210

Discovering Generalized Profile-Association Rules for the Targeted Advertising of New Products

San-Yih Hwang
Department of Information Management, National Sun Yat-Sen University, Kaohsiung, 804, Taiwan,
syhwang@mail.nsysu.edu.tw

Wan-Shiou Yang
Department of Information Management, National Changhua University of Education, Changhua, 500, Taiwan,
wsyang@cc.ncue.edu.tw

We propose a data-mining approach for the targeted marketing of new products that have never been rated or purchased by customers. This approach uncovers associations between customer types and product genres that frequently occurred in previous transaction records. Customer types are defined in terms of demographic attribute values that can be aggregated through concept hierarchies; product types can be generalized through product taxonomies. We use generalized profile-association rules (GP association rules) to identify the advertising targets for a given new product. In addition, we propose two algorithms—GP-Apriori and Merge-prune—to mine GP association rules and develop a value-based targeted advertising algorithm to select prospective customers of a new product on the basis of the discovered rules. We evaluate the proposed approach using both synthetic data and library-circulation data.

Key words: data mining; profile-association rules; generalized profile-association rules; fractional 0–1 knapsack problem; greedy algorithms; targeted advertising; recommender systems
History: Accepted by Amit Basu, Area Editor for Knowledge and Data Management; received November 2002; revised December 2003, September 2004, March 2005, September 2005, December 2005, July 2006, August 2006; accepted October 2006. Published online in *Articles in Advance* November 19, 2007.

1. Introduction

Businesses have long recognized the value of targeting profitable customer segments through advertising. With the emergence of the Internet as a low-latency, low-cost channel, customer solicitation has reached unprecedented levels, and both Web page banners and e-mails are used widely for advertising. Although advertising on the Internet costs little, its efficacy has been constantly questioned. Many studies suggest that customers tend to ignore banners through a phenomenon called "banner blindness" (Benway and Lane 1998), and unsolicited commercial e-mails (i.e., spam) can create legal problems that may impose substantial costs on both providers and e-mail users (http://www.spamlaws.com). Today, the conventional wisdom, which recommends targeting a few good customers with large incentives, seems more important than ever (Cheyne and Ritter 2001).

Traditional approaches to targeted advertising analyze a historical database of transactions and use statistical tools to identify those customers most likely to respond to an advertisement of a given product. New technologies also have led to recommender systems that can identify potential customers (Hayes 1994) according to customer preference ratings either explicitly provided by the customers or implicitly inferred from previous transaction records, web logs, or cookies.

The first type of recommender system, the content-based approach (Loeb and Terry 1992), characterizes recommendable products according to a set of content features and customers according to an analogous feature set. Customers whose interests are similar to the content profile of the product are then targeted. To establish a more accurate content profile for the product, its detailed description must be parsable (e.g., as text) so that the set of content features can be extracted by information extraction or summarization techniques (Mooney and Roy 2000). However, the content-based approach clearly is inappropriate for products whose profile is not electronically available. Furthermore, because a customer's content features are derived purely from the products in which he or she has shown an interest, this approach probably cannot offer recommendations for novel products.

Another type of recommendation technique, the collaborative approach (sometimes called the social-based approach), remedies this problem by considering customers' interest profiles (Shardanand and Maes 1995). Specifically, the collaborative approach looks for similarities among customers by observing the ratings they assign to products in a small training set. Nearest-neighbor customers are those who exhibit the strongest similarity to the target customer and act as recommendation partners for the target customer. Therefore, products that appear within the profiles of recommendation partners should be advertised to the target customer. Although the utility of this approach has been demonstrated in many applications, it has several limitations, such as its inability to advertise either newly introduced products that have yet to be rated by customers or products to a new customer who has not provided any rating data (Mooney and Roy 2000). For a survey of recommendation techniques, see Huang et al. (2004) and Adomavicius and Tuzhilin (2005). More discussion on the relevant work can be found in Appendix 1 in the Online Supplement to this paper (available at http://joc.pubs.informs.org/ecompanion.html).

Products can often be categorized by an existing classification scheme, and customers with similar demographic characteristics often demonstrate similar preferences for certain product types. Therefore, aggregation information about both customer demographics and products should enable targeted advertising to recommend new products to new customers. We investigate advertising of new products in an environment with the following features:

1. The advertised products have not been sold before and therefore no transaction records pertain to them, which makes the collaborative approach inapplicable.
2. Both existing and new customers can be targeted. This feature makes both the content-based and the collaborative approaches inapplicable because neither can provide a new customer (without ratings) with an accurate recommendation.
3. A standardized (usually industry-recognized) product taxonomy exists that helps determine the similarity between products.
4. Customer demographics, such as the name, address, gender, highest level of education, occupation, and family socioeconomic status, are available. Some of these attributes have concept hierarchies that offer different levels of aggregation.

These features exist in many application domains, such as membership stores that must conduct targeted marketing for newly introduced products, online literature databases that seek to recommend new articles to their members, and libraries or bookstores that seek to promote new books to their patrons.

With the preceding restrictions in mind, we develop a data-mining approach for advertising new products (§§2–4). Our approach starts by identifying associations between the types of customers and product types that frequently appear in the transaction database, which we call GP association rules. Using the set of GP association rules, we develop an algorithm for identifying the characteristics of prospective customers and the probabilities of their liking a new product, information that is crucial to the successful online promotion of new products. We evaluate the performance of the proposed approach using both empirical data obtained from National Sun Yat-Sen University (NSYSU) library circulation information and synthetic data (§5).

2. The Problem

We attempt to identify a set of strong associations between types of customers and genres of products that frequently appear in a transaction database and then generate a list of potential customers for a new product on the basis of the discovered associations. Appendix 2 in the Online Supplement summarizes our notation. There are n demographic attributes in domains $D_1, \ldots, D_n$, each of which contains primitive and aggregated literals pertaining to a particular attribute. Let P be the set of product items and C be the set of product categories. Each product item in P must belong to one or more categories in C, and a taxonomy $H(C)$ is a tree whose set of nodes are C and whose set of links represent an "is-a" relationship. An aggregation hierarchy on the ith demographic attribute, denoted $H(D_i)$, is a tree whose set of nodes is D_i. A link in $H(D_i)$ represents an is-a relationship.

Let T be a set of purchase transactions $t = (d_1, d_2, \ldots, d_n, p_1, p_2, \ldots)$ where d_i is a leaf in $H(D_i)$ and represents the ith demographic attribute value of the customer engaging in the transaction, and $p_j \in P$ is the jth product that the customer has purchased. To mine the GP association rules, we group transactions by the same customer, which results in a new type of transaction, called a demographic product transaction. That is, the oth demographic product transaction can be represented as a set $t_o = \{d_{o,1}, d_{o,2}, \ldots, d_{o,n}, c_{o,1}, c_{o,2}, \ldots, c_{o,m}\}$, where $d_{o,i}$ is a leaf in $H(D_i)$ that represents the ith demographic attribute value of the oth customer, and $c_{o,j}$ is a leaf node in $H(C)$ that represents the jth basic category of the product items that the oth customer has purchased. Unless otherwise stated, when we refer to a transaction, we mean a demographic product transaction.

Because our goal is to identify the associations between customer demographic types and product categories, we must convert the demographic values and basic product categories presented in each transaction into demographic types and general product categories, respectively, which results in an extended transaction (Srikant and Agrawal 1995). We include all

Table 1 Sample Transactions

Transaction	Age	Gender	Products
t_1	30	Male	Regular coffee, green tea
t_2	40	Female	Black tea
t_3	40	Male	Decaf coffee, green tea
t_4	60	Male	Decaf coffee, regular coffee

demographic types of each demographic value and all product categories of each product item that appear in the transaction without duplication. Therefore, the oth transaction can be translated into the extended transaction $t'_o = \{d'_{o,1}, d'_{o,2}, \ldots, d'_{o,n'}, c'_{o,1}, c'_{o,2}, \ldots, c'_{o,m'}\}$ ($n' \geq n$, $m' \geq m$), where $d'_{o,i} \in D_1 \cup D_2 \cup \cdots \cup D_n$ ($1 \leq i \leq n'$) and $c'_{o,j} \in C$ ($1 \leq j \leq m'$). The transaction t_o supports a demographic type $X \subseteq D_1 \cup D_2 \cup \cdots \cup D_n$ if $X \subset t'_o$, where t'_o is the extended transaction of t_o. Similarly, t_o supports a product category c if $c \in t'_o$.

In turn, a GP association rule results from the form $X \to c$, where $X \subseteq D_1 \cup D_2 \cup \cdots \cup D_n$ and $c \in C$. The support of the rule $X \to c$ is the proportion of transactions that support both the demographic type X and the product category c. Moreover, the confidence of the rule $X \to c$ refers to the fraction of transactions that support c among those that support X. Therefore, given a set of transactions T, several demographic aggregation hierarchies $H(D_1), H(D_2), \ldots, H(D_n)$ (each represents the generalization of one demographic attribute), and one product taxonomy $H(C)$, to mine GP association rules from transaction data, requires discovering all rules that have support and confidence no less than the user-specified minimum support ($\mathrm{Min}_{\mathrm{sup}}$) and minimum confidence ($\mathrm{Min}_{\mathrm{conf}}$). These rules are called valid GP association rules.

For the four transactions in Table 1, consider as well the concept hierarchy for the demographic attribute age and the product taxonomy in Figure 1. The first transaction t_1: {30, male, regular coffee, green tea} can be extended to {30, young, adult, male, regular coffee, coffee, green tea, tea, beverage} by consulting the age hierarchy and product taxonomy.

Consider the following three GP association rules from Table 1:

R_1: {male} → regular coffee.
R_2: {middle-aged male} → coffee.
R_3: {adult} → coffee.

From our previous discussion, we can determine that the supports for R_1, R_2, and R_3 are 0.5, 0.25, and 0.75, respectively, and their respective confidences are 0.67, 1.0, and 0.75. If we let $\mathrm{Min}_{\mathrm{sup}}$ and $\mathrm{Min}_{\mathrm{conf}}$ be 0.3 and 0.7, respectively, R_3 is the only valid GP association rule.

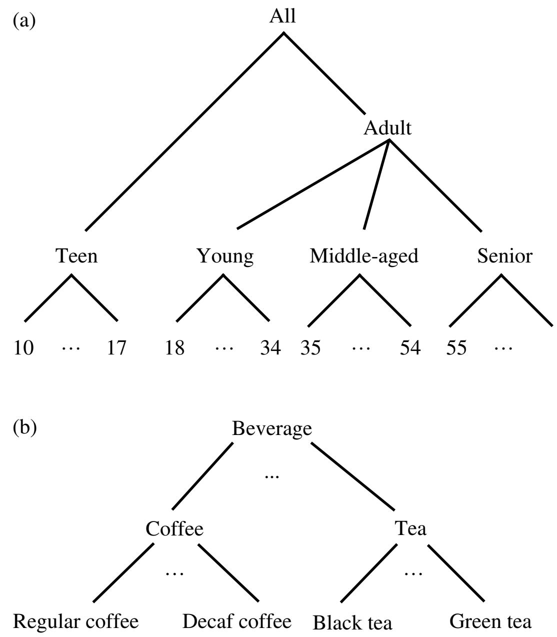

Figure 1 (a) Concept Hierarchy of Age and (b) Product Taxonomy

3. Identifying GP Association Rules

In this section, we describe a method to locate associations between (generalized) products and (generalized) values of the relevant demographic attributes. As noted by Agrawal and Srikant (1994), the bottleneck in finding association rules lies in enumerating frequent item sets. We develop two algorithms that generate frequent item sets comprising both product categories and demographic types: GP-Apriori and Merge-prune.

A demographic product item set contains several demographic attribute values and one product category. If a demographic product item set contains k demographic attribute values, of the form $(d_{i_1}, d_{i_2}, \ldots, d_{i_k}, c)$, where $d_{i_j} \in D_{i_j}$ ($1 \leq j \leq k$) and $c \in C$, it is called a demographic product k-item set. We define a demographic product item set as frequent if it is supported by at least a user-specified number of transactions. We seek to identify all frequent demographic product item sets, given a transaction database, several demographic concept hierarchies, and a product taxonomy.

3.1. GP-Apriori Algorithm

The first approach is based on Srikant and Agrawal (1995) for mining generalized association rules. We modify the means to generate candidate item sets. Let $Cand_k$ denote the collection of candidate demographic product k-item sets and F_k denote the collection of frequent demographic product k-item sets. $Cand_{k+1}$ emerges from joining F_k and F_k in a way similar to the *Apriori* candidate-generation algorithm (Agrawal and Srikant 1994, Srikant and Agrawal 1995), except that the k common attributes must include one product category (c) and the other $k-1$ demographic attribute values (from $(d_{i_1}, d_{i_2}, \ldots, d_{i_k})$).

We first extend each transaction $t_o = \{d_{o,1}, d_{o,2}, \ldots, d_{o,n}, c_{o,1}, c_{o,2}, \ldots, c_{o,m}\}$ in T by adding all ancestors of $d_{o,i}$ $(1 \le i \le n)$ from the concept hierarchy of the ith demographic attribute and all ancestors of $c_{o,j}$ $(1 \le j \le m)$ from the product taxonomy. The set of extended transactions is ET, and scanning it produces the frequent demographic one-item sets FD_1 and the frequent product one-item sets FP_1. If an item is in neither FD_1 nor FP_1, it will not appear in any frequent demographic product item set and is therefore useless; we delete all useless items in every transaction of ET to reduce its size. The set of candidate demographic product one-item sets $Cand_1$ is $FD_1 \times FP_1$. We scan ET again to find the set F_1 of frequent demographic product one-item sets from $Cand_1$.

A subsequent step (e.g., pass k) is composed of two steps. First, we use the candidate-generation function to generate the set $Cand_k$ of candidate demographic product item sets by joining two frequent demographic product $(k-1)$-item sets in F_{k-1} on the basis of their common $k-2$ demographic attribute values and product category. Second, we scan ET and quantify the support of candidates in $Cand_k$. The set F_k includes the demographic product item sets in $Cand_k$ with minimum support. This algorithm is called "GP-Apriori" because it is an extension of the Apriori algorithm for finding GP association rules; its pseudocode is in Appendix 3 in the Online Supplement. Note that numerous optimization strategies, such as *AprioriTid, stratifying*, and *estimates* (Agrawal and Srikant 1994, Srikant and Agrawal 1995), are not included in the pseudocode because they are secondary to this work and can be selectively adopted for more efficient implementation according to the specific application.

3.2. Merge-Prune Algorithm

If there are k demographic attributes the GP-Apriori algorithm needs to perform up to k iterations, each requiring one pass over the database of extended transactions. The candidate set at each iteration can be large, so computation is substantial. To remedy both problems, we improve the GP-Apriori algorithm with two techniques, *merge* and *prune*, resulting in a new algorithm Merge-prune. The merge technique partitions the set of demographic attributes into a set of disjointed attribute groups, and the concept hierarchies of attributes in the same group are merged into a lattice structure. Let $\{A_1, A_2, \ldots, A_h\}$ be a group of h demographic attributes with corresponding concept hierarchies, $H(D_1), H(D_2), \ldots, H(D_h)$, where D_i is the domain of A_i. A node in the merged lattice is a tuple $(d_1, d_2, \ldots, d_h)$, where d_i is a node in $H(D_i)$ and has h parent nodes—$(\text{parent}(d_1), d_2, \ldots, d_h)$, $(d_1, \text{parent}(d_2), \ldots, d_h), \ldots, (d_1, d_2, \ldots, \text{parent}(d_h))$—such that parent$(d_i)$ is the parent of d_i in $H(D_i)$. Now consider two demographic attributes A and B with concept hierarchies in Figure 2. The parent nodes of (a_{11}, b_{21}) in the merged lattice are (a_{11}, b_2) and (a_1, b_{21}).

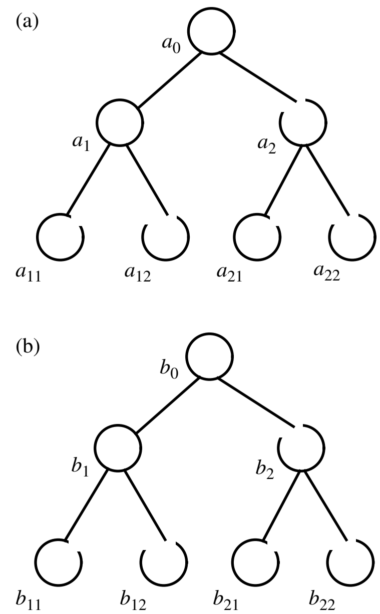

Figure 2 **Concept Hierarchies of Demographic Attributes *A* and *B***

After being merged, attributes of the same group are collectively regarded as an attribute during computation of frequent item sets. The merge technique generalizes the GP-Apriori algorithm; when each demographic attribute constitutes a group, the merge technique has no effect, and the entire algorithm is reduced to GP-Apriori. By properly tuning the size of the partition, the merge technique reduces the number of passes across the database and limits the size of the candidate set. For a given transaction database, the ideal number of attributes in each group is a nondecreasing function of the size of the main memory. We leave exploration for the optimal size of the partition as future work.

The prune technique further reduces the size of the candidate set generated at each iteration. Recall that the first candidate demographic product item set ($Cand_1$) of GP-Apriori is the Cartesian product of the set of frequent demographic one-item sets FD_1 and the set of frequent product one-item sets FP_1. However, we need not join every pair of elements d and c from FD_1 and FP_1 if we recognize that the combination of d and c cannot form a frequent demographic product item set. Such awareness is possible if certain information can be recorded when we scan the database to obtain FD_1 and FP_1. Let U be any set of nodes at some higher levels of demographic lattices that cover all primitive demographic values and V be any set of nodes at some higher levels of the product

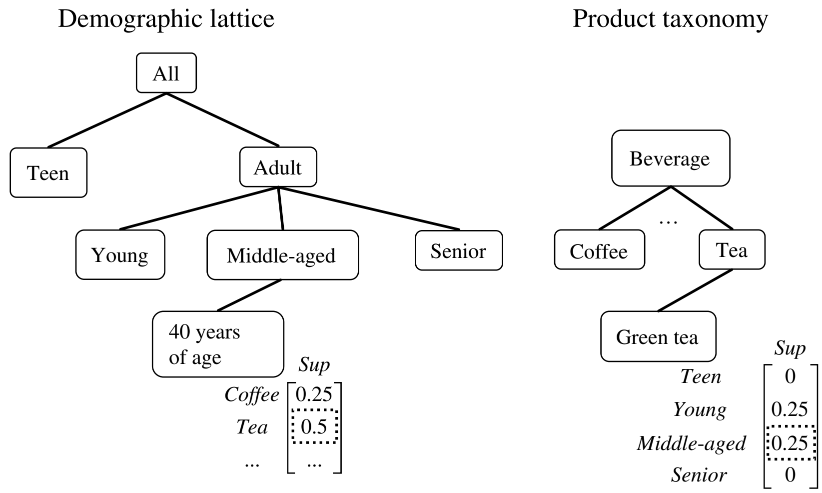

Figure 3 Graphical Illustration of Early Pruning

taxonomy that cover all product items. For each element c in FP_1, we maintain an array of demographic supports, each of which records the ratio of extended transactions that involve c and a demographic type $d' \in U$. Similarly, for each element d in FD_1, we maintain both its support and an array of product supports, each of which records the ratio of extended transactions that involve d and a product category $c' \in V$. Now consider an element d in FD_1 and an element c in FP_1. Let $d' \in U$ be equal to d or the ancestor of d in a demographic lattice and $c' \in V$ be equal to c or the ancestor of c in the product taxonomy. The combination of d and c is frequent only if both the support of (c, d'), denoted $c.\text{sup}[d']$, and the support of (c', d), denoted $d.\text{sup}[c']$, are no less than Min_{sup}. In Figure 3, we provide an illustration of the related notation using the transaction data in Table 1 and the concept hierarchy and product taxonomy in Figure 1, where (40 years of age, green tea) is frequent only if both (40 years of age).sup[*tea*] and (green tea).sup[*middle-aged*], as surrounded by dashed lines, are no less than Min_{sup}.

To extend our preceding example, let $U = \{\text{teen}, \text{young}, \text{middle-aged}, \text{senior}\}$ and $V = \{\text{coffee}, \text{tea}, \ldots\}$. After scanning the transaction data once and computing FD_1 and FP_1, we can identify the information for the demographic attribute value "40 years of age" and the product category "green tea" as in Table 2.

In addition, let $\text{Min}_{\text{sup}} = 0.3$. Obviously, 40 years of age $\in FD_1$ and green tea $\in FP_1$. To determine whether (40 years of age, green tea) may be frequent, we determine whether both the product support of 40 years of age on tea (tea $\in V$ is the ancestor of green tea) and the demographic support of green tea on middle-aged (middle-aged $\in U$ is the ancestor of 40 years of age) are no less than Min_{sup}. We note that (40 years of age, green tea) does not pass the test, so we can exclude (40 years of age, green tea) from $Cand_1$.

The same idea can be applied to the generation of candidate sets in subsequent iterations; we provide pseudocode for Merge-prune in Appendix 4 in the Online Supplement.

4. Targeted Advertising Based on GP Association Rules

Given a new product, the set of discovered GP association rules should identify the subset of customers for whom the promotion might be effective. However, for a given product p, there could be a large number of rules with consequent matching p, and some antecedents of these rules could be very broad. Therefore, it is inappropriate simply to choose those rules with higher confidence and target customers whose demographics match the antecedents of these rules. Consider R_1: age 30–50 years $\rightarrow$ coffee with confidence $= 0.7$ and R_2: age 30–50 years $\rightarrow$ decaf coffee with confidence $= 0.5$. To promote a brand of decaffeinated coffee, it makes sense to use R_2 and assume that half of the customers aged 30 to 50 years like decaffeinated coffee. However, when it comes to the target identification of a regular coffee product, the importance of R_1 must not be overemphasized because many transactions that support R_1 could come from those that support R_2. A more useful rule in such a case is R_1': age 30–50 years $\rightarrow$ nondecaf

Table 2 Information About Demographic Value "40 Years of Age" and Product Category "Green Tea"

40 years of age	Green tea
Support = 0.5	Support = 0.5
Product support on coffee = 0.25	Demographic support on young = 0.25
Product support on tea = 0.5	Demographic support on middle-aged = 0.25

coffee. Unfortunately, the confidence of R_1' might be unknown and hence estimated only from the set of related valid rules. In our example, we can conclude that the confidence of R_1' is at least 0.2 because there may exist transactions that support both R_1 and R_2. Techniques for more accurately estimating the confidence of R_1' therefore need to be developed.

We call a valid GP association rule R_1: $X' \to c_1$ a product-ancestor of another valid rule R_2: $X'' \to c_2$ if $X' = X''$ and c_1 is an ancestor of c_2 in the product taxonomy. Conversely, R_2 is called a product-descendant of R_1. Using the product-descendant relation (which is a partial order), we can form a hierarchy on the set of discovered GP association rules. Let the children of a product category c in the product taxonomy be $c_1, c_2, \ldots, c_{m_c}$. Without loss of generality, assume a rule R: $X \to c$ has k immediate product descendants $X \to c_1, X \to c_2, \ldots, X \to c_k$ $(k \le m_c)$ in the rule hierarchy. We want to derive a means of estimating the confidence $\mathrm{Conf}(R_j)$ of R_j: $X \to c_j$ $(k < j \le m_c)$ to recommend a product of category c_j. Assume that on average each transaction that supports c also supports β_c of its child product categories in the product taxonomy, where β_c $(\beta_c \ge 1)$ is the overlapping factor of c. Therefore, we have the estimation

$$\mathrm{Conf}(R)\beta_c = \sum_{i=1..m_c} \mathrm{Conf}(R_i) = \sum_{i=1}^{k} \mathrm{Conf}(R_i) + \sum_{j=k+1}^{m_c} \mathrm{Conf}(R_j).$$

Equivalently,

$$\sum_{j=k+1}^{m_c} \mathrm{Conf}(R_j) = \mathrm{Conf}(R)\beta_c - \sum_{i=1}^{k} \mathrm{Conf}(R_i).$$

Assuming that the confidence of the rule R_j: $X \to c_j$ $(k < j \le m_c)$ is proportional to the support $\mathrm{Sup}(c_j)$ of c_j, we can estimate the confidence of R_j as

$$\mathrm{Conf}(R_j) \approx \left(\mathrm{Conf}(R)\beta_c - \sum_{i=1}^{k} \mathrm{Conf}(R_i)\right) \cdot \mathrm{Sup}(c_j) \Big/ \sum_{i=k+1}^{m_c} \mathrm{Sup}(c_i), \quad k < j \le m_c.$$

This estimated confidence of R_j derived from R can be regarded as the value (or usefulness) of R with respect to advertising a product from c_j to a customer of X. Specifically, we define a value function of R: $X \to c$ with respect to a child product category c_i of c $(1 < i \le m_c)$, which we denote $\mathrm{Value}(R, c_i)$, as

$$\mathrm{Value}(R, c_i) = \left\{ \begin{array}{l} 0 \quad \text{if } 1 \le i \le k \\ \left(\mathrm{Conf}(R)\beta_c - \sum_{j=1}^{k} \mathrm{Conf}(R_i)\right) \mathrm{Sup}(c_i) \Big/ \sum_{j=k+1}^{m_c} \mathrm{Sup}(c_j), \\ \quad \text{if } k < i \le m_c \end{array} \right\}.$$

Note that $\mathrm{Value}(R, c_i)$ is defined as 0 for $1 \le i \le k$, because the existence of a more specific GP association rule $X \to c_i$ eliminates the values of the rule R: $X \to c$ for promoting a product of category c_i. Here, β_c can be computed empirically as the average number of child product categories of c that appear in a transaction that supports c. Such a computation can be conducted when computing the support of c using either GP-Apriori or Merge-prune, and the overhead costs are negligible.

The value function $\mathrm{Value}(R, c_{ij})$ of R: $X \to c$ with respect to a grandchild c_{ij} of c (i.e., c_{ij} is the jth child of the c_i of c in the product taxonomy) can be similarly derived as

$$\mathrm{Value}(R, c_{ij}) = \left\{ \begin{array}{l} 0 \quad \text{if } 1 \le i \le k \\ \mathrm{Value}(R, c_i)\beta_{c_i}\mathrm{Sup}(c_{ij}) \Big/ \sum_{k=1}^{m_{c_i}} \mathrm{Sup}(c_{ik}), \\ \quad \text{if } k < i \le m_c \end{array} \right\},$$

m_{c_i} and β_{c_i} are the number of children and the overlapping factor of c_i (the parent of c_{ij}) in the product taxonomy, respectively.

The value of a rule with respect to a product category at a lower level can be similarly induced. For completion, we also define the function $\mathrm{Value}(R, c)$ of R: $X \to c$ with respect to c as simply the confidence of R, $\mathrm{Conf}(R)$.

Consider the five GP association rules in Table 3 and assume that coffee has three child product categories in the product taxonomy: regular coffee, espresso, and decaf coffee, whose supports are 0.3, 0.1, and 0.2, respectively.

Note that R_2 is a product-descendant of R_1 and R_5 is a product-descendant of R_3. Suppose that we are promoting a brand of regular coffee and the overlapping factor β_{Coffee} is 1.3. In this case, there are four rules related to the recommendation of the new product: R_1, R_3, R_4, and R_5. Because the consequent of R_5 is regular coffee, $\mathrm{Value}(R_5, \text{regular coffee}) = 0.4$. R_4 has no product-descendants, so $\mathrm{Value}(R_4, \text{regular coffee}) = (0.6 \times 1.3)[0.3/(0.3 + 0.1 + 0.2)] = 0.39$. As R_3 has a product-descendant R_5 that has already been used, $\mathrm{Value}(R_3, \text{regular coffee}) = 0$. Finally, taking into account the product-descendant R_2 of R_1, $\mathrm{Value}(R_1,$

Table 3 **GP Association Rules for Coffee**

Rule	Demographic	Product	Confidence
R_1	Age 20–50 years	Coffee	0.7
R_2	Age 20–50 years	Decaf coffee	0.5
R_3	Age 20–30 years	Coffee	0.6
R_4	Male, age 20–50 years	Coffee	0.6
R_5	Age 20–30 years	Regular coffee	0.4

regular coffee) $= (0.7 \times 1.3 - 0.5)[0.3/(0.3 + 0.1)] \approx 0.303$.

For a product p, we first locate the leaf product category c_p in the product taxonomy such that $p \in c_p$. We then identify the set of valid GP association rules whose consequents are equal to or ancestors of c_p in the product taxonomy. We call this set of rules the matching rule set of p. Now we can define a partial order on the matching rule set of p based on the demographic parts (i.e., antecedents) of the rules. Specifically, we call a rule R_1: $X' \to c_1$ a demographic-ancestor of another rule R_2: $X'' \to c_2$ if $c_1 = c_2$ and $X' \supset X''$; conversely, R_2 is a demographic-descendant of R_1. Based on the demographic-descendant relation, we can define a lattice on the matching rule set. Consider a rule R: $X \to c$ in the matching rule set of p with k immediate demographic-descendants ($X_1 \to c$, $X_2 \to c, \ldots, X_k \to c$) in the lattice. We define the applicable demographic domain of R with respect to p, denoted ADD(R, p), as $X - X_1 - X_2 - \cdots - X_k$. Then, ADD$(R, p)$ represents the set of customers chosen as targets for advertising p when R is selected. Note that the demographic domain X_i $(1 \le i \le k)$ is excluded from ADD(R, p) because, when deciding the degree to which a product $p \in c$ should be promoted to customers with demographics X_i, we should consult the more specific rule $X_i \to c$; we consult R only when we need to decide whether to target a customer in $X - X_1 - X_2 - \cdots - X_k$ to promote p.

Now consider the same five GP association rules in Table 3 for a brand of regular coffee p. The matching rule set of p is $\{R_1, R_3, R_4, R_5\}$. Note that R_3 and R_4 are demographic-descendants of R_1. Thus, ADD$(R_1, p) =$ {female, age 30–50} (i.e., {age 20–50} − {age 20–30} − {male, age 20–50}). We summarize the ADD( ) and Value( ) of each rule in Table 4.

Now suppose we are given a product p and a subset M_p of the matching rule set of p, where each rule R in M_p is associated with a nonzero value Value(R, c_p) and an applicable demographic domain ADD(R, p). However, the applicable demographic domains in M_p may not be mutually disjoint. In our previous example, ADD(R_4, p) and ADD(R_5, p) are not disjoint, and it is not clear what value a male customer between 20 and 30 years of age should receive. To remedy this problem, we fragment the demographic domain covered by M_p into a set of mutually disjoint segments.

Table 4 Detailed Information of GP Association Rules for a Brand of Regular Coffee *p*

Rule	Demographic	Product	ADD(*Rule*, *p*)	Value(*Rule*, regular coffee)
R_1	Age 20–50	Coffee	Female, age 30–50	0.303
R_2	Age 20–50	Decaf coffee	N/A	N/A
R_3	Age 20–30	Coffee	Age 20–30	0
R_4	Male, age 20–50	Coffee	Male, age 20–50	0.390
R_5	Age 20–30	Regular coffee	Age 20–30	0.400

Table 5 Disjoint Segments and Their Value for a Brand of Regular Coffee *p*

Segment	SegValue(Segment, *p*)
Female, age 30–50	0.303
Male, age 30–50	0.390
Female, age 20–30	0.400
Male, age 20–30	0.395

The value of each segment s with respect to a given product p, denoted SegValue(s, p), is the average of Value(R_i, c_p), where $s \subseteq$ ADD(R_i, p) and $R_i \in M_p$. From our preceding example, we can identify four segments in Table 5.

Now we identify a target set of size N for advertising a given new product p, where N is specified by the user according to budgetary constraints and other factors. Let Seg$_p$ be the set of disjoint segments with respect to p. The problem of selecting the set of targeted customers can be formulated as

$$\underset{Y_s,\, s\in \mathrm{Seg}_p}{\text{Maximize}} \sum_{s\in \mathrm{Seg}_p} \mathrm{SegValue}(s, p) Y_s |s|$$

$$\text{subject to} \sum_{s\in \mathrm{Seg}_p} Y_s |s| \le N; \quad 0 \le Y_s \le 1.$$

Y_s is the ratio of customers in segment s to be added to the target set for advertisements of p. This LP is a special case of the fractional 0–1 knapsack problem and can be efficiently solved by a greedy algorithm that incrementally selects the segment with the highest value until N customers are chosen. The pseudocode of this algorithm *Value-based-targeted-adv* appears in Appendix 5 in the Online Supplement.

The problem of selecting advertising targets can be further complicated by additional constraints. For example, to advertise a set of products as Web page banners, the set of products becomes a set of ads and the set of customer segments becomes a set of customer groups. More precisely, let the set of customer segments be $\{s_1, s_2, \ldots, s_m\}$ and the set of products be $\{p_1, p_2, \ldots, p_n\}$. v_{ij} denotes the value of s_i with respect to p_j as computed previously, and Z_{ij} denotes the display probability of p_j given a customer segment s_i. Also, the desired display probability of p_j, determined on the basis of its advertising expense, is w_j $(1 \le j \le n)$, and the probability that customers in segment s_i show up is u_i $(1 \le i \le m)$. The model thus can be formulated as

$$\underset{Z_{i,j},\, 1\le i\le m,\, 1\le j\le n}{\text{Maximize}} \sum_{i=1}^{m} \sum_{j=1}^{n} v_{ij} u_i Z_{ij}$$

$$\text{subject to} \sum_{j=1}^{n} Z_{ij} = 1, \quad 1 \le i \le m;$$

$$\sum_{i=1}^{m} u_i Z_{ij} = w_j, \quad 1 \le j \le n;\ 0 \le Z_{ij} \le 1.$$

This LP is on a much larger scale. Various techniques have been proposed for solving this problem efficiently (Langheinrich et al. 1999, Nakamura 2002, Tomlin 2000), concise descriptions of which we provide in Appendix 1 in Online Supplement. Our main contribution here is identifying how to compute v_{ij} from the set of valid GP association rules such that existing techniques can be applied rather than proposing yet another efficient algorithm for solving the banner-allocation problem.

5. Evaluation

In this section, we apply our approach to synthetic and empirical data. Synthetic data enable us to compare the performance of our algorithms in discovering GP association rules for various operating scenarios. From our application to the circulation data of a university library, we determine the effectiveness of the discovered rules for the promotion of new books.

5.1. Performance Evaluation Using Synthetic Data

5.1.1. Generation of Synthetic Data. We generated transactional data and concept hierarchies of demographic attributes and products similarly to Agrawal and Srikant (1994) and Srikant and Agrawal (1995). We first created a product hierarchy and several demographic attribute hierarchies. Each node in the hierarchy is associated with a weight indicating the likelihood that it will be picked up. We then constructed a product pool containing a set of potentially frequent product item sets and a demographic product pool composed a set of potentially frequent demographic product item sets. To form a transaction, we equiprobably picked a potential product item set from the product pool and determined a set of demographic attribute values from a matching item set in the demographic product pool.

To create a single product hierarchy and the product pool, we used the product parameters in Table 6.

Table 6 **Parameters for Generating Synthetic Data**

Product parameters	
PP	Average size of an item set of potentially frequent products
PI	Number of item sets of potentially frequent products
PN	Number of product nodes
PF	Average degree of fan out of each internal node in the product hierarchy
Demographic parameters	
DI	Number of item sets of potentially frequent demographic products
DN	Total number of demographic attribute values
DK	Number of demographic attributes
DF	Average degree of fan out of each internal node in a demographic concept hierarchy
Transaction parameters	
TN	Number of transactions
TS	Average transaction size

For each internal node in the product hierarchy, we determined the number of children using a Poisson distribution with mean *PF*. We assigned children to the root, to the nodes at the next level, and so on, until we reached all *PN* product nodes. Each product node in the hierarchy also had a weight associated with it, and we equiprobably assigned a weight to each leaf node so that the sum of the weights for all leaf nodes was 1. The weight of each internal node was the sum of the weights of its children. Finally, we normalized the weights of all nodes in the product hierarchy so that the sum of the weights for all nodes was 1.

We generated a pool of *PI* item sets of potential frequent products. For each item set of potentially frequent products, we first determined its size using a Poisson distribution with mean *PP*. Then we populated the set by tossing a *PN*-sided weighted coin, where the weight of each side was the probability of choosing the associated product item. Each product item set in the product pool had a weight equal to the product of the weights of all its constituent product items. Again, the weights of item sets in the product pool were normalized so that the sum of the weights was 1.

To create demographic hierarchies and the demographic product pool, we used the demographic parameters in Table 6. We constructed the *DK* demographic attribute hierarchies just as we did the product hierarchy by assuming that each hierarchy had DN/DK nodes. Next, we generated the pool of *DI* item sets of potentially frequent demographic products. A demographic product item set in the pool was generated by picking one node from each demographic hierarchy and one product node at the second level of the product hierarchy according to the weights assigned to them. The weight of a demographic product item set was equal to the product of the weights of its constituent items. We normalized the weights of all demographic product item sets in the pool that had the same product value so that the sum of their weights was equal to 1.

We created transactions on the basis of the transaction parameters in Table 6. Given the product pool and the demographic product pool, we generated *TN* transactions as follows: We sampled from a Poisson distribution with mean *TS* to determine the size of a transaction. We then picked a product item set from the product pool with probability being its respective weight. Each nonleaf product value in the item set was specialized to a leaf product by tossing an m-sided weighted coin (where m is the number of children of the node in the hierarchy) consecutively until we reached a leaf product. This method is iteratively performed until the transaction contains the desired number of product values. The demographic attribute values of the transaction were determined

Table 7 Parameter Values

Min_{sup}	*PI*	*PN*	*PP*	*PF*	*DK*	*DI*	*DN*	*DF*	*TN*	*TS*
0.05	80	1,000	6	5	6	200	1,000	10	10,000	20

as follows: We first located the product type *c* at the second level of the product hierarchy that had the maximal number of product values in the transaction. All demographic product item sets in the demographic product pool for which the product value was *c* became candidates for selection. We randomly chose one of these candidates according to their weights. Each nonleaf demographic attribute value in the selected demographic product item set was specialized to a leaf demographic attribute value in the same way as that for product values, which provided the demographic attribute values of the transaction.

We performed the experiments on a 400 MHz Celeron personal computer with 256 MB of main memory running Microsoft Windows 2000. The parameter values are in Table 7.

5.1.2. Performance of GP-Apriori. We first examine the performance of GP-Apriori under the base settings in Table 7, except that the total number of transactions *TN* varies between 2,000 and 10,000. As we show in Figure 4, the running time is almost linearly proportional to the number of transactions. This result coincides with previous reports (Agrawal and Srikant 1994), which is rational because GP-Apriori follows the same approach as the Apriori algorithm.

5.1.3. Impact of Merging and Pruning on the Merge-Prune Algorithm. We next examine the impact of the merging technique in the Merge-prune algorithm. Six demographic attributes were merged in four different ways, $(1,1,1,1,1,1)$, $(2,2,2)$, $(3,3)$, and (6), which partition the demographic attributes into six, three, two, and one groups, respectively. Note that the merging pattern $(1,1,1,1,1,1)$ induces no attribute merge, resulting in an algorithm that is basically GP-Apriori with a prune technique. We summarize their relative performances in Figure 5. Note that the figure does not show the running time for

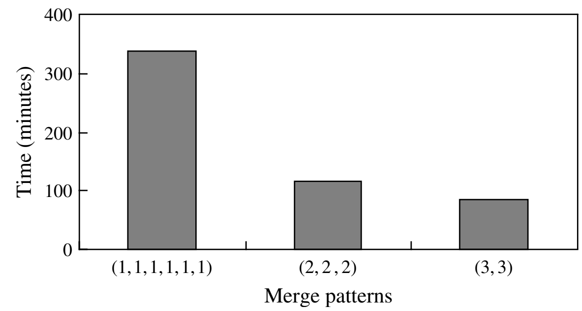

Figure 5 **Impact of Merging for Different Attribute Merging Patterns (with Pruning)**

the single group (6), which is much longer because it results in a huge demographic attribute lattice and many candidate item sets. We observe that $(3,3)$ incurs the least computation overhead, because it requires the database to be scanned only three times. Comparing the execution time for merging pattern $(1,1,1,1,1,1)$ and that of GP-Apriori (as shown in Figure 4) reveals that Merge-prune with merging pattern $(1,1,1,1,1,1)$ has a slightly longer running time than GP-Apriori, so the overhead of the prune technique for GP-Apriori more than offsets the cost gain from its reduction in the size of candidates.

We next investigated the impact of the pruning technique by varying the number of transactions from 2,000 to 10,000 and merging the demographic attributes into two groups $(3,3)$. Figure 6 shows the execution time with and without the pruning technique. Although the pruning technique requires extra time to calculate an array of supports for each frequent item set, the total execution time is still lower than without pruning because of the significantly smaller candidate set. This set of experiments demonstrates how the effectiveness of both merging and pruning techniques contributes to the performance of the Merge-prune algorithm, such that it is superior to GP-Apriori.

5.2. Empirical Testing and Results

We tested our approach for the targeted advertising of library books using the circulation data of the library

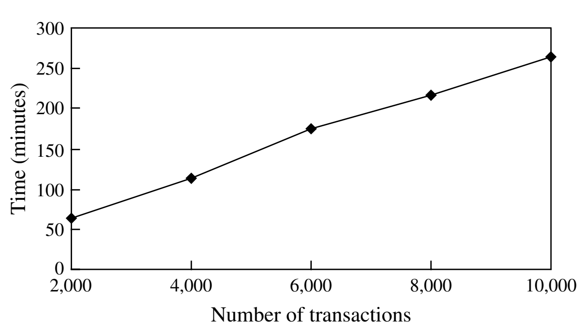

Figure 4 **Running Times of GP-Apriori**

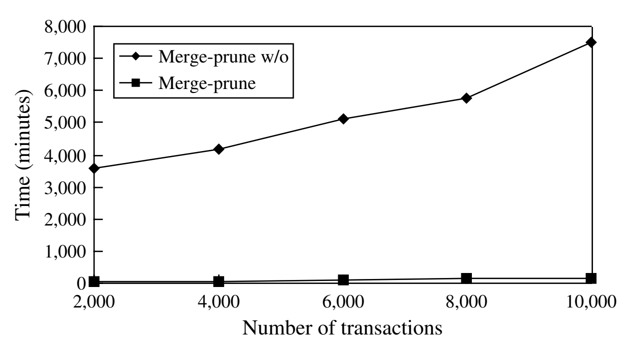

Figure 6 **Running Times for Merge-Prune With and Without Pruning**

for a small university, NSYSU, with approximately 7,000 students, and its library uses an online public access catalog system from INNOVATIVE. Circulation data were collected between May 1, 2003, and May 1, 2004. Because most patrons are students, we tracked only books issued to this group. Over the one-year period, we accumulated 62,568 book transactions associated with 5,631 patrons. The system records the following patron attributes: identification number, name, gender, address, birthday, degree, program, work unit, and academic status. The system also records several book-related attributes, including call number, ISBN, subjects (key words), authors, and location.

5.2.1. Identifying Relevant Patron Attributes. We first identify patron attributes that are highly relevant to the types of books issued. To focus our attention on student patrons, we initially chose five patron attributes: gender, address, birthday, degree, and program. A chi-square test revealed that gender, degree, and program are highly correlated with the books issued. However, senior librarians believed that gender was related to the books issued because the distribution of male and female students varies between programs and degrees; their hypothesis was confirmed by another chi-square test. We therefore chose degree and program in the next step for finding GP association rules.

5.2.2. Generating Patron–Book Rules. We formulated a concept hierarchy of three levels for each degree and program, with eight nodes in the degree hierarchy and 38 in the program hierarchy. The entire degree concept hierarchy and a partial program concept hierarchy appear in Figure 7.

Also, we adopted the first three levels of the Chinese classification scheme for the hierarchy of Chinese books and the first two levels of the Congressional classification scheme for the hierarchy of western books. The numbers of leaf nodes in the Chinese and western book hierarchies were 1,000 and 526, respectively.

We used five-way cross-validation for our experiments. Specifically, we uniformly partitioned the set of books that appear in patrons' transactions into

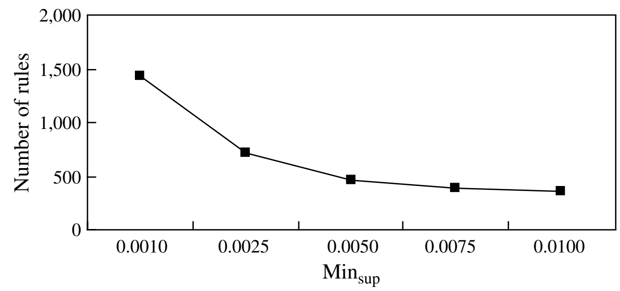

Figure 8 **Number of Valid GP Association Rules for Various** Min_{sup} **Values** ($\text{Min}_{conf} = 0.1$)

five subsets of equal size. In each trial, we kept four subsets of books in these transactions as the training data set and used the other subset as the test data set. We uncovered valid GP association rules from the training data set using GP-Apriori, which we chose because of its simplicity. The small number of demographic attributes in this experiment makes GP-Apriori as efficient as Merge-prune in this instance. Finally, we applied the previously described approach, as well as other approaches, to choose a set of patrons for advertising each book in the test data set.

Figure 8 shows the average number of GP association rules for various Min_{sup} values when Min_{conf} is fixed at 0.1. As we expected, the total number of valid GP association rules decreases monotonically with increasing Min_{sup} (Agrawal and Srikant 1994).

A small value of Min_{sup} leads to many less-than-useful rules. In contrast, a larger Min_{sup} yields fewer rules, which may result in some product (p) whose matching rule set is empty (i.e., $M_p = \varnothing$), thereby preventing targeted advertising. A Min_{sup} value therefore achieves total coverability if every product has a nonempty matching rule set. Table 8 lists the total coverability for various Min_{sup} values. We finally chose $\text{Min}_{sup} = 0.0025$ and $\text{Min}_{conf} = 0.1$ for our subsequent experiments, for which we obtained a moderate number of rules (710) and total coverability.

5.2.3. Effectiveness of the Patron–Book Rules for Targeted Advertising. To evaluate the effectiveness of the valid GP association rules for targeted advertising, we compare the following four targeted advertising

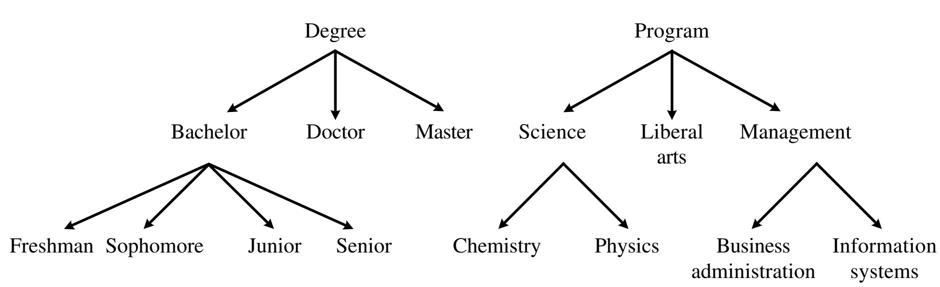

Figure 7 **Concept Hierarchies of Degree and Program**

Table 8 **Total Coverability for Various Min_{sup} Values**

Min_{sup}	Total coverability
0.0010	Yes
0.0025	Yes
0.0050	No
0.0070	No
0.0100	No

approaches:

1. Value-based, as described in §4.
2. Confidence-based, which chooses rules in descending order of confidence and selects patrons whose demographics match the antecedents of these rules.
3. Empirical, which employs rules involving the second level of program and the first level of book hierarchies.
4. Random, where each patron has the same probability of being chosen.

The third approach is that traditionally adopted by NSYSU librarians for recommending new library books. For example, for each of the 10 categories at the first level of the Chinese classification scheme, librarians maintain a list of colleges, computed periodically, whose students may be more interested in books in that category, and new library books are promoted on the basis of these lists. In our experimental data set, the associations between categories at the first level of the book-classification scheme and colleges form a subset of the valid GP association rule set. The empirical approach emulates the traditional approach and is similar to the confidence-based approach except that only rules in the subset are consulted.

For each book p in the test set, we applied each approach to identify a fixed number of patrons (called the target set). The hit rate for p is the ratio of the number of patrons in the target set who had been issued p to the total number of patrons who had been issued p. To show the performance of the various methods, Figure 9 plots the cumulative gains chart, which consists of several lift curves and a baseline curve. The hit rates of the first three approaches are regarded as lift curves, whereas that of the random approach serves as the baseline curve. The greater the area between the lift curve and the baseline, the better the model. The x-axis shows the maximum number of targeted customers (N), and the y-axis shows the average hit rate of 10 randomly selected books in the test data set across five trials.

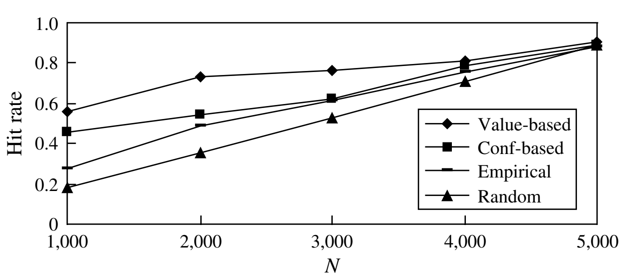

Figure 9 **Cumulative Gains Charts**

Figure 9 indicates that the hit rate of all four approaches increases with the size of the target set, and when $N = 5{,}000$ (close to the total number of patrons), all four approaches exhibit almost the same hit rate. The empirical approach used by NSYSU librarians consistently exhibits a higher hit rate than the random approach at smaller N. The confidence-based approach has a better hit rate than the empirical approach, which demonstrates the usefulness of valid GP association rules. Finally, the proposed value-based approach has the best performance, which we attribute to the deliberate confidence estimation of some useful rules derived from valid GP association rules. The increases in the performance of the value-based approach in four experiments ($N = 1{,}000$–$4{,}000$) are all statistically significant (p-value < 0.05) compared with the confidence-based approach.

Overall, our experiments demonstrate the effectiveness of targeted marketing using GP association rules in the promotion of new library books. Such success is due to the strong correlation between the book-classification hierarchy and the books issued to patrons. We find from our circulation data that, on average, 58% and 71% of the Chinese books issued to a patron can be classified in a single category at the second and first levels of the book hierarchy, respectively. In other words, patrons exhibit highly skewed behavior in the categories (as classified by book-classification schemes) to which their issued books belong. Thus, the success of the proposed approach depends on the existence of such a product hierarchy. For some types of products, such as books and music albums, the inherent product hierarchies are very likely to exhibit this property. However, for other types of products, such as electronic appliances and cars, further analysis should identify the usage-related product hierarchies before the proposed approach could be adapted effectively.

6. Conclusions

We have proposed a novel approach to mining GP association rules intended for targeted advertising of new products that have no associated transaction records. This approach starts with an identification of valid GP association rules, for which we have developed two algorithms: GP-Apriori and Merge-prune. The identified GP association rules indicate a short list of prospective customers for a given new product, for which we have developed a comprehensive algorithm. Applications show that Merge-prune outperforms GP-Apriori, and demonstrate the effectiveness.

Our work might be extended in several directions. First, the optimal size of the partitioning demographic attributes for the Merge-prune algorithm has yet to be determined and might be a function of the memory size or the fan out of each node in the associated hierarchies. Second, the effectiveness of the targeted advertising approach is influenced strongly by the availability of a product taxonomy in which customers demonstrate skewed behavior in their transactions, so systematic approaches for identifying such hierarchies need to be explored. We plan to evaluate the accuracy of the proposed algorithms and apply the algorithms to other product types. Third, we focus on selection of prospective customers for a given new product, though the discovered GP association rules could be used to recommend new products to a given customer (i.e., personalization). The GP association rules, which incorporate aggregated information about both customer demographics and product attributes, allow for recommendations of (new) products to new customers. However, due to the complex is-a relationships exhibited in the aggregation hierarchies, not all matching GP association rules are of equal interest to a given customer. We are conducting further empirical studies into interest measures tailored for personalization.

Acknowledgments

The authors thank Chai-Wen Tsai and Chin-Yuan Hung for implementing the approaches proposed in this paper. The authors also thank the anonymous reviewers who made insightful comments that greatly improved the quality of this paper. This work was supported in part by the National Science Council of Taiwan under Grant NSC-95-2752-H-110-005.

References

Adomavicius, G., A. Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowledge Data Engrg.* **17** 734–749.

Agrawal, R., R. Srikant. 1994. Fast algorithms for mining association rules. *Very Large Data Bases Conf., Santiago de Chile, Chile*, 478–499.

Benway, J. P., D. M. Lane. 1998. Banner blindness: Web searchers often miss "obvious" links. *Internetworking* 1.3, http://www.internettg.org/newsletter/dec98/banner_blindness.html.

Cheyne, T. L., F. E. Ritter. 2001. Targeting audiences on the Internet. *Comm. ACM* **44** 94–98.

Hayes, L. 1994. Using computers to divine who might buy a gas grill. *Wall Street J.* (August 16) B1–B4.

Huang, Z., W. Chung, H. Chen. 2004. A graph model for e-commerce recommender systems. *J. Amer. Soc. Inform. Sci. Tech.* **55** 259–274.

Langheinrich, M., A. Nakamura, N. Abe, T. Kamba, Y. Koseki. 1999. Unintrusive customization techniques for web advertising. *Comput. Networks* **31** 1259–1272.

Loeb, S., D. Terry. 1992. Information filtering. *Comm. ACM* **35** 26–70.

Mooney, R. J., L. Roy. 2000. Content-based book recommending using learning for text categorization. *ACM Conf. Digital Libraries, San Antonio, TX*, 195–240.

Nakamura, A. 2002. Improvements in practical aspects of optimally scheduling web advertising. *Internat. World Wide Web Conf., Honolulu, HI*, 536–541.

Shardanand, U., P. Maes. 1995. Social information filtering: Algorithms for automating "word of mouth." *Internat. Conf. Human Factors Comput. Systems, Denver, CO*, 210–217.

Srikant, R., R. Agrawal. 1995. Mining generalized association rules. *Very Large Data Bases Conf., Zurich, Switzerland*, 409–419.

Tomlin, J. A. 2000. An entropy approach to unintrusive targeted advertising on the Web. *Comput. Networks* **33** 767–774.