Connection Network and Optimization of Interest Metric for One-to-One Marketing

Sung-Soon Choi and Byung-Ro Moon

School of Computer Science and Engineering
Seoul National University
Seoul, 151-742 Korea
{sschoi,moon}@soar.snu.ac.kr

Abstract. With the explosive growth of data in electronic commerce, rule finding becomes a crucial part in marketing. In this paper, we discuss the essential limitations of the existing metrics to quantify the interests of rules, and present the need of optimizing the interest metric. We describe the construction of the connection network that represents the relationships between items and propose a natural marketing model using the network. Although simple interest metrics were used, the connection network model showed stable performance in the experiment with field data. By constructing the network based on the optimized interest metric, the performance of the model was significantly improved.

1 Introduction

The progress in modern technologies made it possible for finance and retail organizations to collect and store a massive amount of data. In consequence, it has attracted great attention to identify systems that explain the data, particularly in the data mining area. An early representative problem is the "market-basket" problem [1]. In the problem, we are given a set of items and a collection of transactions each of which is a subset (basket) of items purchased together by a customer in a visit. The objective is "mining" relationships between items from the baskets data.

First of all, the attraction of the problem arises in a great variety of applications [2]. A typical example (from which the problem got its name) is the customers' shopping behavior in a supermarket. In the example, the items are products and the transactions are customer purchases at the checkout. Determining what products customers are likely to buy together is useful for display and marketing. There are many other applications with different data characteristics. Some examples are student enrollment in classes [2], copy detection (identifying identical or similar documents or web pages) [3] [4], clustering (identifying similar vectors in high-dimensional spaces) [5] [6], etc.

With the rapid growth of the electronic commerce (e-commerce) field, the problem becomes increasingly important. Various solutions to the problem were used for marketing in the e-commerce field [7] [8]. Among them, collaborative filtering (tracking user behavior and providing recommendations to individuals

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 1998–2009, 2003.
© Springer-Verlag Berlin Heidelberg 2003

based on the similarities of their preferences) has been playing as one of the core marketing strategies [9] [10] [11].

In this paper, we suggest a genetic approach to optimize the interest metric that measures the relationships between items. The optimized metric is used to construct a weighted relationship network of the items. Then, we propose a natural marketing model using the network. Experimental results showed that our model was more stable and better than collaborative filtering.

The rest of this paper is organized as follows. In Sect. 2, we describe interest metrics introduced so far that measure the degree of connection between items. We also give a brief description of methods for one-to-one marketing. In Sect. 3, we present our theoretical view of the interest metrics described in Sect. 2 and mention the need of optimizing the interest metric. In Sect. 4, we explain the connection network that represents the relationships among the items, and describe the marketing model based on the network. The genetic framework for optimizing the interest metric is provided in Sect. 5. Section 6 provides the experimental results on real-world data sets. Finally, we make our conclusions in Sect. 7.

2 Preliminaries

2.1 Interest Metrics of Rules

Recently, there were many studies to find the meaningful rules between items based on the interest metrics that measure the degree of connection between items. A rule is denoted by $X \to Y$, for two item sets X and Y, which means that the occurrence of X implies the occurrence of Y.

Agrawal *et al.* [1] proposed the interest metrics called *support* and *confidence* and used them to build rules between items. In particular, they called such rules *association rules*. The support of an item set X is defined to be the number of transactions in which the item set occurs (the number of occurrences of X). For a rule $X \to Y$, the confidence of this rule is the fraction of transactions containing Y among those containing X. In order that a rule $X \to Y$ becomes an association rule, the support of the item set $X \cup Y$ and the confidence of the rule $X \to Y$ must exceed given thresholds θ_s and θ_c, respectively.

Note that the support of the item set $X \cup Y$ corresponds to the number of transactions in which the item sets X and Y occur together (the number of co-occurrences of X and Y). And the confidence of the rule $X \to Y$ corresponds to the number of co-occurrences of X and Y over the number of occurrences of X. We denote the number of occurrences of an item set X by $n(X)$ and the number of co-occurrences of item sets X and Y by $n(X,Y)$. Then, from the definitions of support and confidence, we have

$$sup(X \to Y) = n(X,Y) \text{ and } conf(X \to Y) = \frac{n(X,Y)}{n(X)} . \qquad (1)$$

The association rule $X \to Y$ is said to *hold* if and only if

$$sup(X \to Y) = n(X,Y) > \theta_s \text{ and } conf(X \to Y) = \frac{n(X,Y)}{n(X)} > \theta_c \,. \qquad (2)$$

Cohen *et al.* [12] pointed out a problem of association rules which require high support: Most rules with high supports are obvious and well-known and it is the rules of low supports that provide interesting new insights. Besides, several recent papers mentioned that it is not reasonable to use confidence as the interest measure of rules [2] [13] [14]. In this context, a number of different metrics to quantify "interestingness" or "goodness" of rules were proposed [15]. Among them are *gain* [16], proposed by Piatetsky-Shapiro [17], *variance* and *chi-squared value* [18], *entropy gain* [18] [19], *gini* [19], *laplace* [20] [21], *lift* [22] (also known as *interest* [2] or *strength* [23]), *conviction* [2], and *similarity* [12].

Except *similarity*, Bayardo and Agrawal [15] expressed the definitions of all these metrics with the supports of the related item sets and the confidence of the rule. It is also possible to express *similarity* in the same way. This indicates that all these metrics can be described with the numbers of occurrences and co-occurrences of the related item sets. If we denote by T the set of all transactions, the *laplace*, *gain*, Piatetsky-Shapiro's metric (*p-s*), *conviction*, *lift*, and *similarity* values for a rule $X \to Y$ are expressed as follows:[1]

$$\begin{aligned}
laplace(X \to Y) &= \frac{n(X,Y)+1}{n(X)+k} \,, \\
gain(X \to Y) &= n(X,Y) - c \cdot n(X) \,, \\
p\text{-}s(X \to Y) &= n(X,Y) - \frac{n(X) \cdot n(Y)}{|T|} \,, \\
conviction(X \to Y) &= \frac{|T| \cdot n(X) - n(X) \cdot n(Y)}{|T| \cdot (n(X) - n(X,Y))} \,, \\
lift(X \to Y) &= \frac{|T| \cdot n(X,Y)}{n(X) \cdot n(Y)} \,, \text{ and} \\
similarity(X \to Y) &= \frac{n(X,Y)}{n(X)+n(Y)-n(X,Y)} \,. \qquad (3)
\end{aligned}$$

The rest of the above metrics are also able to be expressed in the same way.

It is difficult to come up with a single metric among the above metrics. It is, however, clear that all the metrics consider only the numbers of occurrences and co-occurrences of item sets. This fact is utilized in modeling the interest metric and optimizing it. And note that these metrics numerically evaluate the relationships between item sets. So, these metrics can be used in quantifying the strengths of the connections between item sets.

[1] k in *laplace* is an integer greater than 1 and c in *gain* is a fractional constant between 0 and 1.

2.2 One-to-One Marketing

Personalization is a sharply growing issue in modern marketing. It helps boost customers' loyalty by providing the most attractive contents to each customer or by locating the most proper set of customers for an arbitrary advertisement [24]. As mentioned before, personal information and huge activity logs are accumulated due to the progress of modern technologies. These data implicitly contain valuable trends and patterns which are useful to improve business decisions and efficiency.

Diverse data mining tools to discover implicit knowledge hidden in a large database have been studied with various tools including neural networks, decision trees, rule induction, Bayesian belief networks, evolutionary algorithms, fuzzy sets, clustering, association rules, and collaborative filtering [25] [26]. Among them, collaborative filtering, which tracks user behavior and makes recommendations to individuals on the basis of the similarities of their preferences, is known to be a standard for personalized recommendations [10] [27]. Most of the personalized marketing tools including collaborative filtering utilize customer profiles [7] [8] [28] [29] [30].

These strategies often have the data-sparsity problem, which greatly undermines the quality of recommendations, as it is in general difficult to collect customers' personal information or preferences [31] [32].

3 Rule Space and Optimized Interest Metric

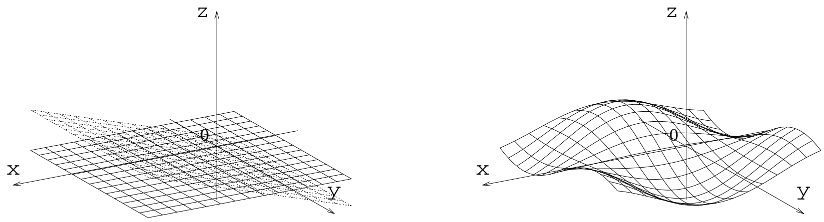
We saw that $sup(X \to Y) = n(X,Y)$ and $conf(X \to Y) = \frac{n(X,Y)}{n(X)}$ for a rule $X \to Y$ in the previous section. Now if we replace $n(X)$, $n(Y)$, and $n(X,Y)$ with independent variables x, y, and z, respectively, the necessary and sufficient condition for an association rule $X \to Y$ to hold are as follows:

$$z > \theta_s \quad \text{and} \quad \frac{z}{x} > \theta_c \, . \quad (x > 0, y \geq 0, z \geq 0) \tag{4}$$

This means that the values of x, y, and z for the rule $X \to Y$ to hold correspond to the points in three dimensions that satisfy the above inequality (4). At this time, the rule space consisting of x, y, and z axes is partitioned by the plane corresponding to the support condition, $f_s(x,y,z) = z = \theta_s$, and the plane corresponding to the confidence condition, $f_c(x,y,z) = \frac{z}{x} = \theta_c$. (e.g., Figure 1(a))

It was mentioned that, for a rule $X \to Y$, the metrics introduced in the previous section are able to be described with $n(X)$, $n(Y)$, and $n(X,Y)$. Therefore, it is also possible to express all the metrics of the previous section as the formulas of x, y, and z. For example, $conviction(X \to Y)$ for a rule $X \to Y$ can be described as follows:

$$conviction(X \to Y) = \frac{|T| \cdot n(X) - n(X) \cdot n(Y)}{|T|(n(X) - n(X,Y))} = \frac{|T|x - xy}{|T|(x - z)} \, . \tag{5}$$



(a) Two linear planes (b) A free plane

Fig. 1. A rule space partitioned by planes

After all, given a threshold θ, each of the metrics mentioned in the previous section is a plane, $f(x, y, z) = \theta$, which partitions the rule space. The selection of interest metrics allows the space partition of a fixed shape regardless of the characteristics of the data set. We suspect that the optimal shapes of the partition planes are different depending on the data sets. In this context, we guarantee the degree of freedom for the partition plane to the utmost, then find the optimized plane for the given data set, namely the optimized interest metric. (Figure 1(b))

To maximize the degree of freedom for the plane, it is desirable to assume that the plane $f(x, y, z) = \theta$ is a free plane. In this case, however, the search space becomes so huge that the learning time increases excessively. We restrict the shape of the plane to help perform the learning. We set a model of $f(x, y, z)$ as follows:

$$f(x, y, z) = (a_x x^{e_x} + b_x)(a_y y^{e_y} + b_y)(a_z z^{e_z} + b_z)$$

where

$$\begin{cases} 0 & < a_x, a_y, a_z \leq 10 \\ -1 & \leq e_x, e_y, e_z \leq 1 \\ 0 & \leq b_x, b_y, b_z \leq 10 \; . \end{cases} \tag{6}$$

We use a genetic algorithm to search the optimal coefficients and exponents of $f(x, y, z)$ for the data set. The details of optimization are described in Sect. 5.

4 Personalized Marketing Model Using Connection Networks

In the previous section, we made a model of the metric $f(X \to Y)$ $(= f(x, y, z))$ that evaluates the strength of a rule $X \to Y$. Intuitively the value of $f(X \to Y)$ indicates the strength of the connection between the item sets X and Y. The optimized metric is used to measure the strength of connection between item sets. Here, we only consider the case that each item set has just one item. Then,

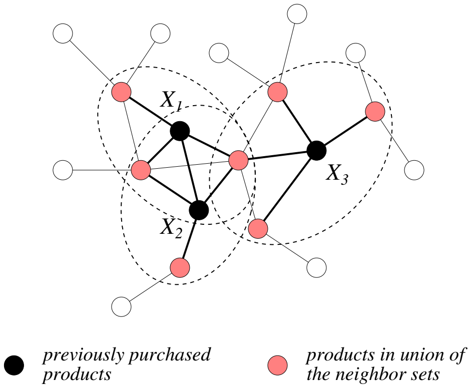


Fig. 2. A connection network

the value of $f(X \to Y)$ represents the strength for the item X to imply the item Y.

Now we are ready to construct a connection network. We set a vertex for each item. We put an arc from the vertex X to the vertex Y with the weight $f(X \to Y)$. We have a directed graph $G = (V, A)$, where V is the vertex set and A is the arc set.

We perform one-to-one marketing using the connection network. Suppose that a customer purchased the products $X_1, X_2, \ldots, X_k$ so far. Let $N(X_i)$ be the set of neighbor vertices of X_i in the connection network ($1 \leq i \leq k$). We define a score function for recommendation, $s : V \longrightarrow \mathbb{R}$, as follows ($\mathbb{R}$: the set of real numbers):

$$s(Y) = \begin{cases} \{\sum_{1 \leq i \leq k} f(X_i \to Y)\}/(a_k k^{e_k} + b_k) & \text{, if } Y \in \bigcup_{1 \leq i \leq k} N(X_i) \\ 0 & \text{, otherwise .} \end{cases} \tag{7}$$

We recommend the products of high scores to the customer. The value of the score function $s(Y)$ for a product Y is proportional to the weight sum of the arcs from the previously purchased products to the product Y. Figure 2 shows an example connection network. We divide the sum of weights by a function of k (the number of previously purchased products). This prevents the recommendations from flowing in upon a few customers that purchased excessively many products before. a_k, e_k, and b_k are optimized in the following ranges,

$$0 < a_k \leq 10, \quad -1 \leq e_k \leq 1, \quad 0 \leq b_k \leq 10, \tag{8}$$

and the details are described in the next section.

Such a recommendation strategy is different from the existing ones based on the customers' profiles in that it performs recommendations just by the relationships between products regardless of the customers' profiles. We need not handle the customer profile vectors of high dimensions nor quantify each field

create initial population of a fixed size;
do {
  choose *parent1* and *parent2* from population;
  offspring = crossover(*parent1*, *parent2*);
  mutation(*offspring*);
  replace(population, *offspring*);
} **until** (stopping condition);
return the best individual;

Fig. 3. The outline of the genetic algorithm

of the customer profiles. We merely optimize the quantitative representation of topology among items and perform recommendation on the basis of this. The computational cost for the recommendation is significantly low compared with existing recommendation strategies. In addition, when reflecting new data into the established network, the cost of updating the network is also fairly low. Furthermore, such a strategy is less sensitive to the data sparsity problem that greatly undermines the quality of recommendations for the existing ones.

5 Genetic Algorithm

The selection of an interest metric and a score function has a great effect on the quality of recommendations. The problem is to find the best set of coefficients related to the interest metric and score function that maximizes the *response rate*, which is defined to be

$$\textit{response rate} = \frac{\#\ \textit{of purchases}}{\#\ \textit{of recommended items}}\ . \tag{9}$$

We used a steady-state genetic algorithm to optimize the nine parameters related to the interest metric (in Sect. 3) and the three parameters related to the score function (in Sect. 4). Figure 3 shows the outline of the genetic algorithm. The details are described in the following.

- **Encoding:** Each solution is a set of 12 real values. A solution is represented by a chromosome; a chromosome is a real array of 12 elements. Figure 4 shows the structure of chromosomes. Each element of the array is called a gene and we restrict the range of each gene as mentioned before.
- **Initialization:** We set the population size to be 100. For each gene in a chromosome, we randomly generate a real number in the restricted range.
- **Parent Selection:** The fitness value F_i of chromosome i is assigned as follows:

$$F_i = (R_i - R_w) + (R_b - R_w)/4 \tag{10}$$

a_x	e_x	b_x	a_y	e_y	b_y	a_z	e_z	b_z	a_k	e_k	b_k

Fig. 4. The structure of chromosomes

where

$$R_w : \text{the response rate of the worst,}$$
$$R_b : \text{the response rate of the best, and}$$
$$R_i : \text{the response rate of chromosome } i.$$

Each chromosome is selected as a parent with a probability proportional to its fitness value. This is a typical proportional selection scheme.

- **Crossover:** We use traditional one-point crossover.
- **Mutation:** We randomly select each gene with a low probability (P=0.1) and perform the non-uniform mutation in which the perturbation degree decreases over time. Non-uniform mutation was first proposed by Michalewicz [33].
- **Replacement:** We replace the inferior of the two parents if the offspring is not worse than both parents. Otherwise, we replace the worst member of the population. This scheme is a compromise between preselection [34] and GENITOR-style replacement [35].

6 Experimental Results

We conducted experiments with two different types of data sets to evaluate the performance of the optimized interest metric and the marketing model using the connection network.

First, we conducted experiments with a massive amount of purchase data set from June 1996 to August 2000 of a representative e-commerce company in Korea. In this data set, the items are products and a transaction is a customer's purchase of one or more items with a time stamp. We first divided the whole data set into two disjoint sets in terms of dates. Then we predicted the purchases of customers in the latter set, based on the data of the former set.

In this experiment, a number of different recommendation models were used: collaborative filtering (CF), a few plain connection networks (PCNs), and optimized connection network (OCN). As mentioned before, collaborative filtering is a proven standard for personalized recommendations [9] [10] [27]. PCNs are our early recommendation models (commercialized by Optus Inc.) in which we construct the networks based on the interest metrics mentioned in Sect. 2 (such as *laplace*, *gain*, and so on) and then recommend attractive products based on the prescribed thresholds. OCN is the recommendation model in which we construct the network on the basis of the optimized interest metric and then recommend the products according to the score-function values.

Table 1. Comparison of experiments with purchase data set

Dates	CF	PCN-*laplace*	PCN-*gain*	PCN-*p-s*	OCN
May 31	.0052 ($\frac{26}{5018}$)	.0032 ($\frac{14}{4310}$)	.0046 ($\frac{23}{5003}$)	.0030 ($\frac{15}{5011}$)	.0052 ($\frac{26}{4994}$)
June 7	.0061 ($\frac{32}{5210}$)	.0034 ($\frac{15}{4411}$)	.0046 ($\frac{24}{5183}$)	.0043 ($\frac{23}{5390}$)	.0087 ($\frac{45}{5162}$)
June 14	.0054 ($\frac{30}{5513}$)	.0028 ($\frac{13}{4565}$)	.0046 ($\frac{25}{5415}$)	.0045 ($\frac{25}{5602}$)	.0092 ($\frac{52}{5634}$)
June 21	.0063 ($\frac{36}{5695}$)	.0032 ($\frac{18}{5742}$)	.0053 ($\frac{29}{5448}$)	.0041 ($\frac{23}{5627}$)	.0099 ($\frac{58}{5859}$)
June 28	.0062 ($\frac{37}{5947}$)	.0034 ($\frac{17}{4937}$)	.0051 ($\frac{31}{6094}$)	.0043 ($\frac{25}{5767}$)	.0088 ($\frac{55}{6229}$)
July 5	.0055 ($\frac{34}{6138}$)	.0027 ($\frac{15}{5459}$)	.0046 ($\frac{28}{6135}$)	.0044 ($\frac{28}{6343}$)	.0090 ($\frac{55}{6117}$)
July 12	.0041 ($\frac{26}{6327}$)	.0043 ($\frac{27}{6289}$)	.0068 ($\frac{41}{6036}$)	.0058 ($\frac{38}{6582}$)	.0084 ($\frac{52}{6192}$)
July 19	.0043 ($\frac{28}{6571}$)	.0063 ($\frac{40}{6314}$)	.0089 ($\frac{56}{6324}$)	.0084 ($\frac{55}{6511}$)	.0084 ($\frac{54}{6458}$)
July 26	.0038 ($\frac{26}{6879}$)	.0069 ($\frac{56}{8111}$)	.0088 ($\frac{59}{6695}$)	.0085 ($\frac{59}{6982}$)	.0112 ($\frac{79}{7036}$)
August 2	.0032 ($\frac{23}{7130}$)	.0069 ($\frac{49}{7067}$)	.0078 ($\frac{58}{7427}$)	.0078 ($\frac{56}{7190}$)	.0079 ($\frac{58}{7372}$)
August 9	.0027 ($\frac{20}{7386}$)	.0045 ($\frac{29}{6385}$)	.0053 ($\frac{39}{7315}$)	.0053 ($\frac{39}{7330}$)	.0061 ($\frac{43}{7047}$)

Table 2. Comparison of experiments with mobile-content data set

CF	PCN-*laplace*	PCN-*gain*	PCN-*similarity*	OCN
.0241 ($\frac{4886}{203054}$)	.0256 ($\frac{5215}{203379}$)	.0269 ($\frac{5522}{205061}$)	.0247 ($\frac{5096}{206474}$)	.0291 ($\frac{5535}{190417}$)

For experiments, we selected eleven pivot dates over the weeks from May 2000 to August 2000. For all the models, the data before the pivot date were used as the training set; the data after the date were used as the test set. In the case of OCN, the training set was further divided into the real training set and the validation set, to optimize the interest metric and the score function.

Table 1 shows the response rates (and the numbers of purchases over the numbers of recommended products) of CF, PCNs, and OCN, respectively. We omitted the experimental results of PCNs based on *conviction*, *lift*, and *similarity* since their performances were not so comparable. Here, the numbers of recommendations were adjusted to be comparable for fair comparison. PCNs showed comparable performance with CF although the networks were constructed based on the general metrics. We consider this to be an evidence of the suitability of the connection network as a recommendation model. The optimized model, OCN, significantly improved all the PCN models. OCN showed on average 76 % and 40 % better results than CF and PCN-*gain* (the best among PCNs), respectively.

Next, we conducted similar experiments with a massive amount of Internet contents access data from August 2001 to January 2002 of a representative contents service company in Korea. In this data set, the items are contents provided by the company and a transaction is composed of the contents that a customer used in a certain period of time. We divided the whole data set into two disjoint sets in terms of dates in the ratio of three to one.

Table 2 shows the response rates of CF, PCNs, and OCN, respectively. The response rate is the number of uses (hits) over the number of recommended contents as in the parentheses. In the table, the experimental results of PCNs based on *p-s*, *conviction*, and *lift* were omitted since their performances were not so comparable. Similarly to the experiments with the purchase data set, PCNs showed comparable performance with CF, and OCN performed significantly better than the other models.

7 Conclusion

We discussed the essential limitations of existing rule-interest metrics and raised the need of optimizing the interest metric. We proposed a novel marketing model using connection networks. We maximized the performance of the proposed model by constructing the network with the optimized interest metric.

Although the suggested method performed impressively, we consider that there remains room for further improvement. More elaborate modeling of the interest metric and the score function is a candidate. Other function models such as neural network and relevant optimization are also worth trying.

The optimized interest metric and the connection network model is not restricted to the personalized marketing only. We believe that they are applicable to various other problems; so far, we found that they are applicable to a few practical problems such as e-mail auto-response system and personalized search engine.

Acknowledgement. The authors thank Yung-Keun Kwon for insightful discussions about data mining techniques. This work was partly supported by Optus Inc. and Brain Korea 21 Project. The RIACT at Seoul National University provided research facilities for this study.

References

1. R. Agrawal, T. Imilienski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pages 207–216, 1993.
2. S. Brin, R. Motwani, J.D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. In *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pages 255–264, 1997.
3. N. Shivakumar and H. Garcia-Molina. Building a scalable and accurate copy detection mechanism. In *Proc. Third Int'l Conf. Theory and Practice of Digital Libraries*, 1996.
4. A. Broder. On the resemblance and containment of documents. In *Proc. Compression and Complexity of Sequences Conf. (SEQUENCES'97)*, pages 21–29, 1998.
5. R. O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley InterScience, New York, 1973.
6. S. Guha, R. Rastogi, and K. Shim. CURE – an efficient clustering algorithm for large databases. In *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pages 73–84, 1998.

7. D. Greening. *Building Consumer Trust with Accurate Product Recommendations.* LikeMinds White Paper LMWSWP-210-6966, 1997.
8. C.C. Aggarwal, J.L. Wolf, K. Wu, and P.S. Yu. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. In *Proc. of the fifth ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 201–212, 1999.
9. D. Goldberg, D. Nichols, B.M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Comm. of the ACM*, 35(12):61–70, 1992.
10. J.A. Konstan et al. Grouplens: Applying collaborative filtering to usenet news. *Comm. of the ACM*, 40:77–87, 1997.
11. H.R. Varian and P. Resnick. CACM special issue on recommender systems. *Comm. of the ACM*, 40, 1997.
12. E. Cohen et al. Finding interesting associations without support pruning. *IEEE Trans. on Knowledge and Data Engineering*, 13(1):64–78, 2001.
13. C. Silverstein, S. Brin, and R. Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data Mining and Knowledge Discovery*, 2:69–96, 1998.
14. C. Silverstein, S. Brin, R. Motwani, and J.D. Ullman. Scalable techniques for mining casual structures. In *Proc. 24th Int'l Conf. Very Large Data Bases*, pages 594–605, 1998.
15. R.J. Bayardo Jr. and R. Agrawal. Mining the most interesting rules. In *Proc. of the fifth ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 145–154, 1999.
16. T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama. Data mining using two-dimensional optimized association rules: Scheme, algorithms, and visualization. In *Proc. of the ACM SIGMOD Int'l Conf. on Management of Data*, pages 13–23, 1996.
17. G. Piatetsky-Shapiro. *Discovery, Analysis, and Presentation of Strong Rules.* AAAI/MIT Press, 1991. Chapter 13 of Knowledge Discovery in Databases.
18. S. Morishita. On classification and regression. In *Proc. First Int'l Conf. on Discovery Science – Lecture Notes in Artificial Intelligence 1532*, pages 40–57, 1998.
19. Y. Morimoto, T. Fukuda, H. Matsuzawa, T. Tokuyama, and K. Yoda. Algorithms for mining association rules for binary segmentations of huge categorical databases. In *Proc. of the 24th Very Large Data Bases Conf.*, pages 380–391, 1998.
20. P. Clark and P. Boswell. Rule induction with CN2: Some recent improvements. In *Machine Learning: Proc. of the Fifth European Conference*, pages 151–163, 1991.
21. G.I. Webb. OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, 3:431–465, 1995.
22. International Business Machines. *IBM Intelligent Miner User's Guide*, 1996. Version 1, Release 1.
23. V. Dhar and A. Tuzhilin. Abstract-driven pattern discovery in databases. *IEEE Trans. on Knowledge and Data Engineering*, 5(6), 1993.
24. R. Dewan, B. Jing, and A. Seidmann. One-to-one marketing on the internet. In *Proc. of the 20th Int'l Conf. on Information Systems*, pages 93–102, 1999.
25. M.S. Chen, P.S. Han, and J.Yu. Data mining: An overview from a database perspective. *IEEE Trans. on Knowledge and Data Engineering*, 8(6):866–883, 1996.
26. M. Goebel and L. Gruenwald. A survey of data mining and knowledge discovery software tools. *SIGKDD Explorations*, 1:20–33, 1999.

27. J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proc. of the 22nd Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 230–237, 1999.
28. V. Venugopal and W. Baets. Neural networks and their applications in marketing management. *Journal of Systems Management*, 45(9):16–21, 1994.
29. S. Park. Neural networks and customer grouping in e-commerce: A framework using fuzzy ART. In *Proc. Academia/Industry Working Conf. on Research Challenges '00*, pages 331–336, 2000.
30. Y.K. Kwon and B.R. Moon. Personalized email marketing with a genetic programming circuit model. In *Genetic and Evolutionary Computation Conference*, pages 1352–1358, 2001.
31. M.J. Culnan. "How did they get my name ?": An exploratory investigation of customer attitudes toward secondary information use. *MIS Quarterly*, 17(3):341–363, 1993.
32. J. III Hagel and J.F. Rayport. The coming battle for customer information. *Havard Business Review*, 75(1):53–65, 1997.
33. Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolutionary Programs.* Springer, 1992.
34. D. Cavicchio. *Adaptive Search Using Simulated Evolution.* PhD thesis, University of Michigan, Ann Arbor, MI, 1970. Unpublished.
35. D. Whitley and J. Kauth. Genitor: A different genetic algorithm. In *Rocky Mountain Conf. on Artificial Intelligence*, pages 118–130, 1988.