Optimized Interest Metric of Rules and One-to-One Marketing Using Connection Networks

Sung-Soon Choi and Byung-Ro Moon
School of Computer Science and Engineering,
Seoul National University, Seoul, 151-742 Korea
{irranum,moon}@soar.snu.ac.kr

With the explosive growth of the electronic commerce data, rule finding became a crucial part in marketing. A number of different metrics to quantify "interestingness" or "goodness" of rules, including *support* and *confidence* [1], have been proposed [3]. We suspect that the most suitable interest metics for different data sets are different one another. In this context, we find the optimized interest metric for the given data set.

We found that most of the proposed metrics can be expressed with respect to supports. Based on this observation, we express the degree of the connection from an item X to an item Y, $f(X \to Y)$, as the function of support values of X, Y, and $X \cup Y$. We replace those support values with independent variables x, y, and z, respectively, and then we set a model of $f(x,y,z)$ $(=f(X \to Y))$ as follows:

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

We use a genetic algorithm to search the optimal coefficients and exponents of $f(x,y,z)$ for the data set.

Now we construct a connection network based on the optimized metric. 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, $s : V \longmapsto R$, as follows (R is the set of real numbers.): if $Y \in \bigcup_i N(X_i)$, $s(Y) = \{\sum_{1 \leq i \leq k} f(X_i \to Y)\}/(a_k k^{e_k} + b_k)$, otherwise, $s(Y) = 0$.

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 1 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 into a few customers that purchased excessively many products before.

This work was supported by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University (SNU) and Brain Korea 21 Project. The RIACT at SNU provided research facilities for this study.

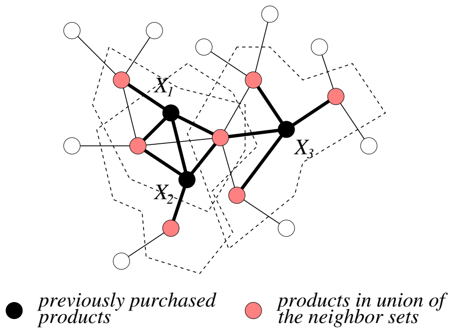

Figure 1: A connection network

Such a recommendation strategy is different from the existing ones based on the customers' profiles (for example, collaborative filtering [2]) in that it performs recommendations just by the relationships between products regardless of the customers' profiles. Experimental results with field data showed that the connection network model is fairly good not only in the computational cost but also in the recommendation quality.

References

[1] R. Agrawal, T. Imilienski, and A. Swami. Database mining: A performance perspective. *IEEE Trans. on Knowledge and Data Engineering*, 5(6):914–925, 1993.

[2] 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.

[3] 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.