Discovering association strength among brand loyalties from purchase history

著者	Hamuro Yukinobu, Katoh Naoki, Yada Katsutoshi
journal or publication title	ISIE 2001 : 2001 IEEE International Symposium on Industrial Electronics Proceedings
page range	114-117
year	2001-06
権利	(C) 2001 IEEE. Reprinted, with permission, from HAMARO Yukinobu, KATOH Naoki, YADA Katsutoshi, Discovering association strength among brand loyalties from purchase history \n, 06/2001. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of Kansai University's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.
URL	http://hdl.handle.net/10112/6925

Discovering Association Strength among Brand Loyalties from Purchase History

Yukinobu Hamuro[1], Naoki Katoh[2], Katsutoshi Yada[3]

[1] Dept. of Business Administration, Osaka Sangyo Univ., Daito, Osaka 574-8530 JAPAN
[2] Dept. of Architecture and Architectural Systems, Kyoto University, Kyoto 606-8501 JAPAN
[3] Faculty of Commerce, Kansai University, Suita 564-8680 JAPAN

Abstract

Analyzing purchase history of customers enables us to discover valuable knowledge that is helpful for developing effective sales promotion. In this respect, we shall introduce a new notion, *association strength* among brand loyalties, which is defined for every ordered pair of brands. If the association strength between loyalties of brands A and B is high, it represents that purchase of brand A is highly correlated to that of brand B. Conventional method for discovering associative purchasing is usually applied for one purchase opportunity (one receipt), i.e., it reveals how often two commodities are purchased at the same time. On the other hand, we are interested in discovering relationship among customers' loyalties to certain brands or manufacturers by investigating long-term purchase history of customers. By computing association strengths from customers' purchase history of drugstore chain in Japan, we could produce several interesting rules that will be useful for sales promotion planning.

1. Introduction

According to the rapid development of modern computer technologies, there has been much progress made in automating daily office work. This, in turn, has resulted in the accumulation of a huge amount of business data into databases. However, it does not seem that most companies can make full use of such accumulated data for strategic planning of their future business.

On the other hand, *knowledge discovery in databases* or *data mining* has become an active research area in which new technologies or methodologies are sought to automatically extract meaningful knowledge from business data [1, 7, 8].

One of the fundamental techniques in data mining is *association rule*. A typical application of association rule is to discover rules of associative purchasing by analyzing data accumulated from POS(point of sales) terminals. For supermarkets or drugstores, customers purchase several items at one time. By investigating receipts issued by POS terminals, we can find how often two commodities are purchased at the same time. This information, for instance, helps to construct effective location of commodities in a store.

If purchase history of customers are available, we can discover much more valuable knowledge that can be helpful for constructing effective marketing strategy [2, 3, 5, 4]. For instance, we can measure brand loyalty, or for customers who are loyal to a certain brand we can find to what brands in other commodity categories such customers have loyalty. Such information can be used in promotion sales planning such as effective cross-sell [2, 6].

In this respect, we shall introduce a new notion, *association strength* among brand loyalties, which is defined for every ordered pair of brands. If the association strength of brand A to brand B is high, it represents that loyalties of brands A and B are highly correlated. Conventional method for discovering associative purchasing is usually applied for one purchase opportunity (one receipt), i.e., it reveals how often two commodities are purchased at the same time. On the other hand, we are interested in finding correlation between loyalties of two brands. If customers with high loyalty to a brand A in category C are usually loyal to other brand B in category D, it indicates that there exists some reason behind such phenomenon. Such correlation provides us with useful customer knowledge. For example, if both brands are made by the same manufacturer, we can infer that there may exist high loyalty to the manufacturer. If manufacturers of brands A and B are different although the manufacturer of brand A produces brand B' which belongs to the same category as that of brand B. This indicates that the manufacturer of brands A and B' should carry out effective sales promotion so as to increase the number of customers who purchase both A and B'. In this manner, by computing

0-7803-7090-2/01/$10.00

association strengths of brand loyalties, we can measure relationships of brand powers among commodities that belong to various commodity categories by which manufacturers can understand which brands are correlated or not correlated with which brands of other manufacturers.

We have carried out computational experiments by using customers' purchase history of drugstore chain in Japan in order to observe whether we can produce interesting rules that will be useful for developing effective sales promotion.

The organization of this paper is as follows. Section 2 rigorously defines association strength, and Section 3 reports computational experiments. Section 4 concludes the paper and mentions future research.

2. Association Strength among Brands in Purchase History

Suppose that there are two commodity categories, say, laundry detergent (category $\mathcal{L}$) and dish-washing detergent (category $\mathcal{D}$) and that there are three manufacturers M_1, M_2, M_3 each of which produce one brand in both categories. Let A_i and B_i for each $i = 1, 2, 3$ denote the brands in categories $\mathcal{L}$ and $\mathcal{D}$ respectively which are produced by manufacturer M_i. Suppose also that there are three customers who bought brands A_i and B_i in their purchase history (Table 1 shows the number of items of each brand purchased by customers).

Table 1: An example used to explain association strength

	category $\mathcal{L}$				category $\mathcal{D}$			
	A_1	A_2	A_3	total	B_1	B_2	B_3	total
customer 1	10	2	3	15	20	2	2	24
customer 2	15	4	1	20	10	3	3	16
customer 3	5	14	11	30	2	20	15	37
total	30	20	15	65	32	25	20	77

From this table, we observe that customers 1 and 2 have rather high loyalty to brands A_1 and B_1 while customer 3 is not loyal to any of these brands. For customer 1, assuming that the numbers of items of categories $\mathcal{L}$ and $\mathcal{D}$ which he/she purchases are the same (notice from Table 1 that such numbers are different, i.e., 15 $\mathcal{L}$ items and 24 $\mathcal{D}$ items), let us consider the conditional probability $P_1(B_1 \mid A_1)$ representing how many items of brand B_1 the customer purchases per purchase of one item of brand A_1. Here the subscript 1 of P_1 stands for the customer number. It is natural to regard that $P_1(B_1 \mid A_1)$ is independent of A_1 since the only information available is Table 1. That is, $P_1(B_1 \mid A_1) = P_1(B_1)$ which is equal to his/her brand share in category D ($= 20/24$). Now reflecting that the total number of items of categories $\mathcal{L}$ and $\mathcal{D}$ are 15 and 24, 10 items of brand A_1 out of 15 category $\mathcal{L}$ items are purchased by customer 1, we consider that the customer 1 purchased $10 \cdot 24/15$ ($= 16$) $\mathcal{D}$ items in association with the purchases of 10 $\mathcal{L}$ items. Among 16 $\mathcal{D}$ items, $16 \cdot P_1(B_1) = 13.33$ items is expected to be of brand B_1. Thus, it is reasonable to consider that 13.33 B_1 items have been purchased in association with brand A_1.

Seeing the last row Table 1, you may naturally consider that for the average customer the conditional probability $P_{ave}(B_1 \mid A_1)$ is equal to $P_{ave}(B_1) = 32/77$. However, it does not seem to suitably reflect the overall purchase behaviour because there is a positive correlation of the purchase of A_1 and that of B_1 as seen from Table 1. Then, what is the best estimated value of $P_{ave}(B_1 \mid A_1)$? We shall define it as the the value x^* that minimizes the squared error from $P_i(B_1)$ weighted by the purchase frequency, i.e., x^* is computed as the solution of

$$\min_x F(x) = \sum_i a_i(A_1) \frac{a_i(\mathcal{D})}{a_i(\mathcal{L})} (P_i(B_1) - x)^2,$$

where i ranges over all customers, $a_i(A_1)$ and $P_i(B_1)$ denote the number of items of brand A_1 purchased by customer i and the brand share of customer i for B_1, and $a_i(\mathcal{L})$ (resp. $a_i(\mathcal{D})$) denotes the ratio of the number of $\mathcal{L}$ (resp. $\mathcal{D}$) commodity purchased by customer i to the total sales volume of commodities of $\mathcal{L}$ (resp. $\mathcal{D}$). It is easy to see that

$$\begin{aligned} x^* &= \frac{\sum_i a_i(A_1) \frac{a_i(\mathcal{D})}{a_i(\mathcal{L})} P_i(B_1)}{\sum a_i(A_1) \frac{a_i(\mathcal{D})}{a_i(\mathcal{L})}} \\ &= \frac{\sum_i P_i(A_1) a_i(\mathcal{D}) P_i(B_1)}{\sum_i P_i(A_1) a_i(\mathcal{D})} \\ &= \frac{\sum_i P_i(A_1) a_i(B_1)}{\sum_i P_i(A_1) a_i(\mathcal{D})} \end{aligned}$$

Such x^* is denoted by $\hat{P}_{ave}(B_1 \mid A_1)$ (expected conditional purchase probability of B_1 with respect to A_1).

Association strength of A_1 on B_1 (denoted by $as(A_1, B_1)$) is defined as

$$as(A_1, B_1) = x^* / share(B_1),$$

where $share(B_1)$ denotes the share of B_1 in category $\mathcal{D}$. From Table 1, $as(A_1, B_1) = 0.622/0.416 = 1.50$. In the same manner, We can compute $as(A_i, B_j)$ for all i and j, and we can also can compute $as(B_i, A_j)$ for all i and j. Intuitively, if $as(A_i, B_j) > 1$, brand loyalties

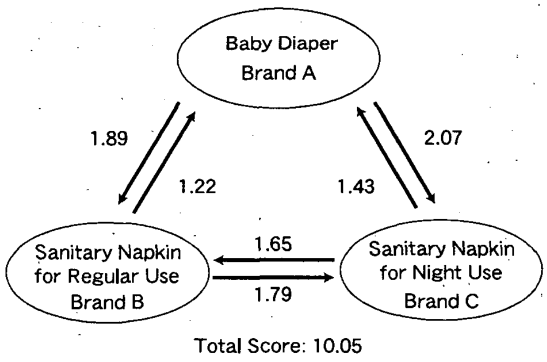

Fig. 1: A triple of brands that attain the highest score

of A_i and B_j are positively correlated. otherwise it is negatively correlated.

We can extend the notion of association strength to a k-tuple of brands which are all belong to distinct commodity category. It is defined simply as the sum of association strengths for all ordered pair of brands.

3. Computational Experiments

In order to observe whether we can obtain interesting rules by computing association strengths, we perform computational experiments using POS data accumulated in drugstore chain in Japan [3]. Our experiments used 252,761 customers and 750 categories. The average number of brands per category is 15.

Because the space limit, we only give two interesting results.

(1) For every triple of categories and for every triple of brands each from distinct category, we have computed association strengths. One of the brand triples that attained highest score is illustrated in Fig.1. The corresponding three categories are baby diaper, sanitary napkin for regular use and sanitary napkin for night use. Brands shown in the figure are those whose prices are relatively low and frequently sold at discount price. We may see that customers who are loyal to these brands are loyal to their prices, but not to brands.

(2) We did a similar experiment for 5-tuple of categories as shown in Table 2. a, b and c in parentheses stand for manufacturers. Three tuples of brands are those which attain highest scores among all combinations of brands in this set of five categories. As seen from Table 2, the second tuple indicates that there are certain fraction of customers who are loyal to brands all from manufacturer b. The other two tuples indicate that manufacturers a and c are not strong in category of liquid laundry detergent. This may be useful information for these two manufacturers.

4. Conclusion and Future Work

In this paper, we have introduced a new notion, association strength among brand loyalties. From computational experiments using real POS data of drugstore chain, we found interesting rules that can be useful in future promotion sales planning. We are planning to carry out further computational experiments using various POS data in order to see how association strengths among brand and manufacturer loyalties can be utilized in sales promotion planning of real business.

References

[1] R. Agrawal, T. Imielinski, and A. Swami, *Database mining: A performance perspective*, IEEE Transactions on Knowledge and Data Engineering, 5 (1993), pp. 914-925.

[2] Katsuki Fujisawa, Yukinobu Hamuro, Naoki Katoh Takeshi Tokuyama, and Katsutoshi Yada, Approximation of Optimal Two-Dimensional Association Rules for Categorical Attributes Using Semidefinite Programming, *Proc. of 2nd Symp. on Discovery Science*, LNAI 1721, Springer-Verlag, 148-159, 1999.

[3] Y. Hamuro, N. Katoh, Y. Matsuda and K. Yada, Mining Pharmacy Data Helps to Make Profits, *Data Mining and Knowledge Discovery.* Vol.2, No.4, (1998), pp.391-398.

[4] Hamuro, Y., Katoh, N., and Yada, K., Discovering Interpretable Rules that Explain Customer Brand Choice Behavior, INFORMS-KORMS Seoul 2000, pp.561-568, June 2000.

[5] Ip, E, Yada, K., Hamuro, Y., and Katoh, N., A Data Mining System for Managing Customer Relationship, Proceedings of the 2000 Americas Conference on Information Systems, August 2000, pp.101-105.

[6] Kitts B., Freed D. and Vrieze M., Cross-sell: A Fast Promotion-Tunable Customer-item Recommendation Method Based on Conditionally Independent Probabilities, Proc. KDD 2000, pp. 437-446, Aug. 2000.

[7] G. Piatetsky-Shapiro (Editor), Knowledge Discovery in Databases, AAAI Press, 1991.

[8] R. Uthurusamy, U.M. Fayyad, and S. Spangler, *Learning Useful Rules from Inconclusive Data*, In [7], pp. 141-157, 1991.

[9] Yada, K., Hamuro, Y. and Katoh, N., The Discovery of Customer Loyalty from Newcomers, Proceeding of 2000 MIS/OA International Conference, pp.185-189, June 2000.

Table 2: An example of manufacturer loyalty

Powder Laundry Detergent	Liquid Laundry Detergent	Fabric Conditioner	Dish-washing Detergent	Rinse in Shampoo	Total Score
A(a)	B(b)	E(c)	H(a)	K(a)	24.7
B(b)	D(b)	F(b)	I(b)	L(b)	24.2
C(c)	B(b)	G(c)	J(a)	M(a)	23.5