

Essays on Database Marketing

著者	Kholod Marina Viktorovna
学位授与機関	Tohoku University
学位授与番号	経博(経営)第63号
URL	http://hdl.handle.net/10097/50704

氏名(本籍地) ^{ホーロド} ^{マリーナ} Kholod Marina
^{ヴィクトロブナ} Viktorovna

学位の種類 博士(経営学)

学位記番号 経博(経営)第63号

学位授与年月日 平成21年9月3日

学位授与の要件 学位規則第4条第1項該当

研究科、専攻 東北大学大学院経済学研究科(博士課程後期3年の課程)
経済経営学専攻

論文題目 Essays on Database Marketing
(データベース・マーケティングにおけるエッセイ)

博士論文審査委員 (主査)
教授 照井伸彦 教授 佃良彦

論文内容の要旨

PART 1 MARKET BASKET ANALYSIS OF POS DATA

The first essay presents a real-world application of market basket analysis using POS data, from a collection of retail convenience stores in Japan. The use of a recent technique in data mining, namely association rule analysis, demonstrates to be very efficient to discover valid and interesting purchase behavior from this type of data. Association rule algorithm is used to identify market baskets - affinity across product items - and the use of virtual items in association rules enriches such type of analysis. The resulting framework and analysis can be relied on to optimize store layout, maximize profitability, enhance assortment decisions, as well as reveal cross-selling and complementary effects for a large number of products.

The goal of this study is to conduct market basket analysis using point-of-sale (POS) data from convenience stores. This kind of analysis might shed light on consumer shopping behavior.

This essay is organized as follows. Section 2 briefly reviews the background on using

data mining techniques to predict consumer buying behavior. Section 3 presents the methodology of analysis, implements it and discusses the reported results. Section 4 makes the interpretation of the association rules, points out how to pick up the interesting rules and demonstrates some managerial implications. Section 5 summarizes and concludes this essay.

As for the managerial application of the results that we obtained, we can mention the following:

- *New substitutes and compliments can be discovered*

For example, as substitutes we can give the example of rules with VOLVIC Natural Mineral Water (See Thesis, Table 2). Their lift is negative, which means that a consumer who purchased GLICO Mild Café au Lait, KIRIN Green Tea has no tendency to purchase VOLVIC Natural Mineral Water and vice versa. The reason is that the antecedent and consequent is both drinks and they might satisfy the same need so that the consumer considers them as substitute goods.

- *Store layout and stock management*

For complementary products: Place products that have a strong purchasing relationship close together to take advantage of the natural correlation between products. Alternatively, place such products far apart to increase traffic past other items. For example, this tactic can be applied to the rule 'if lemon then rice' (See Thesis, Table 2).

For substitutes: shop should always provide at least one of the products of the association pair, for example, one of such products like coffee and tea, should be always present in the shop.

- *Coupons and discounting*

Changing a coupon to feature a second product that a customer might buy would increase sales at no additional promotion expense or the future discount promotions can be determined. As well as for the compliments, for example, discounts for the coffee may increase demand for sugar. In our database there is a quite interesting rule if 'ASAHI Super Dry and Fried Chicken then NATORI Mackerel' (See Thesis, Table 5), which can be a good example of such couponing.

- *Timing and cross-marketing*

If the data about the purchases of the concrete individual is available, then it is possible to assume such rules as "when people who have purchased pet food are three times more likely to purchase pet toys in the time period one-to-three months after the pet food was purchased". The results satisfying such application are beyond the interest of this paper.

- *Extension of life cycle of the product by bundling with another product*

If there is an association between certain products then the manufacturer can produce a product similar to the one the lifecycle of which is closed to the end. In such way the life cycle of the product can be “extended”. This idea can be applied to all interesting rules found in our analysis.

In this essay standard association rules analysis, including virtual item concept was illustrated. This is an interesting technique to discover regularities in purchase behavior that may exist on different levels of abstractions (though in this essay the product affinities on the individual product level are mostly considered). Concept hierarchies (product taxonomies or categories of the products) can also be of great interest to the retailer, but expressed at a higher conceptual level they carry less specific information and they ultimately determine the usefulness and richness of the results.

Virtual items provide a good addition to the association rule discovery. Virtual items are used to include other information than just product related information, such as customer related and store related characteristics and so on. Including these virtual items into the discovery process enables us to find differences in the purchase behavior between different types of consumers or stores. Obviously, the collection of such information is crucial to carry out such analysis. Due to the increasing use of loyalty cards and systematic collection of consumer- related information, these kinds of analysis are gradually becoming possible.

Market basket mining is a useful method of discovering customer purchasing patterns by extracting associations or co-occurrences from convenience stores’ transactional databases. But these databases contain huge amounts of information which sometimes cannot be easily handled by the existing software. Besides, the support- confidence framework often generates too many rules, so it is necessary to judge their strength. Some of the obtained rules can be deceptive, so we will have to verify them by computing the benchmark confidence value and the lift ratio.

In this essay the strength of the rules was judged according to three measures (support, confidence, lift), and so the richness of the results was achieved in this way. This approach provides the full information on each brand purchased from the convenience store.

Convenience stores have their own core categories, with a high degree of purchase incidence among category purchasers and a high degree of loyalty (purchasers who buy the category most often at a convenience store) for core categories like newspapers, cold drinks, coffee/café au lait and cigarette packs. A market basket analysis indicates that

drinks and meals tend to be purchased with something, while newspapers are more likely to be purchased alone.

The obtained results shed light on consumer profile as well. One would suggest that the convenience store has a clear niche. The traditional convenience store shopper profile is presented by young males who read newspapers and smoke. As merchandise shopping becomes less frequent, the shopper is more likely to be female.

Regarding the limitations of this analysis, we can point out the following.

First, for the purpose of analyzing consumer behavior in the retailing industry we chose POS data. Though for studying existing customers, data from the time the customers were acquired, data describing their current status and behavioral data should be included.

Second, it is necessary to have a large number of real transactions to get meaningful data and the data's accuracy is compromised if all of the products do not occur with similar frequency.

Third, because the primary goal of the methodology is to create a stable model that will work at any time of year and also in the future, it is very useful to have data which does not all come from one time of a year, because building a model on data from a single period of time increases the risk of getting the results which are not generally true.

PART 2 BRAND VALUE MEASUREMENT FROM SCANNER DATA WITH CONSUMER HETEROGENEITY

It is widely recognized that brands have value and it is important to measure it. Yet, there are a lot of shortcomings in the approaches used to measure this value. Most commonly the variety of methods to measure Brand Value can be divided into two main approaches: from the brand equity standpoint and from the customer equity standpoint. Calculation of Brand Value from the brand equity standpoint is the topic of central interest in this essay.

In our work we use a brand equity approach to measure Brand Value, as this particular approach reflects one of the key topics of microeconomics—consumer behavior and consumer choice modeling. Through consumer choice modeling we tackle the problem of brand valuation in this work.

We explain the concept and main models which are used in evaluating the brand equity in Chapter 2. We evaluate brand through consumer choice modeling in Chapters 3 and 4. Based on such evaluations, in Chapter 5 we discuss about brand performance which sheds light on the performance of a firm.

For a consumer, brand choice means what type of products/brands to select given a set of alternatives in present market, in order to maximize his self-interests (utility). The utility assigned to a brand by consumer is the brand equity (value) which is measured in this research. The brand value at the level of individual consumer is measured by hierarchical multinomial logit model and the parameters estimation relies on the MCMC technique. The hierarchical model which is applied combines some determinants of product utility such as price and physical features of the product. In this section, the conceptual consumer choice model discussed by Tybout and Hauser (1981) is adopted. We apply similar framework to that presented in Kamakura and Russell, referred as K-R, (1990). However there are some extensions to be pointed out in our model.

We propose a probit model (i.e. the random term in the utility equation is normally distributed) rather than the logit specification as in K-R.

We added to the set of explanatory variables in the intrinsic part of the utility equation the demographic variable Z , which relates the utility of a consumer to his total number of purchases. The same approach was used to allow for heterogeneity in the equation of market response parameters in β .

The introduction of the cross effect between D_j and Z_h through the term $D_j Z_h \zeta$ in the first hierarchy level equation (See Thesis, Equation 4.4), which might catch the type of heterogeneity in consumer behavior, which cannot be explained through D_j and Z_h taken in additive manner. This specification constitutes a natural extension of the K-R model.

In the K-R framework heterogeneity is considered across segments of consumers, while in our approach we consider heterogeneity across consumers or level of individuals. Hence, we switch from a finite mixture to a continuous one (i.e. t tends to h , in size).

Based on this framework the utility is decomposed into two components, the first is the intrinsic utility to the brand and here the physical features of the product are included and the second is the part of the utility that can be explained by situational factors such as price and short-term advertising.

Brand Value can be used as a diagnostic tool for the performances of Brand as a whole, as well as its tangible and intangible components. Simple ordering of brands in terms of market share does not shed light on the reasons of certain performance of the brand.

We think that Brand Value is an alternative approach, allowing the manager or analyst to have alternative measures of brand performance, which account for underlying determinants of market share.

BV is adjusted for the situational constraints (such as short-term price and display),

BIV presents the measure which adjust for not only situational constraints, but also for physical features of the product. BIV can be used as a measure which provides information about non-physical (non-product) variables such as advertising and power of the channel) in building a strong brand. BIV measures the intangible part, which is difficult for competitors to hurt. From this standpoint, BIV can be more important than BV itself.

The main results are reported as follows:

Table: Brand Value Measurements (with quantities scaled to their means)

	BTV	BIV	BY	Choice Share
Brand 1	-0.571	1.405	0.835	0.382
Brand 2	0.299	0.210	0.509	0.241
Brand 3	0.823	-1.189	-0.366	0.133
Brand 4	0.310	-0.817	-0.506	0.122
Brand 5	-0.211	0.061	-0.150	0.069
Brand 6	-0.652	0.330	-0.321	0.053

BV = BIV+BTV

Note: BV, BTV and BIV are scaled to sum to zero across all brands.

As for the intuitive meaning of BV, BIV and BTV, it becomes clear from the

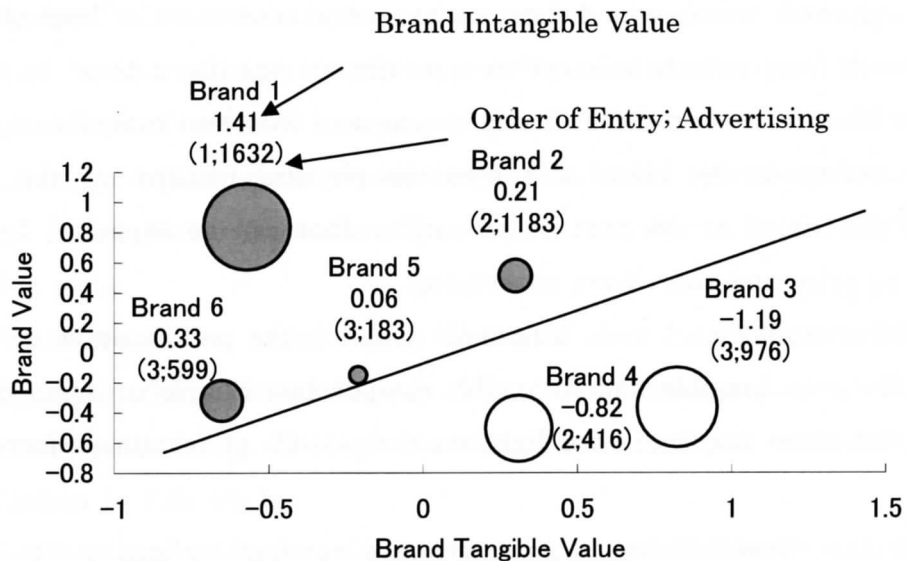


Figure: Brand Value vs. Brand Tangible and Brand Intangible, Order of Entry to the Market and Cumulative Advertising Exposures

following Figure.

The conclusion which can be drawn from this Figure is that overall brand perceptions as evaluated by BV are related to consumer evaluations of physical features of the brand, measured by BTV. The Brand 3 (P&G. Ultra Ariel) has the most desirable physical features but it has just third highest BV, after Brand 1 (KAO Attack) and Brand 2 (LION High Top), which is negative, as well as its BIV, which means that market performance of P&G. Ultra Ariel is weaker than expected on the basis of its objectively measured attributes. The same can be said about Brand 4 (P&G. Lemon Cheer). It is a surprising result, because normally brands which have desirable features are supposed to have high Brand Value, and brands with poor physical features – low Brand Value. But some historical facts can shed light on this fact (See Appendix of the Thesis). P&G. Ultra Ariel was rival brand with KAO Attack and LION Top. The difference between P&G. brands (Ultra Ariel and Lemon Cheer) and KAO and LION brands is that BIV of P&G brands is negative (below the bisecting line) and it contributes to negative BV of those brands. The situation is absolutely opposite in case of KAO Attack and LION High Top, which are the strongest brands in the market, while KAO Beads and LION Dash have positive intangible value, but their negative tangible value contributes to their negative whole value. Actually KAO brands have the lowest tangible values. But it is only due to the intangible part of KAO brands they are ranked as No.1 and No.3 in terms of their Brand Values. KAO Attack was the first brand which entered the compact laundry detergent market in 1987. Following KAO, LION (in 1988) and P&G (in 1990) entered compact laundry detergent market. So, it is quite natural that values of KAO Attack and LION Top brands are high as their intangible values contribute more to the whole value than tangible value do, which can be explained by the fact that KAO Attack was first in making the brand compact and abandoning the use of non-ecological salts of phosphoric acid. So the consumer associated the product as ecologically friendly, cost effective (thanks to the “one spoon is enough” principle), efficient product (use of bioenzymes – blue grains – for strong dirt) and the first product of such type in the market. Since that time KAO Attack retains No.1 market share. From our analysis we can see that it does so thanks to its intangible value, based on the above-mentioned associations of consumer. While its tangible value brings it only the 5th place (See Thesis, Table 9), hence the consumer associations based on perceptions and distortions of those perceptions play bigger role in the evaluation of KAO Attack than the direct perceptions of its physical features.

This positive image about KAO helps Brand 6 (KAO Beads) to keep 2nd-high intangible and 4th-high tangible values.

The line bisecting the Figure separates the brands with positive and negative BIVs. Brands with positive BIV (shown as the vertical deviation of the brand from the dividing line), located above this line, have greater BV than expected from measuring their attributes (KAO and LION Brands). Brands assigned a negative intangible value (P&G Brands) are located below this line, which means that their market performance is weaker than expected from measuring their physical attributes. Managerially, this suggests that the values of these 6 brands are primarily based upon how the brand names are associated by the consumer.

To understand the intuition behind BTV and BIV we can compare Brand 5 (LION Dash) and Brand 6 (KAO Bio Beads). Both brands have close BTV, but too radically different BIV. KAO Bio Beads is assigned higher intangible value, while LION Dash is assigned lower BIV. And the result is that LION Dash has a larger share (0.069) than KAO Bio Beads (0.053), even though they have the same price (0.919).

As we can see, in general, high BIV contributes to high BV, but brands with high value need not have high intangible value. A good example is Brand 5 (LION Dash). Although it has a positive BIV, its overall value is very close to what we would expect from its physical attributes. From the managerial standpoint, it means that the value of LION Dash brand is based upon its product performance.

In the case of our results from the Japanese market, the conclusion that we can draw from the Figure is that overall quality perceptions which is measured by BV are not related to consumer evaluations of physical features of the brand as measured by Brand Tangible Value, but related to consumer associations of the brand and distortions of those associations measured by Brand Intangible Value. Brands, which have positive brand name associations (such as KAO Attack), have high Brand Value, in contrast to brands with poor associations (such as P&G Ultra Ariel), which have low or negative Brand Value. It is exactly the opposite conclusion from the American brands evaluated by K-R.

In this work we made behaviorally based measures of Brand Equity for individual consumers. The residual approach that we used measures the value assigned by consumer to the brand from scanner data, which is data on choice behavior of the consumers in the market.

Brand Value is a measure of intrinsic utility or value of a brand to consumers, after the situational factors were removed. Brand Value contains both tangible and intangible parts.

BIV, though a very useful tool, should be used carefully. First, because it is a residual measure of utility and its validity clearly depends upon the way in which Brand Value is

defined in the choice model. Second, because it is based on a set of physical features used to measure the tangible product.

The disadvantage of this approach is that it is most appropriate for brands characterized with a predominance of product-related attribute associations because it is unable to distinguish between different types of non-product-related attribute associations. Consequently, its diagnostic value for strategic decision making in other cases is much more limited.

Speaking generally, this approach takes a “static” view of brand equity. It contrasts strongly with a “process” view, where consumer response is defined in terms of perceptions, preferences, and behaviors with respect to a variety of marketing activities. A useful extension of our work would be the development of such a “process” view. This can provide new tools for studying the determinants of brand equity, thus contributing to the consumer-based brand equity framework which emphasizes the importance of employing a range of research measures and methods to fully capture the multiple potential sources and outcome of brand equity.

論文審査結果の要旨

本論文では、市場データを用いてマーケティング戦略を展開するデータベース・マーケティングに関する二つの問題を扱っている。第一のトピックであるマーケットバスケット分析は、店舗内商品間で事前に想定される代替補完関係に拠らず、大量購買データから同時購買確率をすべての組み合わせに関して高速評価することにより顧客の商品併買関係を検出し、クロスセリングのプロモーション効率を高めようとするマーケティング分析ツールである。論文では、まずマーケットバスケット分析が現代マーケティングにおいて果たしうる役割とその具体的方法の理論的展開について、文献研究により整理して論述している。さらに実店舗の POS データを利用して、コンビニエンスストアの購買行動に関する実証分析を行っている。そこでは、いくつかのアルゴリズムによる感度分析を経て、意味ある併買関係を検出し、店舗内のレイアウトなどのインスタ・プロモーション、プロモーション収益性評価、アソートメント決定の最適化、クロスセリングの可能性について、マネジリアルな意味づけを導いている。

第二のトピックは、ブランド価値に関する研究である。現代のマーケティングにおいて、ブランドの管理と育成は大きな潮流のひとつであり、1990年代にアーカーらによって提唱されたブランド

価値およびその測定方法の研究は、これらの基礎を構成する重要な問題である。これまでブランドの無形資産として定義されるブランド価値を測定する様々なアプローチが提案されてきているが、本論文では、消費者の市場行動データにより価値を測定するアプローチをとる。具体的には、消費者の市場行動をブランド選択モデルで規定し、ブランドの効用から各種マーケティング戦略の短期的効果を差し引いた残りとして、長期的価値としてのブランド価値を抽出する。その際、ベイズ統計によるマルコフ連鎖モンテカルロ法を駆使した新しい分析モデルを展開し、消費者個別のブランド価値評価を可能としている。さらに、消費者パネルデータを用いた実証分析では、消費者個別に測定されるパラメータ推定値から得られる非集計情報から、先行研究には見られない実務上興味深い結論も得ている。

より踏み込んだ考察と分析がさらに必要である部分もあるが、これらは今後の課題であり、論文の水準を低めるものではない。

以上により、本論文は博士（経営学）の学位を授与するに値する論文であると認定する。