

A Category Role Aided Market Segmentation Approach to Convenience Store Category Management

Yongjie Ye¹, Shuihua Han

Abstract

This paper presents an innovative market segmentation model which is driven by category-role, for the first time, to support category management in chain CVSs. The usefulness and applicability of this study is illustrated by means of an empirical study. The derived results are discussed and compared with the existing works.

Keywords: Category Management, Market Segmentation, Chain Convenience Stores Management

Introduction

In recent years, category management (CM) (Basuroy, Mantrala, & Walters, 2001; Morgan, Kaleka, & Gooner, 2007) gradually gained popularity in retailers to increase their core competitiveness, maximize the profits and ensure long-term healthy customer relationship. CM (Dussart, 1998; Heller, 2012; Zufryden, 1987) is a business fundamental which uncovers a lot of untapped potential to explore. It aims to analyze consumer purchasing behavior and stock the products consumers are most likely to buy. More specifically, the defined categories can be used to target the consumer groups to gain a better understanding of their needs.

Although it attracts increasing attention and some initial results have been obtained (Dussart, 1998; Kamakura & Kang, 2007; Trivedi, 2011), applying conventional CM methods to chain Convenience Stores (CVS) management it is still a challenging task due to its distinguished features. First, CVSs are often distributed in a variety of areas and each store has impulsive consumers. Therefore, different CVSs may have to provide different products to meet the diverse requirements of their consumers. Second, due to the mobility of consumers, it is difficult to employ common CM attributes (e.g. location, population, age) to determine the categories. Third, CVSs are often small and have very limited storage space, such that it cannot provide large number of product categories. This leads suppliers to invest most of their time and effort on their large retailers, but not small chains CVSs.

To overcome these difficulties, this paper proposes an innovative market segmentation model, which for the first time employs a category role (CR) to support CM in chain CVSs. This new method takes three CR dimensions (importance to retailers,

¹ Corresponding author. Email: fjyeyongjie@126.com.

consumers, and marketplace) into account to segment the market. Initially, the historical transaction data provides a full view of the market. Such data are then used to derive a global category index (CIX) for each store. CVSs which share similar CIXs are clustered into the same group by using a new similarity measure, named HCsim() and an improved weighted fuzzy K-mean clustering algorithm (named WFKM). Based on the obtained clustering results, the retailer can design varied CM and marketing strategies for different CVSs clusters.

The remainder of this paper is organized as follows: In Section 2, a novel method, the CR-based market segmentation model, is proposed to cluster the CVSs. In addition, an improved clustering algorithm which employs HCsim() is also introduced in this section. The applicability and utility of the proposed method is demonstrated in Section 3 via an empirical study. The final section concludes this paper and points out further work directions.

CR-based market segmentation approach

Category role

A key tenet of CM is that the retailer should decide on the role each category plays in the overall portfolio. According to cross-category quantitative analysis (Dussart, 1998; Kamakura & Kang, 2007), the typical four CRs are:

- Destination category: Make the store the primary category provider. When consumers would like to purchase products in a certain category, the store is their main choice. It takes 5% - 10% of categories in the store.
- Routine category: Make the store one of the preferred category providers. This category generates solid profits for retailers, while meeting the customers' daily various needs. It takes 50% - 70% of categories in the store.
- Occasional/seasonal category: Make the store a major category provider when consumers would like to buy a given occasional/seasonal product. It takes 10% - 15% of categories in the store.
- Convenience category: This category helps the store to provide convenience to consumers. It takes 10% - 15% of categories in the store.

CRs are essential to maintain a consistent strategic and tactic plan, which provides cohesive market intents across price, promotion and assortment. It is pointed out that the same category may act as a different CR in different stores (Trivedi, 2011), so that customized category strategy should be implemented in different market segments.

Category index (CIX)

Cross-category quantitative analysis (Basuroy, et al., 2001; Dussart, 1998; Kamakura & Kang, 2007) has been widely used in the assignment of CRs. In this method, three dimensions (as shown in Table 1) should be taken into consideration and they are described as below:

Table 1: Three dimensions of CR

Three dimensions	Measures
The importance to consumers	Annual expenditure; purchase frequency; purchase volume

The importance to retailers	Sales volume; sales revenue; sales gross prot; efficiency plateau
The importance to marketplace	Average growth rate; purchase trends; market share changes trends

Due to the availability of dimension measures, five segmentation attributes are selected to aggregate a new category index (CIX):

$$CIX_i = \omega_F F_i + \omega_N N_i + \omega_S S_i + \omega_R R_i + \omega_G G_i \quad (1)$$

Where CIX_i is the index value of category i , and F_i, N_i, S_i, R_i , and G_i represent the average sales frequency, average sales volume, average sales revenue, average gross profit and average growth rate of category i , respectively. $\omega_F, \omega_N, \omega_S, \omega_R$ and ω_G represent the corresponding attribute weights and $\omega_F + \omega_N + \omega_S + \omega_R + \omega_G = 1$.

A new similarity measure (HCsim())

Based on the features of CVSs category data, coupled with both the similarity between category structures and the similarity in data space, a revised category similarity measure, HCsim(), is defined as:

$$HCsim(s_j, s_k) = \frac{\sum_{i=1}^L \frac{p_{jk}(c_i)}{1 + |cix_{ji} - cix_{ki}|}}{L} \quad (2)$$

Where $L = |C'_j \cup C'_k|$ is the total number of categories which store j and k sell, the $p_{jk}(c_i)$ represents the ratio of common sub-categories or products in category c_i that shared by store j and k , cix_{ji} is the CIX values of category i in store j .

This similarity measure has the following properties:

- The greater similarity value indicates the two CVSs are more similar to each other.
- $0 \leq HCsim(s_j, s_k) \leq 1$, if and only if two CVSs are identical, then $HCsim(s_j, s_k) = 1$. In contrast, when $HCsim(s_j, s_k) \rightarrow 0$, store j and k are totally different.
- This measure is symmetric, i.e. $HCsim(s_j, s_k) = HCsim(s_k, s_j)$.

This distance function differs from the traditional distance functions. Its value reflects the similar dimensions (categories, in this case) between any two stores and is less affected by dissimilar categories, thus, the more categories shared by two stores, the more similar the two stores are, and vice versa. It matches the natural human-being perception well.

Clustering

Clustering is a commonly used technique in market segmentation (Liu, Kiang, & Brusco, 2012; Shin & Sohn, 2004). In this work, an improved weighted fuzzy K-means clustering algorithm (WFKM) which employs previously proposed HCsim() similarity measure is proposed herein to cluster CVSs. The new CIX, which considers consumers, retailers, and marketplace, is also employed to help cluster CVSs (as shown in Figure1).

Ultimately, by analyzing the clustering results help to gain a better understanding of market segments and consumers. It also provides solid evidence for the implementation of CM and market strategy in CVSs.

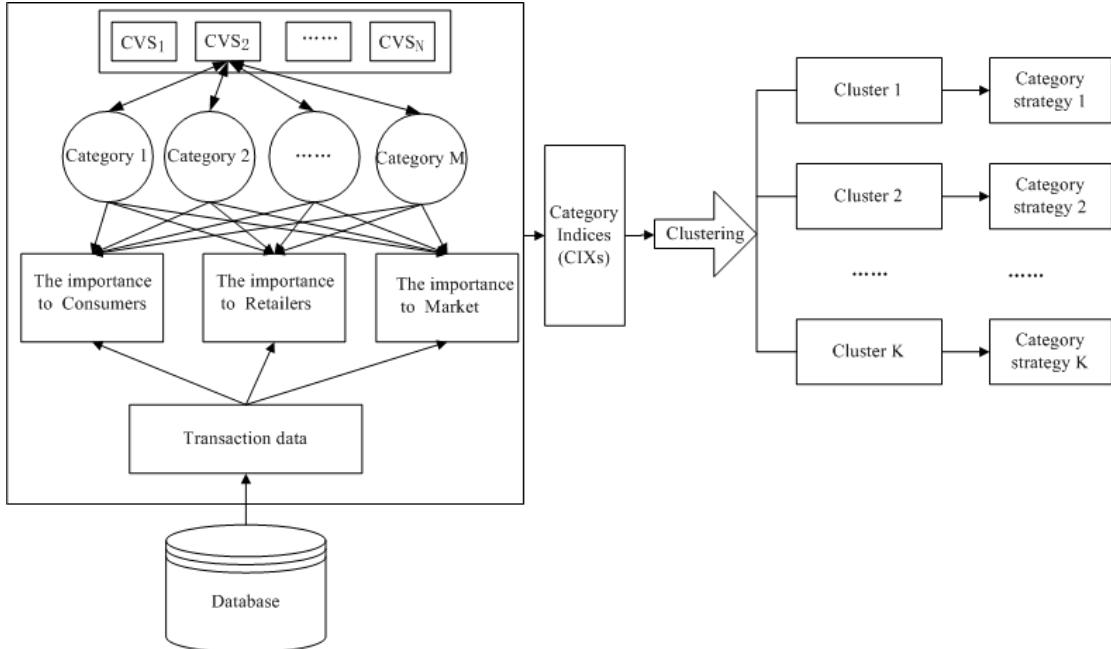


Figure 1: CR-based market segmentation model

Data and empirical results

An empirical study is conducted to verify the proposed model in this section. A dataset from a gas station convenience store in China is used. It contains transaction records that are collected from 82 petrol chain CVSs between January 2009 and June 2009 in Guangdong province, China. The dataset contains 2 category levels: 21 major categories and 95 sub-categories, and 3,456 products.

Data preprocessing

Initially, noisy, error and missing data was removed from the dataset. In this case, 7 out of 82 CVSs (which only contained transaction data in January and February) were filtered out. Meanwhile, the *Fast food* category that does not contain transaction data was also removed. Thus, a total of 20 major categories and 75 CVSs remain. The five segmentation attributes used in the proposed model (i.e. F, N, S, R and G) are derived on a monthly basis. To start with, normalization is performed on these attributes by using the Min-Max method, in which the value of each attribute is mapped onto a range of [0, 1].

The clustering analysis of CVSs market segmentation

In order to avoid biases, the entropy method is selected to determine the weights of the five attributes, and the obtained results are shown in Table 2.

Table 2: Attribute weights

Attributes	F	N	S	R	G

Weights	0.2022	0.1834	0.1870	0.2137	0.2130
---------	--------	--------	--------	--------	--------

Based on Table 2 and collected transaction data, the CIXs (which indicates the global contribution of a given CVS) of different CVSs can be calculated, by the use of Equation (1). The results are listed in Table 3.

Table 3: CIXs of different CVSs

Stores #	category1	category2	category3	...	Category20
1	0.3895	0.5226	0.4591	...	0.6574
2	0.3129	0.5377	0.4729	...	0.7216
3	0.3447	0.5383	0.4844	...	0.7256
...	
75	0.3059	0.3658	0.3356	...	0.4094

For clustering analysis, it is crucial to select the number of clusters, K. In the literature, three evaluations, namely partition coefficient(Bezdek†, 1973), XB coefficient (Xie & Beni, 1991) and fuzzy entropy(Kosko, 1986), are often employed to assess the clustering performance. For the partition coefficient, the greater index value indicates better performance, whilst for the XB coefficient and fuzzy entropy, the smaller index value indicates better performance. In this work, based on the collected dataset, these three indexes are calculated by using different K values, respectively. The obtained results are depicted in Figure 2. With the increase of K, the index value of partition coefficient gradually decreases; the index value of fuzzy entropy increases at first, reaches the peak index value when K = 3, and then slowly decreases; the index value of the XB coefficient has a downward trend when K is in [2, 4], and then tends to become stable when K is in [4, 10]. Guided by Figure 2, this work chooses K as 4 in an effort to achieve optimal clustering results.

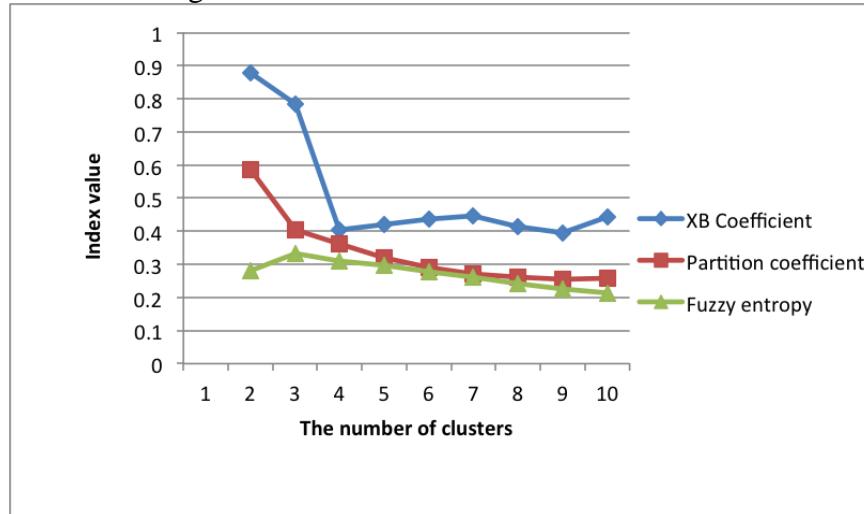


Figure 2: The changes of the indicators of effectiveness

Results and discussion

Based on Table 3 and the selected K value, we get the clustering results. The CVSs are grouped into four market segments (Cluster1, Cluster2, Cluster3 and Cluster4). The detailed results are listed in Table 4.

In terms of CIX, Cluster2 (0.4579) > Cluster1 (0.3780) > Cluster3 (0.3514) > Cluster4 (0.3364). This indicates that CVSs in Cluster2 outperform other CVSs in the market, while CVSs in Cluster4 performs the worst in the market. In addition, Cluster2 only consists of 8 CVSs, which is much less than other clusters. Therefore, further investigations on these CVSs in Cluster2 are necessary to help improve other CVSs' performance.

For category roles, the destination category achieves a high level of consistency. Except Cluster2, the destination category all consists of Drinks and Sweets categories. Since these categories contribute most towards the store performance. It is natural to pay more attention on their sub-categories and provide more variety of products for consumers to choose. Also, frequent and diverse promotion offers on these categories are necessary to gain more products for the CVSs. For Cluster2, Tobacco appears in the destination category, this is quite different from other market segments. One possible reason may be that their consumers include more smokers than other clusters. Further investigations on their smokers-consumers may help this segment to develop valuable market strategies.

The routine category mainly contains ready-to-eat categories (e.g. Sweets, Crisps/snacks, Biscuits, Milk, Hot meals, etc), this indicates the clustering results is meaningful, since the derived CR well reflects the reality. Such categories are consists of routine products in daily life. Due to the very limited storage space in chain CVSs, the included products in routine categories should be carefully selected. Performance analysis on single product should be conducted frequently, those products perform worst should be removed from the shelf. In addition, the products in routine category are often provided in other competitors as well. Therefore, when identifying the price of such products, the prices from other competitors need to taken into consideration.

In the seasonal category, the included categories are quite diverse; none of the market segments consist of the identical categories. Moreover, for these categories, the consumer demands are not stable and normally are short-term. Especially for seasonal categories, various promotion offers can be provided according to festivals or seasons. For example, in Cluster3, Ice-cream appears in the Seasonal category; therefore, for those CVSs in Cluster3, it is a good idea to provide promotions on Ice cream in summer.

The convenience category mainly contains Newspaper/magazine and Ice-cream; this once again reveals the reliability of the proposed clustering algorithm. This is because, it is well known that categories such as Newspaper/magazine and Ice-cream are low profit, but essential to consumers. The goal of these categories is to provide convenience to consumers, so that the consumers can buy all they need in the store in one go. Consumers tend to be less sensitive to price in this category, and care more about the product really meets their requirements. Therefore, promotions in convenience category may not be very attractive to consumers.

Table 4 The clustering results

Type of market	Cluster1	Cluster2	Cluster3	Cluster4
# of CVSS	23	8	23	21
Average CIX	0.3780	0.4579	0.3514	0.3364
Convenient category	Family planning, Ice cream Newspaper/magazine	Ice cream, Newspaper/magazine,Toy/gift,	Family planning,Toy/ gift, Newspaper/magazine,	Ice cream, Household good, Newspaper/magazine,
Seansoal category	Pharmacy, Wine, Toy/gift,	Wine, Family planning, Pharmacy,	Household good, Pharmacy, Ice cream	Wine, Family planning, Toy/ gift,
Rountine category	Crisps/snack, Biscuit, Hot meal, Milk, Tobacco, Engine oil, Household good, Stationery, Health & beauty, Bread, Fruit, Automobile supplies	Sweets, Crisps/snacks, Biscuit, Hot meal, Milk, Household good, Health & beauty, Bread, Engine oil, Fruit , Stationery, Automobile supplies	Crisps/snacks, Hot meal, Biscuit, Milk, Engine oil, Stationery, Bread, Tobacco, Automobile supplies, Health& beauty, Wine, Fruit ,	Crisps/snacks, Hot meal, Biscuit, Stationery , Milk, Engine oil,Health & beauty, Bread, Automobile supplies , Tobacco, Fruit , Pharmacy
Targeted category	Drinks, Sweets	Drinks, Tobacco	Drinks, Sweets	Drinks, Sweets

Conclusion

In this paper, we have proposed an innovative market segmentation model for CVSs. Starting from the three dimensions of CR, each category in different CVSs get a derived CIX (which indicates the global contribution of a given category to CVS). In addition, by the joint use of an innovative similarity measure (HCsim()) and an improved weighted fuzzy K-means clustering algorithm (WFKM), CVSs with similar CIXs are grouped into the same market segment. The applicability and utility of the proposed clustering model is demonstrated via an empirical study on a CVSs dataset provided by a gas station convenience store in China. By using the CR model, the current retail market is divided into four clusters. And they show a high level of consistency in destination category, while performance quite differences in other CRs. Moreover, the cluster results helps to define appropriate market strategies for different categories in different market segments.

Although the proposed approach is promising, much may be done through further research. Firstly, the applicability of the proposed CR model can be further improved, so that the model can be used in various application domains. Secondly, in order to better simulate the changes of consumers' demand and purchasing behavior, time series analysis will be employed to analyze historical sales data.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (Grant No. 70971112).

References

Basuroy, S., Mantrala, M. K., & Walters, R. G. (2001). The impact of category management on retailer prices and performance: Theory and evidence. *The Journal of Marketing*, 16-32.

Bezdek†, J. C. (1973). Cluster Validity with Fuzzy Sets. *Journal of Cybernetics*, 3(3), 58-73.

Dussart, C. (1998). Category management:: Strengths, limits and developments. *European Management Journal*, 16(1), 50-62.

Heller, A. (2012). *Consumer-Centric Category Management: How to Increase Profits by Managing Categories Based on Consumer Needs*: Wiley.

Kamakura, W. A., & Kang, W. (2007). Chain-wide and store-level analysis for cross-category management. *Journal of Retailing*, 83(2), 159-170.

Kosko, B. (1986). Fuzzy entropy and conditioning. *Information Sciences*, 40(2), 165-174.

Liu, Y., Kiang, M., & Brusco, M. (2012). A unified framework for market segmentation and its applications. *Expert Systems with Applications*, 39(11), 10292-10302.

Morgan, N. A., Kaleka, A., & Gooner, R. A. (2007). Focal supplier opportunism in supermarket retailer category management. *Journal of Operations Management*, 25(2), 512-527.

Shin, H. W., & Sohn, S. Y. (2004). Segmentation of stock trading customers according to potential value. *Expert Systems with Applications*, 27(1), 27-33.

Trivedi, M. (2011). Regional and Categorical Patterns in Consumer Behavior: Revealing Trends. *Journal of Retailing*, 87(1), 18-30.

Xie, X. I., & Beni, G. (1991). A validity measure for fuzzy clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(8), 841-847.

Zufryden, F. S. (1987). New computational results with a dynamic programming approach for product selection and supermarket shelf-space allocation. *Journal of the Operational Research Society*, 201-203.