STUDYING ASSOCIATIVE RELATIONSHIPS AMONG PRODUCT CLASSES IN THE CONTEXT OF WEB RETAILING

Eugenia Y. Huang
Department of MIS, National Chengchi University, Taipei, Taiwan, R.O.C.
huang.eugenia@gmail.com, +886-2-29387348
Chia-jung Tsui
Department of MIS, National Chengchi University, Taipei, Taiwan, R.O.C.
chiajung@nccu.edu.tw, +886-2-29393091 EXT 65105
Huimei Liu
Department of Statistics, National Chengchi University, Taipei, Taiwan, R.O.C.
hliu@nccu.edu.tw, +886-2-29393091 EXT 81024
Wellington K. Kuan
Department of Business Administration, National Chengchi University, Taipei, Taiwan, R.O.C.
kuan-wk@nccu.edu.tw, +886-2-29393091 EXT 81254
Hung-Shin Chen
Department of MIS, National Chengchi University, Taipei, Taiwan, R.O.C.
teiria0317@gmail.com, +886-2-29393091 EXT 65105

ABSTRACT

Most web retailers use a product class hierarchy to organize their products, with first-level product classes occupying the most prominent position in the website layout. By using the customer transaction records of a cosmetic web retailer, this study applied co-occurrence analysis to the data and revealed interesting associative relationships among the product classes. This result shows that with readily available analysis tools, web retailers can assess the effectiveness of its product class layout on web and harness cross sale opportunities with a reorganization display of product classes.

Keywords: Product Classes, Web Retailing, Co-occurrence Analysis

INTRODUCTION

Product class hierarchy is usually represented as a tree structure in which a set of products at a low level are grouped into a more general product class at a higher level. The leaves of the tree indicate physical products, while the root of the tree represents the most general product class. The nodes between the leaves and the root are product classes obtained by joining several nodes at a lower level. Figure 1 illustrates a three-level product class hierarchy of the website investigated in this study, where all products are organized according to product classes and each product class is further divided into subclasses. A structure with more than three levels is rarely seen in e-tailing websites.

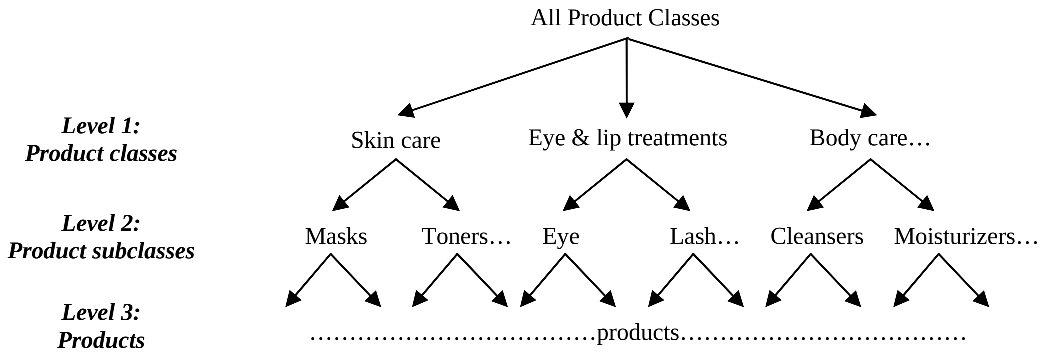

FIGURE 1. A THREE-LEVEL PRODUCT TAXONOMY

Product class hierarchy provides a clear view of product lines offered by a web retailer. In order not to clog up a webpage, only the first-level product classes are displayed upfront. Although product subclasses are often revealed in a drop-down menu when a shopper moves the mouse cursor over the product class to which the product subclasses belong, it is the first-level product classes that are most visible on a webpage and thus play an important role in guiding customers to browse a web retailer's offerings.

Currently, there are no known guidelines for the display of product classes on e-tailing websites. It seems that the display of product classes is more or less in an ad hoc fashion, and often times it is only conveniently based on the time sequence each product class became available on a website. In the case of Amazon.com, for example, the product class of books remains the first among all physical product classes on its navigation menu, despite the fact that its product lines had long expanded into many other varieties. As products proliferate, web retailers become more in need of product class presentation guidelines. For example, what is the best sequence in the navigation menu? Can some product classes be merged to form a more meaningful class? What are the relationships among a web retailer's product classes? To answer these questions, we focus on the initial step of exploring the associative relationships among the product classes of a web retailer. Headquartered in Taiwan, the case company is specialized in selling beauty products online, and its product classes have recently expanded into young fashion in the limited areas of lingerie and accessories. Established in 1999, currently producing approximately NT$2 billion annual revenue, and had entered China and Singapore markets in recent years, the company represents one of a few web retailers which survived the Internet Bubble and boded well in the future.

RELATED WORK

Various techniques can be used to discover association rules. The challenges of mining association rules were first discussed by Agrawal and his colleagues (1993) when they analyzed a large database of customer transactions. They aimed to generate all significant association rules among products that satisfy the user-specified minimum criteria. The criteria usually involve "support" and "confidence". Support is calculated as the percentage of all customers who purchased both antecedent product(s) and consequent product(s), while confidence as the percentage of customers who purchased antecedent product(s) also purchased consequent product(s). An example of an association rule is "9% of all customers buy both makeup removers and mascaras. 50% of the customers who buy makeup removers also buy mascaras."

Association rule mining can be applied to study the associative relationships of a web retailer's product classes. However, it is not appropriate due to the following reasons. First, association rule mining mainly deals with physical products rather than product classes. For example, there was much effort (Agrawal & Srikant, 1994; Liu, Zhai, & Pedrycz, 2012) addressing the scalability issue in the discovery of *all* association rules for a large number of physical products. Second, even if there was rare attempt applying associations mining to product classes (Lawrence, Almasi, Kotlyar, Viveros, & Duri, 2001), a long list of rules discovered nonetheless makes it difficult to interpret.

In order to address these issues and to study the associative relationships among product classes in a less complicated way, this study attempted co-occurrence analysis. Co-occurrence analysis, to be described in more detail later, was shown in literature (Spence & Owens, 1990; Tsui, Wang, & Fleischmann, 2011) to be effective in exploring complex relationships among multiple objects. In the following sections, we first describe how we applied co-occurrence analysis on the transaction records of the web retailer. Then, we report the results and demonstrate that complex relationships can be clearly represented on a diagram. Finally, we conclude by discussing the contributions of our study to the theoretical development and practices of retailing website organization.

METHODOLOGY

Data

This study utilizes sample data from the web retailer, which offers a comprehensive product line of beauty products. The data include transaction records of 1,832 customers, which were randomly selected. Among them, over 90% are female and 70% are in their 20s. We used the transaction records from 2005 to 2009 to study the associative relationships of the web retailer's first-level product classes.

There were seven product classes offered by the web retailer before 2007. During 2007 and 2008, three more product classes were added. Table 1 shows the total ten product classes. The numbers before the product classes indicate the sequence from left to right displayed horizontally on the retailer's website. The three new product classes are the last three (c08–c10) indicating that they were simply appended to the right side of existing product classes without considering a possible better presentation of the ten product classes on the website.

In order to study their relationships, to see how new classes associate with the previous classes, and to identify the potential changes of relationships incurred by the three new product classes, the data were divided into two periods: 2005–2006 and 2008–2009, so that the first period contains the seven product classes and the later period includes the seven product classes plus the three new ones.

TABLE 1. THE TEN PRODUCT CLASSES

Product class
c01.Skin Care
c02.Eye & Lip Treatments
c03.Body Care
c04.Makeup
c05.Fragrance
c06.Hair Care
c07.Aromatherapy
c08.Medical Cosmetic
c09.Lingerie
c10.Accessories

Co-occurrence Analysis

Co-occurrence of words has been used in various fields, such as computational linguistics (Burgess & Lund, 1997) and information retrieval (Smadja, 1993), to study the relationships among words. For example, Spence and Owens (1990) used co-occurrence to evaluate the strength of word association. They found that related pairs of nouns co-occur significantly more often than unrelated pairs. Their finding suggests that co-occurrence frequency indicate the strength of association.

Co-occurrence analysis is also employed to uncover the structure of specific academic fields. Objects reflecting unique characteristics in an academic field, such as keywords, classification codes, or authors, are analyzed and mapped onto a two-dimensional space in order to represent the structure of the field. For example, Ding, Chowdhury, and Foo (2001) utilizes the co-occurrence frequency of keywords, provided by the SCI and SSCI databases, to study the structure of the field of Information Retrieval during the period of 1987–1997. Another stream of research, called author cocitation analysis (ACA), observes co-occurrence frequency of authors in a paper's references to identify specialties within a discipline. For example, White and McCain (1998) applied ACA to visualize specialties of information science discipline on a two-dimensional map.

In co-occurrence analysis, the classification step usually utilizes multidimensional scaling (MDS), a multivariate data analysis. Other multivariate data analyses such as factor analysis and cluster analysis were also used. In this paper, we use both hierarchical clustering and multidimensional scaling, as previous research (Kruskal, 1977) found that applying clustering and MDS separately to the same data results in data representations that together offer greater insight into the structure underlying the data and can detect more subtle relationships than either analysis used alone. The difference between the two is that MDS provides a spatial representation of the data, while hierarchical clustering provides a tree representation.

In this paper, co-occurrence analysis is applied to the transaction records of individual customers to study the relationships of the product classes. As the data are separated into two periods, we construct a co-occurrence matrix with each row and column representing a product class for each period. For the period 2005–2006, there are seven product classes so the constructed co-occurrence matrix is seven by seven. For the other period 2008–2009, the number of product classes was increased to ten, so the matrix is ten by ten. The value in each cell of a matrix represents the number of customers who purchased from the respective pair of product classes during that period. We then apply hierarchical clustering and MDS separately to each co-occurrence matrix in order to visualize the structure of the data and to explore the relationships among the product classes for the two periods.

RESULTS

Our hierarchical clustering on the co-occurrence matrix for the first period 2005–2006 produces a dendrogram (Figure 2), where vertical lines show joined clusters and the position of the lines on the scale from 1 to 25 indicates the distance at which clusters are merged. By inspecting the dendrogram, we identify two broad clusters. Table 2 summarizes the membership of the two clusters. Cluster 1 includes four product classes, which are skin care, makeup, hair care, and body care, while Cluster 2 consists of three product classes, which are fragrance, aromatherapy, and eye & lip treatments. By and large, Cluster 1 is more of body care, while Cluster 2 is about scents and other treatments.

Subsequently, multidimensional scaling on the same matrix generates a two-dimensional plot with Stress of .08. Stress is a “badness-of-fit” measure and the value .08 is considered as good to fair (Kruskal, 1964). We further implement the clustering schedule obtained via the hierarchical clustering result within the MDS plot (Figure 3). In general, product classes that are deemed similar (joined earlier in the clustering schedule) are also located close to each other on the MDS plot.

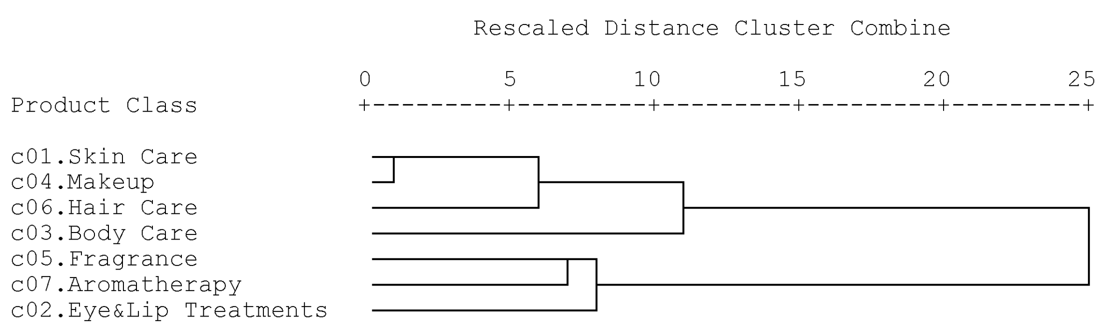

FIGURE 2. THE CLUSTERING RESULT OF THE SEVEN PRODUCT CLASSES FOR 2005–2006

TABLE 2. MEMBERSHIP OF THE TWO CLUSTERS FOR 2005–2006

Cluster	Product classes
1	c01.Skin Care c04.Makeup c06.Hair Care c03.Body Care
2	c05.Fragrance c07.Aromatherapy c02.Eye & Lip Treatments

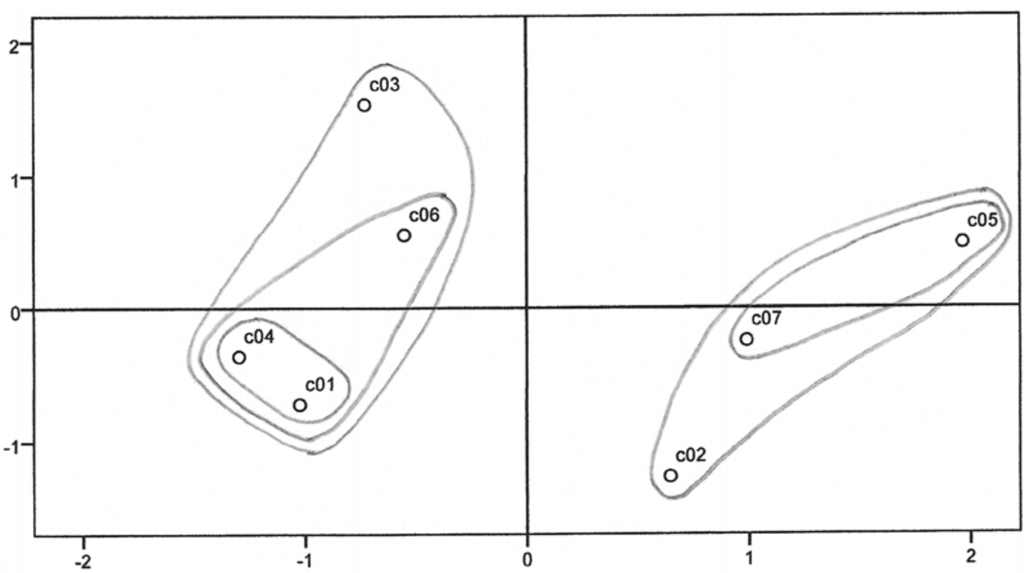

FIGURE 3. THE MDS RESULT WITH THE CLUSTERING SCHEDULE FOR 2005–2006

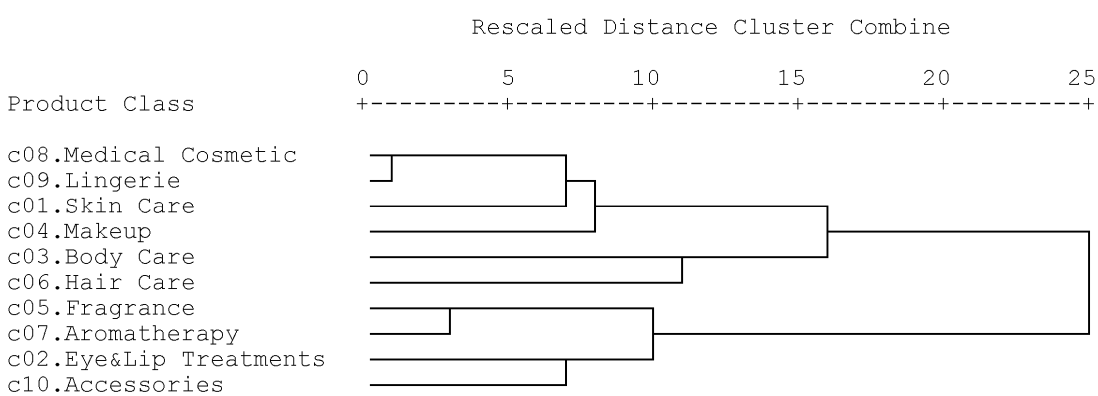

FIGURE 4. THE CLUSTERING RESULT OF THE TEN PRODUCT CLASSES FOR 2008–2009

The above analysis procedure is repeated for the second period, 2008–2009. First, the hierarchical clustering on the co-occurrence matrix produces a dendrogram (Figure 4). Two broad clusters can also be identified by inspecting the dendrogram. Table 3 summarizes the membership of the two clusters where the three new product classes (c08–c10) are highlighted in bold. Then, MDS on the same matrix generates a two-dimensional plot with Stress of .14, which is considered acceptable. We further implement the clustering schedule obtained via the hierarchical clustering result within the MDS plot (Figure 5).

TABLE 3. MEMBERSHIP OF THE TWO CLUSTERS FOR 2008–2009

Cluster	Product classes
1	**c08.Medical Cosmetic** **c09.Lingerie** c01.Skin Care c04.Makeup c03.Body Care c06.Hair Care
2	c05.Fragrance c07.Aromatherapy c02.Eye & Lip Treatments **c10.Accessories**

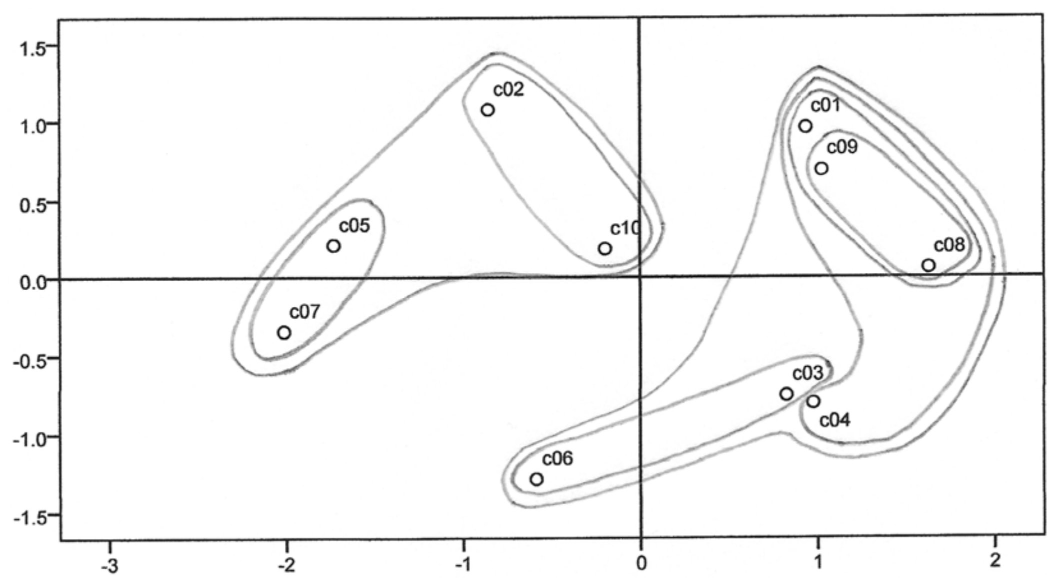

FIGURE 5. THE MDS RESULT WITH THE CLUSTERING SCHEDULE FOR 2008–2009

Comparing Table 2 and Table 3, we can tell that the relationships among the seven product classes remain the same across the two time periods. Skin care, makeup, hair care, and body care are in Cluster 1 while fragrance, aromatherapy, and eye & lip treatments in Cluster 2. As for the three new product classes, which are highlighted in bold in Table 3 of the second period, medical cosmetics and lingerie are grouped with Cluster 1, and accessory is with Cluster 2. Currently, on the e-tailer's website, these three new product classes are being displayed together by simply appending to the previous product classes. Our result suggests that a reorganization display of product classes may be necessary.

DISCUSSION AND CONCLUSION

This paper applies co-occurrence analysis to the customers' transaction records in order to discover the associative relationships among the product classes. Such discovery is essential because when product classes tend to be purchased together by more customers, there are not only opportunities to optimize the display of product classes on websites, but also opportunities for product recommendation, promotion, extension, and cross-marketing. The association of product classes is based upon actual purchase patterns. Thus, marketing attempts among the product classes that show closer associative relationships should be more effective than among those that are more distant. Although the immediate contribution of the discovery of associative relationships lies in redesigning website product class displays, its impact is actually more profound when integrated into product recommendations and cross-selling.

Although our study is based upon the sample transaction records of a particular web retailer, it effectively demonstrates a promising approach in discovering the overall associative relationships among first-level product classes and shows the value of such discovery. It will be interesting to follow up with an experiment to redesign the presentation of the product classes on the retailer's website based on our results and evaluate its performance. We have little doubt of its applicability to a wider range of product types and different web retailers. Yet, universal applicability requires further studies using data from other web retailers encompassing various product types.

ACKNOWLEDGEMENT

This research is partially funded by National Science Council in Taiwan under project number NSC 100-2410-H-004-026-MY2 and a grant from National Chengchi University, Taiwan, R.O.C.

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993). *Mining Association Rules between Sets of Items in Large Databases.* Paper presented at the ACM SIGMOD Conference on Management of Data, Washington, DC.

Agrawal, R., & Srikant, R. (1994). *Fast Algorithms for Mining Association Rules.* Paper presented at the 20th VLDB Conference, Santiago, Chile.

Burgess, C., & Lund, K. (1997). Modelling Parsing Constraints with High-dimensional Context Space. *Language and Cognitive Processes, 12*(2-3), 177-210. doi: 10.1080/016909697386844

Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management, 37*(6), 817-842.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1-27.

Kruskal, J. B. (1977). The relationship between multidimensional scaling and clustering. In J. Van Ryzin (Ed.), *Classification and Clustering* (pp. 17-44). New York: Academic Press.

Lawrence, R. D., Almasi, G. S., Kotlyar, V., Viveros, M. S., & Duri, S. S. (2001). Personalization of Supermarket Product Recommendations. *Data Mining and Knowledge Discovery, 5*(1), 11-32. doi: 10.1023/a:1009835726774

Liu, X., Zhai, K., & Pedrycz, W. (2012). An improved association rules mining method. *Expert Systems with Applications, 39*(1), 1362-1374. doi: 10.1016/j.eswa.2011.08.018

Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, *19*, 143-177.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research, 19*(5), 317-330.

Tsui, C.-j., Wang, P., & Fleischmann, K. R. (2011). *Evaluating two automatic methods for classifying information technology concepts.* Paper presented at the International Conference on Information Systems (ICIS), Shanghai, China.

White, H. D., & McCain, K. W. (1998). Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science, 49*(4), 327-355.