Association for Information Systems
AIS Electronic Library (AISeL)

Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics

Special Interest Group on Decision Support and Analytics (SIGDSA)

Customer Visit Segmentation Using Market Basket Data

Anastasia Griva
Department of Management Science and Technology, Athens University of Economics and Business, Athens Greece,
an.griva@aueb.gr

Katerina Pramatari
Athens University of Economics & Business, k.pramatari@aueb.gr

Cleopatra Bardaki
ELTRUN, Department of Management Science and Technology, Athens University of Economics & Business, Greece,
cleobar@aueb.gr

Follow this and additional works at: http://aisel.aisnet.org/sigdsa2016

Recommended Citation

Griva, Anastasia; Pramatari, Katerina; and Bardaki, Cleopatra, "Customer Visit Segmentation Using Market Basket Data" (2016).
Proceedings of the 2016 Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics. 6.
http://aisel.aisnet.org/sigdsa2016/6

Customer Visit Segmentation
Using Market Basket Data

Anastasia Griva, Cleopatra Bardaki, Katerina Pramatari

ELTRUN, E-Business Research Center, Department of Management Science & Technology, Athens University of Economics & Business, Athens Greece
47A Evelpidon St. & 33 Lefkados St., 113 62, Athens, Greece

an.griva@aueb.gr, k.pramatari@aueb.gr, cleobar@aueb.gr

Extended Abstract

Basket analytics is a powerful tool in the retail context for acquiring knowledge about consumer shopping habits and preferences. In this paper, we propose a clustering-based artifact that mines customer visit segments from basket sales data. Current research on *customer segmentation* utilizes the *complete purchase history (all shopping visits) per customer* to identify customer groups (e.g. Boone and Roehm 2002; Liao and Chen 2004; Miguéis et al. 2012; Park et al. 2014). They see shoppers as a bulk of all the products they have purchased, regardless of whether this took place in one or more visits and try to segment shoppers based on their behavior. Other researchers utilize market basket analysis to examine the *associations between the products/ items* purchased during a shopper's single visit (e.g. bread → milk), i.e. they look for answers to questions such as 'which products are bought together' (e.g. Agrawal et al. 1993; Tang et al. 2008; Cil 2012). The aforementioned studies overlook the shopping purpose of a single customer visit, because they either examine the entirety of a customer's shopping visits or focus on the association between specific products that customers buy during a visit. However, marketing researchers, who talk about different shopping trip types (e.g. fast refilling trip or major monthly trip) (Bell et al. 2011; Walters et al. 2003), as well as practitioners that coined the term "*shopping mission*" to refer to the intention behind a shopper's visit (ECR Europe, 2010), have stressed the need to comprehend each single customer visit.

Adopting the "Design Science" approach (Hevner et al. 2004), we develop an artifact that employs data mining techniques (clustering) to accomplish segmentation and characterization of shoppers' visits, by examining the product categories a customer purchases during a visit in a physical or web store. The proposed business analytics approach utilizes *market basket data* and performs *customer visit segmentation* by looking at the patterns derived by the purchased product categories within a shopping basket, with a view to give a characterization to this visit. Specifically, the artifact examines the purchased product categories within each shopping basket and generates segments of customer visits. For example, when a basket contains mostly powders, dish-washing, bathroom cleaners, paper rolls, shampoos, body creams, oral care etc., then this basket is classified in the 'detergents and hygiene' visit segment. First, our approach adjusts the product categories tree and the input data and then performs clustering to discover shopping patterns and different segments of customer visits. With a typical retail store, having more than 10,000 SKU's in its assortment, it is rather impossible to identify common patterns at an SKU level and working at a higher level of analysis is required in order to avoid data sparsity problems. We claim that identifying the right product category level, i.e. the right level of analysis in the product taxonomy tree, is crucial to the results of the study. Besides, the main store retail activities (e.g. store replenishment, shelf space allocation, product assortment selection) and the relevant decisions mainly refer to product categories, as the shopper needs are often expressed at the category level (e.g. 'I need to buy milk') rather than at a specific SKU level (e.g. 'I need to buy this specific milk in a 250ml bottle'). In addition, by working at the product category level we ensure that the results are more generic and may also apply to new products of a category.

We demonstrate the utility of the artifact by applying it to a real case of a major fast-moving consumer goods (FMCG) retailer in Greece, in terms of both turnover and number of stores. More specifically, the retail chain has provided point-of-sale (POS) data, from January 2012 to May 2013, from eight representative stores in the Attica region. The stores had common characteristics in pairs; two convenience stores, two supermarkets, two mini-hyper markets and two hypermarkets. We analyzed 36.797.639 records that correspond to 3.973.215 distinct baskets/store visits, and some visits were associated with a cardholder id. Hence, we could identify all the baskets a shopper had purchased through this year-and-a-half history. We analyzed each store separately. Indicatively, in a supermarket store, the data mining results identified shoppers visiting the store to purchase products for "meal preparation". Indeed, these baskets contained product categories related to fresh vegetables, red meat, chicken, white cheese, pasta, eggs, bread, oil, vinegar etc. Alike, we identified baskets containing categories such as milk, baked goods, juice, coffee, tea, cereals, and oral care products. Thus, we can infer that here we have store visits that are related to "breakfast". We noticed in this segment that shoppers purchase during the same visit products that they eat for breakfast (e.g. coffee, cereals etc.) and those that are related with a morning habit i.e. "to wash their teeth" (oral care products). This outcome reveals hidden shopper behavior insights that could be exploited for marketing purposes. Similarly, we spotted shoppers visiting the store to buy their "snacks and beverages". Moreover, by examining the days that shoppers visit the store for "snacks and beverages", we found that Friday and Sunday evening are the prevailing days for this type of visit. In the same spirit, we also identified some other shopping purposes such as visits to purchase "detergents and hygiene" products like powders, dish washing, bathroom cleaners, paper rolls, shampoos, body creams, oral care etc. Other shopping purposes that we identified were visits that contained products to prepare a "toast with packed products", as the dominant categories are packed cheese, packed cold cuts, and packed bakery products. In addition, we found out that shoppers enter the store to purchase "toast with serviced products", to buy products to prepare a "light meal" and also to make stock-out visits.

Apart from its apparent theoretical contribution, the proposed approach extracts knowledge that may support several decisions to satisfy the specific needs and preferences of different shoppers. On the one hand, the proposed approach can evolve to a tool for designing new marketing campaigns and promotions for products that share the same shopping visit segment. On the other hand, the behavioral segmentation and characterization of customer visits may prescribe a new redesigned store layout where products of the same segment are positioned in nearby store's aisles and shelves. Moreover, this customer segmentation approach may be the foundation stone of a recommendation system for real time purchases in retail stores and other decision making processes in a retail store. Apart from gaining knowledge to design marketing actions, the store manager could further utilize the extracted knowledge to reengineer the replenishment strategies by ordering groups of products based on the identified visit segments.

Further research may address some limitations of this study. We can use more complex interaction data derived from alternative technological means (e.g. RFID, beacons) from other retail contexts to evaluate and validate the proposed approach. For instance, data that indicate the products a customer puts in his RFID-enabled shopping cart during a shopping visit in a grocery store. It would also be a challenge to use different interaction data of the same retailer and compare the results. For instance, we can examine the visit segments derived via combining different interaction data of the same retailer. In addition, we can identify the selling gaps via comparing the visit segments stemming from data of products in the shopping carts and products that are finally purchased. Finally, since omnichannel retailing is emerging, future research could compare the different customer visit segments resulting from the data of a physical and a web store. As another example, following this approach one could analyze the data captured from a fitting room in a fashion retail store (enabled with RFID-technology) to compare product preferences with actual product purchases. All in all, new data capture capabilities enabled by ubiquitous technologies and the internet-of-things, paired with emerging customer behaviors open-up a new challenging field for researchers and practitioners alike to apply the findings of this and pertinent research for better understanding and serving the customer

Keywords: Customer visit, Customer Segmentation, Retail Analytics, Shopper Behavior, Clustering

References

Agrawal, R., Imieliński, T. and Swami, A. 1993. "Mining association rules between sets of items in large databases", in *ACM SIGMOD international conference on Management of data*, Washington DC,

New York, USA, pp. 207-216.

Bell, D. R., Corsten, D., and Knox, G. 2011a. "From Point of Purchase to Path to Purchase: How Preshopping Factors Drive Unplanned Buying," *Journal of Marketing* (75:1), pp. 31–45.

Boone, D. S., and Roehm, M. 2002. "Retail segmentation using artificial neural networks," *International Journal of Research in Marketing* (19:3), pp. 287–301.

Cil, I. 2012. "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Systems with Applications* (39:10), pp. 8611–8625.

ECR Europe 2010. "The Consumer and Shopper Journey Framework". https://www.ecrireland.ie/uploadedfiles/shopper/projects-ecr-eu/the-consumer-and-shopper-journey-framework-.pdf

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS Quartely* (28:1), pp. 75–105.

Liao, S., and Chen, Y. 2004. "Mining customer knowledge for electronic catalog marketing," *Expert Systems with Applications* (27:4), pp. 521–532.

Miguéis, V. L., Camanho, A. S., and Falcão e Cunha, J. 2012. "Customer data mining for lifestyle segmentation," *Expert Systems with Applications* (39:10), pp. 9359–9366.

Park, C. H., Park, Y.-H., and Schweidel, D. A. 2014. "A multi-category customer base analysis," *International Journal of Research in Marketing* (31:3), pp. 266–279.

Tang, K., Chen, Y. L., and Hu, H. W. 2008. "Context-based market basket analysis in a multiple-store environment," Decision Support Systems (45:1), pp. 150–163.

Walters, Rockney G and Maqbul Jamil (2003), "Exploring the Relationships between Shopping Trip Type, Purchases of Products on Promotion, and Shopping Basket Profit," *Journal of Business Research*, 56 (1), 17–29.