



18th International Conference on Knowledge-Based and Intelligent
Information & Engineering Systems - KES2014

An analysis of the effects of customers' migratory behavior in the inner areas of the sales floor in a retail store on their purchase

Yoshihiro Ohata^a, Asako Ohno^{b,*}, Takahiro Yamasaki^b, Kin-ichiroh Tokiwa^b

^aGraduate School of Engineering, Osaka Sangyo University, Nakagaito 3-1-1, Daito, Osaka 574-8530, Japan

^bFaculty of Engineering, Osaka Sangyo University, Nakagaito 3-1-1, Daito, Osaka 574-8530, Japan

Abstract

Recently, the analysis of in-store customer behavior has been garnering increasing attention as a marketing strategy of retail stores. In this paper, we focus on the effects of customers' migratory behavior in the inner areas on the resulting purchase amount. We examine the difference between the shopping path of high-volume customers and that of low-volume customers in the inner side of the sales floor to investigate how purchase amount differs when the shopping path contains a longer migration in the inner areas or a larger number of transitions from the outer areas to the inner areas.

© 2014 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of KES International.

Keywords: shopping path analysis; migratory behavior; retail stores; marketing strategies; sales floor; inner area; decision tree

1. Introduction

Price competition and layout optimization have been implemented as marketing strategies in retail stores to increase sales. Further, the analysis of in-store customer behavior has received increasing attention in recent years.

For example, there are some studies aiming to investigate how customers' behaviors affect their purchases by analyzing customers' shopping path and POS data. Customers' paths are generally recorded by using Radio Frequency Identification (RFID), which is a wireless identification system that uses radio wave receivers and IC tags attached to shopping carts.

In this study, we divide the sales floor of retail stores into an inner side and an outer side, and focus on the shopping path in the inner side of the sales floor. We examine the difference between a high-volume (HV) and low-volume (LV) customer's shopping path on the inner side of the sales floor, and utilize this knowledge to investigate how the customer's behavior affects purchases.

Further, we build a decision tree, the explanatory variables of which denote customer behavior on the inner side of the sales floor and the objective variable of which is one of two customer types: HV or LV. With the decision tree,

* Corresponding author. Tel: +81-72-875-3001; Fax: +81-72-870-8189.

E-mail address: ohno@eic.osaka-sandai.ac.jp

we examined whether we can discern an LV or HV customer path without using their respective purchase amounts. The paper is organized into four sections as follows. Section 2 discusses the background and extant research. Section 3 provides a detailed explanation of our approach and the results of our analyses. Finally, Section 4 summarizes our results.

2. Background

Prior to the introduction of shopping path analysis, point of sales (POS) data analysis and market basket analysis had been the mainstay of customer behavior analysis practices. However, it is difficult to predict or anticipate the purchase decision making process of customers by using approaches that only analyze purchase results. Therefore, a method focusing more on a customer’s shopping behavior than his/her purchase result, using RFID, has been receiving more attention recently.

RFID technology is a technique that exchanges information via wireless communication between RFID receivers and RFID tags containing identification information. To record the shopping path, RFID tags are attached to the shopping carts and RFID receivers are placed on the display shelves as shown in Figure 1. In this way, data such as customer ID, time, and location information are stored in the database for each round made by the customer.

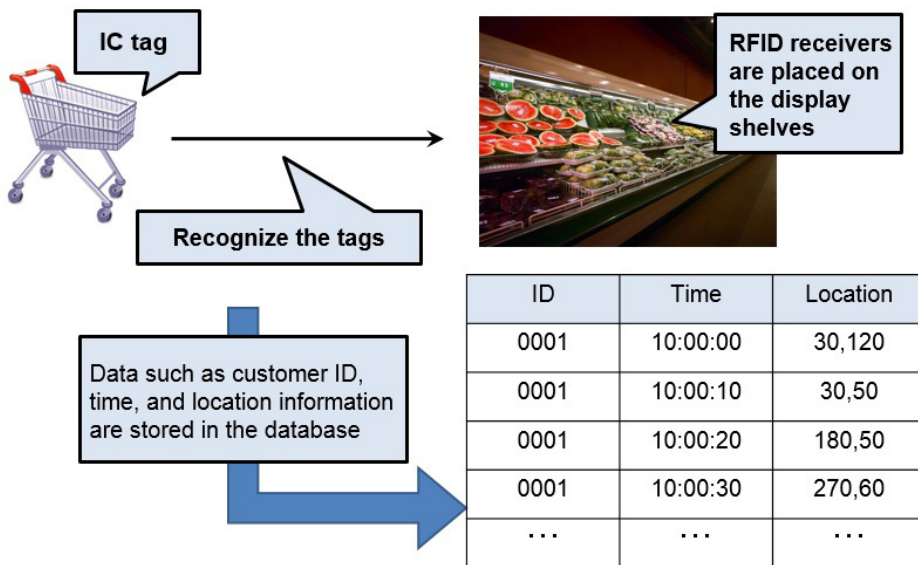


Fig. 1. RFID system

Further, organizing a series of location information into a string enabled us to apply various kinds of string analyses to the shopping path¹.

Hui et al.² decomposed customers’ shopping path into some components: the optimal shortest path calculated by applying traveling salesman problem and the deviations from the optimal path according to travel distance and traveling order to explain various characteristics of the customers’ purchase behavior.

Fujino et al.³ used POS data and RFID data to analyze and investigate the relationship between purchase quantity and customer behavior. They reported that in many cases, customer paths that include more areas located in the central part of the sales floor result in higher purchase quantities.

3. Our approach

3.1. Overview

Our approach focuses on the migratory behavior within a shopping path. We divide the sales floor of a retail store into its inner and outer side. Further, we classify customers as high-volume (HV) and low-volume (LV) customers according to the purchase amounts of their shopping paths.

We examine the differences between the shopping paths of HV and LV customers on the inner side of the sales floor, and utilize this knowledge to investigate how the customers' behaviors affect their purchases. Here, the term "purchase" can be divided into the following subclasses: purchase amount and quantity within all areas, only inner areas, and only outer areas.

Further, we build a decision tree, the explanatory variables of which denote customer behavior on the inner side of the sales floor and the objective variable of which is one of two customer types: HV or LV. Using the decision tree, we examine whether we can discern an LV or HV customer path without using their purchase amounts.

Figure 2 shows the store map used in this study. The sales floor consists of a number of areas, each of which sell different categories of products, such as "Vegetable" and "Fish." Prior to the analysis, we divide the sales floor of a retail store into the inner and outer sides. Generally, fresh foods are placed on the outer side, and processed foods are placed on the inner side of the sales floor. Tables 1 and 2 show the names of the areas and the English characters representing each of the corresponding areas. For instance, "F" represents "Fish" while "B" denotes "General foods." To treat a customer's behavior as a string, we represent the sales areas in Figure 2 as its corresponding English characters shown in Tables 1 and 2.

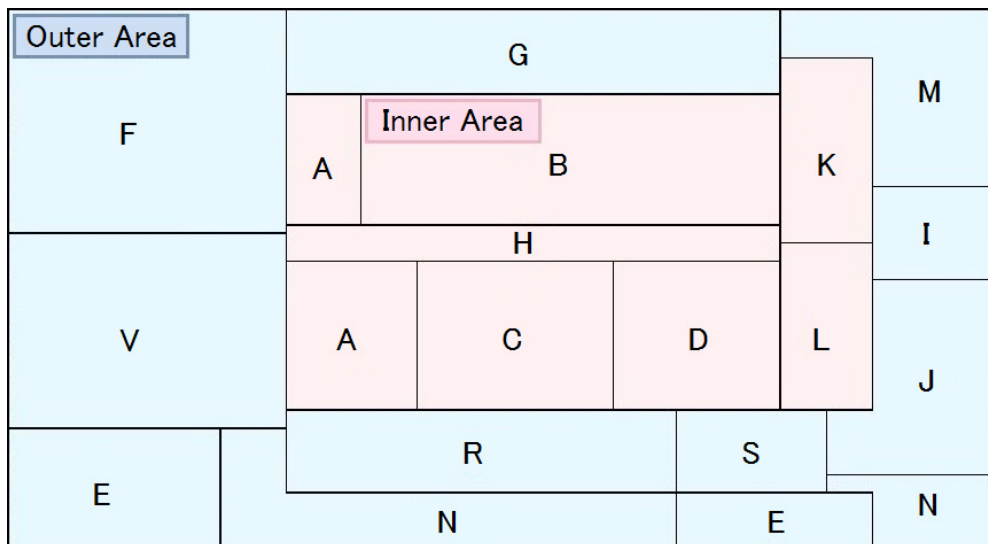


Fig. 2. Store map

3.2. Test data

In this study, we use the shopping path and purchase data recorded in a retail store in Japan between May 11 and June 15, 2009. For data cleansing before the analysis, we compare the shopping path against the corresponding purchase data to check if all the areas contained in the purchase data are also present in the shopping path. We exclude the corresponding data pair from the experimental data if there are any areas contained in the purchase data but not in the shopping path.

Table 1. Inner areas

| Area | Description |
|------|--|
| A | Non-Food Items |
| B | General Foods (Canned, Boxed, Jarred Foods, and Condiments) |
| C | Snacks |
| D | Alcohol |
| H | Center Isle |
| K | Frozen Food |
| L | Drinks/Beverages |

Table 2. Outer areas

| Area | Description |
|------|-----------------------|
| E | Entrance |
| F | Fish |
| G | Grocery |
| I | Western Dairy |
| J | Japanese Dairy |
| M | Meat |
| N | Non |
| R | Register |
| S | Sale |
| V | Fruits and Vegetables |

In this manner, we obtained 5,950 shopping paths with their corresponding purchase data, of which we used the top 100 and the bottom 100.

3.3. Assay method

In this study, we define a customer's consecutive visits to areas belonging to the inner side of the sales floor as a "migration." Prior to the analysis, we performed procedures to extract substrings that denoted migrations from a shopping path.

We substituted every English letter representing one of the outer areas shown in Figure 2 with the symbol "o" in a shopping path. Then, we substituted consecutive "o"s with a single 'o.'

This process is followed because we use the symbol "o" not to represent customer's visits to the outer area but as delimiters to extract migration substrings from the shopping path. Further, we define a substring between a pair of "o"s as a "migration." The length of a migration substring is defined as "migration length" and the number of the substrings extracted from a shopping path is defined as "migration quantity." In this paper, we aim to explain the relationships between a customer's migratory behavior and the resulting purchase by using the following four variables: migration length, migration quantity, purchase amount, and purchase quantity.

As the first step in the analysis, we constructed scatter plots with the purpose of investigating the relationships among the four variables for both HV and LV customers. To create the scatter plots, we normalized the four variables. Applying normalization enables us to plot the four variables using the same criteria, that is, above or below average, to identify and categorize HV and LV customers based on the pattern of their migratory behaviors.

To investigate the relationships between customer behavior and the purchase in greater detail, we create graphs using variables such as customers' visiting rates or purchase rates for each of the areas.

We use WEKA (Waikato Environment for Knowledge Analysis)⁴ to build a decision tree, the explanatory variable of which is migration quantity/length and the objective variable of which is one of the two customer types: HV or LV.

3.4. Results

Figures 3 and 5 show an example of the scatter plots. The horizontal axis of the graph depicts the normalized migration quantity of a single purchase, and the vertical axis depicts the normalized purchase amount; the shaft axes show the averages of each factor.

Figures 4 and 6 show the area-by-area purchase rates. The top-right graph in Figures 4 and 6 shows the purchase rates that are extracted from the customers whose migration quantity and purchase amount are above average and who belong to the first quadrant in Figures 3 and 5. Similarly, the top-left graph in Figures 4 and 6 shows the purchase rates that are extracted from the customers whose migration quantity is below average and purchase amount is above average, and who belong to the second quadrant in Figures 3 and 5.



Fig. 3. Migration-quantity and purchase amount (HV)

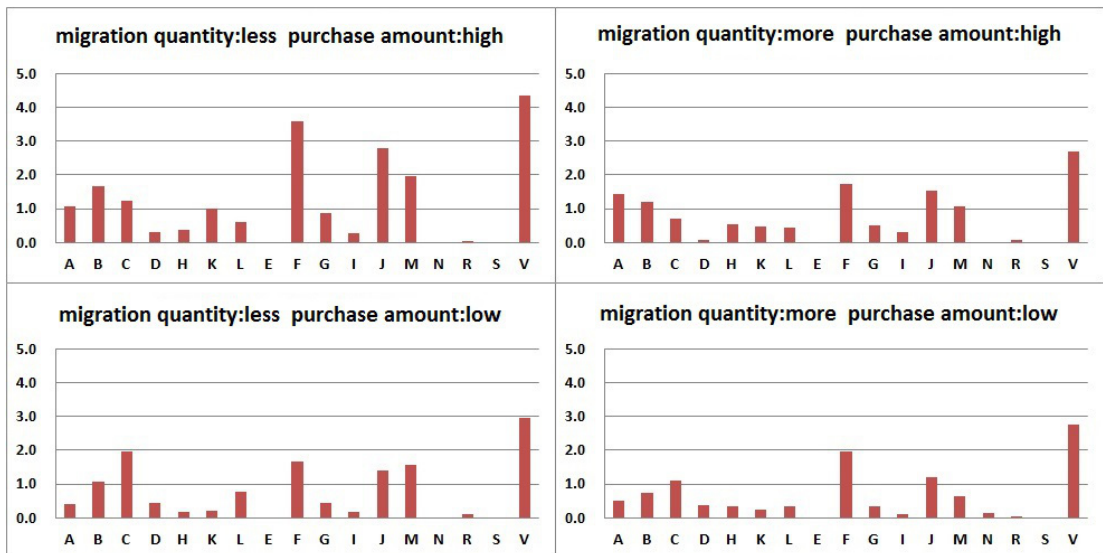


Fig. 4. Area-by-area purchase rate (HV)

Figure 7 shows the decision tree that we build. The nodes that denote branch conditions are represented as circles, and the squares denote the classified groups.

There is a pair of numbers in parentheses in each of the sections: the number on the left side represents the number of data items that met a branch condition, whereas the other number represents the number of misclassified data items.



Fig. 5. Migration-quantity and purchase amount (LV)

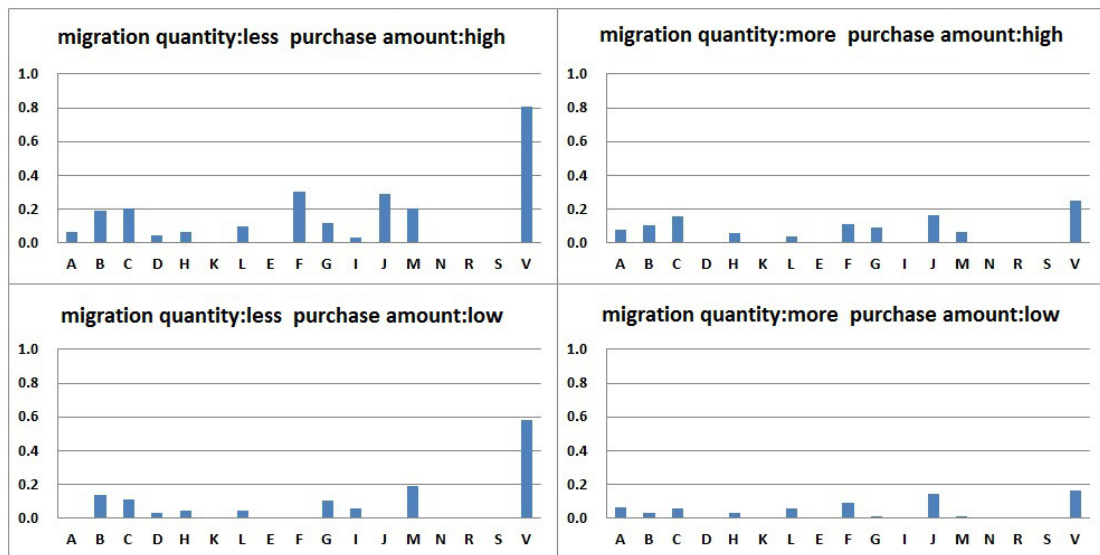


Fig. 6. Area-by-area purchase rate (LV)

3.5. Discussion

We examined if HV and LV customers have distinct trends or share a common habit. We summarized the behavioral tendencies of HV and LV customers as follows.

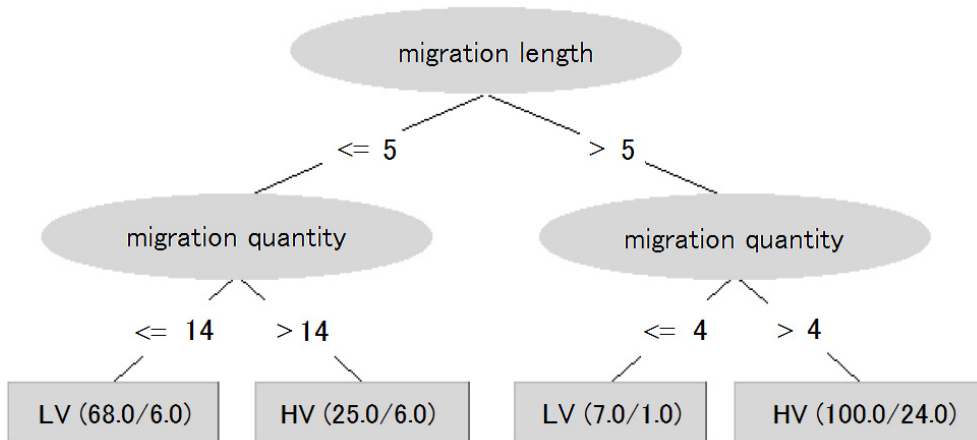


Fig. 7. An example of the trained decision tree to identify and categorize HV and LV customers

[HV]

- The area visited the most by HV customers is Area B.
- HV customers buy two or more goods in Area V every time they visit this area.

[LV]

- LV customers with a low migration quantity show a higher purchase rate.
- As the migration length shortens, the resulting purchase rate on the inner sides increases. However, the purchase rate for the whole sales floor has no relation with the migration length.

Further, regardless of HV or LV, customers with a low migration quantity tend to possess a higher purchase rate.

As for customer classification using the decision tree, we were able to classify 77.5% of the customers correctly when using migration quantity and migration length as explanatory variables.

4. Conclusion

We analyzed the customers' shopping paths in the inner side of the sales floor and investigated how the behaviors of HV and LV customers affect their purchases. We found that customer behavior on the inner side of the sales floor differed between HV and LV customers. In contrast, the purchase rates of both HV and LV customers tend to decline when they show migratory behaviors on the inner side of the sales floor. Further, by using the migration length and migration quantity extracted from the shopping paths as the explanatory variables of a decision tree, the tree distinguished between HV and LV customers with an accuracy of about 80%.

Acknowledgements

The authors would like to thank Prof. Yada for providing the data used in this study.

References

1. Yada K. Path analysis in a supermarket and string analysis technique. *Proceedings of the Institute of Statistical Mathematics*. 2008; 56: 199–213 (in Japanese).
2. Hui SK, Fader P, Bradlow E. The traveling salesman goes shopping: the systematic deviations of grocery paths from TSP-optimality. *Marketing Science*. 2009; 28:3 566–572.
3. Fujino T, Kitazawa M, Tahahashi M, Yamada T, Yamamoto G, Yoshikawa A, Terano T. Simulating customers' behaviors in a supermarket. *Proceedings of the 3rd Symposium of Technical Committee on Social Systems*. 2013; 3: 125–128 (in Japanese).
4. Weka 3: data mining software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>, Retrieved: June 1, 2014.