ソシオネットワーク戦略ディスカッションペーパーシリーズ　ISSN 1884-9946
第 52 号　2017 年 11 月
RISS Discussion Paper Series
No.52　November, 2017

A Framework of ASP for shopping path analysis

Katsutoshi Yada, Kei Miyazaki, Keiji Takai, Kohei Ichikawa

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

The Research Institute for Socionetwork Strategies,
Kansai University
Joint Usage / Research Center, MEXT, Japan
Suita, Osaka, 564-8680, Japan
URL: http://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

A Framework of ASP for shopping path analysis

Katsutoshi Yada, Kei Miyazaki, Keiji Takai, Kohei Ichikawa

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

The Research Institute for Socionetwork Strategies,
Kansai University

Joint Usage / Research Center, MEXT, Japan

Suita, Osaka, 564-8680, Japan

URL: http://www.kansai-u.ac.jp/riss/index.html

e-mail: riss@ml.kandai.jp

tel. 06-6368-1228

fax. 06-6330-3304

A Framework of ASP for shopping path analysis

Katsutoshi Yada, Kei Miyazaki and Keiji Takai
Faculty of Business and Commerce, Kansai University
3-3-35 Yamate-cho, Suita
Osaka, Japan
e-mail: {yada, miyakei, takai}@kansai-u.ac.jp

Kohei Ichikawa
Graduate School of Information Science, NAIST
8916-5 Takayama-cho, Ikoma
Nara, Japan
e-mail: ichikawa@is.naist.jp

***Abstract*— In this paper, we propose an ASP system for shopping path analysis and a cloud based system with the necessary analysis service for marketing strategies that combines sales data with shopping path data. This paper explains a recommendation system based on position information or an application that calculates a basic marketing indicator using shopping path data by introducing a framework of ASP for shopping path analysis. Because the proposed ASP service is built on a cloud based system, the users can easily access the service through the web based interface and perform large-scale data processing using the computational resources of the cloud based system at low cost.**

***Keywords*— ASP; shopping path; marketing; recommendation system**

I. Introduction

This paper proposes an ASP system supporting shopping path analysis using MUSASHI [1], a data marketing tool that we have been developing, and an expansion for MUSASHI, the Cloud-MUSASHI [2] for ASP. MUSASHI extracts items and combines scripted single function commands such as aggregate calculations, making various processes feasible. This is a commonality and traditionally used method in UNIX, and features combining command groups and quickly processing large-scale data, which enables a flexible analysis system.

In order to analyze large-scale data at a low cost, we propose Cloud-MUSASHI, a cloud-based analysis system expansion of MUSASHI. By using Cloud-MUSASHI, it is possible to provide datamining as a low cost service. Many areas desire this type of system that provides cloud-based marketing services.

In recent years, one path data showing the purchasing process that has seen a lot of attention in the marketing area is shopping path data [3]. By integrating shopping path data and sales data, numerous marketing models are proposed [3]–[6]. However, as far as we know, there has been no proposal of a cloud-based system for analyzing this kind of shopping path data. In the highly competitive sales industry, the necessity for utilizing shopping path data in marketing strategies is high, and a cloud system that can be operated at low cost is suited for providing a marketing service.

The aim of this paper is to analyze shopping path data from pursuing the movement of customers in a store, form a cloud-based service system providing useful insights into in-store marketing strategies and to investigate the ability to discover useful knowledge from the raw data. This paper is structured as follows: First, we will explain the architecture of the ASP system forming the base of the marketing system proposed in this paper. Next, we introduce the ASP system for shopping path analysis. Then we introduce a recommendation system based on position information as an example of a real application and outline future problems.

II. Related research

Many studies have been conducted in the past on customer in-store motion. These studies indicated basic indices of in-store motion or the heterogeneity of the motion of customers [7]. However, these use protocols to extract consumer's cognition processes from interviews, and observers utilized data of the routes of individual consumers [8]. The data in these studies was highly biased due to observers directly intervening, and since little data was collected and at such great cost, the examination of in-store motion of consumers was insufficient.

Studies clarifying the relationship between the type of customer shopping path and effects of promotions, explain the motivation for current customer shopping path research. The table 1 shows the differences in studies investigating the effect on the relationship between the type of customer shopping path and effects of promotions. Kahn and Schmittlein [9] attempted to clarify the relationship between the type of customer shopping path and effects of cracker promotions. From the results of the experiment, the effects of leaflets or coupons differed according to the type of customer shopping paths; however, the effect of in-store displays were not statistically significant. They classified the type of customer shopping paths based on sales amount of each purchase. Walter and Jamil [10] classified the customer shopping paths based on the quantity of each purchase and investigated the effect on promotions however, no statistically significant difference was seen. Anic and Radas [11] classified the type of customer shopping paths based on sales amount and the quantity of each purchase, and statistically clarified the difference of the effect on promotions. Common points in these studies is that the criteria for classifying the types of customer shopping paths did not use the actual movement of the customers, but the shopping trends such as sales amount and quantity. As stated above, as well as having a large bias in the data for in-store movement of customers and there is little data.

TABLE I. TYPE OF CUSTOMER FLOW-LINE AND THE EFFECT ON PROMOTIONS

Authors (year)	Subject of experiment	Criteria for classification of shopping path	Effect on promotions
Kahn and Schmittlein (1992)	Supermarkets in America	Sales amount of each purchase	Partial significance
Walter and Jamil (2003)		Quantity of each purchase	Insignificant
Anic and Radas (2006)		Sales amount and quantity of each purchase	Significant

Due to the advancement of IoT technology, such as RFID, we are moving towards a large turning point in investigating customer motion in stores. Recently, studies, such as Sorensen [12], Larson et al. [13], Yada et al. [6], attached RFID tags to shopping carts and gathered a large amount of customer shopping paths, and are able to scientifically analyze in-store motion of customers. Larson et al.[13] de several classification types using clustering on the large amounts of customer shopping path data gathered using RFID tags. Hui et al. [4], using the Travelling Salesman Problem approach, clarified the characteristics of customer purchasing trends based on divergence from the optimal path. Takai and Yada [14] clarified the relationship between stationary time and purchasing actions at numerous sales points in store.

Customer shopping path studies are gradually accumulating more research, and its implementation in actual in-store marketing is highly anticipated. However, analyzing this type of large-scale customer shopping path data, must integrate various data, and will incur large costs and time spent on customizing this system to each business. In order to utilize customer shopping path data in businesses, a high-speed, low-cost system and related network is required to process large-scale, varied data.

III. A FRAMEWORK OF ASP FOR SHOPPING PATH

This paper integrates the proposed shopping path data or sales data, where the analysis system is assumed to be accessed by staff at retailer's HQ or stores, wholesalers that purchase products, marketing staff and manufacturers that produce each product and many other related personnel and is expected to be operable at low cost. It is therefore preferable to have this type of system formed with ASP. ASP is an internet service that provides users with analysis services. This paper introduces MUSASHI and MUSASHI-CLOUD, important components in the proposed ASP system, and explains the overall structure of the ASP system.

A. MUSASHI-CLOUD

The main parts of ASP proposed in this paper are formed from the MUSASHI command group. MUSASHI (Mining Utility and System Architecture for Scalable Historical data) has strong preprocessing for extracting useful data from large data sets, and is an efficient and flexible data mining platform that uses large-scale data written in XML [1].

MUSASHI is equipped with command groups that efficiently handles data structures written in XML as in Figure 1, similar to Plain Text. The XML format for MUSASHI is designed so that it can represent large-scale table dataset in an efficient manner. The root element named <xmltbl> has two elements, <header> and <body>. The metadata related to the target dataset is stored in the header element, while the large raw table dataset is stored as simple space separated values in the body element. Because the metadata of the target dataset is stored as a rich XML document, it can represent any variety of data. We basically store the title of the dataset, the field names of the table dataset, and some comments for the dataset in the header element. On the other hand, the simple space separated values for the raw table dataset do not include redundant XML tags and it can be handled by the MUSASHI's analysis tools efficiently.

Table 2 is an example of MUSASHI's command groups. Similar to traditional UNIX systems, MUSASHI utilizes pipes and redirection to form a system combining these command groups, allowing it to quickly and flexibly carry out a large amount of analyses [1]. For example, *xtcat* and *xtcut* in Table 2 are *cat* and *cut* of MUSASHI version respectively. Because many commands of MUSASHI are designed similar to the UNIX commands, the developers can easily learn and start to develop data mining systems using MUSASHI commands.

Figure 2 shows an example script for analyzing sales data shown in Fig. 1. The script calculates the daily sales amount and sales volume of each store. The first command extracts only four columns of data required for this analysis. The second command calculates the amount of sales and sales volume per store and day. The last command converts the results from the XML format to the CSV format. MUSASHI allows users to combine the provided command-line programs for their analysis purpose in this kind of intuitive manner.

```
<?xml version="1.0" encoding="euc-jp"?>
<xmltbl version="1.1">
<header>
<field no="1" name="customer"></field>
<field no="2" name="store"></field>
<field no="3" name="date"></field>
<field no="4" name="item"></field>
<field no="5" name="quantity"></field>
<field no="6" name="sales"></field>
<field no="7" name="unit_price"></field>
</header>
<body><![CDATA[
10001 101 20100102 bread 2 400 200
10001 101 20100102 curry 1 200 200
10001 101 20100103 bread 1 110 110
10001 101 20100103 milk 1 120 120
10002 102 20100102 tea 2 300 150
10002 102 20100102 noodle 3 600 200
10002 102 20100103 rice_ball 1 105 105
]]></body>
</xmltbl>
```

Fig. 1. Example of XML data structure

TABLE II. EXAMPLES OF MUSASHI COMMANDS

Command	Description
xtagg	calculates summation, average, etc
xtcal	calculates between columns
xtcat	concatenates files
xtcut	selects requested columns
xtjoin	joins rows of two files on common fields
xtsel	selects rows where match given conditions
xtsort	sorts rows
xtuniq	omits repeated rows

```
xtcut -f store,date,sales,quantity -i input.xt |
xtagg -k store,date -f sales,quantity -c sum |
xt2csv -F -o result.csv
```

Fig. 2. Example of MUSASHI script

MUSASHI-CLOUD distributes large-scale data using several servers running MUSASHI in the cloud, and allows analysis to be carried out via a web interface [2]. It provides a web-based data mining environment as a service. The service allows users to execute data mining processes in the cloud environment through the web interface.

In the large-scale data mining processes, I/O performance becomes more serious problem than processing power of the machine. No matter how high-performance machine we use, a single standalone machine cannot increase the I/O performance. To increase overall I/O performance, we have to leverage multiple machines and increase the number of the I/O channels physically. Therefore, data processing needs to be distributed over multiple difference machines. MUSASHI-CLOUD is designed to handle the large-scale data processing by distributing the data processing over multiple machines in a cloud base system.

Figure 3 shows the overview of the architecture of MUSASHI-CLOUD. The MUSASHI-CLOUD consists of a frontend node and multiple computational nodes. All datasets provided by the users are redundantly transferred and stored on each computational node in advance. By duplicating the datasets, the I/O processes are multiplexed over the computational nodes, and this increases the I/O performance physically.

The frontend node provides a web interface for the users. The requests from the users are received at the frontend node and rapidly distributed into the computational nodes. On each computational node, the specified MUSASHI scripts from the users are executed for the datasets. And then, all results from the computation nodes are aggregated back to the frontend node and returned to the users through the web interface. This service allows users to perform large-scale data mining processes in the cloud computing resources.

B. *ASP for shopping path*

In this paper, we propose an ASP system for shopping path analysis that provides information gained from important suggestions for marketing strategies by integrating shopping path data and sales data. The ASP front end server receives analysis requests from each user through a WEB interface, makes an analysis script formed from MUSASHI's commands, and executes them on each calculation node [2].

If this analysis is carried out daily for marketing strategies, many requests are generated from temporary projects. MUSASHI-CLOUD, operating on the cloud, can flexibly handle calculation resources depending on the calculation load in line with analysis requests. Therefore, when there are few analysis requests, operating costs are also reduced, and the analysis system proposed in this paper is suitable for MUSASHI-CLOUD.

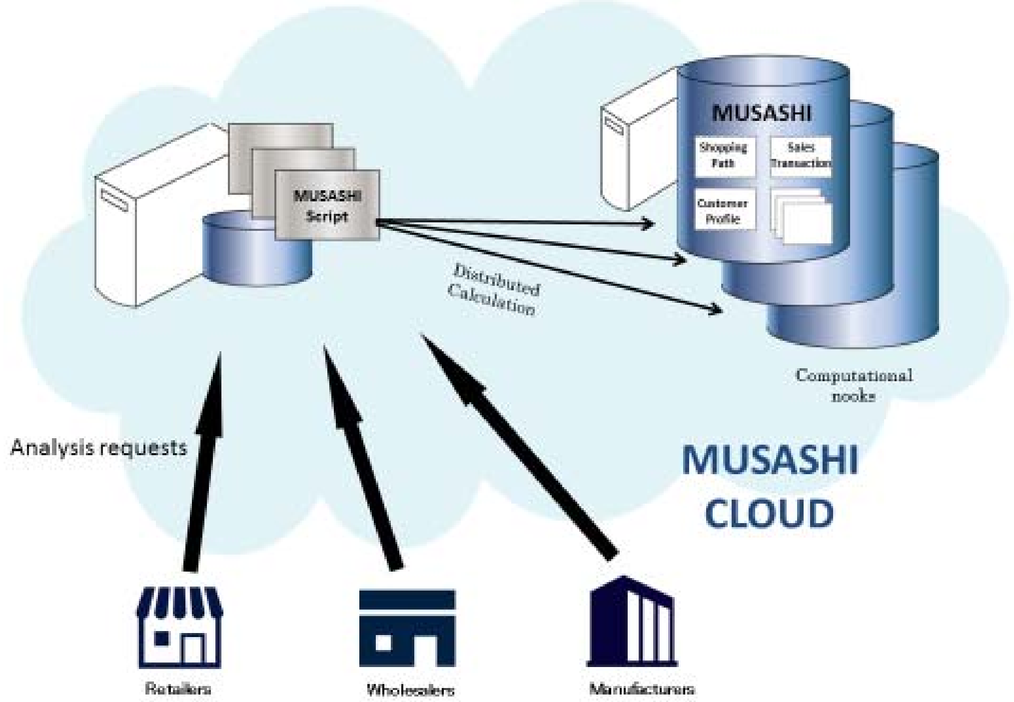

Fig. 3. Overview of MUSASHI-CLOUD

C. Interface of the ASP system for shopping path

The proposed system automatically forms the necessary program depending on the user's analysis requirements and quickly processes large-scale data through a distributed processing system. Here we introduce the main interface of the proposed system through analysis applications. First, the user selects an analysis from the analysis menu on the selection screen (Fig. 4). This system has a user management function that handles the accessible analysis menu through a user group. It can provide an appropriate analysis menu from the position or aim of the user.

Next, select the target data for analysis (Fig. 5). The proposed system can simultaneously handle data from several businesses, and different access authorities can be set for each user group. An analyst can implement the same analysis on numerous allowed data sets. In addition, this analysis menu requires differing parameters. The user can freely set parameters such as period, target customers or target sales.

Fig. 4. Analysis selection screen

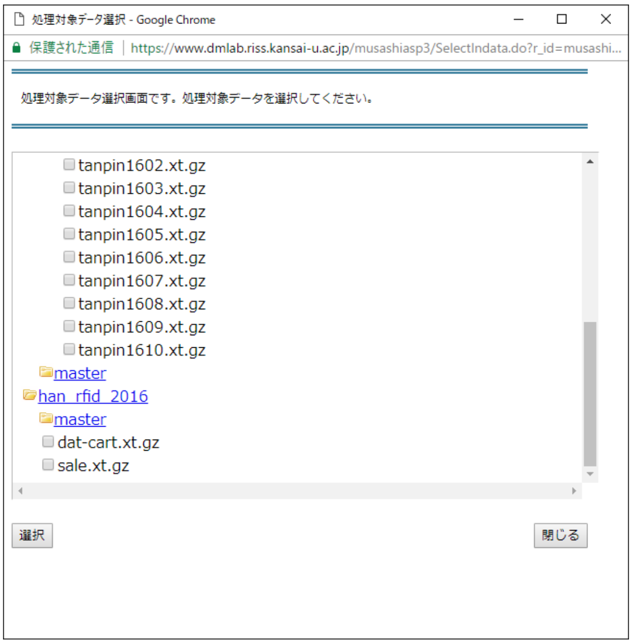

Fig. 5. Pop-up window when selecting target data

IV. MARKETING APPLICATION IN THE ASP SYSTEM FOR SHOPPING PATH

The ASP system for shopping path proposed in this paper not only includes basic marketing indices, but also includes an analysis system for sales promotion based on purchasing models. This section introduces a system for calculating basic indices such as visiting ratio at each area in the store and a recommendation system based on positions.

A. A system for calculating basic indices of shopping path analysis

The most basic function of the ASP system for shopping path provides basic indices regarding shopping path data. The proposed system calculates basic indices such as visiting ratio, stationary time and purchasing probability for each store. For example, Fig. 6 shows the visitor rate for each area in the store. Each user can extract the required data via a web interface.

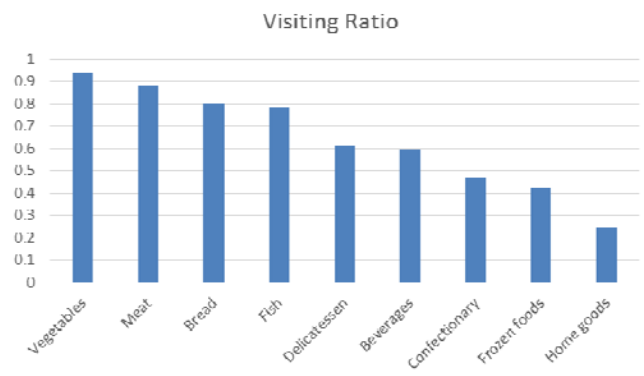

Fig. 6. Visiting ratio for each store

In the proposed system, detailed parameters are set depending on each user's preferences, and basic indices can be calculated for each customer group and each hour. Figure 7 shows the average stationary time for each age group at the vegetable and fresh fish areas. From this, we understand that younger customers spend more time in the vegetable section whereas, the fish section shows a trend in longer average stationary time with older customers. By calculating basic indices from these varied conditions, useful information for an effective marketing strategy can be acquired.

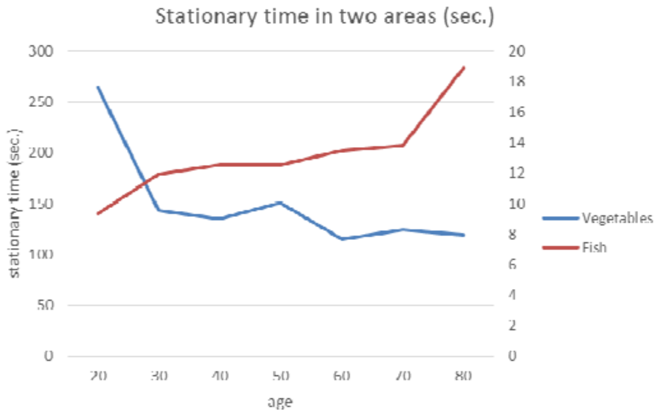

Fig. 7. Average stationary time by age group in the fish and vegetable sections

B. Recommendation system based on position

We implemented a recommendation system based on on-line shopping of numerous small business past purchasing history information. As this effectively promoted sales, it is essential for modern business. Since many of these use past purchasing history or personal attributes, it is extremely rare to recommend services or products based on customer's position information. So and Yada et al proposed a recommendation system based on position information using shopping path data [15].

Their proposed system calculated the shopping path data for a supermarket based on the stationary time of customers in each section, and from the similarity in stationary time and purchasing, recommended specific sections to each customer [15]. Let the stationary time of customer u, in section i, on day d, be T^{uid}, then t_j^{uid} is the stationary time of customer u, in section i, on day d, at time j. Since the stationary time is systematically biased depending on the customer and section, then normalized store ratings, b_u, b_i, can be obtained for a customer, u, in a section, i, using the baseline predictor [16]. Let the stationary time of customer u in store i be r_{ui}, then the baseline predictor, b_{ui}, for r_{ui} is calculated using the following formula [17].

$$b_{ui} = \mu + b_u + b_i \quad u \in U, i \in I \tag{1}$$

From this, the store rating, $\hat{r}_{ui}$, normalizing for customer or store bias is given by;

$$\hat{r}_{ui} = \frac{r_{ui} - b_{ui}}{|b_{ui}|} \quad u \in U, i \in I \tag{2}$$

Then, the Pearson correlation coefficient between $\hat{r}_{ui}$ normalized for similarities between customer u and customer v, $s(u, v)$, can be defined with the following formula [10].

$$s(u,v) = \frac{\sum_{i \in I}\left(\hat{r}_{ui} - \bar{\hat{r}}_u\right)\left(\hat{r}_{vi} - \bar{\hat{r}}_v\right)}{\sqrt{\sum_{i \in I}\left(\hat{r}_{ui} - \bar{\hat{r}}_u\right)^2}\sqrt{\sum_{i \in I}\left(\hat{r}_{vi} - \bar{\hat{r}}_v\right)^2}} \quad u, v \in U, u \neq v \tag{3}$$

This similarity can extract the number of people, k, similar to customer u.

The proposed system can implement a distributed processing system in the cloud for this calculation, and has the capacity to handle peak times at large department stores. The precision of actual recommendations is reported to be higher than conventional recommendation systems using sales history data [15].

V. Conclusion

In this paper, we propose an ASP system for shopping path. The proposed system is formed in the cloud and can flexibly respond to the analysis demands of numerous users, allowing for low operation costs. The proposed ASP system uses shopping path data and includes a recommendation system based on in-store position information of customers or basic indices that support effective sales promotions.

However, many problems remain in the proposed system. The current proposed system handles data taken from a position information tracking system based on RFID for shopping path data. It cannot handle chronological data obtained from a customer tracking system using video monitoring. The recommendation system based on position information outlined in this paper can calculate store recommendations; however, methods for informing the customer, such as an interface for sending information to the customer's cell phone, have yet to be developed. We will develop various applications for a more practical system.

Acknowledgment

This is a product of research which was financially supported by the Kansai University Fund for the Promotion and Enhancement of Education and Research, 2014. "Research of cloud based data-mining platform for using big data."

References

[1] K. Ichikawa, K. Yada, and T. Washio, "Development of data mining platform MUSASHI towards service computing," in *Proceedings - 2010 IEEE International Conference on Granular Computing, GrC 2010*, 2010, pp. 235–240.

[2] K. Yada and K. Ichikawa, "Decision support system for policy making during a financial crisis," in *2011 IEEE International Conference on Granular Computing*, 2011, pp. 756–760.

[3] S. K. Hui, P. S. Fader, and E. T. Bradlow, "Path Data in Marketing: An Integrative Framework and Prospectus for Model Building," *Mark. Sci.*, vol. 28, no. 2, pp. 320–335, 2009.

[4] S. K. Hui, J. J. Inman, Yanliu Huang, and J. Suher, "The Effect of In-Store Travel Distance on Unplanned Spending: Applications to Mobile Promotion Strategies," *J. Mark.*, vol. 77, no. 2, pp. 1–16, 2013.

[5] N. Sano and K. Yada, "The influence of sales areas and bargain sales on customer behavior in a grocery store," *Neural Comput. Appl.*, vol. 26, no. 2, pp. 355–361, 2014.

[6] K. Yada, "String analysis technique for shopping path in a supermarket," *J. Intell. Inf. Syst.*, vol. 36, no. 3, pp. 385–402, 2011.

[7] B.-D. Kim and K. Park, "Studying patterns of consumer's grocery shopping trip," *J. Retail.*, vol. 73, no. 4, pp. 501–517, Dec. 1997.

[8] J. W. Payne and E. K. E. Ragsdale, "Verbal Protocols and Direct Observation of Supermarket Shopping Behavior: Some Findings and a Discussion of Methods," *Adv. Consum. Res.*, vol. 5, pp. 571–577, 1978.

[9] B. E. Kahn and D. C. Schmittlein, "The Relationship between Purchases Made on Promotion and Shopping Trip Behavior," *J. Retail.*, vol. 68, no. 3, pp. 294–315, 1992.

[10] R. G. Walters and M. Jamil, "Exploring the relationships between shopping trip type, purchases of products on promotion, and shopping basket profit," *J. Bus. Res.*, vol. 56, no. 1, pp. 17–29, Jan. 2003.

[11] I.-D. Anic and S. Radas, "The relationships between shopping trip type, purchases made on promotion, and unplanned purchases for a high/low hypermarket retailer-evidence from the croatian market," *Evid. from Croat. Mark. Econ. Trends Econ. Policy*, vol. 107, pp. 27–45, 2006.

[12] H. Sorensen, "The science of shopping," *Mark. Res.*, vol. 15, no. 3, pp. 30–35, 2003.

[13] J. S. Larson, E. T. Bradlow, and P. S. Fader, "An exploratory look at supermarket shopping paths," *Int. J. Res. Mark.*, vol. 22, no. 4, pp. 395–414, 2005.

[14] K. Takai and K. Yada, "A framework for analysis of the effect of time on shopping behavior," *J. Intell. Inf. Syst.*, vol. 41, no. 1, pp. 91–107, 2013.

[15] W. T. So and K. Yada, "A Framework of Recommendation System Based on In-store Behavior," in *In Proceedings of the 4th Multidisciplinary International Social Networks Conference (MISNC '17)*, 2017, vol. 33, pp. 1–4.

[16] Y. Koren and R. M. Bell, "Advances in Collaborative Filtering - Hledat Googlem," in *Recommender Systems Handbook*, 2015, pp. 77–118.

[17] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative Filtering Recommender Systems," *Trends Hum.-Comput. Interact.*, vol. 4, no. 2, pp. 81–173, 2011.