Character String Analysis and Customer Path in Stream Data

Katsutoshi Yada
Faculty of Commerce, Kansai University
E-mail: yada@ipcku.kansai-u.ac.jp

Abstract

This purpose of this study is to propose a knowledge-discovery system that can abstract helpful information from character strings representing shopper visits to product sections associated with positive and negative purchasing events by applying character string parsing technologies to stream data describing customer purchasing behavior inside a store. Taking data that traced customers' movements we focus on the number of times customers stop by particular product sections, and by representing those visits in the form of character strings, we propose a way to efficiently handle large stream data. During our experiment, we abstract store-section visiting patterns that characterize customers who purchase a relatively larger volume of items, and are able to show the usefulness of these visiting patterns. In addition, we examine index functions, calculation time, and prediction accuracy, and clarify technological issues warranting further research. In the present study, we demonstrate the feasibility of employing stream data in the marketing field and the usefulness of the employing character parsing techniques.

1. Introduction

Thanks to technological advances and a lowering of implementation costs, radio frequency identification, commonly known as RFID, has come to be used in a variety of businesses. In 2005, the Ministry of Economy, Trade, and Industry conducted an experiment entitled the "Future Japanese Store Project." In the experiment, shopping carts equipped with RFID devices were used to grasp customers' behavior within stores by gathering data on customer behavior and purchasing activity at the store counter. In this project, data on customers' movements within the store was gathered electronically, and thus it was possible to obtain detailed data on customer purchasing behavior within the store, something which had previously remained totally unknown. The focus on using RFID to gather detailed data on customer behavior within the store is a trend observed not just in Japan but in Europe and America as well.

Until now, in order to understand consumer behavior in the retail industry, historical data on customers, like point-of-sale data (POS data) has traditionally be used. Using such data, one can determine which customer purchased what and where, and that data can then in turn be analyzed in greater detail. For example, in the field of marketing, Guadagni and Little [3] and Gupta [4] proposed consumer purchasing behavior models using such data. More recently, in order to handle large volumes of data, data mining was conducted in many industries [5] [14], and this was helpful for improving sales promotion activities or brand strength. However, while customer purchasing history data is able to record the purchasing results for a given customer, it is not able to shed any light on how customers moved through the store or how they came to purchase them. In previous studies, in other words, the route traced by customers within a store was treated as a form of black box, and only the data on resulting purchases was made the subject of subsequent analysis.

Progress in RFID technology in recent years brought a complete about face to that situation. In particular, in marketing applicability studies on RFID technology, the greatest emphasis was placed on providing RFID devices for customers or their carts, and analyzing customer routes within the store by tracing their movements and determining their behavior [11]. Tracing customers' movements within a store makes it possible to have a better understanding of what and why customers make purchases than is the case when simply noting the product purchases, as was the case with previous marketing studies. There have been very few studies based on customer data that describes customer movements within the store. The reason for this is that until now it was exceedingly difficult to obtain such data. Accordingly, customer movement data obtained using RFID will be a springboard for new avenues of research in the field of marketing.

978-0-7695-3503-6/08 $25.00 © 2008 IEEE
DOI 10.1109/ICDMW.2008.41

Among studies that have employed RFID-based customer movement data analysis, there is a study by Larson et al. [10]. They employed a clustering method that improved on the k-means algorithm, and thereby discovered a number of customer groups. By exploring these customer groups in data, they were able to suggest a number of hypotheses. However, until now there have been no studies of applied implementation or research focusing on classification issues or abstraction of characteristics based on customer movement data.

In the retail industry, those targeting customers for a given marketing strategy need to grasp these characteristics and understand purchasing behavior. Accordingly, application studies that focus on classification problems and characteristic abstraction, and not just clustering, are thought to have important business implications.

The RFID data used in the present study is typically referred to as stream data or a data stream. Stream data is data in which changes in a subject are recorded electronically and continuously over time. In the distribution and communications fields, there is a tremendous need to obtain useful information based on such data. Moreover, such data has attracted the attention of many researchers as an important domain of application for data mining. However, because the volume of data tends to be huge and because the data tends to be unstructured, it is difficult to directly apply methods that target the sort of tabular data that in past studies were largely ignored.

We introduce knowledge expressions in the form of character strings for stream data including information about customer movements, and have proposed the adoption of EBONSAI [7] [15], a character parsing application used in the field of business. In other words, by abstracting information on the paths that customers trace within a store and expressing that information in the form of character strings, we thought to implement rule-based abstraction using existing character string parsing algorithms. The application of this existing technology to a new field not only demonstrates the usefulness of that technology but also clarifies new technological issues at the same time. In this study, by applying this approach to actual stream data, we hope to lay open discussions of technological issues and the feasibility of applying it to stream data to which character parsing methods are applied.

2. Analysis of Customer Movements and Character Strings

2.1. Analysis of Customer Movements and Character Strings

Customer movement analysis is a store management method that makes it possible to improve the efficiency of store layout design and sales promotional plans by analyzing the routes that customers take within a store. Figure 1 shows the movement of a customer within a store superimposed over the store layout. The paths of customer movements and their directions are shown using linked lines with arrows. Moreover, sections where a customer stops are shown as nodes, whereas red nodes indicate locations where the customer purchased something. As can be seen from the figure, customers move in extraordinarily complex manners when doing their shopping.

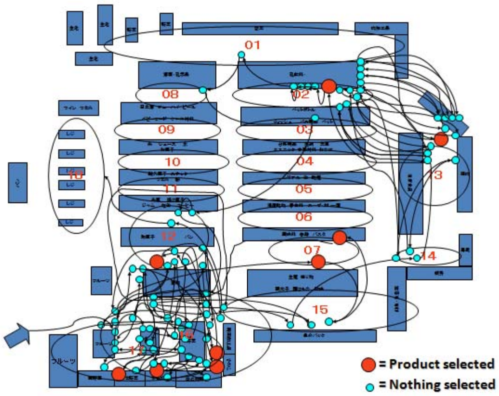
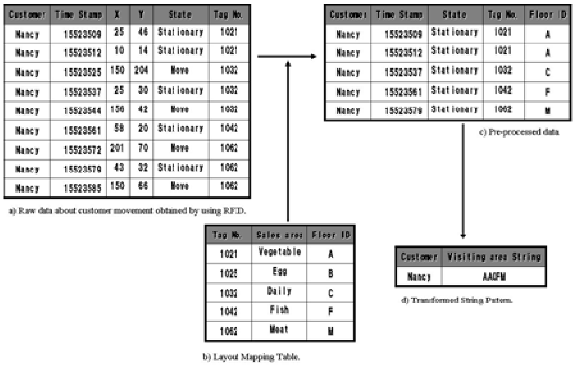


Figure 1. An example of customer movement data.

A particularly important influence on purchasing behavior is the rate at which a customer stops in a particular section of the store; in other words, what is key is whether the customer actually passes by and stops in any given section. This is expressed in the data as product section stops. Naturally, there are cases where customers stop by a section but elect not to buy anything. Whether the customer chose to buy something can be easily determined by comparing the movement data with the purchase data. This information is particularly important to those in the store merchandizing industry. For this reason, when looking at customer movement data in this paper, we focus on customer stops at product sections and abstract characteristics of the routes taken within the store by the customer.

Because it is difficult to process the stream data obtained by RFID as is, some additional processing measures must be undertaken. For that reason, in this paper, we employ character strings as knowledge representations that can be used to analyze customer movement data. We shall explain this transformational process with reference to Figure 2. 2a shows the raw data obtained using RFID. The data includes a wide variety of items, including RFID tag number, shopping cart state, and acceleration in the X and Y directions as a function of time and customer ID. This raw data is transformed using the layout mapping table shown in 2b. This layout mapping table has been provided with floor section IDs by joining the RFID tags with the store location points. Each RFID record is transformed into a character that uniquely identifies each floor. At this point, this narrows the data to one thing, namely, which section within the store the customer is currently located. Then, by linking up the succession of floor IDs based on the order in which the customer visits different sections of the store, we obtain a character string pattern like that shown in 2d. For example, if we use the mapping table, we can express the store-section visiting pattern for the customer identified as Nancy in Figure 2 as "AACFM."



Customer	Time Stamp	X	Y	State	Tag No.
Nancy	15523509	25	46	Stationary	1021
Nancy	15523512	10	14	Stationary	1021
Nancy	15523525	150	204	Move	1032
Nancy	15523537	25	30	Stationary	1032
Nancy	15523544	156	42	Move	1032
Nancy	15523561	58	20	Stationary	1042
Nancy	15523572	201	70	Move	1062
Nancy	15523579	43	32	Stationary	1062
Nancy	15523585	150	66	Move	1062

a) Raw data about customer movement obtained by using RFID.

Tag No.	Sales area	Floor ID
1021	Vegetable	A
1025	Egg	B
1032	Daily	C
1042	Fish	F
1062	Meat	M

b) Layout Mapping Table.

Customer	Time Stamp	State	Tag No.	Floor ID
Nancy	15523509	Stationary	1021	A
Nancy	15523512	Stationary	1021	A
Nancy	15523537	Stationary	1032	C
Nancy	15523561	Stationary	1042	F
Nancy	15523579	Stationary	1062	M

c) Pre-processed data.

Customer	Visiting area String
Nancy	AACFM

d) Transformed String Pattern.

Figure 2. RFID data and product-section visiting pattern strings.

2.2. Purpose of This Research

The purpose of this research is to propose a knowledge discovery system that can abstract useful information from character strings representing store-section visiting patterns for both positive and negative purchasing events. This is accomplished by applying character string parsing technologies on stream data pertaining to customer purchasing behavior within the store. At the time that we devised this system, we made use of a previously existing system known as EBONSAI. EBONSAI [5] [15] is a time series analysis technique adapted from the BONSAI character parsing approach employed for the genome project. Up to now, EBONSAI had been used for time series analyses of sales data, web log data, and the like, but it had never been used for the kind of stream data that is generated by RFID. In this paper we hope to demonstrate that it can be applied to the kind of stream data found in the field of marketing. We shall do this by clarifying technological issues, showing the method's usefulness, and applying it to character parsing for this kind of stream data.

2.3. EBONSAI

EBONSAI is an adaptation of the BONSAI character parsing system that was originally developed in the field of molecular biology. It is a system whereby positive and negative events are expressed as character strings, and using those partial character strings or partial sequences highly refined decision trees are generated. We shall begin by first explaining the BONSAI algorithms that form the core of the EBONSAI system.

Let P be positive data set, N be negative data set, and |P| and |N| be the numbers of records in P and N, respectively. Given a substring α, let p_T and n_T be the numbers of records containing α in p_T and n_T, respectively, and let p_F and n_F be the numbers of records not containing α in P and N, respectively. Defining entropy function ENT(x,y) in the following manner,

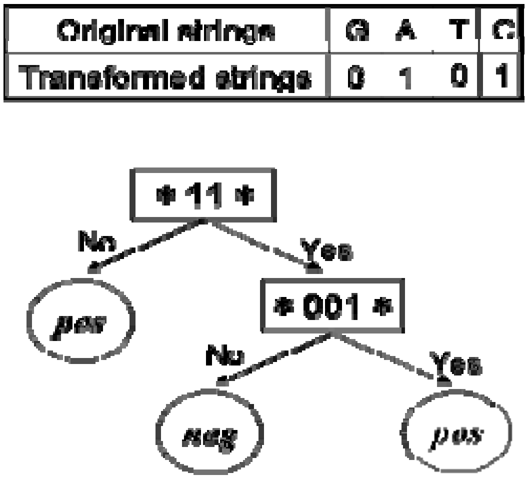
$$\mathrm{ENT}(\mathrm{x},\mathrm{y}) = \begin{cases} 0 & \mathrm{x}=0 \text{ or } \mathrm{y}=0 \\ -\mathrm{x}\log\mathrm{x} - \mathrm{y}\log\mathrm{y} & \mathrm{x},\mathrm{y} \neq 0 \end{cases} \quad (1)$$

we define in the following expression the entropy obtained after classifying the original data into two subsets depending on whether data contains α as a substring or not.

$$\frac{\mathrm{p_T}+\mathrm{n_T}}{|\mathrm{P}|+|\mathrm{N}|}\mathrm{ENT}\left(\frac{\mathrm{p_T}}{\mathrm{p_T}+\mathrm{n_T}},\frac{\mathrm{n_T}}{\mathrm{p_T}+\mathrm{n_T}}\right) + \frac{\mathrm{p_F}+\mathrm{n_F}}{|\mathrm{P}|+|\mathrm{N}|}\mathrm{ENT}\left(\frac{\mathrm{p_F}}{\mathrm{p_F}+\mathrm{n_F}},\frac{\mathrm{n_F}}{\mathrm{p_F}+\mathrm{n_F}}\right). \quad (2)$$

We compute α which minimizes this value. Namely, we choose α for which the information gain is maximized. After partitioning the original data based on α, BONSAI continues to proceed in a recursive manner.

Like BONSAI, EBONSAI incorporates an alphabet indexing mechanism. This mechanism is achieved by substituting the smallest possible character string for a given characteristic character set for positive events. This makes it possible to abstract high-level rules that can interpret relatively small character strings while reducing the search space. From the total alphabet set Σ, we convert the original character string using the mapping (image) ϕ for the smallest collection of letters generated randomly Γ, and in the above-mentioned manner generate a decision tree. Next, we search until the neighborhood of ϕ cannot be further refined, and output a decision tree that has the greatest discrimination. By using the appropriate alphabet listing, it is possible to refine the classification and simplify one's hypothesis.



Original strings	G	A	T	C
Transformed strings	0	1	0	1

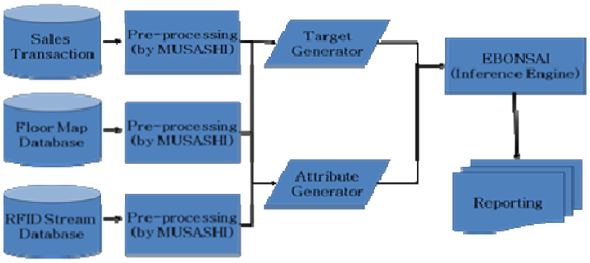
Figure 3. Example of EBONSAI output.

EBONSAI functions are easy to appreciate by looking at the output. Figure 3 gives an example of EBONSAI output. Based on the upper mapping table, EBONSAI converts the four given character strings into 1s and 0s. For positive events that have been converted, it is possible to check whether they conform with character strings abstracted from the root of the decision tree. For example, we follow along the "yes" arrows in the case that a character string of "11" is included, and follow along the "no" arrows for cases where it is not. In this way, by using only a few converted character strings, EBONSAI can generate decision trees having a relatively simple predictive ability. Because EBONSAI was for the most part used for purchasing pattern character strings, it can be applied to character string data comprising 100 character strings or more. In addition, other areas where EBONSAI was improved are described below.

- In business, in order to handle a number of cause-and-effect relations, it is necessary to contend with a various attributes simultaneously. For this reason, EBONSAI is able to employ a number of character string attributes. In addition, just like general decision tree algorithms, EBONSAI can handle category attributes and numerical attributes in one model simultaneously, and not just character string attributes.
- EBONSAI can handle data structured in the form of a table described using XML, and if used in conjunction with the MUSASHI open source platform, a viable system can be easily constructed.

2.4. System Overview

Figure 4 shows a concept diagram of the knowledge discovery system employing RFID which we developed. Three databases were used for the raw data, and each of these is associated with a preprocessing system. The preprocessing systems handle data in XML form, and then transfer this data to the target generator and attribute generator in the next stage. Next, the data is combined and a classification model is constructed based on the mining engine. All of this was put together using the MUSASHI open source platform for data mining [6].



Sales Transaction
Pre-processing (by MUSASHI)
Target Generator
EBONSAI (Inference Engine)
Floor Map Database
Pre-processing (by MUSASHI)
Attribute Generator
Reporting
RFID Stream Database
Pre-processing (by MUSASHI)

Figure 4. Overview of the knowledge discovery system for customer movement data.

We will now explain all of these major subsystems in greater detail. This system employs three databases. The first database houses data on customer purchasing history, and includes information on customer ID, purchase price, product information, and the like. The second database contains store layout information. This database contains a database of products together with RFID sensor location information. This makes it possible to track customer position information and the sections in the store where purchased products were obtained. The third database contains RFID sensor logs. By pooling all of these databases together, it is possible to determine how a given customer moved within the store, as well as where the products that a customer purchased were located within the store.

Next, let's examine the target attribute generator. What we have developed here is a system that, by pooling various databases together, can construct classification models for customers. Accordingly, it is necessary to generate target attributes to be subject to classification from the above-mentioned databases. Using the RFID sensor log database and the purchasing history database, this component generates attributes freely defined by the user. For example, we might want to contemplate the ideal customer for a particular shop or the characteristics of buyers of a particular product. In the same way, we can prepare components that generate explanatory attributes that use classification models from the above database.

From this, we can derive explanatory attributes relating to purchasing information on a particular product or a product category together with information on customer movement within the store. For example, using data on the customer's movements within a store, we can generate a sequence attribute that elucidates the order in which a customer visited various product sections of a store.

Finally, the mining engine can construct a classification model based on target attributes and explanatory attributes. The mining engine for this system was developed using EBONSAI as its foundation. For that reason, decision trees can be output using numerical figures, categories, and string attributes that were generated by the above described databases.

3. Experimental Results

3.1. Explanation of the Data

We will now demonstrate how this system can be used with actual customer movement data, and will perform an experiment on rule abstraction. We used customer movement data gathered at a mid-sized super market in Japan. In this project, the shopping carts that customers used were equipped with RFID receivers, and each product section had RFID tags. This made it possible to track customer movements within the store precisely. The experiment was conducted in September 2006. In addition to passenger movement data, floor layouts and purchasing history data were also gathered. The floor layout within the store was divided into seven sections. Each of those sections had subsections, and in total there were 17 subsections.

The purpose of the analysis in this experiment was to use the system proposed in this paper to clarify what characterized the movements of customers who bought a relatively large number of items. In this case, our data was somewhat restricted, as we simply measured the number of purchased items at the time the customer visited the store. However, we did not consider purchasing power (the total amount a particular customer spent per month on purchases), nor did we consider the intervals between shopping trips or frequency with which customers shopped. For our clustering method we used k-means, and we defined customers purchasing a relatively large number of items as "high-volume" (HV) customers, with the rest of the customers being deemed "low-volume" (LV) customers. The average number of items purchased by HV customers was 19 per store visit; the same average for LV customers was 7.86.

In this experiment, we used two kinds of attributes, numerical attributes and character string attributes, and these were output from the component that generates explanatory attributes. In terms of character string attributes, we used two kinds of product section visiting pattern strings, those for product sections and those for product subsections.

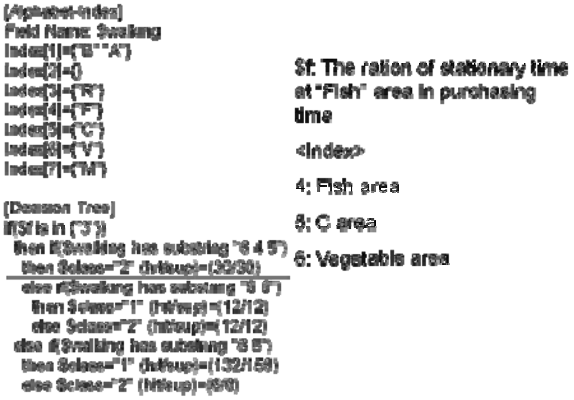
In addition, for numerical attributes we used component ratios comprising the time that a customer stayed in each section x, where x was a value from 1 to 7, relative to total time spent in the store. Thus, if customer i remained in section x t_{it} seconds, then the component ratio r_{ix} of time spent by customer i in area x was expressed as follows:

$$r_{ix} = \frac{t_{ix}}{\Sigma\, t_{ix}} \qquad (3)$$

For the component ratios of time spent in each area, we used a model having seven attributes.

3.2. Rule Abstraction and Interpretation

When we built a classification model using EBONSAI, we obtained extremely simple results like those shown in Figure 5. In the figure, class 1 represents LV customers, and class 2 represents HV customers. Attribute f is the component ratio that a customer spent in the fish section; product section visiting pattern character string 4 represents the "fish section"; 5 represents the "general goods section"; and 6 stands for the "vegetables section."



```
[Alphabet-Index]
Field Name: $walking
Index[1]={"B" "A"}
Index[2]={}
Index[3]={"R"}
Index[4]={"F"}
Index[5]={"C"}
Index[6]={"V"}
Index[7]={"M"}

[Decision Tree]
if($f is in ("3"))
  then if($walking has substring "6 4 5")
   then $class="2" (hit/sup)=(30/30)
   else if($walking has substring "6 5")
    then $class="1" (hit/sup)=(12/12)
    else $class="2" (hit/sup)=(12/12)
  else if($walking has substring "6 5")
   then $class="1" (hit/sup)=(132/158)
   else $class="2" (hit/sup)=(6/6)
```

$f: The ration of stationary time at "Fish" area in purchasing time

<Index>

4: Fish area

5: C area

6: Vegetable area

Figure 5. Rules abstracted using EBONSAI.

We obtained three rules. The first rule states that if the percentage of time spent in the fish section exceeded 10%, and the customer goes from the vegetable section to the fish section followed by the general goods section, then the customer was an HV customer. The second rule states that if the amount of time spent in the fish section exceeded 10%, and in the visiting pattern there was a movement from the vegetable section to the general goods section, then the customer was an LV customer. Finally, the third rule states that if the percentage of time spent in the fish section was less then 10%, and there was a character string indicating movement from the vegetable section to the general goods section, then the customer was an LV customer.

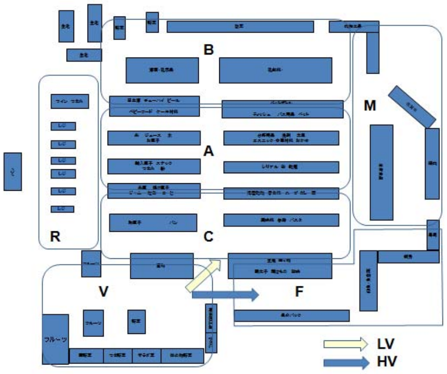


Figure 6. Patterns of movement between store sections by HV and LV customers.

What is characteristic of the rules that we obtained is that the area visited by the customer following the vegetable section (in other words, the general gods or fish section) determined the number of goods that the customer would purchase. In other words, this decided whether the customer was an HV or an LV customer (see Figure 6). According to specialists, supermarket fish sections are extremely important for attracting patrons' loyalty. In the past, shopkeepers have traditionally used a simple rule of thumb whereby the longer a customer spent in the fish section, the more items they would buy. However, in our experiment, we determined that what is really mattered was the pattern of customer movement between product sections.

First off, because target attributes essentially classify the number of items purchased by the customer on that day, then HV customers are expected to visit more sections of the store. On the other hand, LV customers, who purchase relatively fewer items, probably visit only those product sections containing the items that they initially targeted for that day. However, in this experiment, there was no significant difference between the probability of an HV or LV customer visiting the fish section. When we discussed these discoveries with shopkeepers, it seems they felt that most HV customers get some form of stimulation in the vegetable section, and may be deciding whether or not to buy some fish. In their opinion, the decision to opt for some fish while in the vegetable section is something that happens only after they have entered the store. For small retail stores, how customers are led to move from the vegetable section to the fish section is an important consideration when designing the store layout.

3.3. Usefulness of Character String Parsing Method and Technological Issues

Studying the above experimental results, we feel that analysis of visiting patterns to product sections using the character parsing analysis method has led to important discoveries when compared to methods used in the past. Knowledge represented using character strings is able to express far richer information on visiting patterns compared to typical tabular data. In fact, the patterns derived from our experiment were able to abstract characteristics of customer movement patterns within the store. In this way, we feel that the usefulness of the character string parsing method is quite high not just for purchasing history data but also for stream data obtained in marketing.

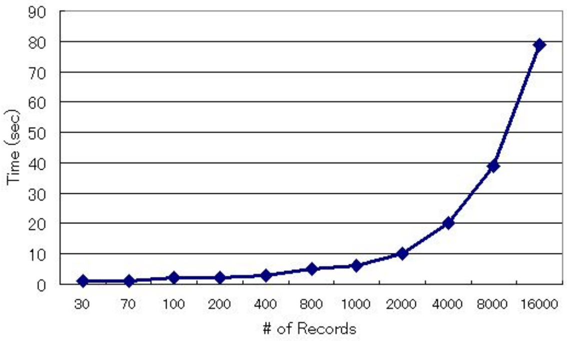


Figure 7. Calculation time as a function of data size.

We would now like to discuss the usefulness of and technological issues surrounding our proposed method with respect to prediction accuracy and calculation time. Figure 7 shows the relation between data size and calculation time needed to abstract rules using EBONSAI. Data size and calculation time are almost proportional, and we expect that an increase in record size in the future is not likely to be a major obstacle.

Next let's compare EBONSAI with other techniques from the standpoint of prediction accuracy. For our evaluation index, we used *overall accuracy* [13], defined as a percentage of the correctly classified positive events out of the total number of events. Figure 8 shows an average of cross verification (10 fold) for the various methods. The other techniques do not employ character strings for product section visiting patterns, and only employ seven attributes comprising the component ratios of time spent in a given product section. As can be seen in Figure 8, EBONSAI was approximately 4% more accurate than other methods. This seems to show that classifying character string attributes to describe visiting patterns, as was done in EBONSAI, provides useful information.

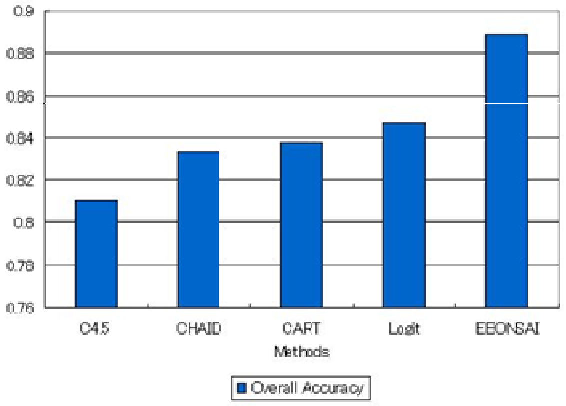


Figure 8. Comparison of EBONSAI's prediction accuracy with other methods.

Lastly, let us explore the indexing functionality contained within EBONSAI. The indexing used in EBONSAI has the effect of not only reducing search time but also making it easier to interpret the meaning of rules. For example, EBONSAI can be applied to identify brand switching patterns, such as when a manufacturer launches a new product or when a user switches between products in the same category or having the same flavor. In such cases, using indexing it is often sufficient to simply substitute a single letter to indicate this switch. As a result, the rules that are abstracted are simplified, and it becomes far more feasible to interpret rules as a result. However, in the case of customer movement analysis, indexing did not perform well from the standpoint of rule interpretability. The reason for this is that a number of product sections were substituted for using one character, and it was difficult to work out any special meaning in the grouping. In fact, in discussions with specialists, it seems that indexing could invite confusion. With customer movement analysis, because most specialists are interested in discovering special routes that may exist, in that kind of analysis they either do not use indexing, or they use a rather large index size.

However, increasing the index size generally results in higher calculation times. Figure 9 shows the relation between calculation time and index size, which is an EBONSAI parameter. In this experiment, by subdividing product sections into more detailed subsections, subdivided the store into 17 total subsections. The index size is a parameter in EBONSAI that the user can specify. The default index size is 2. If we look at matters from the standpoint of prediction accuracy, in most cases, we can obtain high accuracy using an index size of 2 or 3. Moreover, a general trend is that the smaller the index size, the less time required to perform calculations. In the graph shown in Figure 9, the calculation time increases up to an index size of 7 and onwards, but thereafter there is no increase beyond that point. The reason this happens is that subsections that were visited extremely infrequently are included in the calculations. And so from the standpoint of calculation time, even if the index size is made rather large, the system can cope. However, because the maximum index size for EBONSAI is 9 at present, in the future, the system will need to be modified so that it can handle larger index sizes. Moreover, in the case of shops where the customer can visit various product sections, increases in index size can incur an extremely large calculation time. Consequently, future research will need to look into alternative approaches other than indexing to reduce search times.

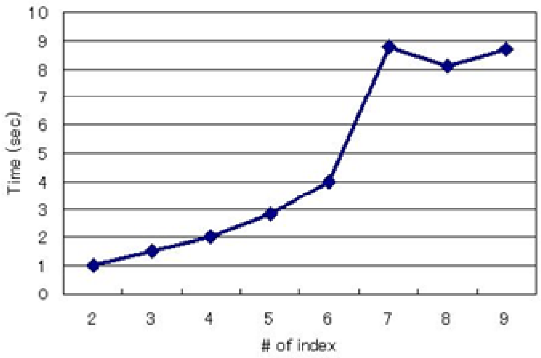


Figure 9. Calculation time as a function of indexing.

4. Conclusions

In the present study we sough to discover information on customer purchasing behavior by applying existing character string parsing techniques and applying them to stream data describing customer movements and obtained using RFID. In looking at customer movement data we chose to focus on visits that customers made to each product section. By then expressing product section visiting patterns in terms of character strings, we sought to efficiently handle large volumes of stream data. We found that HV customers, who purchase a relative large number of items tended to move from the vegetable section to the fish section. While hypotheses obtained this way are extremely novel and rich in their implications, we suffered from a rather small sample size, and future studies will be needed. Moreover, through this experiment, we were able to appreciate certain issues pertaining to existing character string parsing techniques.

Nevertheless, there are fundamental problems with applying the character string parsing techniques used in this study. Namely, time series information with respect to visiting patterns largely vanishes. For example, important information like the time spent at a particular product section or the amount of time spent moving from one section to another was not reflected in the character-string based knowledge representation. To resolve such issues, it seems that a fruitful approach might be to introduce graphical data. If graphical data were provided, one would be able to include not only product section visiting patterns but also time series information such as the amount of time spent between sections or at a particular section. We hope to address such issues in the future.

References

[1] Arikawa, S., Miyano, S., Shinohara, A., Kuhara, S., Mukouchi, Y. and Shinohara, T. (1993). A machine discovery from amino acid sequences by decision trees over regular patterns, *New Generation Computing*, Vol. 11, pp. 361-375.

[2] Asai, T., Abe, K., Kawasoe, S., Arimura, H., Sakamoto, H. and Arikawa, S. (2004). Efficient substructure discovery from large semi-structured data, IEICE Trans. on Information and Systems, Vol. E87-D, No.12, pp. 2754-2763.

[3] Guadagni, P. M. and Little, J. D. C. (1983). A logit model of brand choice,calibrated on scanner data, *Marketing Science*, Vol. 2, pp. 203-238.

[4] Gupta, S. (1988). Impact of Sales Promotions on When, What, and How Much to Buy, *Journal of Marketing research*, Vol. 25, pp. 342-355.

[5] Hamuro, Y., Katoh, N., Matsuda, Y. and Yada, K. (1998). Mining Pharmacy Data Helps to Make Profits, *Data Mining and Knowledge Discovery*, Vol. 2, pp. 391-398.

[6] Hamuro, Y., Katoh, N., Yada, K. and Washio, T. (2005). MUSASHI: System for knowledge discovery in large business data, Journal of JSAI, Vol. 20, pp. 59-66. (in Japanese)

[7] Hamuro, Y., Kawata, H., Katoh, N. and Yada, K. (2002). A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History, *Progress in Discovery Science*, LNAI 2281, 565-575.

[8] Hirao, M., Hoshino, H., Shinohara, A., Takeda, M. and Arikawa, S. (2003). A practical algorithm to find the best subsequences patterns, *Theoretical Computer Science*, 292, 465-479.

[9] METI (2005). "Japanese-version Future Store Project" (experimental trial of electronic tags for the realization of futuristic store service), News Release. (http://www.meti.go.jp/english/newtopics/data/n051108e.html)

[10] Larson, J.S., Bradlow, E.T. and Fader, P.S. (2005). An exploratory look at supermarket shopping paths, Int. *J. Research in Marketing*, 22, 395-414.

[11] Sorensen, H. (2003). The Science of Shopping, *Marketing Research*, 15, 30-35.

[12] Shimozono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara, S. and Arikawa, S. (1994). Knowledge acquisition from amino acid sequences by machine learning system BONSAI, Trans. Information Processing Society of Japan, 35, pp. 2009-2018.

[13] Witten, I.H. and Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementation*, Morgan Kaufmann.

[14] Yada, K. (2004). Business application of data mining for marketing field, *Journal of JSAI*, Vol. 19, pp. 376-377. (in Japanese)

[15] Yada, K., Ip, E. and Katoh, N. (2007). Is this brand ephemeral? A multivariate tree-based decision analysis of new product sustainability, *Decision Support Systems*, 44, 223-234.