An Effective Early Detection System for the Potential Marketable Items from the Initial Learning Datasets

Masakazu Takahashi, and Kazuhiko Tsuda

***Abstract*—**It is hard to detect the potential marketable items from the transaction data, all the more distinguish the marketable ones or not among the new released items. Concerning to the conventional extracting method is base on the item time-series analyses of the same category. On the other hand, one of the extraction methods for the issue is to make use of the recommendation algorithms. As for the recommendation functions, although the conventional recommendation system is in common among the Internet businesses, those engines provide the proper results; therefore, there need to improve recommendation algorithms. From those research backgrounds, this paper, we focus on detecting algorithms analyses for the marketable item among the new released items with improvement of the recommendation engine. We have gathered the ID (Identification)-POS (Point-Of-Sales) data from a local grocery store in Japan, and defined the index of the trend leader with the criteria for the day and the sales number. From the analysis of the gathered data, we have extracted trend leaders among the customers. Using the results of the recommendation engine, we are able to make detailed decisions in the following three points: 1) to make appropriate recommendations to the other group member based on the transitions of the trend leaders' preferences; 2) to evaluate the effect of the recommendation with the trend leaders' preferences; and 3) to improve the retail management processes: prevention from the stock-out, sales promotion for early purchase effects and the increase of the numbers of sales.

***Keywords*—**Collaborative Filtering System, Customer Preference, Dual-Directed Recommendation, Recommendation Systems, Service Science and Management Engineering

I. Introduction

There are various factors for the store operation inefficiency in the retail business, especially small to mid size companies in Japan; one of the reasons for the issue is luck of the timing understanding for the items that customers needs. For instance, it comes from the stock out of the items for insufficient understandings. This issue will affect the whole retail industry not only to the whole seller but the manufacturers; therefore, we need to understand both the preferences and the timings for the customers. As for the understanding for the preferences, the collaborative filtering system is one of the extracting algorithms for the preferences, and is adopted among the B2B or B2C industries such as Amazon or e-Bay [1].

This paper, we focus on the recommendation functions improvement to extract the high potential sales items from the trend leaders' activities with the ID-POS data and is composed of the following components; 1) Describe the conventional recommendation systems in the chapter II, 2) Propose the item detection algorithms from the ID-POS data in chapter III, 3) Performances evaluation of the proposed algorithms stored recommendation engine in chapter IV, 4)Algorithms application to the Dual-directed recommendation systems in chapter V, and 5) Conclusion of the paper.

Manuscript received Jan 31, 2009: Revised version received April 28, 2009.

Masakazu Takahashi is a graduate school student with the Graduate School of Business Sciences, University of Tsukuba, 3-29-1, Otsuka, Bunkyo, Tokyo 112-0012, Japan (phone: +81-3-942-6877; fax: +81-3-3942-6829; e-mail: masakazu@ gssm.otsuka.tsukuba.ac.jp)

Kazuhiko Tsuda is a professor with the Graduate School of Business Sciences, University of Tsukuba, 3-29-1, Otsuka, Bunkyo, Tokyo 112-0012, Japan (e-mail: tsuda@gssm.otsuka.tsukuba.ac.jp)

II. Related Works

This section, we focus on the related studies for understanding of the preferences, especially on recommendation functions. Concerning to the conventional recommendation systems are in common among the Internet businesses [2]-[4], the conventional recommendation engines provide the proper results, so that this issue should improve with new criteria in the case of applying to the retail industry. Most of the frequent visited customers of the retail industry are composed from with the diversity as following points; preferences, income, and number of family, therefore, we need to care generating the recommendation information not only with the properly but with the diversity.

A. Preferences Detection Algorithms

The collaborative filtering system is one of the understanding tools for the customer preferences and has following characteristics of the algorithms; Easy to data convert, Affordability for the huge data, such as the data depend system. On the other hand, this system has a tendency influenced for the majority answer and both the items and the customers need to extract the adequate recommendation information as the negative points [5] - [7]. Therefore, a large amount of information related to consumers is required. From this large amount of the information, appropriate characteristics are extracted and the users are categorized into an appropriate segment, on this bases and appropriate product group needs to be identified for long tail business. The collaborative filtering system is one of the techniques for providing suitable items from the customer's preference as the recommendation information among the many candidate items, and forecasts the customer's preferences from both the customer's activities and the purchase record. The feature of this method is to evaluate the customer's preference concerning item information and to generate the group of the customer preference with the similarity [1] [8].

Recommendation information is provides among items with the high evaluation from another customers but is not purchased yet with similar preferences. Therefore, the steps for generating the recommendation information is necessary as follows; (1) correct many purchase history for the customer, (2) Make the group and retrieve the customers who bought the same item, when a certain customer purchased the item, (3) Generate the recommendation information based on the item group among the same group customer. In this method, it doesn't analyze it concerning the content of contents at all.

Therefore, it has the feature that the restriction concerning the recommendation object doesn't exist, and the situation in which only similar information to information evaluated in the past is recommended can be evaded by using other customers' evaluation information. That is, collaborative filtering system is a mechanism that customer's community is generated only from the purchase information without the analysis of contents of the item.

B. Characteristics for the Collaborative Filtering System

Some of the problem issues for the collaborative filtering system are described. Denning et al noted that the collaborative filtering system is required both the large number of customers and the amount of contents. Therefore, the same recommendation result is generated if the items to recommend are little. Moreover, the existing items only to be recommended. This is the critical point to sell the new release items. From those past research, the improvement issues of the collaborative filtering system are summarized as follows [7] [9];

1) Both large number of customers and the amount of contents are required
2) Same recommendation information is generated if the items to recommend are little.
3) The items already bought only to be recommended.
4) Impossible to recommend to those who have preferences different from the ordinary.
5) Cannot prevented from influencing from mis-inputting
6) New item is recommended until someone gives the evaluation even if the item was registered
7) Can not connect the customers with the different ID even though they have the similar contents for without analyzing for the contents,

Concerning to the forecasting for the new items sales number, Nakamura evaluated and classified the characteristics of the new items with the market reflection data such as the ID-POS data [10]. The conventional evaluation methods such as the trial repeat model only indicates the characteristics for the items that purchased repeatedly but, are insufficient for the customer classification or the recommendation from the ID-POS data classification [11].

III. Extraction from the Transactional Data

One of the conventional methods for understanding the timing for the customer needs is the time-series data analysis with the quantitative data such as the POS data. These methods are good for the inventory controlling such as the stable demand from the customer needs because of the sufficient time-series data and able to mark high score ratio for the demand forecasting. On the other hands, it is hard to forecast the demand such as the new items or the items that demand rapidly stood up because of the insufficient of the forecasting base data.

A. Sufficient Recommendation Criteria

To generate recommendation information with understanding both the customer preferences and the timing for the needs, it is required to not assuming the existence of the items; therefore, it comes to consider the existence for the new items. Moreover, it is expected to prevent the stock out to share the demand information among the retail industries; therefore, we referred Dual-Recommendation because of providing the recommendation information not only to consumers but also to manufacturers and retailers [12]. Therefore, this paper proposes the algorithm that detect to the timing of the item needs efficiently with the ID-POS data gathered from the local super market in Japan. Especially, to understand the timing of the item needs, we make use of the concept of the trend leader to forecast the demand of the new item. It is difficult to detect the ability with the trend leader of the item selection that foreseen the fashion in advance from the numerical data such as the ID-POS. Moreover, the candidates for the trend leader include the customer only with short span curiosity, so that an efficient selection method of detecting the trend leader is required.

B. Gathering Data

For the purpose of the detecting the items that the trend leader took, we make use of the ID-POS data gathering from the local grocery store in Japan. TABLE I indicates the gathering data attributes.

TABLE I
GATHERING DATA ATTRIBUTES

Term	2007/05/01～2007/12/31
Data	POS Transaction
	Reward Card Transaction
Category	Super Market (Grocery Store)
Site	Shimane pref., Japan

Based on the above gathered data set, the demography, the number of the transactions, the number of frequent customers and the gross number of the transaction are observed. TABLE II indicates the transaction results from the gathered data. There are almost about 10,000 customers frequently visited the store with the about 70,000 transactions in total observed from the transaction. Averagely, there are about 10 items in one transaction.

TABLE II
TRANSACTION RESULTS

	Transactions	Card Members	Purchased Items
2007/05	69,106	9,766	689,592
2007/06	69,681	9,970	689,686
2007/07	68,853	9,791	658,716
2007/08	69,275	9,832	674,667
2007/09	69,294	10,021	680,102
2007/10	71,403	9,980	690,444
2007/11	67,182	9,872	651,723
2007/12	71,192	10,279	705,600
	555,986	79,511	5,440,530

TABLE III indicates the number of the new registered items during the period. There are 1,922 items registered, among them, 1,738 items sold at least one piece after the release and among 1,008 items have been increased the sales number from the release to 28 days later. Those who purchased 1,008 items amount 9,141 frequent customers.

TABLE III
TRANSACTION ATTRIBUTES

Master Database Items in POS	18,091
New Registered Items	1,922
Include Purchased Items	1,738
Include Increased Items	1,008
Active Reward Card Member	9,141

Fig.1. indicates the relation between the number of Items for the new registered, the purchased items, and the increased items for each month, based on the TABLE III. As for the September, 425 items have registered because of the seasonal change from summer to fall, therefore, there are about twice of items registered among the gathering data.

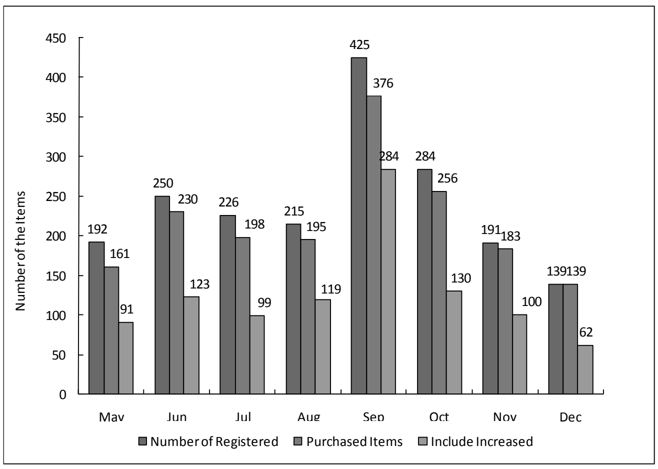

Fig.1 Number of the Registered, the Purchased, and the Sales Increased Items

Fig.2 indicates the percentage of the Items for the purchased items and the sales increased items. From the figure, about 50% of the newly registered items could not increase the sales number. From the figure, it is hard to select the items that will increase the number of sales in the future in advance.

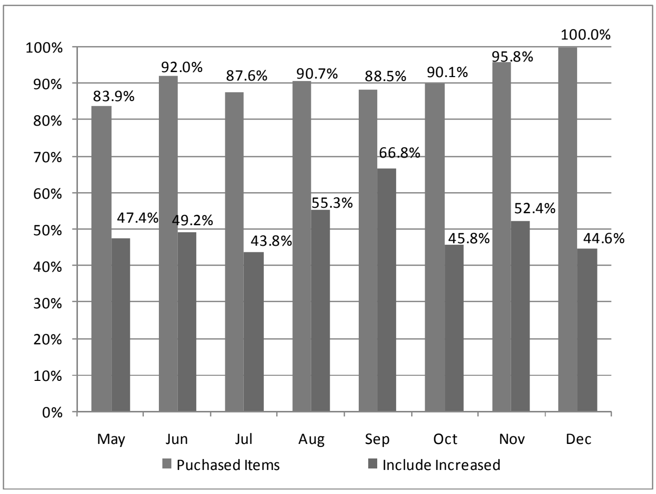

Fig.2 Percent of the Purchased, the Sales Increased

From the observations of the gathering data, only 50% of the new released items could not increase number of the sales, it is complicated to extract the items with the potential sales number because of the combinations of the best-before date of the items and the customer preferences. Therefore, there need some extraction steps from the numerical data.

C. *Extraction Criteria*

The conventional forecasting methods for the new item sales number have been based on the sales results of the similarity. It is good for the new items that have fewer criterions both to avoid the stock out and to achieve the large scale sales with some indexes for forecasting of the maximize sales items from the initial demand. At the same time, this forecasting will good for the inventory controlling to avoid the stock out with the understanding the timing for the customer needs with sharing the item demand information among the retail industries; the whole sellers and the manufactures. Then, this paper focuses on extracting the items that can be maximized sales number with the index of the trend leader. One the other hand, it is difficult to detect the ability with the trend leader of the item selection that foreseen the fashion in advance from the numerical data such as the ID-POS.

We have defined the trend leaders who bought the items at the initial stage were able to maximize the sales amount and defined the index of the trend leaders from those customers' activities. To measure the indexes of the trend leader for each item, we take the following criteria;

· How first purchased items from the release

· How many purchased items that increasing of the sales number

D. *Extraction Algorithms*

To extract the marketable items, we set the trend leaders who bought the items at the initial stage were able to maximize the sales amount and defined the index of the trend leaders from those customers' activities.

As for the extraction steps of the marketable items with the ID-POS data based on the number of sales from the release with the following steps;

1. Item based calculation

Step1 Count the date for the release

Step2 Count amount of sales for 28days from the release

Step3 Count amount of sales for 35days from the release

Step4 Subtract the result of Step3 from the result of Step2

Step5 Extract the items of Step 4 > 0

Extraction with high support items from the customers from the amount of the sales

2. Customer based calculation

Step6 Count the number of the sales

Step7 Correct the purchased date for each item

Step8 Subtract the date for start selling from the purchased date for each item

Calculate the index of the early purchase

Step9 Summation for the reciprocals of the early purchase date

For the calculation of the early purchase function

Step10 Count the items of the step5 among the step 6

Step11 Calculate Step9 / step6

Let N_i as the new item that the customer i purchases and M_i as the items that increased the sales number, the index of the trend leader (ITL) is given as follows;

$$ITL = \frac{\sum_{j \in M_i} \frac{1}{t_j}}{N_i} \quad (M_i \in N_i) \tag{1}$$

E. Trend Leader Extraction

From the above extraction steps, we have obtained the following results. TABLE VI indicates the demography for the Index of the trend leader (ITL) based on the extraction procedure. From the table, the maximum of the ITL was 6.91, and minimum ITL was 0 and 8,707 customers in 9,141 customers became 1 or less for the index. The customer exceeded 6 or more was 3.

TABLE IV

DEMOGRAPHY FOR THE ITL

Index	Customer	Percent
~1	8,707	95.25%
~2	360	3.94%
~3	45	0.49%
~4	20	0.22%
~5	4	0.04%
~6	2	0.02%
~7	3	0.03%
total	9,141	

Fig.3 indicates the relation between the number of the purchased items and the index of the trend leader. Most of the customer purchased less than 50 items and scored the low index of the trend leader. Otherwise, not always scored high index of the trend leader but purchased lot. For example, those who scored more than 4, the number of the purchased items varied wide.

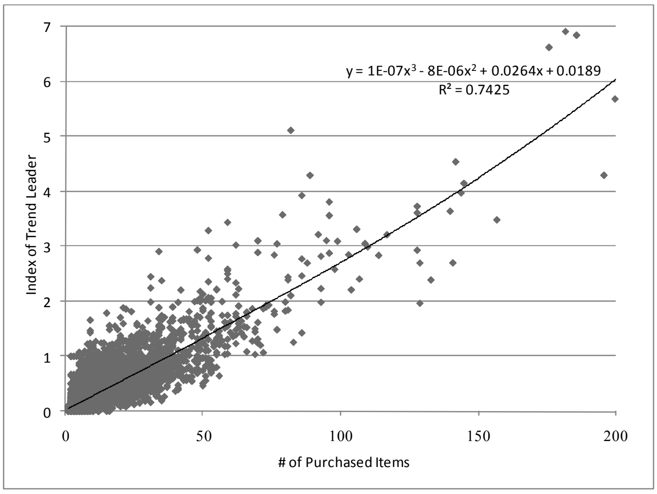

Fig.3 Number of Purchased Item and the ITL

Fig.4 indicates the relation between the ITL and the Area Code for the customers' living location. This area code shows the distance from the store. For example, the more the area code number gets, the more far away from the store. Namely, the customers have the area code of zero are the nearest location from the store.

From the figure, if we set the temporary threshold of the trend leader scored more than two, we can understand the following issues; 1)Except the exceeded score for the trend leader, those who scored high index usually locates the near the store. 2) Almost all the locations have the trend leader.

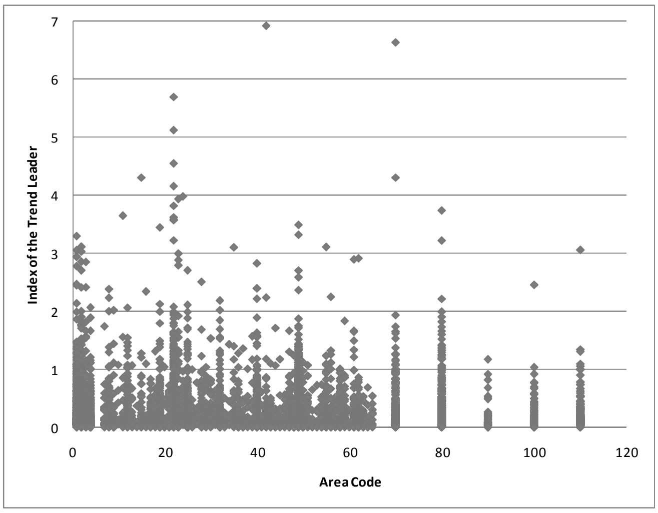

Fig.4 the ITL and Area Code

From the relation between the demography of the ITL and the other attributes, depending on the threshold for the trend leader, we can use the criteria as the key index for the recommendation based on the preference and the timing for the needs.

To understand both the customers' preference and the purchase timing at once, we have set up the dual-directed recommendation system that is able to consider both the new items and the stock outs with the Taste [13]. We have defined the trend leaders who bought the items at the initial stage were able to maximize the sales amount and defined the index of the trend leaders from those customers' activities and put those into the recommendation filters. This recommendation engine based on the functions for the proposed extracting method of the trend leaders is with the efficiency and with the diversity compared to the conventional recommendation information.

IV. Recommendation Engine Evaluations

The index of the trend leader is one of the reference values, so that it is necessary to set the threshold of the index properly according to characteristics of the region, items, customer attributes, and sales policy. Then, index of more than 2 is assumed to be a threshold, both to extract the trend leader and to recommend the items. In this paper, we took Taste for the recommendation engine for the popular recommendation open source software. As for the method for the similarity, the correlation coefficient was originally used, and changed the similarity method into the cosign distance from the correlation coefficient for holding many indexes of the similarity.

A. Similarity Evaluations

As the criteria for the collaborative filtering systems, both the precision and the recall are mostly used for the index of the accuracy and the coverage for the recommendation results [14]. Both the precision and recall relevance were calculated by 10 fold cross-validation in this paper. Fig.5 indicates the relation for both recall and precision based on the validation. This indicates 74 customers scored the index more than 2. This relation indicates the possibility for applying the recommendation criteria change the index of the similarity to the cosign distance from the correlation. Therefore, we take the cosign distance for the recommendation similarity.

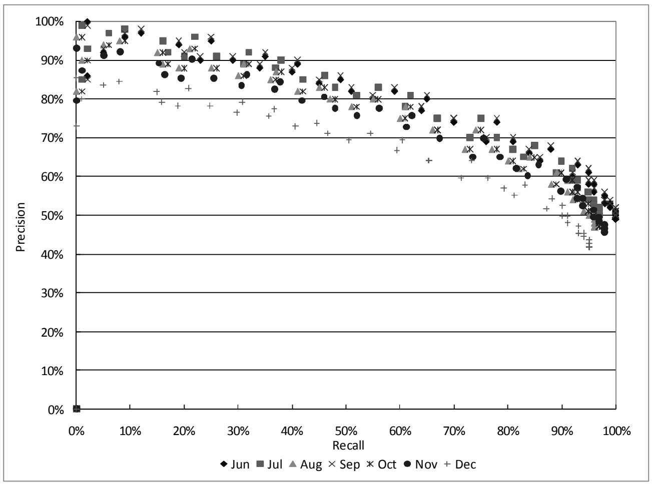

Fig.5 Relation between the Recall and the Precision

B. Collaborative Filtering Systems Evaluations

Fig.6 indicates the demography of the relation for the average increase ratio for the gross purchase number and the customer ID. The value of the maximum scored 4.87 % rise with the customer number of 25 in November. On the other hand, the minimum scored 0.58% rise with the customer number of 29 in July. This graph also indicates the diversity for the preference of the customer even for the high scored index.

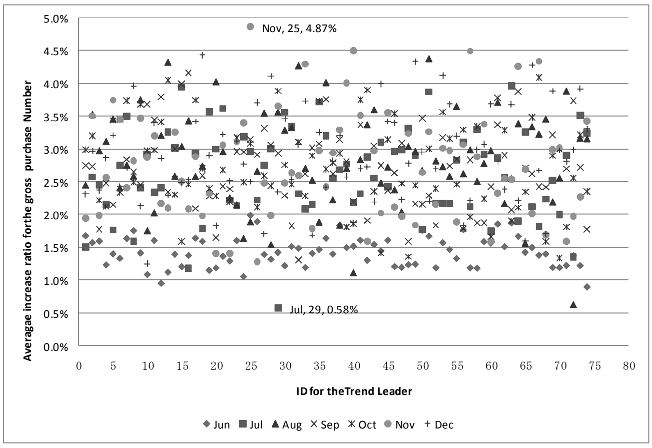

Fig.6 Average Increasing Ratio

Fig. 7 also indicated the time-series of the average increase ratio for the certain customer numbered 12. This customer scored 3.107 for the Index of the trend leader, this score is upper level compared from the demography of the index, besides, and this customer could not keep the high score of the sales increase, averagely scored 2.23% for the series as the result. This is because of the reason for the diversities of the customers' preferences. Therefore, we need to generate dynamic recommendation results.

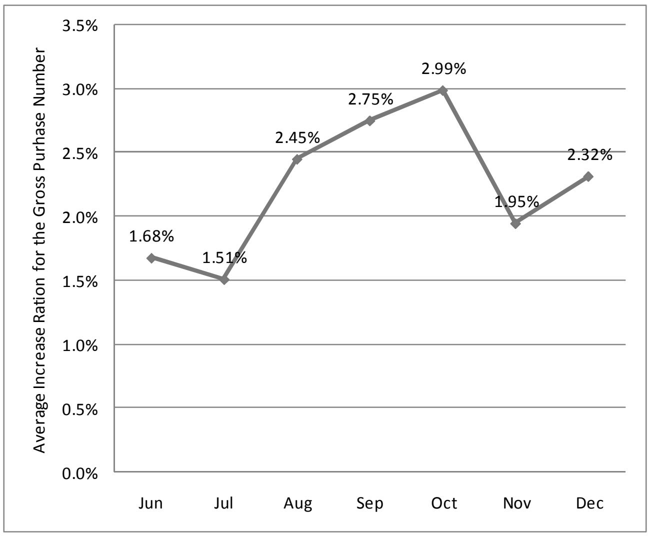

Fig.7 Average Increasing Ratio for the Customer ID of 12

Fig.8 indicates the results for the evaluation between the default recommendation engine and the improved one. The w/TL indicates the recommendation based on the index of the trend leader in the figure. This indicates the average increase ratio for the sales number. From the figure indicates the sale increase at least 1.43% rise with the proposed algorithms.

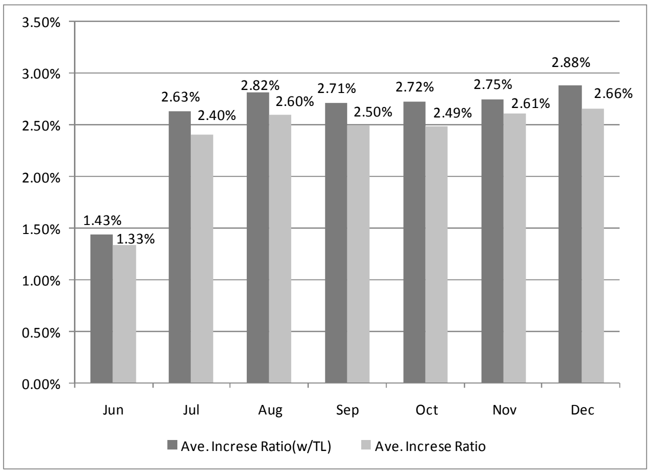

Fig.8 Average Increasing Ratio for the Gross Purchase Number

Fig.9 indicates the average early purchase effect. To recommend the items that the trend leader took are expected the distinct results. This early purchase effect is good for not only recommendation for the right time but ready for the stock out of the items.

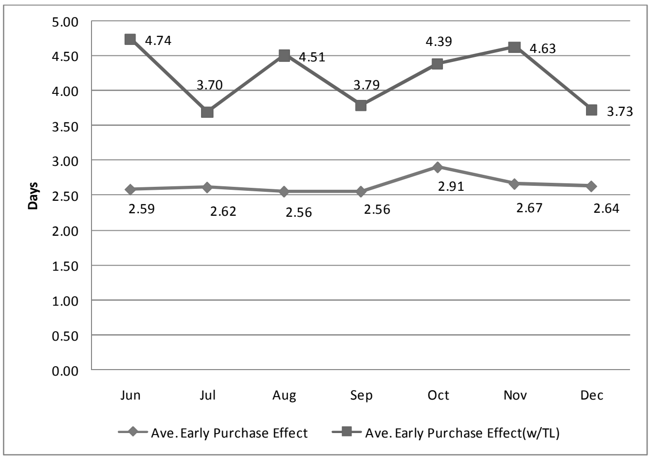

Fig.9 Early Purchase Effect

From the results of the evaluation with the recommendation based on the trend leader, we have figure out the following issues; As the benefits for the recommendation with proposed indexes, detect the items that will increase sales number in the future,

1. Prevention from the stock out
2. Sales improvement for early purchase effect
3. Increase of the sale number

As a result, we have succeeded to extract both the trend leaders from their purchase activities and the customer groups that became the trend leaders' candidates from the ID-POS data.

We found out the items that will become the high potential for the increasing sales number in the future with the proposed indexes. The recommendation information from the trend leaders with high score made increase the sales number and shortened the days until purchasing.

C. Evaluation for the System with the Short Best-before Date Item

This section, we evaluate the recommendation performance with the item of new released milk that obtains the short best-before date at least one week or so. This item is labeled number for 4902081205724 in 13 digits based on the JAN (Japanese Article Number) regulation. This unique code is issued by each company and searched by the GEPIR (the Global Electronic Party Information Register) site [15]. This unique number is composed of the following attributes in case of 13 digits; 1) country code for 2 digits, 2) company code 7 digits, 3) item code for 3 digits, and 4) chick digit for 1 digit. For this unique identification number enable us to identify the transaction of the item. For example, applying to this classification to the item of the simulation, 49 for the country code of Japan, 02008120 for the company code, 572 for the item code, and 4 is for the check digit respectively. With the 74 customers scored more than 2 of the index of the trend leader, we obtain the following results with the 10-fold cross-validation. Fig.10 indicates the relation between the increase ratio for the item and the index of the trend leader.

This indicates the performances of the proposed index. From the graph, we obtained that the more scored of the index of the trend leader, the more scored high increase ratio for the gross sales number except the around scored 5. This result of the exception comes from the preferences of the each customer. This result obtains the slightly different performances with the customer preferences and shows the diversity for the preference of the customers.

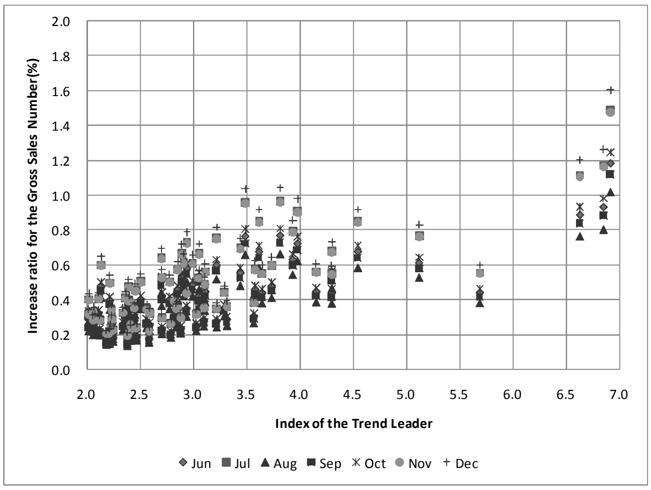

Fig.10 Increase Ratio for the Sales Number (Short Best-before Date Item)

Fig.11 indicates the relation between the early purchase effect and the index of the trend leader. This also indicates the performances of the proposed index with the early purchase effect. From the graph, we obtained the following issue; the more scored of the index of the trend leader, the more scored earlier. This also obtains the slightly different performances with the customer preferences. This means the diversities for the preferences of the customers.

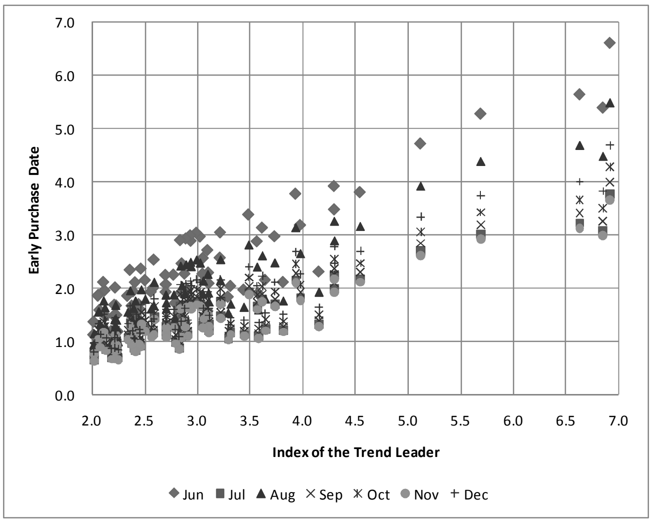

Fig.11 Early Purchase Effect (Short Best-before Date Item)

TABLE V and TABLE VI indicate the recommendation results of the item with the proposed index. As for the increased ratio for the item sales number scored between 0.13% and 1.60% rise wit the average of 0.44% and the early purchase effect marked between 0.62 day and 6.61 days with the average of 1.71 days earlier than the default. The recommendation information from the trend leaders with high score made increase the sales number and early purchased days. We found out the items that will become the high potential for the increasing sales number in the future both the proposed index and the short best-before date item.

TABLE V
RECOMMENDATION RESULTS FOR THE INCREASE RATIO (%)
(SHORT BEST-BEFORE DATE ITEM)

Max	Min	Average	Standard Deviation
1.60	0.13	0.44	0.23

TABLE VI
RECOMMENDATION RESULTS FOR THE EARLY PURCHASE EFFECT (DAYS)
(SHORT BEST-BEFORE DATE ITEM)

Max	Min	Average	Standard Deviation
6.61	0.62	1.71	0.84

D. Evaluation for the System with the Long Best-before Date Item

Next, for the performance evaluation of the system, we took the item of new released pops that obtains the long best-before date at least over month labeled number for 4902179011725 with the same conditions. Fig.12 indicates the relation between the increase ratio for the item and the index of the trend leader with the long best-before date item. From the figure, we found out the item with the long best-before date item could not affect the sales increase. Fig.13 also indicates the relation between the early purchase effect and the index of the trend leader with the long best-before date item. From the both graph, we obtained from the simulation that the more scored of the index of the trend leader, the more scored early purchase effect and the low effect for the increase ratio for the item with the long best-before date item.

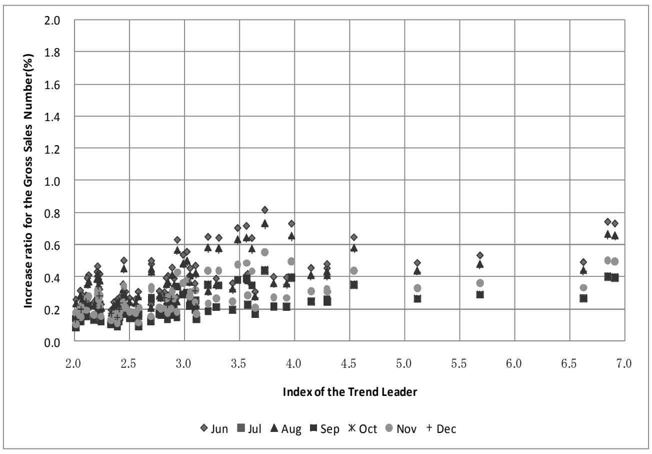

Fig.12 Increase Ratio for the Sales Number (Long Best-before Date Item)

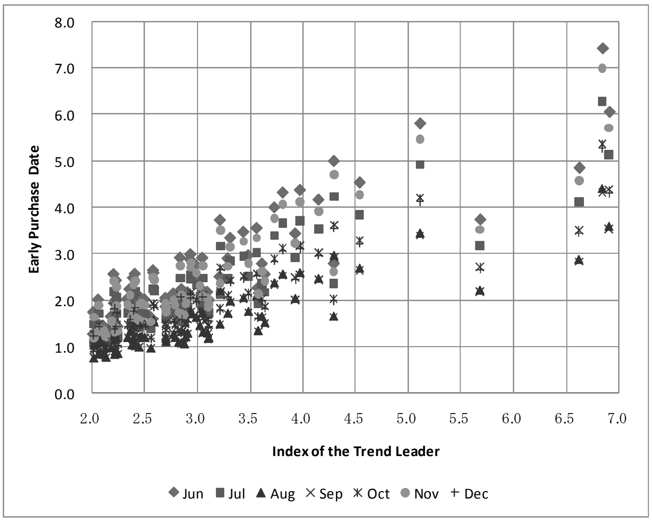

Fig.13 Early Purchase Effect (Long Best-before Date Item)

TABLE VII and TABLE VIII indicate the recommendation results of the item with the proposed index. As for the increased ratio for the item sales number scored between 0.09% and 0.82% rise wit the average of 0.29% and the early purchase effect marked between 0.74 day and 7.44 days with the average of 2.03 days earlier than the default. The recommendation information from the trend leaders with high score made early purchased days effect. We found out the items that will become the high potential for the early purchase effect with both the proposed indexes and the long best-before date item.

TABLE VII
RECOMMENDATION RESULTS FOR THE INCREASE RATIO (%)
(LONG BEST-BEFORE DATE ITEM)

Max	Min	Average	Standard Deviation
0.82	0.09	0.29	0.14

TABLE VIII
RECOMMENDATION RESULTS FOR THE EARLY PURCHASE EFFECT (DAYS)
(LONG BEST-BEFORE DATE ITEM)

Max	Min	Average	Standard Deviation
7.44	0.74	2.03	1.01

As the benefits for the recommendation with the proposed indexes, we have succeeded detect the items that will increase sales number in the future, and extract the customer groups that became the trend leaders' candidates especially to good for the short best-before date items.

E. Recommendation Performances Comparison

This section, we evaluate the recommendation performances. Both TABLE IX and TABLE X indicate the comparison result of the recommendation of the Best-Before date. From the comparison indicates the follows; 1) Concerning to the sales number increase effect, the short best-before date item is obtained better recommendation results than the long best-before date. 2) Concerning to the early purchase effect is not observed the difference between the short and the long best-before date.

TABLE IX
RECOMMENDATION RESULTS FOR THE INCREASE RATIO (%)

Item	Max	Min	Average	Standard Deviation
Short Best-before Date	2.18	0.17	0.56	0.30
Long Best-before Date	0.74	0.08	0.26	0.11

TABLE X
RECOMMENDATION RESULTS FOR THE EARLY PURCHASE EFFECT (DAYS)

Item	Max	Min	Average	Standard Deviation
Short Best-before Date	7.96	0.89	2.39	1.07
Long Best-before Date	8.60	0.98	2.68	1.24

As for the early purchase effect, not only indicates the recommendation to the customers as the purchase hits but the early detection for the commodity items that will become the sales increase effect for the store. This early detection will be good for the inventory controlling such as early ordering. Even for the staff from the retail industry, it is important to indicate even for one day of the early detection for the recommendation items. Because, they are usually controlling and ordering about 1,000 items in one category, and there are mixed frequency for the sales number, that makes hard to find out the potential sales items. Therefore, it is important for the early detection.

Form the evaluation for the recommendation results, in the case of applying to the posed indexes, the selection for the items based on the recommendation results that have the recommendation effect is required. That is, both the attributes for the customer and the recommendation algorithm matched to the customer preference in making use of the index of the trend leader is necessary. From those requirements of the above condition, we have set up the portal site for the dynamic recommendation based on the index of the leader. This function is good for the early detection for the potential sales items.

V. DUAL-DIRECTED RECOMMENDATION

We have set up the portal site with the recommendation functions stored the index of the trend leader as the application for the algorithms for dynamic recommendation. This portal site is named the DAIKOC (Dynamic Advisor for Information and Knowledge Oriented Communities) [16]. This portal site is currently operates for the effective recommendations for the actual retail store in Japan.

A. Concept for the DAIKOC Systems

First, we describe the backgrounds for the site development. Long tail business was referred originally by C.Anderson argued from his book [17] that the products that have low sales volume can collectively make up a market share that exceeds the relatively few current bestsellers, if the distribution channel is large enough. Lots of the recommendation systems in many of the commercial Web sites adopt the collaborative filtering systems. The primary goals of traditional techniques of collaborative filtering systems are improving the accuracy of the recommendation. Although they include many items the users have already known. In consideration of the accuracy, these recommendations appear adequate. But on the contrary, if we consider users' satisfactions, they are not enough adequate because of the lack of discovery. "Preference" is enumerated in one of the factors to decide one's behavior. Therefore, it is important to model the preference for becoming the variety of basic recommendation information. Among the adjustable recommendation information with user's preference, it is important for recommendation information not only fit user's preference but has both the unexpectedness and the surprise. With the proposed algorithms and the application of the DAIKOC system, can cover the customers' preferences with the capricious preference changes.

Long tail businesses usually refer only to the E-businesses of large scale firms, in which they focus on small groups from among so many consumers. Actually, large scale E-commerce traders, such as Amazon and E-bay only sell pre-existing sales stuff. Also, E-business firms, such as Google and Yahoo only provide information search services in the Internet. However, from the viewpoint of producers or retailers of small business firms, there have been few discussions about long tail businesses. Although they were not able to get information of consumers, so far, they have considered it is impossible for small business firms to utilize recommendation information of the long tails of consumers. We believe that, using cutting-edge information and communication technologies, we are able to realize a long-tail business model for small sized firms. Especially in Japan, in order to effectively develop long tail business for small sized firms, all consumers, manufacturers, and retailers must develop their flexible relationships and information sharing. To cope with the issues, we are employing the approach to develop a common portal site to share the long tail business information. DIKOC uses agent-based software technology in order to collect, analyzes, and provide recommendation information for them. We will implement a match-making functionality to bridge the gaps among consumers, manufacturers, and retailers. DAIKOC system will have the following components:

- Databases to store the characteristics of targeted consumers, manufacturers, and retailers;
- Components to gather requirements on new products or services collected from small groups of long-tail consumers;
- Brokerage functions of these requirements among consumers, manufacturers, and retailers;

and

- Generators for dual-directed recommendations focused on temporally changing information of latent consumer groups.

B. *Configurations for the DAIKOC Systems*

Based on the above architectures, we have set up the site. Fig.14 indicates the configuration for the DAIKOC site. This indicates the relation for both the actual purchase and the recommended items to each customer. The vertical axis show the category for the items and the horizontal axis indicates the customer. We can figure out the tendency view for the relation both of the purchased items and the recommended items from this site. This is commonly used for the daily operation of the inventory controlling. As for the numbers in the figure are as follows; 1) Date, 2) Purchased items and Recommended items from the proposed indexes, 3) Clients axis, and 4) Category List. Among the graph of the black dots indicates the purchased items, ten dark dots indicates recommended items but not purchased yet, and the light and white dots shows the other items stocked in the store respectively.

Fig.14 Recommended Information

DAIKOC system has the following components with match-making functionalities to bridge gaps among consumer, retailer, and manufactures. We referred the Dual-recommendation because of providing the recommendation information not only to consumers but also to manufacturers and retailers. The recommendation information system is analyzed from the ID-POS data and can offer recommendation information generated with the trend leaders' algorithms. Not only indicates the whole trend from between the purchased items and there recommended, but this system has search functions for the personal activities that show the display for both the actual purchases and the recommended items of a certain customer. Fig.15 indicates the display for both the actual purchase and the recommended items of a certain customer. For each customer, the system displays both purchased and recommended items. From the actual purchase matrix, we can figure out items, prices, and quantities. The recommended items show the index for the strength of the recommendation, respectively.

As for the numbers in the figure are as follows; 1) Date and Customer ID, 2) Purchased items 3) Recommended items base on the proposed indexes.

Fig.15 Recommended Information

From this recommendation information generates the new knowledge for customer, retailer and manufacture. This information is able to obtain not only the recommendation to the customers but the relation the items with the customers and becomes the hints in time for the ordering, manufacturing and inventory controlling. This will be the information sharing function with the dual-direction among the retail industry.

C. *Tabu-GA in DAIKOC Systems*

The ID-POS data is composed from the customers varied with the following attributes; number of family, income, and food preferences. The conventional recommendation systems based on the similarity of purchase trends, therefore, it is difficult to generate adequate recommendation information and to expand the amount of sales, if there are frequent purchased items. To generate the adequate recommendation information fits with the customer preference from the above the attributes, the recommendation engine of DAIKOC stored the Tabu-GA [18] [19] algorithms inside to make the efficient results to the customer based on the index of the trend leader. This engine is composed both of the tabu search component and the GA (Generic Algorithms) component, with the customer ID stored as the generic information.

Fig.16 indicates the outline for the extracting steps.

1. Set H Empty, H which is a historical memory.
2. Select $x^{now} \in X$ as an initial solution.
3. Choose $selection\_N(x^{now}) \subset N(H, x^{now})$, where $N(H, x^{now})$ is a set in $x \in X$ except in the neighborhoods of H.
4. Select $x^{next} = \max(c(H, x^{now}))$, $x^{next} \in selection\_N(x^{now})$, where $c(H, x^{now})$ is an objective function is a mapping of a set in $x \in X$ except in the neighborhoods of H.
5. Run GA.
6. If a condition of ending is true then end
7. Exchange x^{best} for x^{old} in H. Return to 3.

Fig.16 Tabu-GA Algorithm

Multipronged optimal solutions can be led with this GA algorithm, even though with the conventional GA algorithm can be led one optimal answer. The main idea of the algorithm is that, (a) in each generation, one best individual generated by GA operation is stored into the tabu-lists to inhibit it from selecting specified times, and (b) solution candidates found in the previous generations will become tabu, and thus, the other candidates are explored in order to get better and divergent solutions.

VI. Conclusion

In this paper, a basic research project in relation to the ID-POS transactional data analysis was described. Finding out the formula for extracting the trend leaders among the customers, that which confirm that there is a possibility to make appropriate recommendations to the other group member based on the transitions of the trend leaders' preferences. Result from the simulation, we have succeeded to extract the trend leaders among the customers, that which confirm that there is a possibility to make appropriate recommendations to the other group member based on the transitions of the trend leaders' preferences and confirmed the effect of the recommendation with the trend leaders' preferences. Using the results, we are able to make detailed decisions in the following three points: 1) to make appropriate recommendations to the other group member based on the transitions of the trend leaders' preferences; 2) to evaluate the effect of the recommendation with the trend leaders' preferences; and 3) to improve the retail management processes: prevention from the stock-out, sales promotion for early purchase effects and the increase of the numbers of sales.

And making use of the index, we evaluate with the item both of the long best-before date and the short best-before date, finding out the short best-before date item is better for sales expansion than the long best-before date.

Acknowledgment

This research was with the courtesy support of the local super market in Japan. We express appreciation to those involved.

References

[1] NetPerceptions, "Recommendation Engine White Paper", 2000 Available:http://www.netperceptions.com

[2] G.Linden, B.Smith, and J.York, "Amazon.com Recommendations; Item-to-Item Collaborative Filtering", *IEEE INTERNET COMPUTING*, Jan-Feb, 2003, pp.73-80

[3] L.V.Orma, "Consumer Support Systems", *Communications of the ACM*, Vol.50, No.4, 2006, pp.49-54

[4] Y.Hijikata, "Techniques of Preference Extraction for Information Recommendation", *Journal of Information Processing Society of Japan*, Vol.48, No.9, 2007, pp. 957-965

[5] R.Burke, "Hybrid Recommender Systems: Survey and Experiments", *User Modeling and User-Adapted Interaction*, Vol.12, 2002, pp.331-370

[6] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", *IEEE Trans. on Knowledge and Data Engineering*, Vol.17, No.6, 2005, pp.734-749

[7] J.Herlocker, J.Konstan, L.Terveen, and J.Riedl, "Evaluating Collaborative Filtering Recommender Systems", *ACM Transactions on Information Systems*, Vol.22, No.1, 2004, pp.5-53

[8] J.B.Schafer, J.A.Konstan, and J.Riedl, "E-Commerce Recommendation Applications", *Data Mining and Knowledge Discovery*, Vol.5, 2001, pp.115–153

[9] P.J.Denning and R.Dunham, “The Missing Customer”, *Communications of the ACM*, Vol.50, No.4, 2006, pp.19-23
[10] H.Nakamura, “Marketing of the New Products”, Chuokeizai-Sha, 2001(Japanese)
[11] M.Abe and F.Kondo, “Science of Marketing, POS Data Analysis”, Asakura Publishing, 2005 (Japanese)
[12] M.Takahashi, T.Nakao, K.Tsuda, and T.Terano, “Generating Dual-Directed Recommendation Information from Point-of-Sales Data of a Supermarket”, *Vol.5178 of LNAI*, Springer-Verlag, 2008, pp.1010–1017
[13] Available:http://taste.sourceforge.net
[14] B.Sarwar, G.Karypis, J.Konstan, and J.Riedl, “Application of Dimensionality Reduction in Recommender System”, *ACM Web KDD Workshop*, 2000
[15] Available:http://www.gepir.jp
[16] Available:http://www.daikoc.net
[17] C.Anderson, “The Long Tail: Why the Future of Business Is Selling Less of More”, Hyperion,2006
[18] M.Takahashi and S.Kurahashi, “Tabu Search Algorithms for Multimodal and Multi-Objective Function Optimizations”, *IJCSNS*, Vol.7, No.10, 2007, pp.257-264
[19] F.Glover and M.Laguna, “Tabu Search”, Kluwer Academic Publishers, 1998

Masakazu Takahashi became member of IEEJ (The Institute of Electrical Engineers of Japan), IPSJ (Information Processing Society of Japan), and KES (Knowledge-Based and Intelligent Engineering Systems) International in 2008 respectively. He was born in Tokyo, 1967 and received M.B.A. degrees in Operations Research from University of Tsukuba, Tokyo in 1996. Now he is a doctoral candidate from the Graduate School of Business Sciences, University of Tsukuba. His working experiences were composed both of database business and marketing. He was with the Nikkei Research Inc., a subsidiary of the Nihon Keizai Shimbun Inc., a newspaper company in Japan and with some consulting company from1992 to 2003. During this period, he was a research member from CMHC (Canada Mortgage and Housing Corporation, Canadian Federal Government Agency), was assigned to figure out market potentialities of the Japanese housing market for the Canadian Housing Industries. Now he operates software company specialized in Artificial Intelligence and Operations Research, titled in CEO. His research interests include Optimization, Data Mining and Recommendation Systems.

Kazuhiko Tsuda became a member of IPSJ in 1988 and IEICE (The Institute of Electronics, Information and Communication Engineers) in 1991. He received his BA in 1986 and PhD in 1994, both in engineering, from Tokushima University, Japan. He was with Mitsubishi Electric Corporation from 1986 to 1990 and with Sumitomo Metal Industries Ltd. from 1991 to 1998. He is a professor from the Graduate School of Systems Management, University of Tsukuba since 1998. His research interests include natural language processing, database, and human-computer interaction.