2010 13th International Conference on Network-Based Information Systems

Application of the RFID Data Mining to an Apparel Field

Shigeaki Sakurai
Corporate Research & Development Center
Toshiba Corporation
Kawasaki, Japan
Email: shigeaki.sakurai@toshiba.co.jp

Masanori Sanbe
Distribution System Laboratory
Toshiba TEC Corporation
Izunokuni, Japan
Email: Masanori_Sanbe@toshibatec.co.jp

Katsutoshi Watanabe
Tokyo Main Branch
Toshiba TEC Corporation
Tokyo, Japan
Email: Katsutoshi_Watanabe@toshibatec.co.jp

***Abstract*—This paper proposes a new method that efficiently uses the RFID data collected from apparel shops. This method learns prediction models from the data by using data mining techniques. The models represent relationships between the number of sales in the next term and the actions of customers, such as the number of pick-up, the number of fitting, the number of customers, and so on. It is possible to predict sales volume by applying the present RFID data to the models. This paper verifies the efficiency of the method through numerical experiments based on the RFID data collected from two branches of an apparel company.**

***Keywords*-RFID; Apparel Shop; SVM; Predication of Sales Volume;**

I. Introduction

Owing to the progress of RFID (Radio Frequency IDentification) readers and RFID tags, it is possible to easily bury them in various objects and places. Large amount of RFID data will be collected from sensor networks composed of them in near future. We anticipate that the data can be used to help our decision making in various situations. We have such high needs that speedily access the data and appropriately analyze it. Various methods activating the RFID data have been studied aggressively.

For example, Akahoshi et al. [1] propose a method that accesses the data collected from sensor networks. The method uses the framework of traditional relational database. The sensor networks are composed of apparatuses dynamically set in the networks. Dass and Mahanti [5] propose a method that discovers frequent patterns from data collected in real time. The method combines two strategies: the strategy of wide priority and the one of depth priority. Ihler et al. [7] propose a method that generates an explanatory model of abnormal events. The model is based on the Pearson distribution. Also, they propose a method that statistically infers the parameters of the model from sequential data. Kuramitsu [8] proposes a method that discovers repeatedly observed patterns collected from various sensors. Also, he proposes a method that extracts abnormal patterns. Teng and Lin [12] propose a method that efficiently discovers patterns by expressing the data attached the time with the tree structure. Here, the patterns are related to the time interval between actions and events representing the actions. In addition, Sakurai et al. [10] [11] propose methods that efficiently discover patterns from sequential data, even if the methods are not always limited in the analysis of the data collected from sensor networks. The pattern is a sequential row of events and reflects the interests of users. Here, the interests are described based on time constraints and events related to the interests.

Next, we introduce some application examples of the RFID readers and the RFID tags in real world environments. Wal-Mart [9], Metro group [13], and Marks & Spencer [4] introduce them to efficiently manage items processed in the supply chain. Here, Wal-Mart is the biggest supermarket company in the world and the headquarter is located in the USA. Metro group is one of famous companies in the retail field and mainly performs the business in the Europe. Marks & Spencer is the biggest retailer in the UK. In addition, some experiments in the publisher field and the apparel field are performed in Japan. Mitsukoshi, one of famous department stores in Japan, introduces the RFID readers and the RFID tags at the apparel shop dealing with imported casual items. It operates them in daily business. The application examples show that the RFID readers and the RFID tags can decrease the management cost of the stock.

The examples show the limited success brought by the RFID readers and the RFID tags. We believe that the RFID data collected by the RFID readers and the RFID tags has various possibilities. Also, we believe that the detail analysis of the RFID data can built various services improving our life. On the other hand, it is not possible to establish the only one analysis method dealing with various services. This is because the analysis method depends on target fields and each field has respective needs. It is necessary to focus on a

978-0-7695-4167-9/10 $26.00 © 2010 IEEE
DOI 10.1109/NBiS.2010.17

specific field in order to generate an efficient system based on the RFID data.

In this paper, we select an apparel field from various fields. This is because the apparel field deals with items whose prices are comparatively expensive and which many people buy in daily living. Also, we focus on a method that predicts the number of items ordered in the present term. This is because persons concerned with the apparel field have high needs. This method generates prediction models representing relationships between actions of customers and the number of sales in the next term. It predicts the number by applying actions in the present term to the models. This paper verifies the efficiency of the method based on the RFID data collected from an apparel company.

II. Prediction Method

It is a tough task for managers of apparel companies to decide the number of items ordered in the present term. This is because their mistaken decision leads to the decrease of returns by discount sales of remaining items and chance loss of sales. It is important to appropriately order the items. A system based on RFID readers and RFID tags can manage stock of items in real time. The managers can easily grasp the stock in the present term. Also, the system can trace the movement of items in shops. The movement represents actions of customers to some extent. We can anticipate that the actions are related to the number of sales in the next term. The system may be able to predict the number through the analysis of the RFID data. The stock and the number help the managers to decide the number of items ordered in this term. The system has a great impact on their decision making. Thus, we consider a method that predicts the number of sales in the next term based on the RFID data.

The present RFID system supposes that RFID tags are assigned to items and RFID readers are sets at cash registers, shelves, fitting rooms, and a gate between the backyard and the floor space of a shop. Also, it supposes that sensors are set in gates between the floor space of the shop and the entrances. We can collect the information related to sales, pick-up, fitting, stock, and the number of customers. In addition, it stores the information related to the feature of items such as size, color, and shapes. On the other hand, we may be able to use the external information such as weather, events performed around the shop, and articles related to the fashion. However, we do not know how degree the external information contributes to prediction models regardless of huge workload related to the collection of the external information. In the first research, we try to generate the prediction models without using the external information.

In the case of items dealt with apparel shops, the number of specific items sold in a specific day is at most one or two. Many items are not sold in the specific day. Even if prediction models predict that all items are not sold, they may be able to have high efficiency. However, the models are different from the ones we want and are not useful. It is necessary to gather items to some extent in order to avoid generating the models. Thus, this paper uses a week as the collection unit of items. Also, it uses items which ignore the difference of colors and sizes.

Figure 1 shows an outline of the prediction method. The method generates prediction models based on the RFID data. The data is composed of independent variables and a dependent variable. The independent variables correspond to the number of customers, stock, sales, and so on in a week. The dependent variable corresponds to the number of sales in the next week. The data is generated for each branch of apparel companies. The method applies the models to the RFID data in this week and predicts the number of items sold in the next week. The managers can decide the number of items ordered in this week by referring to the number of stock in this week and the predicted number of sold items.

III. Data Generation

We have three data sets which can use in order to acquire prediction models The sets are collected from three branches of two companies in the period as shown in Table I. However, we are not allowed to disclose results based on the Y company. In the following, we deal with only the sets of the X company.

Data set	Period
X Company: α branch	25th November, 2008$\sim$25th March, 2009
X Company: β branch	25th November, 2008$\sim$25th March, 2009
Y Company: γ branch	18th September, 2008$\sim$18th December, 2008

Table I
Data Collection Period using a RFID system

The branches are stood up at more or less different areas. That is, α branch is located at the outskirts and β branch is located at the center of a city. The floor space of the former one is much larger than the floor space of the latter one. The difference of the floor space leads to the difference of places which RFID readers are set. Each branch cannot collect the same data items. It collects the data items as shown in Table II. In this table, "◯" shows collected data items and "$\times$" shows no collected data items. Here, "shop" and "backyard" represents the number of stock in the floor space and the number of stock in the backyard. "whole" represents the number adding "shop" to "backyard". We use all data items for each branch in order to acquire prediction models, because we do not have knowledge which data items are important.

We note that the data items except the number of customers are collected for each item and the number of customers is collected for each branch. The number of customers in a specific week is the same value for each

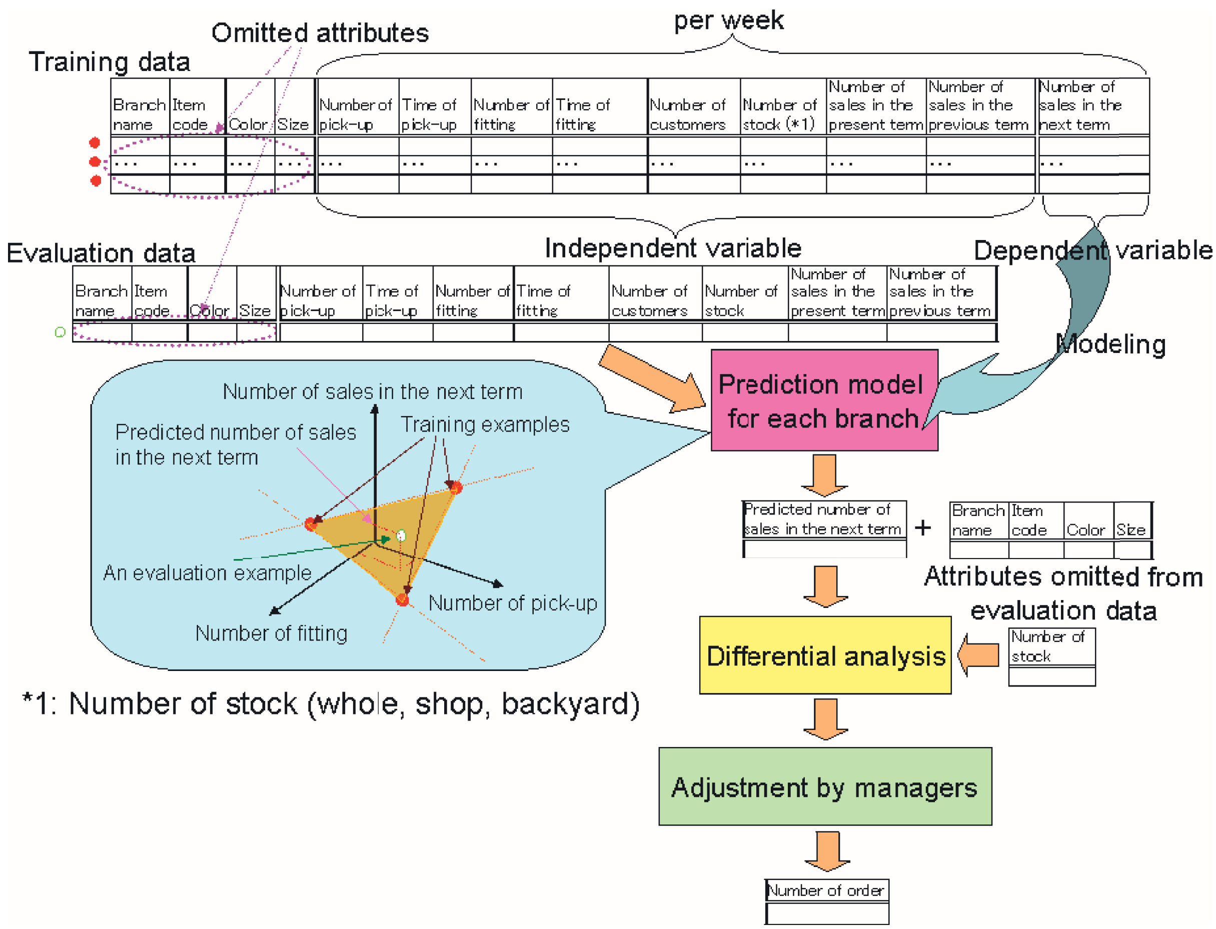


Figure 1. An Outline of a Prediction Method

	Item code	Number of pick-up	Time of pick-up	Number of fitting	Time of fitting	Number of customers	Number of stock (whole)	Number of stock (shop)	Number of stock (backyard)	Number of sales
X company: α branch	○	○	○	○	○	○	○	○	○	○
X company: β branch	○	×	×	○	○	○	○	○	○	○

Table II
DATA ITEMS

item in the same branch. The items with the same item code are gathered per a week and are accumulated. Each training example is composed of accumulated values. In the example, the number of sales in the next week and the number of sales in the previous week are decided by keeping the order of weeks. Also, we note that the RFID system collects the stock data in the decided day of the data collection. That is, the system collects the data on each Monday and each Thursday. The daily stock data is not stored in the system. On the other hands, parts of the data are lacked due to mistaken operations of the system. We can improve these problems by changing the data collection method and realizing stable operations of the system in near future. However, it is necessary for this paper to estimate the numbers of stock in a specific day. Therefore, we estimate them based on the following algorithm and accumulate them per a week.

1) If the stock data is recorded in the target day, the data is regarded as the stock data of the target day.
2) If the stock data is not recorded in the target day, this algorithm searches both the forward side and the backward side of the target day.

a) If the stock data corresponding to each side is searched, this algorithm estimates the stock data of the target day based on the linear interpolation of the searched data.
b) Otherwise, 0 is assigned to the data.

Lastly, the prediction method acquires prediction models from training examples as shown in Table III.

IV. Experiment

A. Generation of Prediction Models based on the SVM

This section mainly uses the SVM (Support Vector Machine) [2] [6] to generate prediction models. The SVM is one of inductive learning methods and many free software packages are distributed. LIBSVM [3] is one of the packages and offers 5 kinds of SVM. We use C-SVC which deals with standard classification problems and epsilon-SVR which deals with regression problems. LIBSBM also offers 4 kinds of kernel functions. The radical basis function is a default kernel function. We use it as a kernel function in the following experiments.

B. Experimental Method

Even if training examples are generated per a week and per a similar item, the number of sales in the next term is 0 in many training examples. The inductive learning methods tend to acquire imbalanced predication models. Really, our preliminary experiments showed that predication models based on the neural network predict the number as 0 in the case of almost training examples. Also, they showed that the prediction models based on the SVM have comparatively small influence for the imbalance. However, the influence exists in the case of the SVM. Thus, this paper verifies effect of a two step-wise generation method based on the combination of C-SVM and epsilon-SVR. The method generates a two-class prediction model based on the C-SVM from all training examples featured by two kinds of classes. That is, one class shows that the number of sales in the next term is 0 and the other class shows that the number is larger than 0. The former class and the latter class are referred as no-sales class and sales class, hereafter. Next, the method extracts training examples whose classes are the sales class. It generates other predication model based on the epsilon-SVM from the extracted training examples. The method is referred as the first two step-wise method. In the prediction phase of the method, the method classifies examples by applying them to the model based on the C-SVM. If the classes are no-sales classes, the numbers of sales in the next term are 0. Otherwise, the method predicts values corresponding to the examples by applying them to the model based on the epsilon-SVM. The values are the numbers of sales in the next term.

In the comparison with the first two step-wise method, we generate prediction models based on only the epsilon-SVR from all training examples. It is referred as the sample method, hereafter. Also, we introduce the second two step-wise method. It generates two models from all training examples based on the C-SVM and the epsilon-SVR. It classifies examples into two classes by using the former model. If predicted classes are sales class, the method applies the examples to the latter model and predicts the numbers of sales in the next term.

On the other hand, we use one of options of the C-SVM. The option can give a weight to training examples included in a specific class. We calculate the ratio of the number of training examples included in the no-sales class to the one included in the sales class. The ratio is used as the weight. We anticipate that the weight can revise the imbalance of the training examples to some extent.

We performed in the preliminary experiments based on the multiple regression analysis. The experimental results showed that the analysis gives the highest relationship between independent variables and a dependent variable in the case that the first day of the analysis is 25th January, 2009. Therefore, we selected the day as the first day in the following experiments.

C. Experimental Results

Figure 2 and Figure 3 show the distribution of training examples and parts of experimental results based on the SVM. The former one and the latter one correspond to the case of α branch and the case of β branch, respectively.

The graph (a) and the graph (b) in each figure show the number of training examples. The former one corresponds to the two-class problem and the latter one corresponds to the regression problem. Also, the graph (c) $\sim$ the graph (h) show experimental results. Each graph shows results without using the middle classes and results with using them. Here, they are classification codes which are assigned to higher level than item codes. In the case that the middle classes are not used, all training examples originally generated in each branch are used in the experiments. In the case that they are used, the training examples are divided based on them and experiments are performed for each divided training examples. In each graph, “Total” shows the case that the middle classes are not used and “AAaa” $\sim$ “AAal” show the case corresponding to each middle class.

The graph (c) shows the validity of prediction models that classifies training examples into the two classes. That is, it shows correction ratios for all training examples C_{total}, training examples assigned the sales class C_{sales}, and training examples assigned the no-sales class C_{nsales}, respectively. Also, it shows the geometric mean of C_{sales} and C_{nsales}. These correction ratios are defined by the following Formula (1) $\sim$ Formula (3). We think that the geometric mean is a better criterion. This is because the mean can evaluate prediction models with two viewpoints:

Independent variable										Dependent variable
Number of pick-up	Time of pick-up	Number of fitting	Time of fitting	Number of customers	Number of stock (whole)	Number of stock (shop)	Number of stock (backyard)	Number of sales in the previous term	Number of sales in the present term	Number of sales in the next term

Table III
A FORMAT OF TRAINING EXAMPLES

the sales class and the no-sales class.

$$C_{total} = \frac{|S_{correct}|}{|S|} \quad (1)$$

$$C_{sales} = \frac{|S_{correct} \cap S_{sales}|}{|S_{sales}|} \quad (2)$$

$$C_{nsales} = \frac{|S_{correct} \cap (S - S_{sales})|}{|S - S_{sales}|} \quad (3)$$

Here, S is the set of training examples, $S_{correct}$ is the set of training examples whose classes corresponds to the classes predicted by the prediction model, and S_{sales} is the set of training examples whose classes are classified into the sales class. Also, $|\cdot|$ is an operator that calculates the number of elements included in the set.

On the other hand, the graph (e) and the graph (g) show results based on the 5-fold cross-validation method in the case of the two-class problem. That is, the method divides training examples into 5 subsets. It acquires a prediction model by referring to 4 subsets and evaluates a remaining subset with the model. The acquisition and the evaluation are repeated by exchanging the evaluation subset. The method can evaluate the efficiency of the model for unknown examples to some extent. In addition, the graph (e) uses weights of the training examples and the graph (g) does not.

The graph (d), the graph (f), and the graph (h) show the average absolute error E_{diff}. The difference is defined by Formula (4).

$$E_{diff} = \frac{\sum_{t \in S} |v_{true}(t) - r_{predict}(t)|}{|S|} \quad (4)$$

Here, $v_{true}(t)$ is a function that extracts the number of sales in the next term from a training example t. $r_{predict}$ is a function that extracts the predicted number of sales in the next term from a training example t and rounds it off to the nearest whole number. The graph (d) shows the case of training based on all training examples. The graph (f) and the graph (h) show the case of the 5-fold cross-validation method. Also, the graph (f) uses weights of the training examples and the graph (h) does not.

D. Discussions

This section discusses the efficiency of the prediction method with 5 viewpoints.

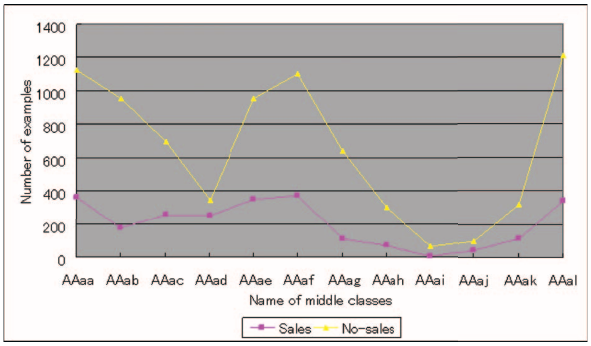
(1) Validity of Prediction Models: We note the results based on the 5-fold cross-validation method. They show that the errors are about 1.0. The prediction models can predict the number of sales in the next term to some extent. However, in the case of the two-class problem, the correct ratios are at most 0.7. They show that it is necessary to revise the prediction models.

Next, we note the difference of the results based on the 5-fold cross-validation method and the results based on all training examples. The results show that the difference is very large in the case of the two-class problem. That is, the former ones are much worse than the latter ones. We think that the SVM acquires prediction models which excessively depend on the training examples. In the experiments, we usually use default settings except the weight parameters. If the settings are adjusted, the SVM can avoid the dependency to some extent. In future work, we will try to adjust them.

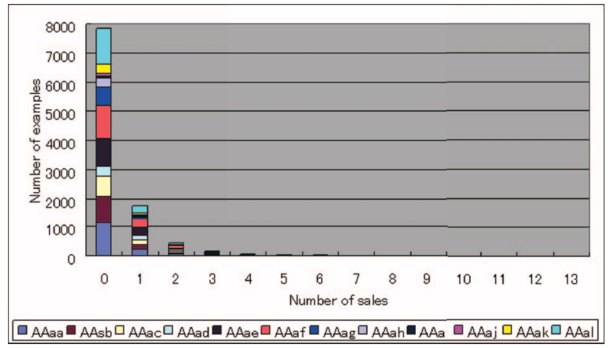
(2) Effect of the Two Step-wise Methods: We note results based on the two step-wise methods and the results based on the simple method. They show that the first two step-wise method is not better than the simple method with the viewpoint of the average absolute errors. Especially, the first two step-wise method gives bad performance in the case of β branch and the use of weighted training examples. In the case of β branch, the number of training examples is comparatively small. In addition, the method extracts small parts of original training examples. The number of training examples is much fewer. The average absolute errors may deteriorate due to the small number. Also, the combination of the first two step-wise method and the weighted training examples doubly reinforce training examples in the sales class. The combination may excessively reinforce the sales class.

On the other hand, the second two step-wise method is comparatively better than the simple method, even if the difference is small. In the case of the RFID data collected from the apparel company, the numbers of sales in the next term are 0 for many training examples. Therefore, the difference of the average absolute errors tends to be small. The absolute error may be not a good criteria. In future work, it may be necessary to consider other criteria instead of the average absolute error.

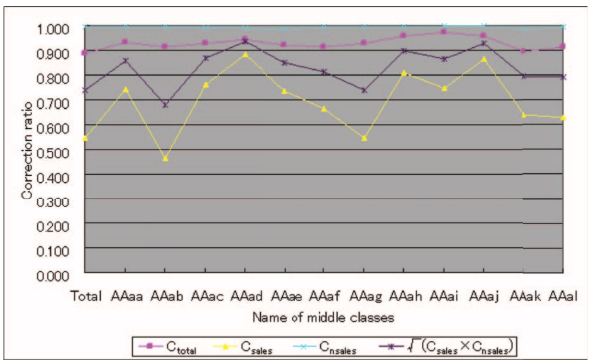
(3) Effect of the Weighted Training Examples: We note the geometric mean in the experimental results. The results show that the C-SVM generates well-balanced prediction models in the case of the weighted training examples. The



(a) Distribution of sales volume (two-class problem)

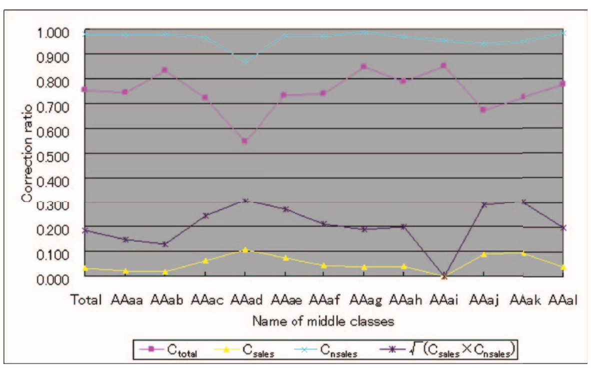


(b) Distribution of sales volume (regression problem)



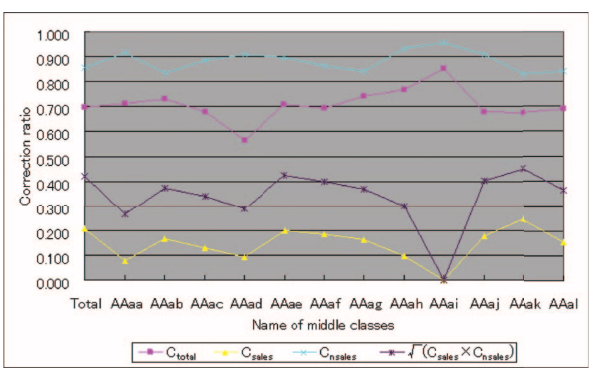
(c) Two-class problem (All data)

(d) Regression problem (All data)



(e) Two-class problem (5-fold cross-validation: No weight)

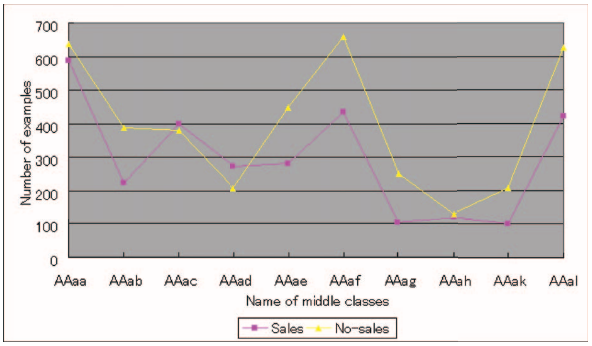
(f) Regression problem (5-fold cross-validation: No weight)



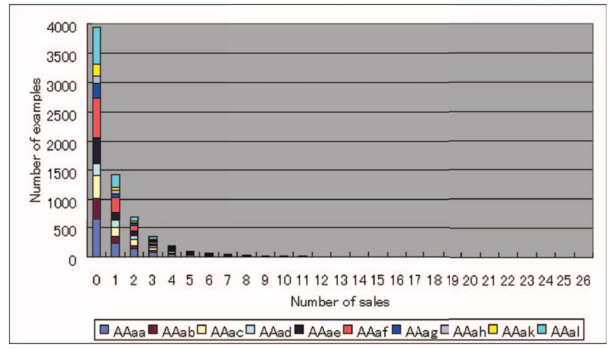
(g) Two-class problem (5-fold cross-validation: Weight)

(h) Regression problem (5-fold cross-validation: Weight)

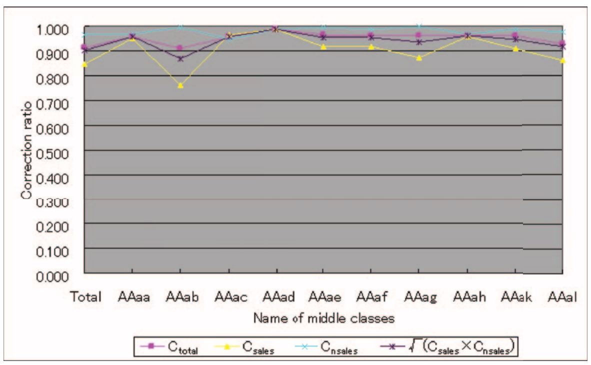
Figure 2. Results for α Branch of X Company



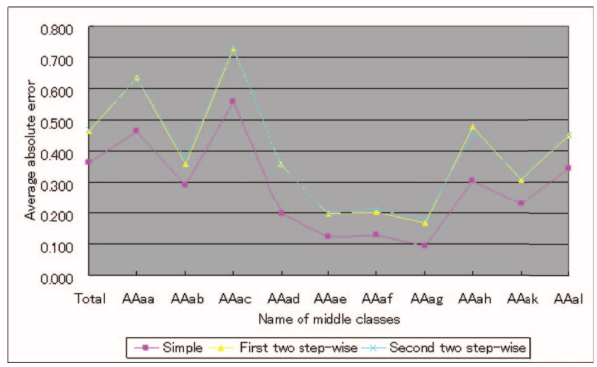
(a) Distribution of sales volume (two-class problem)



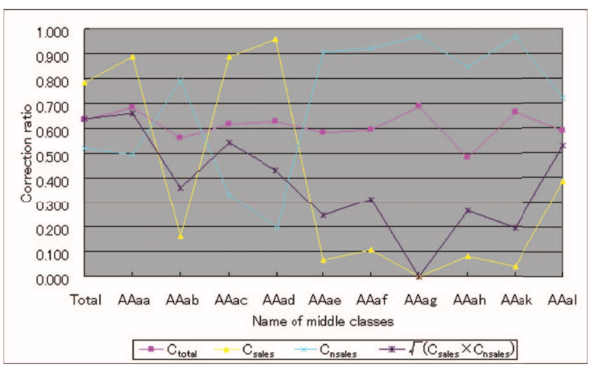
(b) Distribution of sales volume (regression problem)



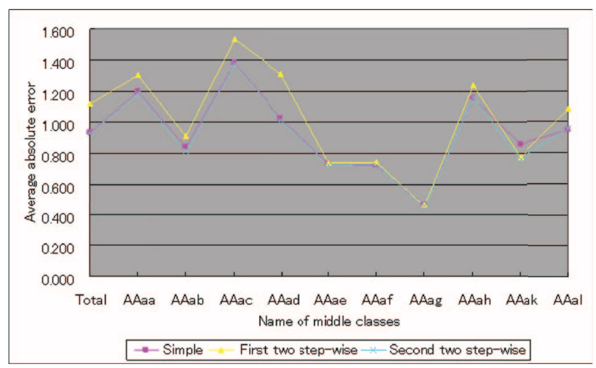
(c) Two-class problem (All data)



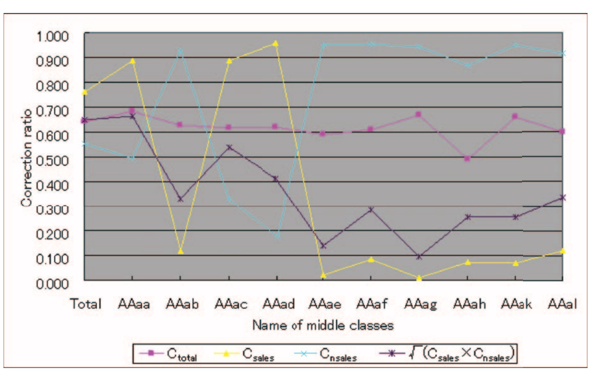
(d) Regression problem (All data)



(e) Two-class problem (5-fold cross-validation: No weight)



(f) Regression problem (5-fold cross-validation: No weight)



(g) Two-class problem (5-fold cross-validation: Weight)

(h) Regression problem (5-fold cross-validation: Weight)

Figure 3. Results for β Branch of X Company

models can appropriately classify the training examples included in the sales class to some extent. The adjustment of weights contributes to the improvement of the prediction models, even if the improvement does not always give a great influence to the improvement of the average absolute errors.

(4) Difference of Branches: We do not show the experimental results in the case of the combination of α branch and β branch due to the limit of pages. In the case of the two-class problem, the results show that the performance based on the combination is similar to the performance which averages the performance based on α branch and the performance based on β branch. We anticipated that the increase of training examples generates better prediction models. However, the results show that the actions of the customers and the sales depend on the branches. It is necessary to generate the models for each branch.

Next, we note the difference of the branches in the case of the average absolute errors. The results show that the error of α branch is smaller than the one of β branch. α branch sells more items than β branch does. The prediction models in the case of α branch tend to predict higher number of sales in the next term. The average absolute errors tend to enlarge. We think that the difference of sales volumes leads to the deterioration of the errors in the case of α branch.

(5) Usage of Middle Classes: We note results of the two-class problem. The results show that the case "Total" gives comparatively higher performance. The use of the middle classes does not always lead to better prediction models. This is because the influence of the decrease gives a greater impact than the influence of the common features. That is, the decomposition of training examples leads to their decrease. The decrease may make the models deteriorate. On the other hand, the items included in the same middle class have common features. The decomposition of training examples based on the middle classes can generate prediction models reflecting the features. The decomposition may give on better effect for the prediction, if many training examples could be gathered.

V. Conclusion and Future Work

This paper proposed a method that predicts the number of sales in the next term based on the RFID data. Also, it evaluated its efficiency based on the data collected from two branches of an apparel company. The experimental results show the possibility of the prediction, even if it is necessary for the prediction models to be revised their performance.

In future work, we will tackle on the improvement of the prediction models. For example, we will try to introduce new independent variables by gathering comments of the persons working in the apparel field. We will try to introduce the external information as independent variables. We will try to collect new RFID data. These improvements of the data can revise the models. Also, we will tackle on the improvement of the prediction method. For example, we are planning to introduce techniques of ensemble learning such as boosting and bagging. We believe that the prediction based on the RFID data can be realized through the improvements.

On the other hand, we will aggressively tackle to establish many methods which the RFID data efficiently activates in various fields.

References

[1] Y. Akahoshi, Y. Kidawara, and K. Tanaka, *A Database-Oriented Wrapper for Ubiquitous Data Acquisition/Access Environments*, Proc. of the Second Intl. Conf. on Ubiquitous Information Management and Communication, pp.25-32, 2008.

[2] B. E. Boser, I. Guyon, and V. Vapnik, *A Training Algorithm for Optimal Margin Classifiers*, Proc. of the Fifth Annual ACM Conference on Computational Learning Theory, pp.144-152, 1992.

[3] C. -C. Chang and C. -J. Lin, *LIBSVM - A Library for Support Vector Machines*, http://www.csie.ntu.edu.tw/˜ cjlin/libsvm/.

[4] J. Collins, *Marks & Spencer Expands RFID Trial*, RFID J., February, 2004 (http://www.rfidjournal.com/article/articleview/791/1/1/).

[5] R. Dass and A. Mahanti, *Implementing BDFS(b) with Diff-Sets for Real-Time Frequent Pattern Mining in Dense Datasets - First Findings*, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp.113-120, 2005.

[6] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, *Support Vector Regression Machines*, NIPS, pp.155-161, 1997.

[7] A. Ihler, J. Hutchins, and P. Smyth, *Adaptive Event Detection with Time-Varying Poisson Processes*, Proc. of the Twelfth ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, pp.207-216, 2006.

[8] K. Kuramitsu, *Finding Periodic Outliers over a Monogenetic Event Stream*, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp.97-104, 2005.

[9] M. Roberti, *Wal-Mart Begins RFID Rollout*, RFID J., April, 2004 (http://www.rfidjournal.com/article/articleview/926/1/1/),

[10] S. Sakurai, Y. Kitahara, and R. Orihara, *A Sequential Pattern Mining Method based on Sequential Interestingness*, Intl. J. of Computational Intelligence, vol.4, no.4, pp.252-260, 2008.

[11] S. Sakurai, Y. Kitahara, R. Orihara, K. Iwata, N. Honda, and T. Hayashi, *Discovery of Sequential Patterns Coinciding with Analysts' Interests*, J. of Computers, vol.3, no.7, pp.1-8, 2008.

[12] V. S. Tseng and K. W. Lin, *Mining Temporal Moving Patterns in Object Tracking Sensor Networks*, Proc. of the 2005 Intl. Workshop on Ubiquitous Data Management, pp.1-8. 2005.

[13] R. Wessel, *Metro Group's Galeria Kaufhof Launches UHF Item-Level Pilot*, RFID J., Swptember, 2007 (http://www.rfidjournal.com/ article/articleview/3624/1/1/).