2016 IEEE International Conference on Cloud Computing and Big Data Analysis

Demographic Transformation and Clustering of Transactional Data for Sales Prediction of Convenience Stores

Xiaojun Zhang, Jisheng Pei, Xiaojun Ye
School of Software
Tsinghua University
Beijing, China
e-mail: {zhang-xj13, pjs07, yexj}@mails.tsinghua.edu.cn

Abstract—**Reliable retail sales prediction of convenience stores (CVSs) can not only help in making correct purchase decision but also in determining which new products to be launched. Therefore, the main aim of this paper is to propose an enhanced method based on clustering and different abstraction forms of data to forecast the retail sales of CVSs. We use customer type proportion to calculate the similarity of stores to improve the accuracy of clustering. We propose the Extraction of Customer Demographic Characteristics (ECDC) from transactional data as a new approach to solve the problem of the lacking of user information. Driven by domain knowledge, sparse transactional data are aggregated and transformed into ECDC using a rule tree built from shopping habits. The effectiveness of ECDC in clustering of transactional data is demonstrated through its resulting higher accuracy of sales prediction. By experimenting our method with several benchmark methods, our proposed method is found to have an optimal accuracy in forecasting the retail sales of CVSs.**

Keywords-sales prediction; convenience store; clustering; customers' demographic characteristics

I. INTRODUCTION

As today's society changes so fast, personal income is higher, working hours are longer, the pace of life is faster and lots of people prefer shopping in the accessible convenience stores (CVSs), rather than supermarkets far from home. Thus, the number of CVSs has increase fast in recent years. According to some investigations, the total number of chain CVSs in Taiwan was more than 10000 in 2014 [1]-[4]. At the same time, the competition among CVSs grows intense, and the growth rate of CVSs in Shanghai was -5% in 2014. The manager of CVSs wants to raise competitiveness through opening more CVSs, reducing or extending the business scale based on the business performance. Consequently, in order to get higher profit, the owner of CVS should be able to make correct local decisions. Meanwhile forecasting could be an importance support of making decisions. Accurate forecasting of retail sales can not only help in determining commodities' type and quantity but also in choosing which to sale from many new commodities.

In real life, there are lots of factors that can influence the retail sales of commodities. Meanwhile relevant references show that the interaction of factors is quite complicate. Li and Wu came up with 102 factors that affect the business performance greatly, which are all CVSs' basic information [5]. But sales prediction just based on CVSs' basic information is not enough, there are many other factors affecting the sales, such as commercial and its competitor's information. The prediction of retail sales is quite a challenge.

The conventional approaches used to forecast the business performance include managers' judgment simply based on their knowledge, questionnaire method, which is not scalable any more for the growing size of CVS business. On the other hand, running existing machine learning models (e.g. statistic models, time series analysis decision tree and neural network [5], [6]) on primitive transactional data usually could not lead a satisfactory sale prediction quality [7]. Even though plenty of transactional data are available as input, the important demographic characteristics hidden in the transactional data are still transparent to the prediction models. In other words, although we have an abundance of data, but we appear to lack the enough insight to fully exploit the information hidden inside for satisfactory sales prediction results.

To be specific, we are facing two major challenges in retail sales prediction.

1) Diversity of CVSs. Affected by noise, data missing and other negative factors, using only one convenience store to forecast sales faces with the problem of information shortage. But if we use amount of CVSs to forecast, we'll be disturbed by the data from inhomogeneous convenience store because of the diversity of CVSs.

2) Lacking of demographic information. Both retail sales prediction and measuring CVSs' similarity are faced with the same problem. Just considering the primitive characteristics is too much of the underlying, and the effect of clustering and prediction is poor. However, characteristics transformation and extraction need CSV background information and user basic information which are hard to get because of data deficient and privacy issues.

Because of the two challenges, if we forecast the retail sales of CVSs using conventional approaches and primitive characteristic data, it will result in the accuracy is far below our expectations. We resolve the above problems by following two approaches:

1) We partition CVSs before we train prediction models for each cluster. In this way, we can leverage and emphasize the influence of the defining characteristics of CVSs that are similar with each other, while reducing the impact of the data from the less similar CVSs in other clusters, which are less relevant for the retail sales prediction.

2) Using the professional concept of demographic characteristics in the field, such as age, gender, profession, we proposes an method to transform primitive characteristic data to professional concept. And then, we can use the background information and user basic information.

Based on the above background, this paper expects to achieve the following research goals:

1) Proposing an approach how to extract customers' demographic characteristics of CVS from transactional data.

2) Partitioning all the CVSs into different sets by clustering using customers' demographic characteristics.

3) Forecasting sales of commodities using data mining with different abstraction forms of input data and finding the best method of sales prediction with highest accuracy.

The rest of this paper is organized as follows. Section 2 gives an overview of our prediction system architecture. Section 3 describes our method to extract customers' demographic characteristics and partition the dataset of CVSs. Section 4 presents our Strategies of forecasting retail sales. In Section 5, experiments with some benchmark methods are used to demonstrate the workability of our proposed method. Some related works are summarized in Section 6. Concluding remarks are made in Section 7.

II. OVERVIEW

In order to approach challenges that we mentioned above and improve the accuracy of prediction, our prediction system consists of two components: CVSs clustering and sales prediction (Fig. 1).

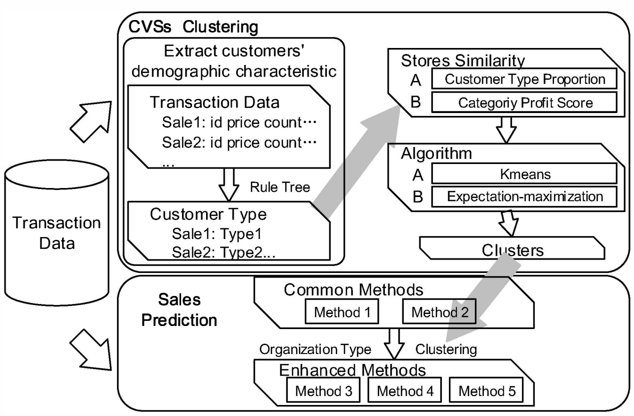

Figure 1. Architecture of prediction system

1) We use different similarity and clustering algorithms to partition the CVSs dataset into clusters. The way of similarity calculation includes customer type proportion and category profit score. The clustering algorithms includes K-Means and Expectation-maximization. Due to the lack of demographic information, we provide a method to extract demographic characteristic from transactional data.

2) When forecasting the retail sales, we propose three enhanced strategies, including various high-level abstraction of the primitive transactional data and building models fter clustering(the clustering take is done in Phase 1).

III. PARTITIONING DATASET USING CLUSTERING

As mentioned in the previous section, in order to take advantage of the common points of similar CVSs and cancel the influence from dissimilar CVSs, we can partition CVSs dataset before prediction and model the sales data for each cluster type C_j (see Fig. 2). Since the cluster is non-overlapping, separate models M_j constructed for each cluster C_j can simply be combined to give the total sales model M,

$$M = \sum(Mj) \tag{1}$$

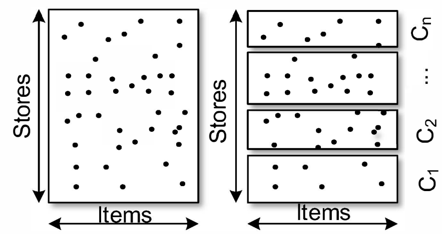

Figure 2. Two levels at which we can model the sales data

Each object in a set is represented by a set of d measurements (attributes) or d-place vector [8]. Choosing appropriate attributes, such as spatial location, sales status, is significant to clustering. Considering that customers of same type would have similar consumer preference, the similarity of the composition of customers between two convenience stores can serve as a good similarity measurement.

A. Extracting Customers' Demographic Characteristic from Transactional Data

When performing tasks including clustering, mining association rules, recommending commodities and so on, using primitive transactional data directly usually could not fully release the predicting power hidden inside these data. In contrast, introducing higher level characteristics about the target scenario in these tasks usually provides more helpful insights. In CVS sales prediction, for example, demographic characteristics serve as an important and more direct clue to the customers' preferences compared to primitive transactional records. However, the collecting of demographic information of customers is not trivial. CVSs usually attempt to gather customers' information through membership card systems. However, most customers of the convenience store have strong liquidity. They usually do not buy a membership card. And on account of the privacy and other issues, it is extremely difficult to get the valuable and complete customer data. In most cases, what we have is large amount of primitive transactional data. Hence, if customers' demographic characteristics can be extracted from the transactional data, it will be very meaningful for various types of analytic tasks. In this way, the hidden information in the primitive transactional data can be revealed and utilized.

In this paper we have data of each transaction. Each sale may have many items I_i, including the information of item id, unit price, item volume and so on. Each customer type T_i has some descriptive customer information along various demographic descriptors (such as age, gender, profession). For each sale, there should be a customer type T_i with corresponding. This is also our goal of this section to find the mapping customer type for each sale.

Different types of customers have different shopping habits, such as shopping time, most often buying item, preferential brand and so on. Because of the correspondence from customer type to shopping habits, it's reasonable back extracting customers' type from each sale.

Fig. 3 shows the method of extracting customers' demographic characteristic from transactional data. The types of customers who consume frequently in the convenience stores and customers' shopping habits are established through consulting the working staffs and professional consults of CVS. Let us denote shopping habits' set by $H = \{H_i \mid i \in 1 \ldots K\}$. Each customer type maps one or more habits. ($T_i \leftrightarrow \{h_{i1}, h_{i2} \ldots\}$). We denote this set of customer types mapping habits by set $M = \{M_i \mid i \in 1 \ldots L\}$, $M_i = \{Ti \leftrightarrow \{h_{i1}, h_{i2} \ldots\}\}$. For a sale, it would be labeled as T_i if it matches the mapping M_i. In order to reduce unnecessary matching, we construct a rule tree using recursion (Algorithm 1). The entire sales are put into the rule tree, spilt into subsets along the shape of rule tree and find their corresponding customer types at the leaf node.

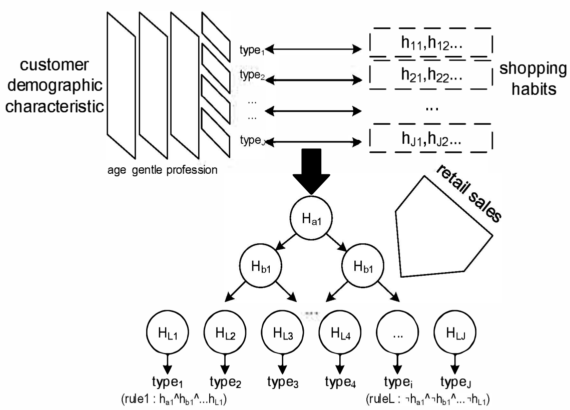

Figure 3. The method of extracting customers' demographic characteristic

Algorithm 1 BUILD-TREE (M)

```
Input: M the set of customer types mapping habits
if M is empty  then return null
if  there is a habit hi appear once time
    then h ← hi  h.type ← mi.type
    else h ← random(habits)
MLeft ← M(mapping has h)
Hright ← M – MLeft
for mapping in Hleft
    do remove h from mapping
h. left ← build_tree(MLeft)
h. right ← build_tree(MRight)
return h
```

In our experiment, we select in total 2000 gas station CVSs as the target. For each store we have daily sales figures for 1399 items of year 2014. According to the suggestion of experts and considering the particularity of gas stations, we came up 14 types of customers, pure smoker, normal smoker, night driver, travel driver, gentle driver, motorbike driver, trip driver, professional driver, house wife, officer, overtime worker, party animal, gentle customer, passerby. Table I shows the mapping from customer type to shopping habit and the customer type distribution after extracting. Fig. 4 shows the rule tree constructed based on the mapping.

TABLE I. THE MAPPING FROM CUSTOMER TYPE TO SHOPPING HABITS

Customer Type	**Shopping Habits**	**Percent**
pure smoker	"cigarette"	0.125
motorbike driver	"motor oil"	0.0744
night driver	"red bull" "00:00-06:00"	0.0151
travel driver	"red bull" "toilet requisites"	0.00044
trip driver	"red bull" "food"	0.04054
professional driver	"red bull"	0.0127
gentle driver	"car supplies"	0.0627
house wife	"grain and oil"	0.03041
officer	"food" "meantime"	0.0242
overtime worker	"food" "closing time"	0.0436
party animal	"wine" "closing time"	0.0278
passerby	"water"	0.000641
normal smoker	"cigarette"	0.162
gentle customer		0.381

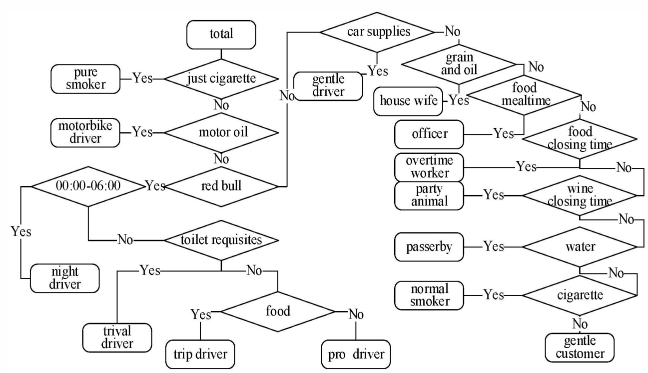

Figure 4. Rules tree

B. Clustering

The purpose of extracting customers' demographic characteristic is to get a picture of the composition of customers of a convenience store, which can be used for clustering. The customer type of each sale is calculated from the rule tree. So we can get the percentage of each customer type, a standardized vector, of each convenience store. This can be clustering attributes of CVSs. In order to verify the effectiveness of our method that how to extract customers' demographic characteristic from transactional data, other method for clustering similarity calculating is put forward.

After calculating the profit of every category of goods, we choose the top six highest profit categories as target. So the descriptive vector of a store consists of the profit of six categories.

In order to minimize the bias that may occur in single clustering algorithm, the clustering of store is carried out with two clustering algorithms. We use weka as our tool for clustering. The clustering algorithms are:

1) K-means clustering. This is a common and robust method of clustering, though not completely deterministic.

2) Expectation-maximization(EM) algorithm. EM finds clusters by determining a mixture of Gaussians that fit a given data set [9]. EM of weka can decide how many clusters to create by cross validation.

IV. STRATEGIES FOR SALES PREDICTION

Many mature approaches could be used for prediction. The prediction algorithms that we use for modelling are linear regression, least square method, decision tree, bagging, and random forests. However, the abstraction form of input data to build a model is critical for improving the accuracy sometimes. Fig. 5 shows common methods and enhanced methods of prediction. The most common methods of modelling data are:

1) Using entire data, all the items of all stores, to build a model (Method 1).

2) For each CVS, using its sales data to build a model. Compairing with Method 1, this method can avoid the interference from the other stores and take full advantage of characteristics of the product itself, but also abandon the similarity of the other stores (Method 2). The main thought of this method is isolating the target store and finding some association by considering commodities' behavior only.

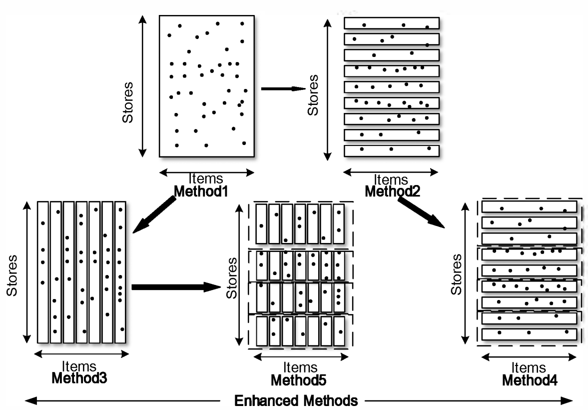

Figure 5. Common methods and enhanced methods to prediction

In most cases, the more data we use, the better model we get. But in the experiment, we find the accuracy of Method 1 is very low. we propose three kinds of enhanced methods:

1) For each item, using its sales data from all stores and some statistical information of the store, such as category sales and store sales, to build a model. There will be a model for each item. Compairing with Method 1, this method can avoid the interference from the other items and take full advantage of the item's information in other stores, but also abandon the similarity of the other items. The main thought of Method 3 is isolating the target item and finding some association by considering store level information of all CVSs. Method 3 is an improvement of Method 1, and another thinking direction of Method 2. (Method 3).

2) For the stores in same one cluster, using its sales data of all items to build a model. This method is like Method 1, but the stores are selected from same cluster. Using this, we can emphasize the influence of characteristics of CVSs that are similar with each other, while reduce the impact of the data from the less similar CVSs in other clusters. But, same with Method 2, combining all the commodities including noisy and irrelevant data is bad for accuracy. Method 4 is an improvement of Method2. (Method 4).

3) For each item, only using its sales data from stores in same cluster. This method is a refinement of Method 3. The improvement way is as same as Method 4 improving from Method 2, putting the similar items into a cluster to achieve high cohesion. Method 5 avoids the interference from the other items and reduces the influence of not similar stores through clustering, but also abandon the similarity of the other items. In other words, Method 5 has same weakness as Method 3, but owns stronger advantage. (Method 5).

V. EXPERIMENT AND RESULT

The experiment data are retail sales of 2000 gas station convenience stores of a Chinese province for year 2014. For each store we have daily sales figures for 1399 items.

In this study, the accuracy of forecast is the performance rating of the forecast and can be expressed as follows:

$$Ai = |F_{ai} - R_{ai}|/R_{ai} \tag{2}$$

where A_i is the forecasting accuracy for sample i, F_{ai} is the forecasting result of the retail sales of sample i and R_{ai} is the real sales of sample i. From expression (2) we can see that the higher accuracy we obtain, the lesser value of A_i.

Obviously, average is not a good measure to reflect the global accuracy level because one high value of A_i may raise the average. We choose two criterions instead of average, weighted arithmetic mean and median. Expression (3) defines weighted arithmetic mean.

$$wx' = (w_1x_1 + w_2x_2+...+w_nx_n)/(w_1+w_2+...+w_n) \tag{3}$$

A. Experiment of Common Methods

Based on what we talked about in the previous section, Method 1 is the most common method, and Method 2 is one kind of abstraction form of input data representing the different direction of thinking. Method 1 is all data model. Method 2 is the model for one store. We choose 20 stores and 100 items to calculate the result. All the experiments use 80% data to train a model and 20% data to test the model. Fig. 6 shows our experiment result of common methods. LR represents linear regression. PLSR represents partial least squares regression. CART represents decision tree. RF represents random forests. We can see that

1) Method 1 and method 2 have almost the same result for both median and weighted arithmetic mean. But the best performance of method 1 (randomforests:0.766,0.620) is better than mehod 2 (bagging:0.832,0.675). So the similar characteristics of commodities in a store are not very stronge to forecast sales of each other. In other words, we can not forecast sales of a commodity precisely just according to its category, price and other commodities' information of its convenience store.

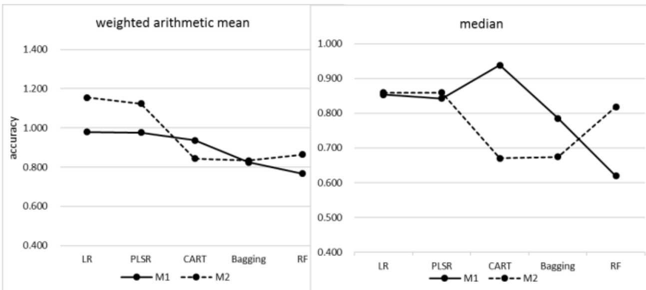

Figure 6. Result of common methods (method 1, method 2)

B. Experiment of Enhanced Methods

We improve the accuracy of prediction by using different algorithms based on method 3, method 4 and method 5. Method 3 models the data of same commodity from different stores. Method 4 models the data of stores that in same cluster. Method 5 models the data of same commodity from different stores that in same cluster. In the experiment, we have different ways to partition the CVSs dataset into clusters. For compairing obviously, we just choose the worst and best performance to display. Table II shows the experiment result of enhanced methods. We can see that they perfom much better than Method 1 and Method 2.

TABLE II. RESULT OF ENHANCED METHODS (METHOD 3, METHOD 4, METHOD 5)

	Weighted Arithmetic Mean				
	LR	***PLSR***	***CART***	***Bagging***	***RF***
M3	0.394	0.400	0.404	0.319	0.289
M4-Worst	0.630	0.629	0.619	0.514	0.482
M4-Best	0.605	0.596	0.613	0.485	0.437
M5-Worst	0.384	0.389	0.357	0.294	0.277
M5-Best	0.373	0.377	0.271	0.217	0.213
	Median				
	LR	***PLSR***	***CART***	***Bagging***	***RF***
M3	0.291	0.299	0.313	0.260	0.266
M4-Worst	0.438	0.441	0.518	0.415	0.311
M4-Best	0.438	0.441	0.517	0.404	0.278
M5-Worst	0.260	0.262	0.298	0.252	0.238
M5-Best	0.214	0.221	0.250	0.200	0.195

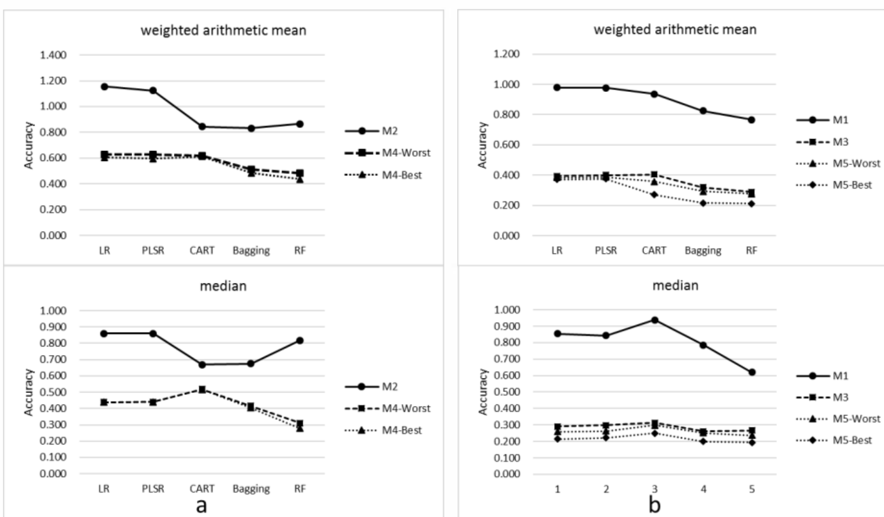

Figure 7. Result of enhanced methods (method 3, method 4, method 5)

Because Method 4 is an improvement of Method 2, we compare them in the Fig. 7a. We can see that the weighted arithmetic mean and median of Method 4 are much lower than Method 2. In other words, clustering improves the accuracy of prediction.

Because Method 3 is an improvement of Method1, and Method 5 is an improvement of Method 3, we compare them in the Fig. 7b. We can see that the weighted arithmetic mean and median of Method 3 are much lower than Method 1. Which means, prediction according to the information of the same commodity in other stores performs very well. Method1 uses all the data to build a model, and Method 3 uses the data of target item in all stores and the store level information to build a model. It is proved that isolating the target item and finding some association by considering itself behavior and store level information of all CVSs is a good approach to improve the prediction performance. Meanwhile, it is proved that prediction after clustering is an effective and feasible approach, because the result of Method 5 is better than Method 3.

Fig. 8c shows the best accuracy of five methods. We can see that the performance of enhanced methods is much better than common methods. Method 5 has the best performance of the total experiment, reaches 0.2 and improving a lot from method 1. It is proved that isolating the target item and clustering before prediction can get the best accuracy.

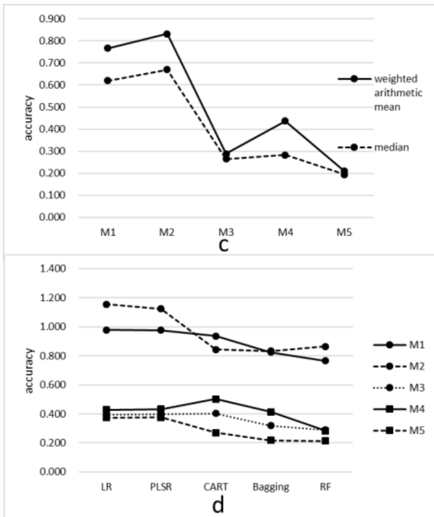

Figure 8. Performance of different methods and algorithms

Fig. 8d shows the best accuracy of different prediction algorithms. We can see that Bagging and Random Forests performs better than other algorithms, especially Random Forests has a stable and excellent outcome.

TABLE III. RESULT OF METHOD 5

Weighted Arithmetic Mean					
	LR	**PLSR**	**CART**	**Bagging**	**RF**
M5-GEO	0.364	0.405	0.371	0.312	0.299
M5-PROFIT-EM	0.384	0.389	0.357	0.294	0.277
M5-PROFIT-KMEANS	0.358	0.363	0.360	0.291	0.276
M5-CUST-EM	0.368	0.378	0.265	0.213	0.210
M5-CUST-KMEANS	0.373	0.377	0.271	0.217	0.213
Median					
	LR	**PLSR**	**CART**	**Bagging**	**RF**
M5-GEO	0.286	0.290	0.316	0.263	0.252
M5-PROFIT-EM	0.260	0.262	0.298	0.252	0.238
M5-PROFIT-KMEANS	0.260	0.263	0.297	0.246	0.234
M5-CUST-EM	0.214	0.221	0.250	0.200	0.195
M5-CUST-KMEANS	0.224	0.234	0.258	0.207	0.200

C. *Experiment About Culstering*

In order to improve the accuracy, method 4 and method 5 model the data clustered. We use two different data attributes, customer type proportion and categoriy profit score, and two different algorithms, k-means and EM, to cluster. At the same time, we have the geography information of CVSs, which partitions the stores into six clusters. The experimental results of accuracy are good or bad represents the clustering method is good or bad. Fig. 9 shows the experimental result of method 4 and method 5. Table III shows the accuracy of method 5. We can see that:

1) Whether using method 4 or method 5, customer type proportion performs better than categoriy profit score. Therefore, to a certain extent, our method extracting customers' demographic characteristic from transactional data is right. (Fig. 9)

2) K-means and EM have almost the same performance.

3) Geography have almost the same performance with clustering of categoriy profit score. In other words, geography has a great influence to CVSs' business performance.

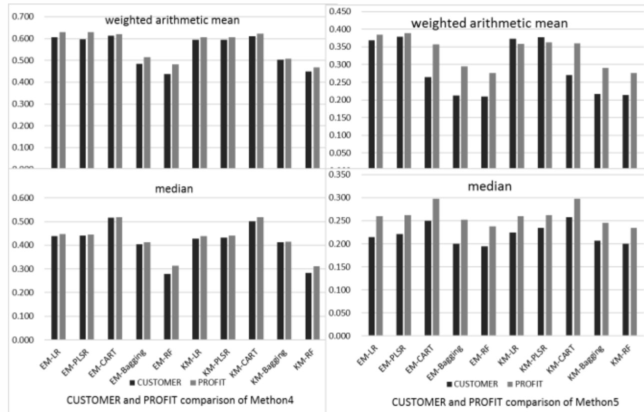

Figure 9. CUSTOMER and PROFIT comparison of method 4, method 5

VI. RELATED WORK

A. *Clustering of Convenience Stores*

Cluster analysis is the process of classifying objects into subsets that have meaning in the context of a particular problem. Clustering has been applied in various fields, including data mining, statistical data analysis, compression and vector quantization. Intrinsic clustering is sub-divided into hierarchical and partitional clustering by the type of structure imposed on the data [8]. The k-means is a very popular algorithm of partitional clustering.

Clustering is most frequently used as a preparatory tool for next analysis. Shu-Hsien et al. [7] divided clients into three clusters by clustering analysis for their recommendation system. Clients' characteristics included purchase preference, purchase behavior and brand endorsers. Mohsin et al. [10] did segmentation of dataset with respect to 'gender' and 'age' for better association rules having higher support and confidence. Mehrdad et al. [11] proposed three effecting factors for similarity function to clustering, 1. Euclidean distance, 2. Lot size, 3. Order concurrency. Philip [12] developed customized geospatial tools using store spatial clustering to support network planning decision.

B. *Business Performance Forecasting of Convenience Store*

Business performance forecasting is a useful tool helping shop owner to make decision. The most often-used methods for forecasting business performance consist of statistic models, time series analysis, and regression models. Statistic models are extracted from large quantity of data, such as Bayesian approach and Markov model [13]-[15]. Time series forecasting is the use of a model to predict future values based on previously observed values. Regression models are often constructed based on certain conditions that must be verified for the model to fit the data well, and to be able to predict accurately. Li and Wu [5] designed an enhanced fuzzy neural network based predictor to forecast the daily average number of visiting customers. They determined the weight of each evaluation factor of neural network through questionnaire. Michael [16] partitioned the dataset across customer types and demographics yields to improved performance of retail sales prediction. To obtain a model for all the stores they rolled up model for each partition together. Virtual Worlds (VWs) have emerged as a new commercial activity. Sales in VWs were predicted by the frequency of visiting and the time spent within VWs' stores [17].

C. *User Demographic Information in Convenience Store Study*

In the field of CVSs', lots of research utilized user demographic information directly. They didn't transform transactional data to user demographic information. Wayne [18] developed a novel product recommender System, which made product recommendation based on matching the users' demographic information extracted from their public profiles with product demographics learned from microblogs and online reviews. Yoshinori Fukue [19] proposed a method to capture the purchase patterns of customers by analyzing purchase history data of E-Commerce Market in time series. They captured the customer data from the CVS E-Commerce site. Shu-Hsien et al. [7] used the customer demographic information for clustering and recommendation. They got the data from questionnaires. Michael [16] used the customer demographics to split the total data and built a model for each demographic descriptor and customer type. They obtained customer information from the retailer and publicly available data sets.

VII. Conclusion

This paper proposes a novel methodology to forecast the retail sales of items in CVSs. We solve the problem that diversity of CVSs through clustering. In our experiment, prediction after clustering (Method 4 and Method 5) is better than prediction without clustering (Method 2 and Method 3). We give three enhanced methods of sales prediction from two different thinking ways, isolating the target store or the target item. In the end, isolating the target item and finding some association by considering itself behavior and store level information of all CVSs to predict sales (Method 3 and Method 5) do better than the other way (Method 2).

This paper also provides a method to extract the latent demographic characteristics of CVS customers underlying the primitive transactional records, and thus addresses the problem of the lacking of demographic information. We find the mapping from customer types to shopping habits through the suggestion of experts. We use the mapping to build a rule tree and get customer types of each sale. In the experiment, we use both customer type proportion and category profit score to cluster. And, the result of customer type proportion is better, which proves our method of extracting customers' demographic characteristic is effective.

Our future research work will be on extracting more users' demographic characteristic from transactional data. In this paper, we only get the users' type information. Some other demographic characteristic such as age and gender should also be extracted from transactional data. Extracting users' demographic characteristic can be used in many other researches, such as association rules mining.

References

[1] Berman B, Evans JR (2003) Retail management: a strategic approach, 9th edn. Prentice-Hall, New Jersey.

[2] Craig CS, Ghosh A, Me LS (1984) Model of the retail location process: a review. J Retail 60:5–36.

[3] Gorucu FB, Geris PU, Gumrah F (2004) Artificial neural network modeling for forecasting gas consumption. Energy Sources 26:299–307.

[4] Zhao FY (2003) Theory of supermarket management. Enterprise Management Publishing House, Beijing.

[5] Li S G, Wu Z M. Business performance forecasting of convenience store based on enhanced fuzzy neural network. Neural Computing and Applications, 2008, 17(5-6):569-578.

[6] Applebaum W (1996) Method for determining store trade areas, marketing penetration and potential sales. J Mark Res 3:124–141.

[7] Liao S H, Wen C H, Hsian P Y, et al. Mining Customer Knowledge for a Recommendation System in Convenience Stores. International Journal of Data Warehousing and Mining (IJDWM), 2014, 10(2): 55-86.

[8] Jain A K, Murty M N, Flynn P J. Data clustering: a review. ACM Computing Surveys (CSUR), 1999, 31(3): 264-323.

[9] Alldrin N, Smith A, Turnbull D. Clustering with EM and K-means. University of San Diego, California, Tech Report, 2003: 261-95.

[10] Riaz M, Arooj A, Hassan M T, et al. Clustering based association rule mining on online stores for optimized cross product recommendation, Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on. IEEE, 2014: 176-181.

[11] Kargari M, Sepehri M M. Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs. Expert Systems with Applications, 2012, 39(5): 4740-4748.

[12] Bermingham P, Hernandez T, Clarke I. Network Planning and Retail Store Segmentation: A Spatial Clustering Approach. International Journal of Applied Geospatial Research (IJAGR), 2013, 4(1): 67-79.

[13] Chen D, Yuan XJ (2004) A Markov model for seasonal forecast of Antarctic sea ice. J Clim 17:3156–3168.

[14] Coelho CAS, Pezzulli S, Balmaseda M (2004) Forecast calibration and combination: a simple Bayesian approach for ENSO. J Clim 17:1504–1516.

[15] Goldberg DE (1989) Algorithm in search optimization and machine learning. Addison-Wesley, Reading.

[16] Giering M. Retail sales prediction and item recommendations using customer demographics at store level. ACM SIGKDD Explorations Newsletter, 2008, 10(2): 84-89.

[17] Krasonikolakis I, Vrechopoulos A, Pouloudi A. Store selection criteria and sales prediction in virtual worlds. Information & Management, 2014, 51(6): 641-652.

[18] Zhao X W, Guo Y, He Y, et al. We know what you want to buy: a demographic-based system for product recommendation on microblogs, Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 1935-1944.

[19] Fukue Y, Masayuki K, Tsuda K. Extracting purchase patterns in convenience store e-commerce market using customer cube analysis, Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin Heidelberg, 2004: 501-508.