Wine-sales prediction by dynamic learning

PANG, Bo

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
8
(開始ページ / Start Page)
15
(終了ページ / End Page)
18
(発行年 / Year)
2013-03
(URL)
https://doi.org/10.15002/00009515

Wine-sales prediction by dynamic learning

Pang Bo
Graduate School of Computer and Information sciences
Hosei University
Email: bo.pang.6w@stu.hosei.ac.jp

Abstract—**Sales prediction is a basis of the enterprises' budget management and it plays a decisive role in the production and sales. This paper proposes a method for prediction of wine-sales at a wine store. A typical prediction system consists of a learning part and an actual prediction part. The learning determines several parameters needed to the prediction. One of known difficult issue is how to cope with unknown changes caused by accidental events. In this paper, to cope with this issue, we examine two methods; static and dynamic learning in the former part, and three methods; template matching, linear regression, and neural networks in the latter part. Experiments using real wine-sales data show that the neural networks with the dynamic learning give us the least prediction error, 23.8% on average. Future work includes the improvement of the method by using other independent time-series data which might operate the prediction.**

Keywords; prediction, template matching, neural network, linear regression, sales

I. INTRODUCTION

The prediction is one of important technologies in economics and natural sciences. People want to provide for and against the future. For example, the weather forecast helps your decision if you should bring an umbrella with you or not. The route prediction of typhoons and the earthquake prediction may save people's life. In economics, people efficiently invest their money in the stock of a company with predicting the stoke-price. A product manager and/or a sales manager have to control the inventory of goods correctly with predicting their stocks.

The data type to handle in the prediction is numeric or categorical. The number of goods is expressed by numeric data while the weather is by categorical data such as "sunny," "cloudy," and "rain." Whatever the data type is, the data is a function of time, t, because it must express a value or a categorical term at each time.

In this paper, we take wine-sales data as an example in order to develop a prediction system. The data are numerical and time-series. Figure 1 shows the sales record of SKU1288 through six years. SKU stands for "Stock Keeping Unit," and "1288" means a kind of the product code corresponding "yellow tail" (Australian wine) for example. This particular prediction can greatly help a wine store because they can adjust the inventory and improve the marketing strategy to gain more profits from the sales. Therefore, the prediction is an indispensable technology for sales and marketing.

Because of the importance of the prediction, many methods have been developed so far. The prediction problem can be normally regarded as the detection of the rules among data in time series. Therefore, this is a so-called mapping problem which can be solved by Template Matching, Linear Regression, Neural Networks, etc.

The Template Matching[1] is used to search for similar time series data to the current data and estimate the future data from the searched data. The Linear Regression[2] and the Neural Networks[3] can be used for analyzing the relationship among the time series data.

In actual sales, the prediction is not only depends on the sales data of the past but also many hidden context, which is not given explicitly in the form of predictive features[4].Some examples of the hidden context are temperature, rainfall and some accidental events in the wine-sales. As shown in Figure 1, the data from the year 2006 through 2008 are much different from that from the year 2009 through 2011 in the amount of sales. There must be some hidden factor between the year 2008 and 2009. This is one of known difficulties in the prediction problems and this paper proposes a prediction system which could solve this issue.

In Chapter II, the prediction system is proposed. In chapter III, after the determination of the parameters to optimize the system, the proposed prediction system is evaluated by using real wine-sales data. In Chapter IV, we concludes the result and states the future study.

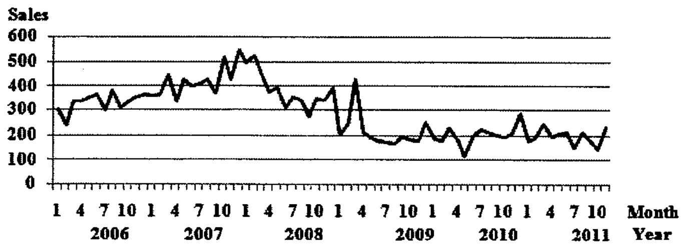


Figure 1. Six-year-sales data of SKU1288

II. METHOD

Figure 2 shows the outline of the proposed system which is composed of three units; a Data Fetch Unit, a Prediction Processing Unit, and a Learning Processing Unit.

Supervisor: Prof. Hiroshi SAKO

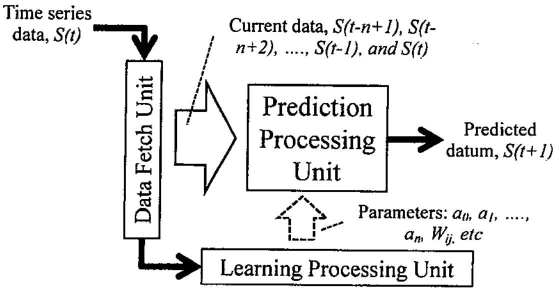


Figure 2. Outline of the proposed method

A. Data fetch unit

The purpose of the Data Fetch Unit(DFU) is to store some range of time series data for predicting the next future datum. The data is shifted into DFU one by one. In the case of SKU1288 in Figure 1, the data unit is monthly sales. It depends on demands and can change it to daily sales or yearly sales. Hereinafter, we denote the data as $S(t-n+1)$, $S(t-n+2)$, ..., $S(t-1)$, and $S(t)$, where t is the current month in the case of monthly data and n is the number of fetched data. For example, if t=7 (July) and n=5 (months), then the fetched data are $S(3)$, $S(4)$, $S(5)$, $S(6)$, and $S(7)$. Usually, the parameter, n is very important in the prediction accuracy. We discuss this issue later in Chapter III.

B. Prediction processing unit

The purpose of the Prediction Processing Unit (PPU) is to analyze the input data from the DFU and to predict the next future datum, $S(t+1)$. For example, if t=7 (July) and n=5 (months), then the system must predict the sales on August, $S(8)$, from the current data, $S(3)$, $S(4)$, $S(5)$, $S(6)$, and $S(7)$ (though they actually includes the past data, but here we call it current data for simplicity). To investigate the rules between the next datum, $S(t+1)$, and the current data, $S(t-n+1)$, $S(t-n+2)$, ..., $S(t-1)$, $S(t)$, Template Matching (TM), Linear Regression (LR), and Neural Networks (NN) are examined.

1) Template matching:This model is based on the assumption that the future datum is also similar if the current data are similar. The current data are scaned on the database in the learing processing unit, and the past data which are similar with the current data are detected by the following matching;

$$t_0 = \operatorname{argmin}\{\sum_{t'} \sum_{i} [S(t'-n+i)-S(t-n+i)]^2\}, \quad (1)$$

where $1 \leq i \leq n$, $t_{min} \leq t' \leq t_{max}$, $S(t'-n+i)(1 \leq i \leq n)$ are the past data at time=t',$S(t-n+i)(1 \leq i \leq n)$ are the current data at the current time=t, and t_0 is the time having a minimum distance. The searching (learning) interval is from t_{min} throuth t_{max}. Then, The next dataum, $S(t+1)$ is expressed by;

$$S(t+1)=S(t_0+1), \quad (2)$$

2) Linear regression:This model is based on the assumption that the future datum, $S(t+1)$, is expressed by a liner equation of the current data;

$$S(t+1)=a_0+ \sum_{i} a_i*S(t-n+i) \quad (3)$$

where $a_0, a_1, \cdots, a_n$ are are constant values determined in advance so as to minimize the total error, E, in the period to learn, $t_{min} \leq t' \leq t_{max}$. E can be defined by;

$$E = \sum_{t'} \{S(t'+1)- a_0-\sum_{i} a_i*S(t'-n+i)\}^2 \quad (4)$$

where $a_0, a_1, \cdots, a_n$ are easily determined by solving $\partial E/\partial a_0=0$, ...,$\partial E/\partial a_n=0$.

3) Neural networks: Though the LR gives you a linear function for prediction, however, some prediction might require a non-linear function, f;

$$S(t+1)=f(S(t-n+1), S(t-n+2), ..., S(t-1), S(t)), \quad (5)$$

which is determined by trained $S(t')$ from t_{min} throuth t_{max}. One of the methods to determine the function is a Neural Networks. The function can be expressed in the form of networks using sigmoid functions and weights W_{ij}. Back propagation can tune the weights so as to minimize the total transform error in the trained data, $S(t')$ from t_{min} throuth t_{max}. Since the range of the output of the neural networks is normally between 0.0 to 1.0, therefore the data $S(t')$ are normalized by;

$$S'(t')=\{S(t')- S_{min}\}/\{S_{max}- S_{min}\}, \quad (6)$$

where $S'(t')$ is the normalized data, S_{min} is set a little below from the minimum value of $S(t')$ and S_{max} is set a little above from the maximum value of $S(t')$ so as to the range between S_{min} to S_{max} covers all data of $S(t')$.

C. Learning processing unit

The purpose of the Learning Processing Unit (LPU) is to deliver the parameters such as the past data, $S(t'-n+i)$ in the TM, the coefficients, a_0, a_1, $\cdots$,a_n in the LR, and weights, W_{ij} in the NN, to the PPU.

Those parameters are determined in advance by using $S(t'-n+i)$. The learning procedures are explained already in previous section. We call this learning static learning. This learning uses the data $S(t'-n+i)$ from a fixed period $t_{min} \leq t' \leq t_{max}$ and it is done in offline mode.

To cope with drastic change caused by hidden factors explained in the introduction, a dynamic learning is proposed here. The kinds of parameters are same as those of the static learning, but the values of parameters are changed dynamically in order to fit with the tendency of the data. The timing of changing parameters is important, but it's very hard to find out the data changes. Therefore, this method calculates optimum parameters at every prediction. The top and end of the searching (learning) interval, t_{min} and t_{max}, must be a function of time. The most simple case is $t_{min}=t-t_{lenght}+1$ and $t_{max}=t$, the searching (learning) always uses previous t_{lenght} number of data before the current time t in order to calculate the parameters by online. We compare the dynamic learning with the static learning in the next chapter.

III. Experiments

In this chapter, we firstly define the prediction error rate, and secondly discuss the optimum number of the fetched data, and then we show the prediction results of the TM, the LR, and the NN using the dynamic learning, which are compared with the results by using the static learning. Moreover, we show the prediction result of a different SKU case and the result of the case added by temperature information.

These experiments use mostly monthly sales-data of the wine, SKU1288, which were recorded from 2006 through 2011. The original data was already shown in Figure 1.

A. Definition of the prediction error rate

In order to investigate the accuracy of the proposed method, we need a quantitative definition. In this paper, the following simple prediction error rate, E_{od}, is defined;

$$E_{od} = |S(t+1) - S_r| / S_r, \qquad (7)$$

Where $S(t+1)$ is the prediction sales-datum, and S_r is a real sales.

B. Optimum number of the fetched data

The number of the fetched data, n is very important because it usually influences the prediction accuracy. Figure 3 shows the variations in the average error rate of the 2011 sales-data against the number, n, when it changes. The experimental condition is that the method is the LR with the static learning. The learning interval is fixed by t_{min}= 1 (Jan.) of 2006 and t_{max}= 12 (Dec.) of 2010.

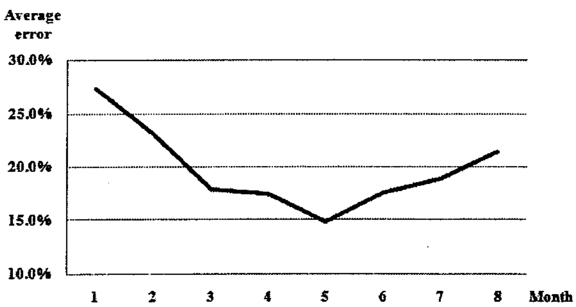
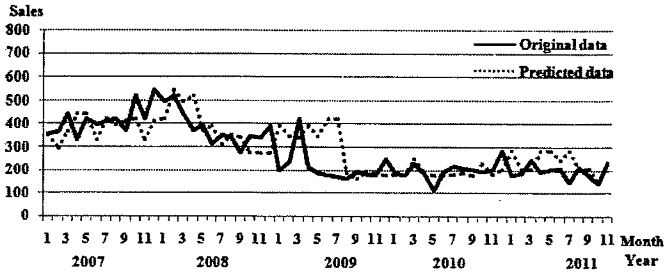


Figure 3. Variations of the prediction error rate

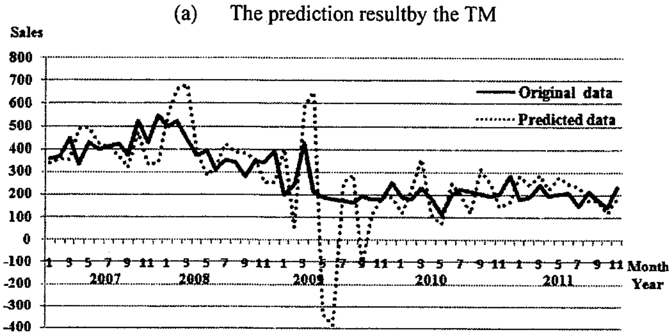
As shown in the figure, the error rate in the case of 5 months is lowest. Therefore, we set n=5 and uses this in the following experiments. We assume that this result is applicable to the other methods and the dynamic learning.

C. Prediction using the dynamicand the static learning

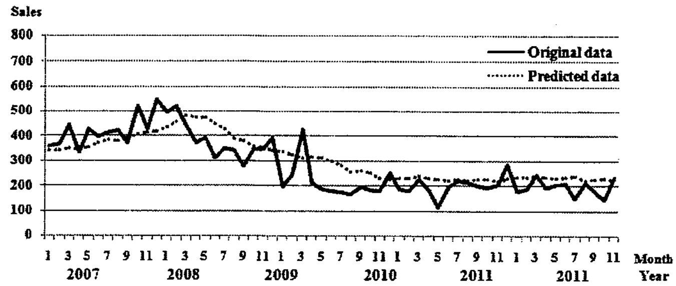
1) Prediction using the dynamic learning: Figure 4 (a) (b) and (c) show the prediction results by the the TM, the LR, and the NN using the dynamic learning under the condition that $t_{min}=t-11$ and $t_{max}=t(t_{lenght}$ is 12), and n=5. Each dotted line in the figure indicates the result obtined by each method. Table 1 and Figure 5 show the avearage error rate of each year by each method. The structure of the NN is 5*10*1 neurons.



(a) The prediction resultby the TM



(b) The prediction resultby the LR



(c) The prediction resultby the NN

Figure 4. The prediction results by the dynamic learning

Table 1. The prediction error in the dynamic learning

Year		2007	2008	2009	2010	2011	Ave.
E_{od} %	TM	14.8	17.5	56.8	14.7	28.6	26.5
	LR	16.1	22.5	115.0	36.0	26.9	43.3
	NN	11.0	18.0	44.9	21.0	24.1	23.8

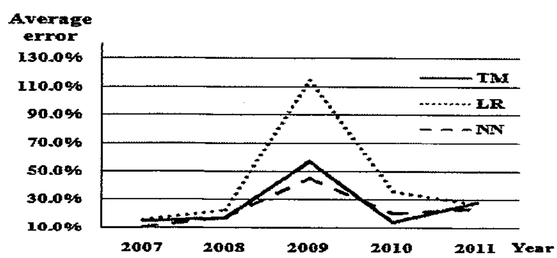


Figure 5. The prediction error in the dynamic learning

2) Prediction using the static learning: Table 2 and Figure 6 showthe avearage error rate of each year by the the TM, the LR, and the NN using the static learning. The condition is;t_{min}=1 (Jan.) and t_{max}=12 (Dec.) of the previous year, and n=5.

Table 2. The prediction error in the static learning

Year		2007	2008	2009	2010	2011	Ave.
E_{od} %	TM	15.0	30.1	77.2	20.2	38.2	36.1
	LR	19.1	66.8	68.7	24.8	29.9	41.9
	NN	18.0	22.8	74.5	23.1	27.2	33.1

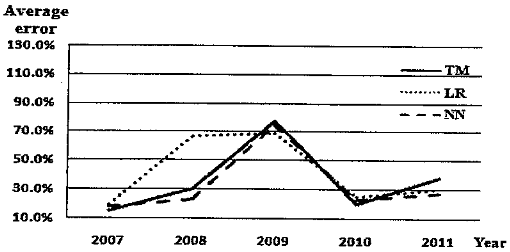


Figure 6. The prediction error in the static learning

3) Consideration: The comparison between the results shows that the dynamic learning gives us more accurate prediction, especially, the prediction in the year 2009 just after some hidden factors changed the sales. The NN gives us a much better prediction results.

D. Another SKU case

The same experiment is done by using another sales data of SKU1334. Figure 7 shows the results by the NN with the dynamic learning. The average error is 17.6%, which is a slight better than that of SKU1288. In this SKU, the optimum number of the fetched data, n=1.

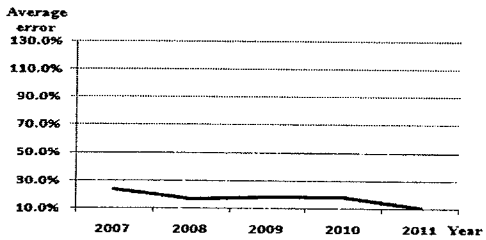


Figure 7. The prediction error(NN)in the dynamic learning (SKU1334)

E. Addition of another information to predict

Here, we attempt to use temperature data as well as the sales data for the prediction. Figure 8 shows the temperature data at the location of a wine shop which gave the sales data.

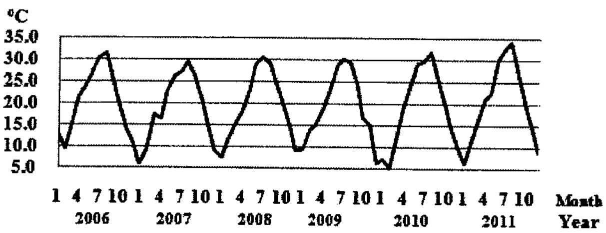


Figure 8. Monthly-averaged temperature

In this experiment, we use the TM with dynamic learning, because the implementation of the method is simple enough for this first attempt. The sales data are normalized by temperature and they are used in the experiment. Table 3 shows the results.

Table 3. The prediction error by the addition of temperature info.

Year	2007	2008	2009	2010	2011	Ave.
E_{od}% Temper. Info.	12.8	16.3	37.8	24.5	25.1	23.3
E_{od}% No info.	14.8	17.5	56.7	14.7	28.6	26.5

F. Summary of experiments

We summarize the results of the experiments as follows;

1) When the number of the fetched data, n is five, the prediction error is minimum in the case of SKU1288.
2) The dynamic learning gives better prediction than the static learning.
3) The NN is the best method to predict, the second is TM, and the last one is the LR.
4) Another SKU data gives similar prediction accuracy, which proves that the results are reliable.
5) The attempt using the temperature data gives a slight better prediction.

IV. Conclusions

This paper proposed a prediction system which could cope with data having big variation caused by some hidden factors. The system consists of three units; a data fetch unit, a prediction processing unit, and a learning processing unit. We examined three prediction methods; Template Matching, Linear Regression, and Neural Networks, and proposed the dynamic learning. Experiments using real wine-sales data show that the Neural Network with the dynamic learning gives the least error rate of the prediction, 23.8 %.

The future work includes; 1) attempts to use other information such as rainfall, holidays, promotions, and economic environment, because the attempt using the temperature has better effect, and 2) mining the useful features detected from the original sales data.

ACKNOWLEDGEMENTS

Special thanks go to an US wine shop which gave us real wine-sales data and some useful comments about prediction.

References

[1] Cuiling lan, Jizheng Xu, Feng Wu and Guangming Shi, "Intra Frame Codingwith Template Matching Predictionand Adaptive Transform," *Proceedings of 2010 IEEE 17th International Conference on Image Processing*, September, Hong Kong, 26-29, 2010.

[2] Kunhui Lin, Qiang Lin, Changle Zhou, Junfeng Yao,"time Series Prediction Based on Linear Regression and SVR,"*Third International Conference on Natural Computation*, 688-691, 2007.

[3] William W. Hsieh and Benyang Tang,"Applying Neural Network Models toPrediction and Data Analysis inMeteorology and Oceanography,"*Bulletin of the American Meteorological Society*, Vol. 79, No. 9, September, 1855-1870, 1998.

[4] Partrick Meulstee, Mykola Pechenizkiy,"Food Sales Prediction:If Only It Knew What We Know,"*2008 IEEE International Conference on Data Mining Workshops*, 134-143, 2008.