An Application of Data Mining for Actual Products in the Market

Onozaki, Tsuneo Leo*

Introduction

This paper discusses the method used in the actual time-series forecasting performed by the author. First, an ordinary method using back propagation neural networks is recalled. Powerful as it is, that method is not quite appropriate for our purpose. Then an alternative method invented by the author is explained.

Let us recall that the first (and most important) phase in analyzing time-series data like market trend is the selection of attributes and variables that are used to explain the change over the time. Typical ones include product attributes and external conditions such as economic trend.

Then, prediction models are generated, based on these attributes and variables, using data mining techniques. The predicted values given by the models are compared with the real data, and the accuracy of each model is calculated to see the validity of the selection of attributes and variables.

There are many methods for time-series predictions advocated by many people (especially ones without using data mining techniques) and this report can't exhaust them in any sense. As mentioned above, we should be going to review two possible approaches based on data mining, namely the sliding window and the author's original one.

---

* Professor at Management, School of Logistics, Ryutsu Keizai University, Ibaraki, 301-8555, Japan.
Phone: +81-297-64-0001 ex.624, FAX: +81-297-64-0011,
E-mail: onozaki@ipc.ryukei.ac.jp

1. Ordinary approach

The usual method used in time-series predictions is called the sliding window. This is conceptually discussed in the other paper (Onozaki, 1998), but will be discussed again here in a more pragmatic context.

1.1 The sliding window

The sliding window is a method of making time-series predictions based on past trend data. This method is powerful if the phenomenon can be thought quasi periodic, i.e., the data values to be predicted are fluctuating in a bounded way.

There are many phenomena of this type: stock value indices are typical examples. The advantage of this method over traditional linear regression is that it takes account in non linear factors.

1.2 Implementation

We can assume that the values to be predicted are fluctuating in a bounded way. This was stated above. Furthermore, the fluctuation can be explained using known factors. That is, we have at hand the attributes that influence the data to be predicted. For example, stock values are known to be sensitive to other economic indices like prime rate and currency exchange rate.

The prediction model is constructed as follows: for the sake of simplicity suppose the auther has eight time points of data and an attribute to explain the data at each time point. The model is fed the first four (Set $d = 4$) tuples of (attribute, data) as the variable vector and the data at the fifth ($= d + 1$) time point as the "teacher". We can proceed in the same fashion with the second to the fifth (I.e., 2 to $d + 1$) tuples as the variable vector and the data for the sixth ($=$ d $+$ 2) time point as the teacher. This continues until we reach the forth time point.

At the completion of the forth round, the model is considered "trained" and the fifth to eighth tuple is given to the model for prediction.

2. The alternative method

2.1 The author's original

The sliding window described in the other paper (Onozaki, 1998) is very powerful to predict phenomena of bounded fluctuations. Note that the sliding window actually performs data interpolation.

However there are situations in which we would like to make predictions for rapidly decaying trends. The sliding window doesn't look very promising for those situations since *the values to be "interpolated" are "out of scope" the experienced fluctuation.*

Thus we have to resort to yet another method.

2.2 Implementation

The auther proposes an alternative to the sliding window method to be used for predicting rapidly decaying trend.

This method comprises two distinct phases. *The first phase* performs clustering of trend patterns in question based on the data value and assign an identification number to each cluster. For the clustering of time-series data, we should use Kohonen self-organization map networks rather than demographic clustering. Kohonen self-organization map networks generate fairly comprehensible results. For example, clusters are described by initial values and damping rate.

The second phase performs a training of classification based on attributes (rather than chronological data) to explain the clustering result. For classification we should adopt back propagation network. It is shown that the network classifies given attribute data to explain the clustering.

Back propagation is good for tasks like making distinctions based on subtle differences in data. However, there is a serious drawback: its training algorithm is extremely processing-power consuming. That is, training back propagation networks requires a long time. Robert Heinline has it, "There ain't no such things as a free lunch."

3. Actual forecasting

As is mentioned in the introduction, the auther would like to analyze cash flow related to personal computers in the market (PCs) by making sales forecast. Since this amounts to making predictions for rapidly decaying trends, the traditional time-series prediction methods using linear regressions are not very applicable.

It should be proceeded as stated in the other paper (Ozozaki, 1998) following the "alternative" method explained above:

Step 1. Segmentize PCs into clusters based on the chronological sales figures.

Step 2. Create classification models to explain the clustering result by "attributes" of the products and other factors.

Step 3. Predict sales figures for new products using the models, with known intrinsic and external factor data.

Step 4. Estimate the cash flow from the predicted sales figures.

4. Noise and missing data handling

A device to handle that inevitable noise can be incorporated into (the usage of) neural networks. Kohonen feature maps are robust to noisy data. For back propagation networks, avoidance of over-training is the key to overcome noisy data. The prediction accuracy for the training should be kept something like 80% (rather than that tempting 100%).

As for missing values, we should use some device suggested in Wittmann and Ruhland (1998): instead of simply filling the missing value by the average of the total population,

1) perform a clustering on the portion of data without missing values,
2) calculate the average for the item corresponding the missing values for each cluster identified in 1),
3) apply the clustering result to the data with missing values,
4) fill the missing value with the average of the cluster to which each datum (with a missing value) belongs.

In contrast to the usual method of filling with the average of the total population, the method by Wittmann and Ruhland (1998) does not bring any noise into the data.

5. Results

The auther gathered PC sales figures from various sources, mainly published reports from domestic and overseas market research firms and MOF, MITI and the other agencies in Japan. Note, however, that all names and figures used in this paper have been modified to be fictional ones and hence any resemblance to existing entities is only a coincidence.

Table 5.1 summarizes the data for two corporates of interest together with the industry trend.

Table 5.1 Sales figure summary

FY 1996	*No. of models*	*Shipped Units*	*Revenue (Million ¥)*
Vendor A	96	2,638,000	896,885
Vendor B	80	1,775,000	517,155
Industry Total		8,099,220	2,673,980
FY 1997		*Shipped Units*	*Revenue (Million ¥)*
Vendor A	100	2,371,500	692,875
Vendor B	90	1,870,500	528,375
Industry Total		7,927,090	2,390,375
1H 1998		*Shipped Units*	*Revenue (Million ¥)*
Vendor A	52	1,277,150	?
Vendor B	48	1,085,700	?
Industry Total		4,095,300	?

At this writing, only the (estimated) result for the first half year is known. Also, no estimation on revenues has been published for that period.

5.1 Product attributes used

As was mentioned in the introduction, it was assumed that the sales figure of each model can be predicted by its attributes (and historical facts). Table 5.2 shows the main attributes and derived ones used to create the prediction models.

Sometimes, we experienced missing values: many models are now sold under "open pricing" policies.

Table 5.2 Main attributes used

Vendor	Identification only
Model	Identification only
Type	Desktop/Notebook/Mobile
Price	Wholesale, estimated
Weight	(Kg)
Dimension- Width	(cm)
Dimension- Depth	(cm)
Dimension- Height	(cm)
Processor Speed	(M-Hz)
Memory-Installed	(MB)
Memory-Max	(MB)
VRAM	(MB)
Display Size	(inch)
HDD	(GB)
CR-ROM	Speed
FAX-MODEM	Speed
Price/Weight	Calculated
Price/Processor-Speed	Calculated
Price/Memory-Installed	Calculated
Price/VRAM	Calculated

To estimate the wholesale price, the auther

1) assumes that the wholesale price is sixty-five percent of the retail price,
2) performed a clustering on all the models using attributes other than the price,
3) then calculate the average of the price for known ones for each cluster,
4) set the price as the average for "unknown" product models.

While this may not add any further information to the known data, it is much better than to either

a) simply discard the data containing missing values

or **b)** put the mean value for the missing item.

Note that

a) would reduce the quantity of information and **b)** may introduce unnecessary noise to the data.

For the discussion of missing value handling, see the interesting article by Wittmann and Ruhland (1998) and references cited there.

5.2 Clustering

Figures 5.1 and 5.2 depict the result of clustering on sales figures of models provided by Vendor A and Vendor B, respectively. It was used Kohonen self organization map neural networks to cluster the data, as most attributes used are of numerical nature. The auther didn't use the data for the fiscal year 1996, since the overall figures for that year are "higher" than that of the fiscal year 1997 (See Table 5.1.) Using the data for 1996 might introduce inexplicable bias in the prediction.

Note that the Kohonen self organization map neural networks give similar (but, of course, not identical) clustering results for the two vendor data.

Figure5.1 Clustering for A 1997 Sales Figures

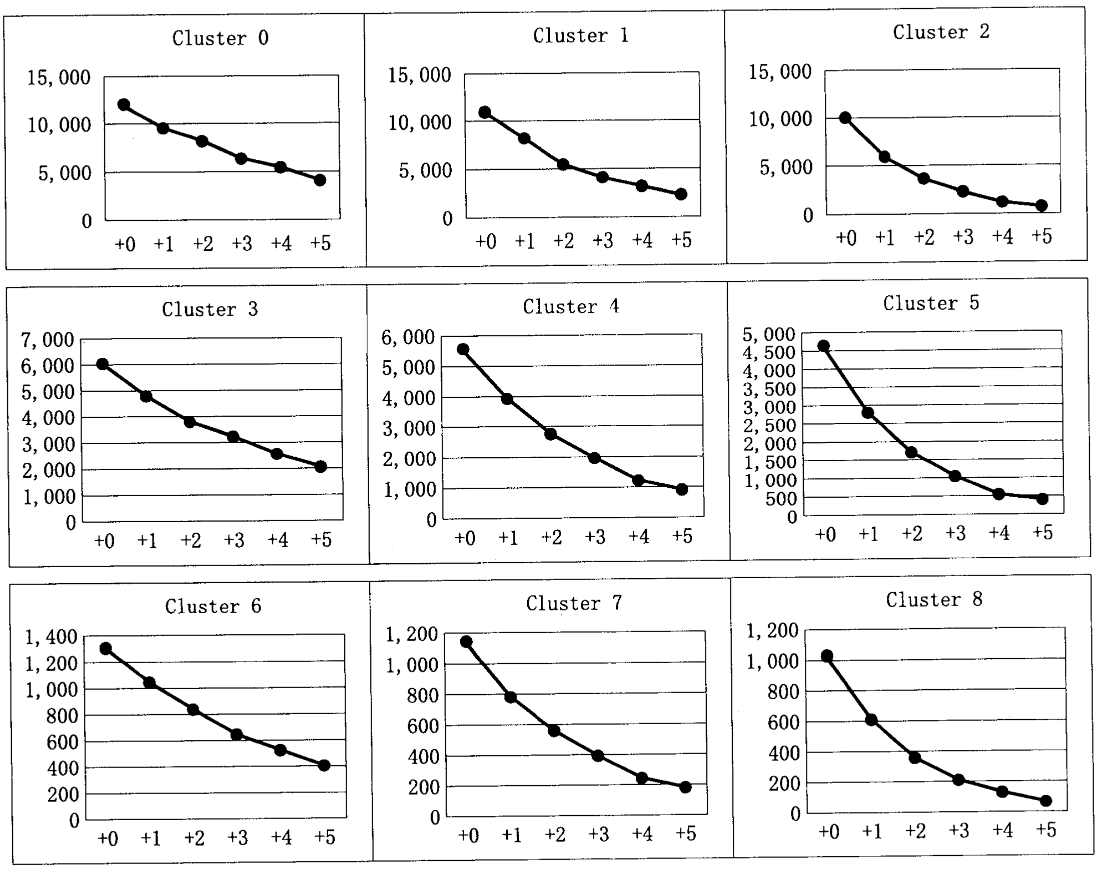
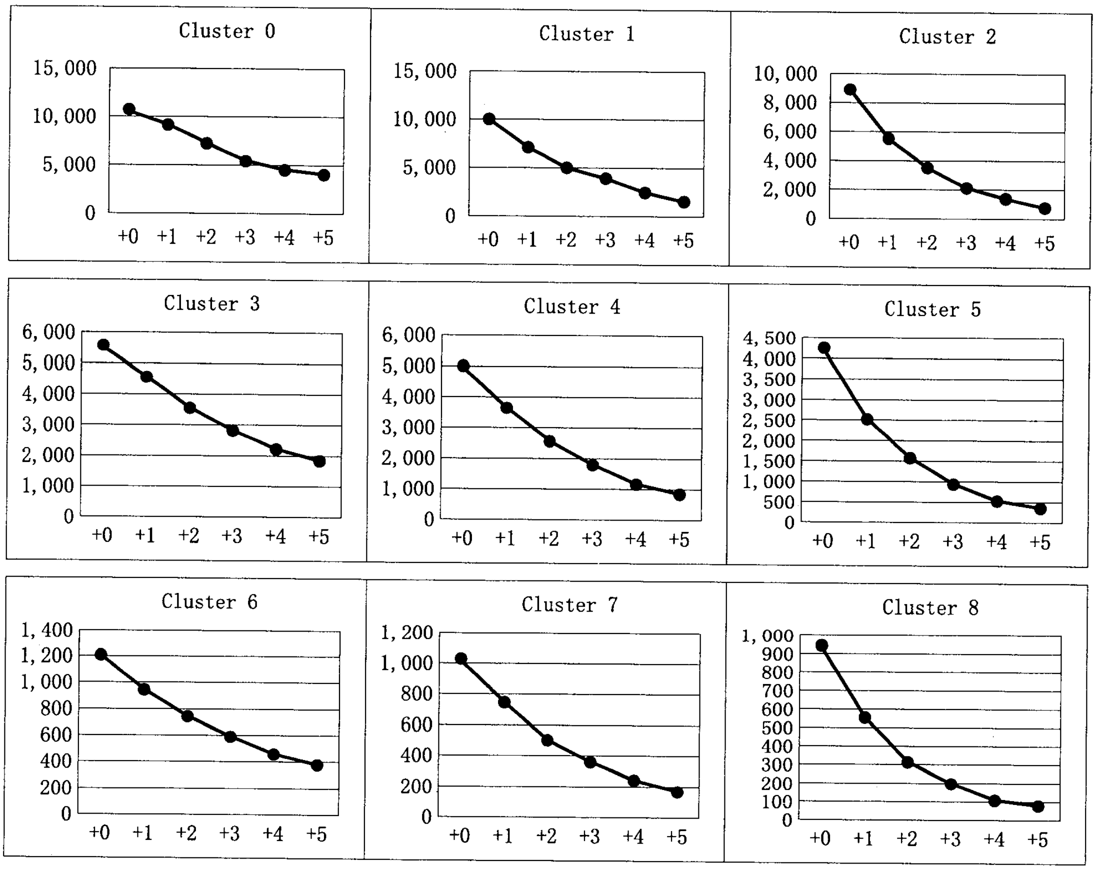


Figure5.2 Clustering for B 1997 Sales Figures



5.3 Classification

Prediction models for the cluster numbers in Figure 5.1 and Figure 5.2, respectively, were generated using the attributes listed in Table 5.2 as input data for the two vendors. It was used back propagation neural networks with one hidden layer for the classifiers. As for the training parameters, the anther specified 90% accuracy and 5% error allowance to avoid over-training. Besides the time-consuming process, back propagation networks have the drawback that the only output they generate are the predicted values: there are no practical way to grasp the "shape" of the models.

5.4 Prediction using the model

The sales figures for the first half are "predicted" using the established models. Tables 5.3 and 5.4 summarizes the prediction result.

Table 5.5 summarizes the comparison between the "actual" figures in table 5.1 and the predicted ones in tables 5.3 and 5.4.

Table 5.3 Prediction for Vendor A, the first half, 1998

Cluster	Shipped Units (K=1,000)	Revenue (Yen Million)
0	407	99,636
1	262	59,387
2	167	29,497
3	157	61,819
4	114	43,056
5	93	30,473
6	9	5,843
7	9	5,689
8	7	3,773
Total	1,225	339,173

Table 5.4 Prediction for Vendor B, the first half, 1998

Cluster	Shipped Units (K=1,000)	Revenue (Yen Million)
0	330	78,098
1	209	45,282
2	168	33,133
3	143	55,493
4	74	27,291
5	56	19518
6	9	4863
7	9	5,096
8	4	2,495
Total	1,002	271,269

Table 5.5 Comparison between actual and prediction

	Shipped units	Predicted units	Prediction error (%)
Vendor A	1,277,150	1,225,000	4.1
Vendor B	1,085,700	1,002,000	7.7

6. Discussion

Table 5.5 shows that the prediction error is less than 10% for the both vendors. For the practical purpose, this is not bad.

It is also predicted revenues, but we have no way to estimate the accuracy for the prediction since no actual figures are available for revenues for the first half of the year 1998.

The auther should study on the accuracy for the prediction related revenues for that period and report it in the near future.

References

Adriaans P., Zantinge, "*Data Mining*. Addison-Wesley," Harlow England 1996

Bender E. A., "*Mathematical Methods in Artificial Intelligence*," IEEE Computer Society Press, Los Alamos, CA 1996

Bigus, J.P., "*Data Mining with Neural Networks*," McGraw-Hill, New York 1996

Bose, N. K. & Liang P., *Neural Network Fundamentals with Graphs, Algorithms, and Applications*. McGraw-Hill, NY 1996

Bourbaki, N., "*Elements de Mathematique*," Topologie Generale Chapitre 10. Deuxieme edition. Hermann, Paris 1961

Breiman, L. & Friedman, J., Ohlsen, R., and Stone, C., "*Classification and Regression Trees*," Wadsworth & Brooks, Pacific Grove, CA 1984

Conway, J. B., "A Course in Functional Analysis," 2nd edition. Springer, Berlin 1990

Fiesler, E. & Beale R. (Editors), "*Handbook of Neural Computation*," IOP Publishing and Oxford University Press, 1997

Hecht-Nielsen, R., "Kolmogorov's mapping neural network existence theorem," *Proceedings of International Conference on Neural Networks III*. IEEE , NY pp11-13. 1987

Jubak, J., "*In the Image of the Brain: Breaking the Barrier between the Human Mind and Intelligent Machines*," Little Brown, Boston 1992

Kohonen, T., "*Self-organization and Associative Memory*," 2nd edition, Springer, Berlin 1988

Kohonen, T., "The self-organizing-map," *Proceedings of the IEEE*, 78(9) : 1464-1480.

Kolmogorov, A. N., "On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition," *Doklady Akademii Nauk USSR*, 114:953-956,1957.

McLelland, J. L. & Rumelhart, D. E., "*Parallel Distributed Processing*," vol.2. MIT Press, Cambridge MA 1986

Mehrotra K., Mohan C. K. & Ranka S. *Elements of Artificial Neural Networks*," MIT Press, Cambridge MA 1997

Minsky M. & Papert S., "*Perceptrons*," 2 nd edition, MIT Press, Boston, MA 1989

Nakhaeizadeh G. & Taylor C. C. (Editors), "*Machine Learning and Statistics*," John Wiley & Sons, NY 1997

Rudin W., "*Principles of Mathematical Analysis*," McGraw-Hill, New York 1964

Rumelhart, D. E. & McLelland J. L., "*Parallel Distributed Processing*," vol. 1. MIT Press, Cambridge MA 1986

Onozaki, T. L., "*Data Minig as an Effective Measuring Method for Corporate Growth*," Ryutsu Keizai Daigaku Shakaigakubu Ronso, Vol.9, No.1, Ryutsu Keizai University, 1998

Schalkoff R. J., "*Artificial Neural Networks*," McGraw-Hill, New York 1997

Wasserman P. D., "*Advanced Methods in Neural Computing*," Van Nostrand Reinhold, NY 1993

Werbos. P., "*The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting*," Wiley, NY 1994

Wittmann T. & Ruhland J., "Intelligent Data Preprocessing for Neuro-Fuzzy Data Mining," *Fuzzy Demonstration-Zentrum*, Dortmund, EFDAN '98, 1998