Data Mining as an Effective Measuring Method for Corporate Growth

ONOZAKI, Tsuneo Leo*

Introduction

Every corporate must keep growing to continue to survive and to "reward" its shareholders for their support. To insure its growth, every corporate should remain competitive in the market. A steady cash flow as the corporate financial background is required to effectuate the competitiveness. In view of this, the competitive advantage of a corporate can be measured in terms of the following:

A. Market Share

B. Revenue

C. Profit

D. Profit Ratios

It is a well known fact that for a corporate to open up a new area or to get a new set of customers requires five times more effort to than to maintain and promote what it currently has. The current article focuses the competitive advantage form the view point of maintaining and promoting. Thus, certain aspects like research and development and related investment problems will not be covered here.

The author is trying to establish a corporate evaluation methodology with cash flow analysis. Since competitiveness should be measured for the near future as well as the past and the present, this study inevitably involves a kind of time-series financial prediction on the cash flow.

The subject of the cash flow study is a certain product that shows rather short life cycles (henceforth PSL), typically half a year. The sales figures of PSLs, when plotted chronologically, show a rapidly decaying feature. (See figure 1.1.)

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

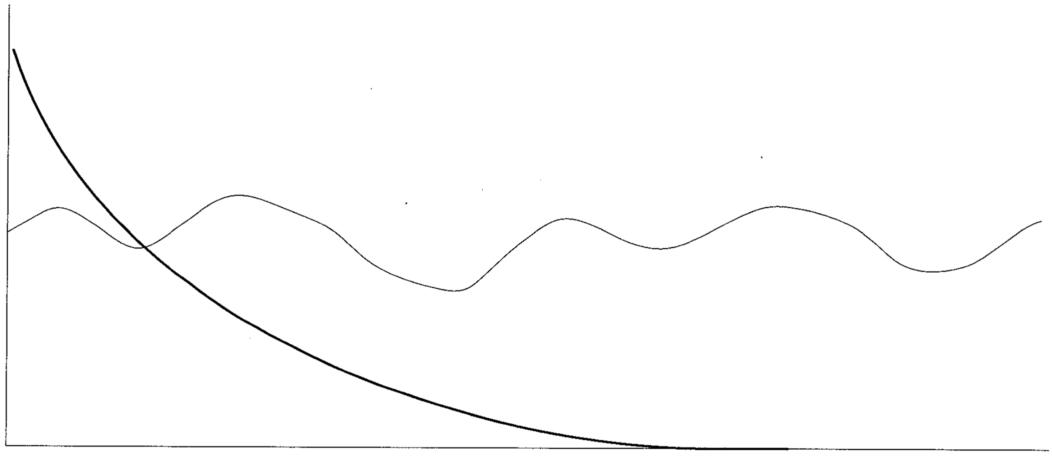


Figure 1.1 A rapidly decaying trend vs quasi-periodic phenomenon

The future cash flow analysis should depend on the sales forecast on the product. Since usual prediction methods like linear regression are not very promising for our case study, the auther resorts to data mining to analyze available financial data and to obtain reasonable estimation for the future trend.

2. Data Mining

Let us recall what is meant by data mining: "Data mining is the efficient discovery of valuable, nonobvious information from a large collection of data." (Bigus 1996), The auther is going to deal with cases where the information necessary to foresee the future lies in the past trend and hence data mining can be used to making predictions. To this end, techniques of data mining like classification and clustering are used. It is briefly reviewed those techniques along with their implementation, following Bigus (1996) and Mehrotra, Mohan and Ranka (1997).

2.1 Classification

Classification is the assignment of each object to a specific "class" based on similarities and differences between the objects. This has been successfully used in areas like pattern recognition. In the business field, classification of customers and/or prospects into "profitable" and "other" ones is of fundamental importance. For example, a card business would like to "reward" profitable (= of high lifetime value) customers with marketing promotions, and at the same time would like to "encourage" unprofitable (= of low lifetime value) customers to leave.

In classification, a "training set" comprising sample patterns that represent all "possible" classes is given, together with class membership information for each pattern. Rules for membership in each class can be deduced using the training set to create a classifier that, in turn, can be used to assign other ("new") patterns to their respective classes based on these rules. Of course, this assumes that the trend in the near future can be predicted using the currently available data.

2.2 Clustering

Clustering groups objects that are similar to each other. In contrast to classification where the identification of classes is *a priori* known, clustering requires grouping together objects in the sample according to their attributes. That is, similarities should be measured by the distance relationships that has to be derived from the sample descriptions. (See figure 2.1.)

In the business field, clustering is mainly used in the field of marketing. Many business corporates have brought into practice the so called "target marketing" by clustering customers into groups based on the spending patterns such as the products they buy or services they use. In that knowledge, marketing promotion can be "targeted" toward customers most likely to purchase. This amounts to so-called the customer relationship management in which each customer is treated according to his/her characteristics and/or idiosyncrasies rather than collectively and uniformly,

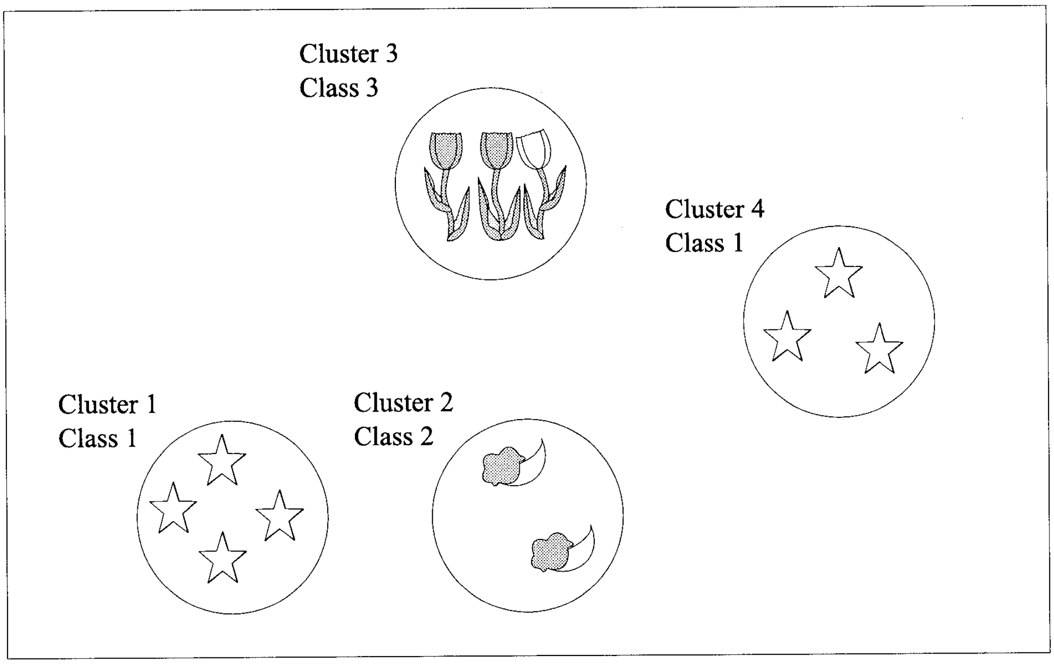


Figure 2.1 Four clusters and three classes in input.

that is mechanically, as one of the "silent majority".

2.3 Value prediction

There are processes for which the outputs corresponding to some input may be known from training data, but the exact form of the mathematical function that generates the outputs from the inputs is not likely to be read off from the data. Value prediction learns or constructs a function that approximately generates the outputs from input data. In other words, value prediction is a kind of function approximation. The "desirable" approximating function should be simple in terms of science, that is, continuous or smooth and at the same time, in terms of performance (= minimum error). Those two constraints sometimes oppose each other, especially in the real world where noise in data is inevitable.

3. Neural Network Models

The implementation of the data mining techniques explained above is usually done using neural networks.

Let us recall that a neural network is a labeled directed graph in which each node performs some computation and each connection transmits a signal from one node to another. (For practical reasons, the number of nodes are finite. This is essential when implementing neural networks on digital computers.) The label ("connection weight" or "weight") on each connection indicates the value multiplied to the signal. Usually, node (or units) are partitioned into subsets called layers. (See figure 3.1.)

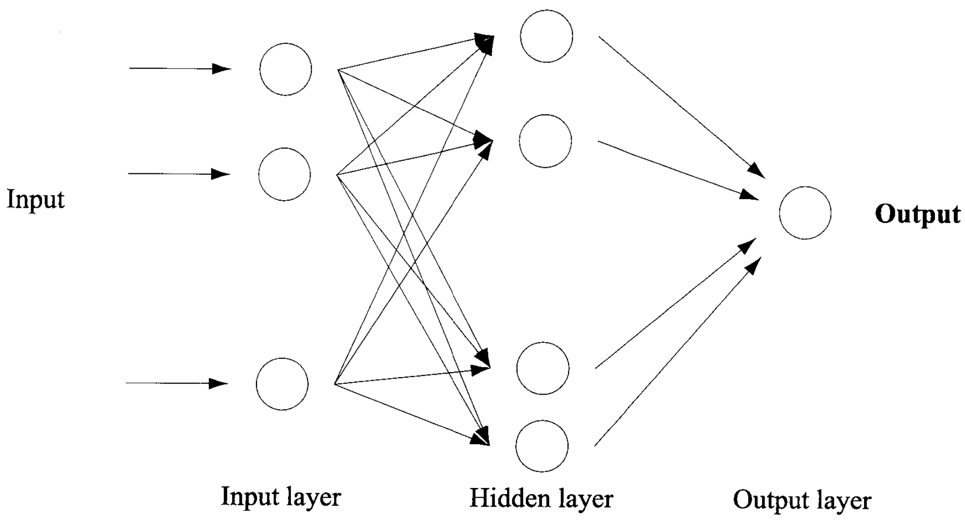


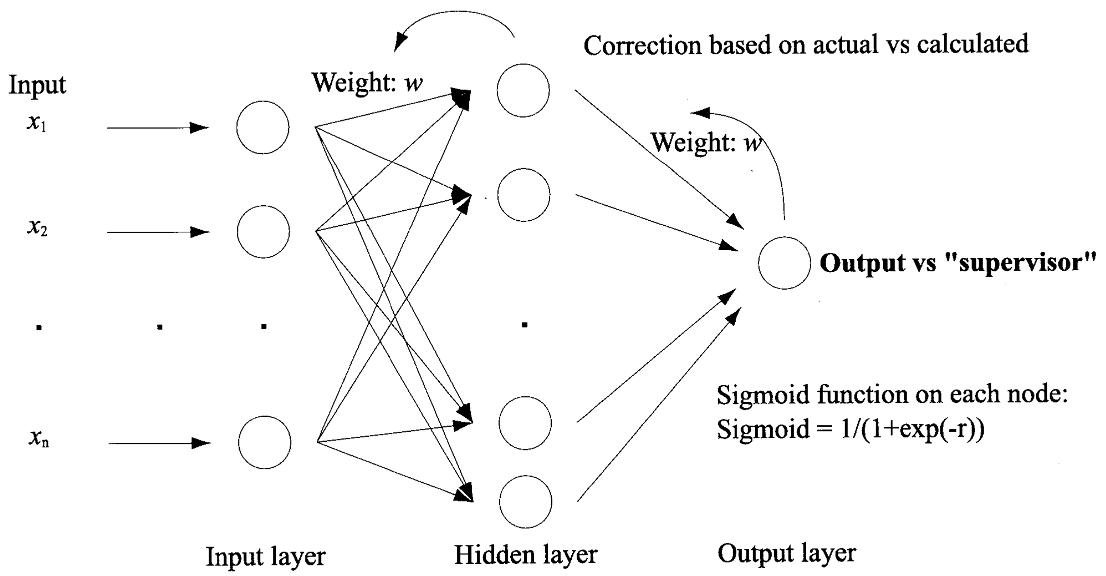
Figure 3.1 A neural network

It is briefly reviewed how those techniques are implemented. No detailed discussion of the theory of neural networks is meant here since it is beyond the scope of this article. An elementary but readable account is given in Mehrotra, Mohan, and Ranka (1997).

3.1 Back propagation networks

A back propagation network is a feedforward supervised learning with the back propagation algorithm. It has three (or more) layers, one for input, one (or more) hidden layer(s) and one for output. (See figure 3.2.)

The input pattern is fed to the input layer. These inputs are propagated through the network until they reach the output units. This forward pass generates the actual or predicted output pattern. The desired outputs are provided ("supervised".) The outputs for actual inputs are subtracted from the desired outputs to yield an error signal. This error signal is passed back through the network (hence the name "back propagation") by calculating the contribution of each unit in the hidden layer(s). The



1. The ith node in the input layer holds a value $x_{p,i}$ of pth input pattern.
2. The net input to the jth node in the hidden layer = $net_j^{(1)}$ = Sum($w_{j,i}^{(1,0)}x_{p,i}$). This includes the threshold with $x_{p,0} = 1$; the connection from the ith input node to the jth hidden layer node is assigned a weight value $w_{j,i}^{(1,0)}$.
3. The output of the jth node in the hidden layer = $x_{p,j}^{(1)}$ = Sigmoid(($w_{j,i}^{(1,0)}x_{p,i}$)).
4. The net input to the output node layer = $net^{(2)}$ = Sum($w_j^{(2,1)}$), including the threshold; the connection from the jth hidden layer node to the output layer node is assigned a weight value $w_j^{(2,1)}$.
5. Output of the output layer o_p = Sigmoid(Sum($w_j^{(2,1)}$)).
6. The desired output of the output layer is dp, and the corresponding squared error is $l_p^2 = |d_p - o_p|^2$.

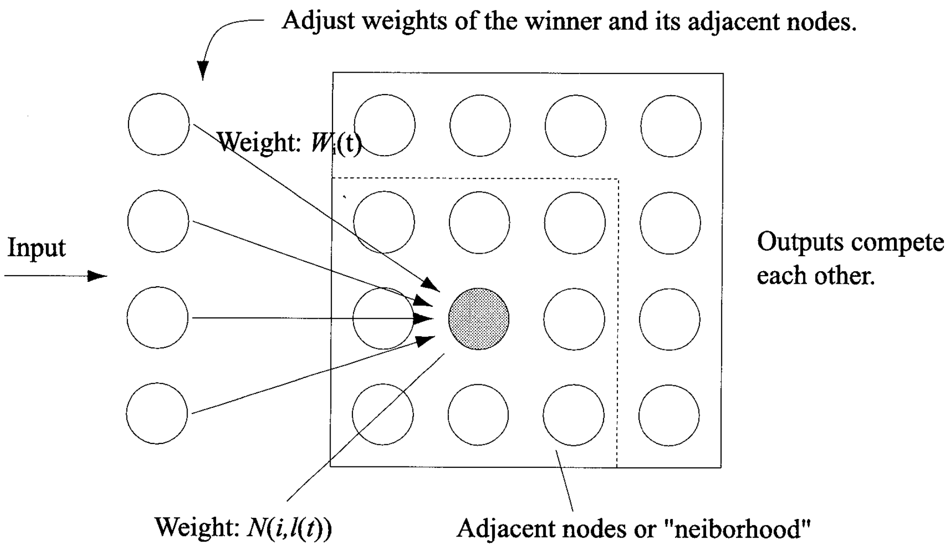
Figure 3.2 A back propagation network connection weights are adjusted and the network has "learned".

For classification problems, the input attributes are mapped to the desired classification categories. The training of the network amounts to obtain the appropriate set of discriminant functions to correctly classify the inputs.

For more details on back propagation networks, see Rumelhart and McLelland (1986).

3.2 Kohonen feature map networks

Kohonen feature map networks (also called self-organizing map networks) are two-layer networks that use a competitive learning principle with a process called self-organization. A feature map network has an input layer fully connected to an output layer. It transforms an input pattern to select an output unit (called the "winner") according to the proximity of the connection weights to the input. The connection weights of the winner and adjacent nodes get updated in the following fashion : . (See figure 3.3.)



The self-organization algorithms consist of two steps: for an input vector x, find the neuron whose activity is maximum. Then, in a defined subset around this maximum $N(i,l(t))$, the weight vectors arc moved in the direction of the input vector x according to the equation

$W_i(t+l) = W_i(t) + a(t)*(x(t) - W_i(t))$ for i in $N(i,l(t))$

$W_i(t+l) = W_i(t)$ otherwise.

The function $l(t)$ controls the width of the neighborhood, and $a(t)$ controls the amplitude of the weight modification. These functions are decreasing over the time t. Iterations of these steps build an organized network, where the weights are ordered and quantify the input.

Figure 3.3 A Kohonen feature map

With Kohonen feature map networks, we can map inputs of many variables into, say, two dimensional square to obtain comprehensible clustering images. The two axes of each square correspond to the two most significant (non-linear) combinations of attributes that characterize the data. In other words, a Kohonen feature map network performs a non-linear dimension reduction.

For more details on self-organizing networks, see Kohonen (1988).

4. Mathematical Background

It needs some mathematics to understand why data mining (with neural networks) works. General notions from elementary set theory, elementary calculus and point set topology such as "countability", "function" and "continuity" are assumed. Rudin (1964) covers most of these in a succinct fashion. It is highly recommended for anyone who wants to (re)study analysis starting from elementary calculus.

The principle of neural networks is to regard phenomenal data as continuous functions on a compact (=closed and bounded) subset of Euclidean spaces and to approximate them with a finite number of "nice and tame" (smooth, say) functions residing in the nodes. It stands to reason to assume the continuity of the phenomena and the compactness of the "space and time" for the phenomena since we are dealing with continuous economic events that show bounded behavior in a finite time duration.

Let us recall why this "simplification" works. The Stone-Weierstrass' Theorem asserts:

The Stone-Weierstrass's Theorem

The set of continuous functions on a compact set has a countable dense subset in the uniform convergence topology.

The auther will not give a proof for Stone-Weierstrass' theorem here. See Bourbaki (1961), Conway (1990) or Rudin (1964) for a systematic account.

Recall the definition of "denseness" in general topology: a set X contained in another set Y is dense in Y if every non-empty open subset of Y meets X. Intuitively speaking, for any element of X, there is at least one element of Y "arbitrarily" close to X. Thus Stone-Weierstrass' theorem asserts that for any continuous function defined on a closed and bounded set, there is at least one function belonging to a countable (= at most "as many as" the totality of the natural numbers) family that approximates it.

The implication of Stone-Weierstrass' theorem in terms of neural networks is that *any continuous phenomenon can be arbitrarily approximated by a neural network consisting of at most a countable number of neurons.*

When using neural networks, things are further reduced to a finite world since it has to perform our tasks on digital computers. Fortunately, experiences show that finite approximation works in most cases. Hence the raison d'etre of data mining.

5. Market Forecast with Data Mining

As is stated in Introduction, the auther would like to analyze cash flow related to PLSs 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 will be proceeded as illustrated below:

Step 1. Segmentize PLSs into clusters based on the chronological sales figures. (See figure 5.1)

Step 2. Create classification models to explain the clustering result by "attributes" of the products and other factors. (See figure 5.2)

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

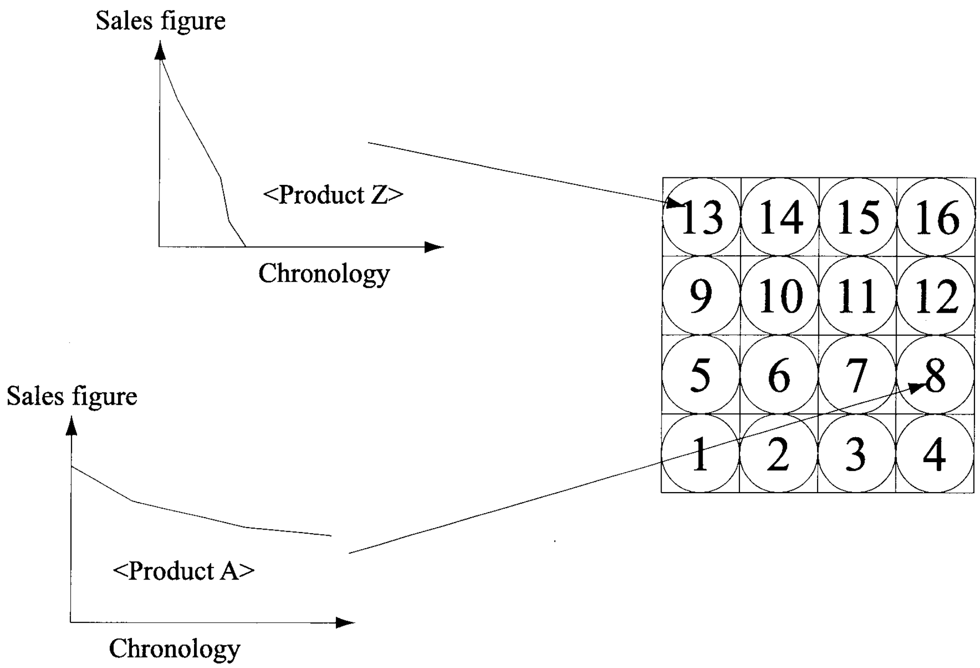


Figure 5.1 Segmentation by Kohonen map network

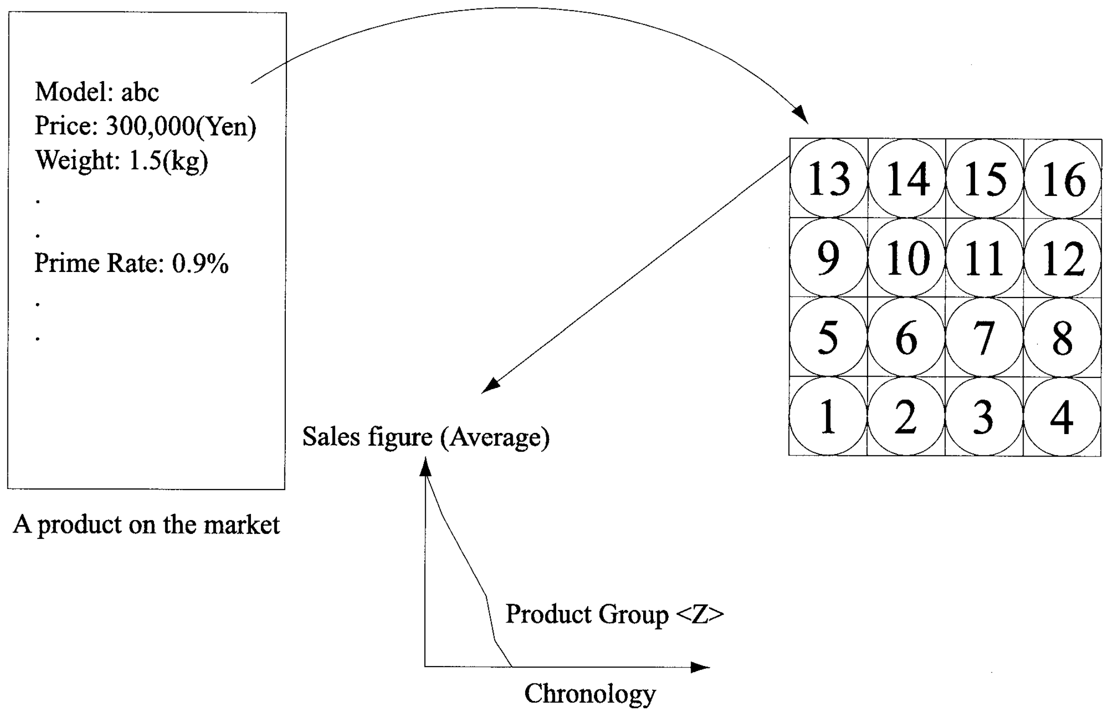


Figure 5.2 Classification of the segmentation result

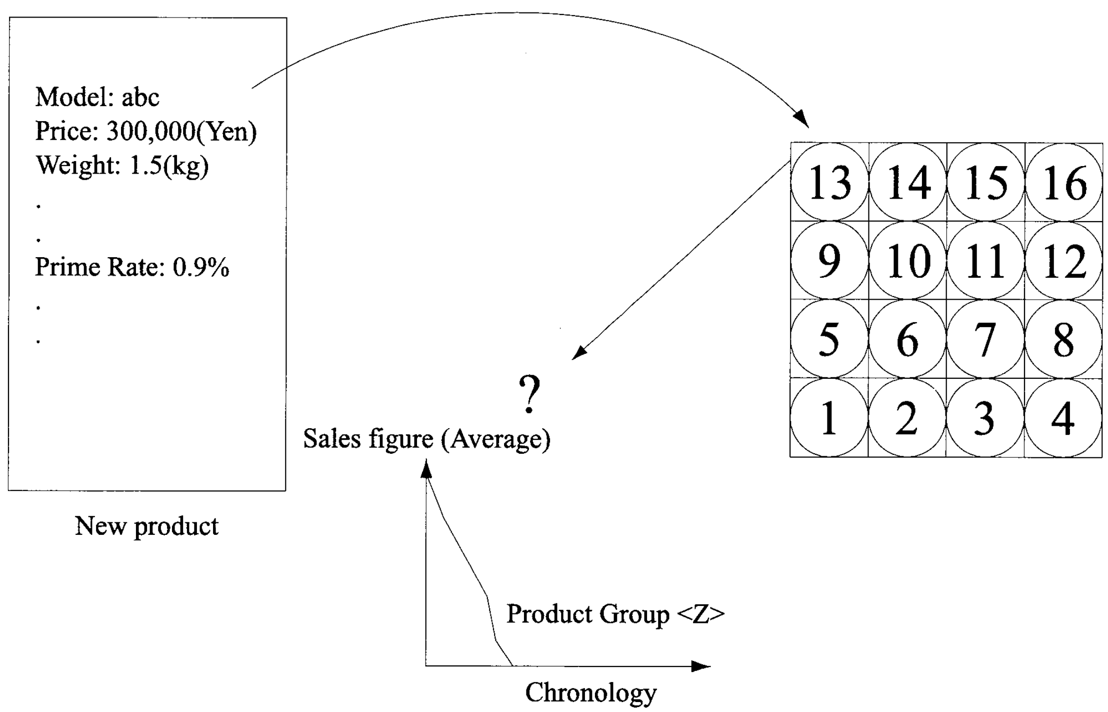


Figure 5.3 Classification application (prediction)

and external factor data. (See figure 5.3)

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

For the step 1., It is used Kohonen feature map networks. As mentioned above, Kohonen feature map networks generate fairly comprehensible results. For example, clusters in figure 5.1 are described by initial values and damping rate.

For the step2., It is adopted back propagation networks. Figure 5.2 shows 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 resource eating. That is, training back propagation networks requires a long time. Robert Heinline has it, "There ain't no such things as a free lunch."

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%).

Perspective

The author is currently studying the followings to create and verify data mining models for measuring corporate growth.

a. Gather product attributes, sales figures and other relevant data such as economic indices.

b. Run the algorithm to create models.

c. Test and verify with the existing data..

d. Apply the models to predict.

Refferences

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

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

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

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

Bourbaki N., "*Elements de Mathematique*," Topologie Generale Chapitre 10.

Deuxieme edition, Hermann, Paris 1961

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

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