A Case-Based Forecasting System

HoonYoung Lee

Research Institute of
Samsung Life Insurance Co., Ltd.

Abstract

Many business forecasting problems are characterized by infrequent occurrences, a large number of variables, presence of error, and great complexity. Because no forecasting models and tools are effective in handing these problems, managers often use the outcomes of past analogous cases to predict the outcome of the current one. They (1) observe significant attributes in describing a case, (2) identify the past cases similar in these attributes to the current case, and (3) predict the outcome of the current case based on those of the analogous cases identified through some mental simulation and adjustment. This process of forecasting can be termed forecasting-by-analogy. In spite of fairly frequent use of this forecasting process in practice, however, it has not been recognized as a primary forecasting tool, nor applied on a regular basis. In this paper, by automatizing this process using computer models, we develop a case-based forecasting system (CBFS), which identifies relevant cases and applies their outcomes to generate a forecast. We demonstrate the effectiveness of the CBFS in terms of its accuracy in predicting the outcome of the current problem based on the similar cases identified. We compare the forecasting accuracy of the CBFS with that of regression models developed by stepwise procedure under varied simulated problem conditions. The CBFS outperforms regression models in most comparisons. The CBFS could be used as an effective forecasting tool.

1. Introduction

In modern business management, managers frequently face prediction or forecasting tasks of extreme complexity, ambiguity, and consequence. To help managers cope with these issues, academicians and practitioners have developed a number of forecasting models and techniques. When the forecasting problems have been inherently well structured and the necessary data readily available, they have produced impressive results (e.g., econometric models and time series analyses). However, it seems problematic when (1) only meager data points are available, (2) too many variables are involved, (3) the relationship among variables is complex, and (4) a large amount of error is involved. Unfortunately, many business forecasting problems (e.g., new product forecasting, business strategy forecasting, and so on) fall into this category.

As a result, managers often use their intuitive judgments to make predictions based on a few past cases available. They identify the most important elements of a problem, retrieve the most relevant past cases in these elements, relate the current problem to them, and make a prediction based on outcomes in the past through some mental simulation and adjustment. This process of forecasting by applying intuitive judgment from analogous past cases to predicting the current problem can be termed "forecasting by analogy."

Analogy has provided a viable means for making predictions and business forecasting. The potential of analogy in prediction and forecasting has been recognized and discussed by many researchers (Burke 1991; Choffray and Lilien 1986; Easingwood 1989; Mahajan and Wind 1988; Mullick et al. 1987; Thomas 1985; Wind, Mahajan, and Cardozo 1981). For example, a new product's performance can be predicted by investigating the performance of analogous products marketed in the past, because products developed and marketed in similar situations are likely to perform similarly in the market (Choffray and Lilien 1986; Mahajan and Wind 1988). Based on the assumption that analogous products would follow similar sales paths and have similar diffusion patterns, for example, Thomas (1985) and Easingwood (1989) showed how the diffusion of a selected,

analogous product could be used as the basis for forecasting the diffusion of the new product. Burke (1991) developed an analogical reasoning system for predicting consumer response to advertising campaigns.

In this paper, based on this forecasting process, we develop a case-based forecasting system (CBFS). A computer simulation study demonstrates the effectiveness of this forecasting system by comparing its forecasting accuracy with that of regression models developed by stepwise procedure. This forecasting system outperforms regression models in most comparisons under varied problem conditions simulated. It is more robust against increasing amount of error, decreasing number of data points, increasing number of variables, and increasing complexity of relationship between independent and dependent variables. Therefore, this forecasting method is effective when (1) variable selection is necessary to apply forecasting tools because there are many independent variables; (2) there are only a few data points in the dataset; (3) an amount of error is expected in the dataset; and (4) there are no appropriate forecasting tools available due to the complex relationship among variables with non-linearity.

In the next section, we discuss the development of a CBFS based on the managers' forecasting by analogy and then investigate the models and methods to implement the process. In Section 3, we validate the potential of the CBFS by demonstrating its predictive accuracy vis-a-vis regression models using simulated data. In Section 4, we conclude by suggesting future extensions of this research.

2. A Case-based Forecasting System (CBFS)

The forecasting process of case-based forecasting systems consists of three subprocesses: (1) identifying key attributes in identifying similar cases to predict the target variable; (2) accessing similarity and retrieving analogous cases; and (3) generating a forecast through combining similar cases selected.

Identifying key attributes is the subprocess of investigating the important attributes or factors which are critical to identifying analogous cases, as well as to predicting the value of the target variable. When predicting the performance of a new product by analogy, for example, this corresponds to the process new product managers use to determine both the attributes which are most effective in locating analogous products and the best predictors to the performance variable.

Similarity judgment and retrieval is the stage of noticing and measuring the similarities between the past cases in the system's database and the current forecasting problem using the elaborated attributes, and retrieving the most similar cases from the database. For example, in predicting the performance of a new product, it corresponds to identifying and retrieving analogous products in the past from the database.

Generating a forecast is the final subprocess of case-based forecasting which proceeds by mapping the outcomes of the selected analogous cases over that of the current problem. Depending on the similarities of retrieved cases to the current problem, their outcomes are adjusted to generate a consolidated forecast.

The development of a CBFS therefore requires models or schemes to conduct these three subprocesses above: to identify key attributes, to estimate the similarity between cases, and to combine some useful analogous cases to generate a forecast on the target variable. In this section, we discuss models and methods to implement these three subprocesses.

2.1 Selection of Key Matching Attributes

To measure similarity between cases accurately, we must first decide which matching attributes to use. Since many variables are used to describe a case, we can omit nonessential attributes which are not useful in predicting the target variable. This allows the system to focus selectively on a few important attributes to explain much of the similarity between cases. As yet, there is no

commonly agreed upon procedure or method for selecting the optimal set of matching attributes for identifying the most relevant cases in a multidimensional space. Recognizing the need for such a procedure, however, we develop a model which could be applied regardless of the data structure (such as multicollinearity or type of dependent variable). The model is derived from Ghiselli's model which represents the correlation between scores on an outside variable and scores on a composite variable composed of *1* to *k* components (see chapter 7 in Ghiselli 1964). The model derived by Ghiselli is represented as follows:

$$\rho_{o\kappa} = \frac{k\,\bar{\rho}_{oj}}{\sqrt{k + k(k-1)\,\bar{\rho}_{jj'}}} \qquad (1)$$

where $\rho_{o\kappa}$ is the correlation between the outside variable (dependent variable) and the composite variable composed of k component variables; ρ_{oj} represents the average of the coefficients of correlation between the outside variable and each component variable j; and $\rho_{jj'}$ is the average of the coefficients of correlation among the component variables (mean intercorrelation). In the equation, even when k goes to infinity, the correlation between the outside variable and the composite variable does not become one, but converges to the ratio between the average correlation between the outside variable and each component variable (ρ_{oj}) over the squared root of the mean intercorrelation of component variables. Moreover, it converges very rapidly with a small increase in k (Ashton 1986; Ghiselli 1964; Hogarth 1978; Libby and Blashfield 1978). It implies that even if we increase the size of the component variable group, the correlation between the outside variable and the composite variable is limited by the mean correlation between the outside variable and each component variable and the mean intercorrelation among component variables, and does not improve after the component variable group reaches 20 or so (Ashton 1986; Hogarth 1978).

Without losing the essence of Ghiselli's model, we have modified it to develop the proposed mathematical programming model which selects the optimal set of variables to maximize the correlation between the target attribute and the set of selected explanatory attributes, corresponding to

outside variable and component variables in Ghiselli's model. The model[1] is represented as follows:

$$Max \quad Z = \frac{\sum_{j=1}^{m} \rho_{ij} x_j}{\sqrt{\sum_{j=1}^{m} \sum_{l=1}^{m} \rho_{jl} x_j x_l}} \qquad (2)$$

$$st. \quad x_j = 1 \text{ or } 0$$

where m is the number of all independent (explanatory) attributes; ρ_{ij} is the correlation between the dependent (target) attribute and explanatory attribute j; ρ_{jl} is the correlation between explanatory attribute j and l. The variable x_j represents attribute j and has a binary value of 0 or 1. In the solution, if variable x_j is 1, then attribute j is selected; otherwise it is left out. We can also select a certain number (k) of optimal matching attributes by simply adding a constraint of $\Sigma_j x_j = k$ to this model. The optimal function value of z represents the correlation between the target attribute and the composite of explanatory attributes selected, and thus the effectiveness of the set of selected attributes in predicting the target attribute.

2.2 Similarity Measure

The most obvious measure of the similarity (or dissimilarity) between two cases is the distance between them. Weighted Euclidian distance has been recognized as one of the most general forms of distance function. The differential weighting of the component dimensions in the multidimensional space is critical to the accurate measure of distance. There are several methods for assigning weights to the matching attributes in this distance function. When no relevant information and tools are available, the equal weighting scheme seems appropriate because it works substantially well and safely enough not to assign weights inappropriately, i.e., assign heavy weights on less important attributes. In equal weighting, each matching feature is equally important in the distance calculation. However,

[1] The complexity of this problem increases exponentially as the number of attributes to select and the total number of attributes increase. We employ heuristics to increase the speed of computation by using the K-opt procedure where the number of maximum batch set of attributes to consider for inclusion and exclusion from the selected set is limited by K (Potvin, Lapalme, and Rousseau 1989). In this problem, we increase K from 1 to m, and stop just before the objective function value changes from increase to decrease. This heuristic is widely used to speed up the exponential searching process. However, it may end up with a suboptimal solution. Another heuristic is to reduce the search set by a preliminary screening of the candidate variables based on their mean correlation with the dependent variable. Using these heuristics, the model can handle a large number of variables.

effectively estimated weights should improve the accuracy of distance and similarity measures.

Using the available cases in the dataset, we can empirically derive the unequal weights using some statistical analyses. One of these methods is to assign weights in proportion to the size of correlations with the target variable (i.e., $w_j = | \rho_j | \div \Sigma_k | \rho_k |$). This weighting scheme, however, does not consider the correlation among the matching attributes. As a slight modification to this method, another weighting scheme can be used, considering both correlation with the target attribute and the correlations among the matching attributes. In this weighting scheme, weights reflect both correlation with the target attribute and the correlations among the matching attributes as follows: $w_j = \left\{ \rho_{dj} + \sum_{l=1}^{m} \rho_{jl} \right\} \div \left\{ \sum_{k=1}^{m} \left(\rho_{dk} + \sum_{l=1}^{m} \rho_{kl} \right) \right\}$, where m is the number of matching attributes, ρ_{dj} is the correlation between the target variable and matching attribute j, and ρ_{jl} is the correlation between matching attributes j and l. Weights are proportioned to their contributions to the objective function in Equation 2. A matching attribute is heavily weighted when it is highly correlated with the dependent target variable but not so correlated with other matching attributes. Even though a matching attribute is highly correlated with the target variable, if it is also highly correlated with other matching attributes, it will not be weighted so highly.

Once the distance between two cases is measured, we derive inter-case similarity from distance. There are various kinds of similarity measures based on the distance between objects. Conceptually, any monotonically decreasing function of distance can be used to relate similarity to distance. Some examples of these monotonic decreasing functions are mathematically and graphically represented in Table 1 and Figure 1 respectively.

Table 1: Mathematical Representation of Functions Relating Similarity to Distance

$$\cdot\ s_{ab} = \begin{cases} \lambda - d_{ab}, & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{-- Linear Transformation 1}$$

$$\cdot\ s_{ab} = \begin{cases} \dfrac{\lambda - d_{ab}}{\lambda}, & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{-- Linear Transformation 2}$$

$$\cdot\ s_{ab} = \begin{cases} \left(\dfrac{\lambda - d_{ab}}{\lambda}\right)^{-1}, & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{-- Nonlinear Transformation 1}$$

$$\cdot\ s_{ab} = \left\{ a^{-\alpha \cdot d_{ab}}, \quad a > 1 \right. \quad \text{-- Nonlinear Transformation 2 (Exponential Decay Function)}$$

$$\cdot\ s_{ab} = \begin{cases} \left(1 - \left(\dfrac{d_{ab}}{\lambda}\right)^3\right)^3, & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{- Nonlinear Transformation 3 (Tricube kernel)}$$

$$\cdot\ s_{ab} = \begin{cases} \dfrac{3}{4}\left(1 - \left(\dfrac{d_{ab}}{\lambda}\right)^2\right), & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{- Nonlinear Transformation 4 (Epanechnikov kernel)}$$

$$\cdot\ s_{ab} = \begin{cases} \dfrac{3}{8}\left(3 - 5\left(\dfrac{d_{ab}}{\lambda}\right)^2\right), & \text{for } d_{ab} \le \lambda \\ 0, & \text{otherwise} \end{cases} \quad \text{- Nonlinear Transformation 5 (Minimum variance kernel}$$

In these examples, the parameter λ is a scale parameter reflecting the tolerable boundary distance from the target case. The cases only within the boundary distance λ are considered to have a certain level of similarity. Accordingly, the cases beyond this limiting distance (λ) are regarded as the ones not similar to the target case. However, we can also include all cases by setting λ to be the maximum distance. The appropriate value of this parameter λ could be determined in terms of variance of distances, the number of cases in the dataset, and so on. These issues are discussed in the next section when we address the problem of determining the set of useful analogous cases.

2.3 Generating a Forecast by Combining Useful Cases

The next step is to accurately predict the target value for a new case by combining the values of the most similar past cases. This process of integration is analogous to that of combining the judgments of a number of experts or forecasts of various models.

There has been considerable research on the issue of combining judgments (Clemen 1989; Winkler 1989). The focus of these studies has been the extraction of appropriate weights for each individual forecast, i.e., the method of combining individual forecasts so as to maximize accuracy. Depending on situational factors, a variety of weighting schemes have been studied. These have

ranged from simple, equal-weighting methods to variable-weighting methods based on arithmetic and geometric measures; from linear to non-linear methods; and from subjective to objective ratings. In combining target values of multiple analogous cases, however, information about similarities between the target and base cases provides guidelines for deciding which cases are more useful in predicting the unknown variable (target variable) for the present situation. The more similar the cases, the more effective they would be in predicting the target value of the current case. Whatever functional form the weighting scheme takes, it should incorporate the information about similarities.

For this purpose, we develop a model based on the distribution of case similarities. In the model, the expected target value (TV_t) of the target case is derived as follows:

$$E\left(TV_t \mid \{S_a\}_{t-1,n}\right) = \sum_{b=1}^{n} P\left(TV_k = TV_b \mid \{S_a\}_{t-1,n}\right) \cdot TV_b = \sum_{b=1}^{n} \left(\frac{S_{tb}}{\sum_{i=1}^{n} S_a} \right) \cdot TV_b \qquad (3)$$

where n is the number of cases selected to generate the overall prediction' S_{tb} is the similarity between the new target case t and the base case b, and TV_b is the predicting (target) value of base b. In the model, the similarity ratio (i.e., similarity of each base case with the new target case over the sum of the similarities of all cases) is used as the case's weight in the combining process. Thus, the combined prediction (predicted target value) on the target value of the current case (TV_t) is represented as a linear combination of the target values of base cases, weighted in proportion to their relative similarities to the current case.

Another important issue is to determine how many cases to combine to generate the system's prediction. There is no clear answer to this question. Intuitively, however, it would seem that if we combine many cases, the combined prediction will have a low variance but a potentially high bias (toward the mean of the target variable). The converse would be true when small numbers of target values are combined. Thus there is a fundamental tradeoff between bias and variance, governed by the number of target values combined. The variance of combined prediction (σ^2) decreases by the

increasing number (n) of target values to combine (e.g., σ^2/n when target values are equally weighted and combined). Increasing n increases the bias because the combined prediction involves more target values of less similar cases, which are more likely to be different from the true value of target case. Similarly, decreasing n increases the variance but tends to decrease the bias.

This tradeoff between bias and variance is similar to the question of how many variables to include in a regression equation. Just as the optimal number of terms in the best linear regression model varies, the optimal number of analogous cases to combine also varies with each target case. This issue is also analogous to the question of how many judgments to combine in contexts where we have multiple expert judges, because each analogous case provides a prediction on the target variable. Hogarth (1978) and Ashton (1986) applied Ghiselli's model and demonstrated its effectiveness in evaluating the validity of combined judgments. According to Hogarth (1978), the validity of combined judgments was represented as a function of the number of experts involved, their mean individual validity, and the mean intercorrelation among their judgments.

We can apply Ghiselli's model to this problem by substituting correlations among judgments with similarities among cases (replacing correlations with similarities in Equation 2). In many instances, correlations and similarities are highly correlated and covary: the higher the correlation, the higher the similarity (see Sjoberg 1980). We can therefore replace the correlations in the model for selecting attributes by similarities, to build a model for determining the number of cases to combine. Thus, the validity (cf. Hogarth 1978) of the combined prediction can be represented as a function of the number of base cases combined, their mean individual similarity with the target case, and the mean similarity among themselves.

Due to the computational complexity, we slightly modify the model so that a similar case always has priority over less similar ones in combination. To exploit the information from the cases in the dataset for this decision, we estimate a parameter p in cross-validation using the cases in the

dataset. The model is represented as follows:

$$
\begin{aligned}
Max \quad SF &= \frac{\sum_{b=1}^{n} s_{tb}\, y_b}{\left(\sum_{b=1}^{n}\sum_{q=1}^{n} s_{bq}\, y_b\, y_q\right)^{p}} \\
st. \quad & (s_{tb} - s_{tq}) * (y_b - y_q) \geq 0 \quad \forall\, b \text{ and } q \\
& y_b = 0 \text{ or } 1 \\
& 0 \leq p \leq 0.5
\end{aligned}
\tag{4}
$$

where n is the number of cases selected to combine, s_{tb} is the similarity between target case t and base case b, and s_{bq} is the similarity between base b and base q. The variable y_b represents case b and has a binary value of *0* or *1*. In the solution, if variable y_b is *1*, then base case b is selected; otherwise it is left out. The constraint $((s_{tb} - s_{tq}) \cdot (y_b - y_q) \geq 0)$ requires that a similar case always have the priority over less similar ones in combination. The system combines the target values of the set of cases maximizing this function value (*SF*).

The decision about how many cases to combine can also be made simply by assigning zero similarities to the cases beyond a certain distance from the target, and combining only the cases within the boundary distance. The parameter λ in the transformation functions in Table 1 represents this boundary distance. The cases within the boundary distance of λ are considered as useful analogous cases with strictly positive similarities, while cases outside this boundary are not useful, and assigned zero similarities. Thus the decision about how many to combine is converted to the accurate determination of this parameter λ. Lambda can be subjectively determined, or estimated through cross-validation analysis using the cases in the system dataset. Cross-validation works by leaving cases out one at a time, measuring its expected value using the remaining cases, and estimating the parameter λ, which minimizes the average squared prediction error as follows:

$$
Minimize\ PSE(\lambda)_{in\ cross\text{-}validation} = \frac{1}{n}\sum_{b=1}^{n}\left\{ TV_b - \sum_{k=1_{k\neq b}}^{n}\left(\frac{S_{bk}(\lambda)}{\sum_{l=1_{l\neq b}}^{n} S_{bk}(\lambda)} \right) \cdot TV_k \right\}^2 \tag{5}
$$

There are a number of other alternatives for selecting the set of most similar cases to use in

prediction. The decision can be made in terms of the distribution of similarities between the target case and base cases in the knowledge base. That is, a certain percentile (i.e., a certain number of standard deviations away from the mean) of most similar cases in the standard normal distribution of similarities can be combined. Another option is to use the cases with similarities greater than a certain threshold level of similarity, or to use a specific number of similar cases (e.g., 5 most similar cases). For these methods, however, the percentage (the number of standard deviations), the level of minimum similarity, or the specific number must be determined. They can also be estimated in terms of cross-validation as discussed above.

3. Validation of the CBFS

A simulation study was conducted to validate the effectiveness of the CBFS. Data were generated to simulate real forecasting problems. We compare the system's forecasting accuracy, reliability, and factors affecting its performance with those of regression models developed by stepwise procedure. We also inspect which provides the more accurate and reliable forecast under which conditions.

3.1 Data Generation and Measurement of Forecasting Errors (MSEs)

Values of independent variables are randomly generated from a uniform distribution between 0 and 1. An error term is also randomly generated from a normal distribution of $N(0,1)$. The dependent variable is created by adding a certain percentage of error to the value generated, using independent variables in three different relationships (i.e., linear and non-linear (square and multiplicative). We varied the total number of independent variables and the correlations of these variables with the dependent values.

There were 60 different types of datasets: 3 (relationship types) x 2 (# of independent variables) x 2 (# of weighted variables) x 5 (percentage of error). One hundred sample datasets were generated for each combination of factors. Thus a total of 6,000 sample sets were generated. Each

sample set consists of 100 data points, which are split in half and by 20% and 80% to create estimation and holdout samples. When they are split by half, each estimation or holdout sample consists of 50 data points. When they are split by 20% and 80%, each estimation or holdout sample consists of 20 and 80 data points respectively. By applying two split methods to each sample set, we made 12000 sample datasets for the simulation study. Table 2 summarizes the factors considered in generating simulation data and split methods to create sample datasets.

Table 2: Summary Table of Factors Considered in Generating Simulation Data

Factors	Relationship	Number of Independent Variables	Number of Independent Variables Weighted	Number of Cases in An Estimation Sample	Percentage (α) of Error Term (ε) Added
Varied Conditions	• Linear ($Y_i=\Sigma^m_{j=1}W_j \cdot X_j+\alpha \cdot \varepsilon$) • Square ($Y_i=\Sigma^m_{j=1}W_j \cdot X^2_j+\alpha \cdot \varepsilon$) • Multiplicative ($Y_i=\Sigma^m_{j=1}W_j \cdot X_j \cdot X_{j+1}+\alpha \cdot \varepsilon$)	15 variables 30 variables	5 variables 15 variables	20 cases 50 cases	10% 20% 30% 40% 50%

Using each estimation sample, we developed regression model by stepwise regression analysis[2] In total, 12000 regression models were developed. We applied these linear regression models (developed using estimation sample) to the holdout sample, and obtained forecasts. We compared these forecasts against their known true values, and measured the squared differences (squared prediction errors). We summed these squared prediction errors across the entire holdout sample in each sample set. Each total prediction error was divided by the number of cases in the holdout sample to compute the average prediction error (mean squared error (MSE)) for each data point of each sample set.

Using each estimation sample, the CBFS also generated forecasts on the target variable of the corresponding holdout sample. These forecasts were compared with their true known values of holdout sample to compute their prediction errors. Next, we computed the mean prediction error (mean squared error (MSE)) for each target variable of each holdout sample in the same way as we

[2]The SAS's default significance levels (i.e., p=0.15 for both entry and staying) are used for stepwise regressions.

did for the regression model. This MSE was compared with that of the regression model.

3.2 Comparison between Regression Models and the CBFS

In most comparisons, the CBFS outperforms regression models developed by stepwise regression analyses. The average and variance of 120 MSEs of each regression model and the CBFS are summarized in Table 3. The average MSE of the regression model is 0.269, which is far greater than the CBFS by at least 0.074. This difference is very significant with F value of 453 in ANOVA.

We also investigated the separate effect of four control factors (relationship, number of independent variables, number of variables weighted, and size of error terms) and the size of estimation sample on each forecasting method. In order to examine the effect of relationship between dependent variable and independent variables, we calculated the average MSEs of each forecasting method for the relationship as shown in Table 3. The CBFS significantly outperforms the regression model in all three relationships. All forecasts are found to be more accurate when the relationship is simple and linear than when it is more complex (with non-linearity). Unlike regression models, however, increasing complexity of relationship (such as going from linear to nonlinear, or from square to multiplicative relationship) does not have a significant impact on the performance of the CBFS. In short, the CBFS appears to be more robust than regression models against increasing complexity of relationship between independent and dependent variables.

The number of independent variables has a significant effect on the performance of all forecasting methods, as summarized in Table 3. This effect is particularly significant to the performance of regression models. The average MSE of regression rapidly increases, while that of the CBFS increases in a relatively small scale. The performance of the CBFS is significantly better than that of regression model as independent variables increase from 15 to 30 (F value in ANOVA increases from 36 to 502). Thus, the CBFS is more effective than regression models developed by stepwise procedure when there are many independent variables.

Table 3: Summary Table of Comparison between Regression and CBFS

Factors Affecting Forecasting Accuracy	Different Forecasting Conditions and Measures of Indicators	Forecasting Methods	
		Regression	CBFS
Relationship	Linear	0.260	0.184
	Square	0.270	0.187
	Multiplicative	0.277	0.188
	Total Increase of MSE	0.017	0.004
	F Values in ANOVA	3.92	1.40
Number of Variables	15 independent variables	0.208	0.179
	30 independent variables	0.330	0.194
	Increase of MSE	0.122	0.015
	F Values in ANOVA	650	49
Number of Weighted Variables	5 weighted variables	0.261	0.177
	15 weighted variables	0.277	0.195
	Increase of MSE	0.016	0.018
	F Values in ANOVA	10	76
Size of Estimation Sample	50 cases in an estimation sample	0.186	0.169
	20 cases in an estimation sample	0.351	0.203
	Increase of MSE	0.165	0.034
	F Values in ANOVA	1247	248
Amount of Error	10% Error	0.058	0.062
	20% Error	0.130	0.101
	30% Error	0.233	0.163
	40% Error	0.371	0.248
	50% Error	0.552	0.356
	Total Increase of MSE	0.494	0.294
	F values in ANOVA	2270	13421
Summary	Average MSE	0.269	0.186
	Variances of MSE	0.268	0.117

Increasing the number of weighted variables among independent variables also has a negative effect on the performance of forecasting, as show in Table 3. This effect, however, appears to be slightly less significant to the regression model (F value of 10 in ANOVA) than to the CBFS (F values of around 70 in ANOVAs).

The size of the estimation sample is critical to both the regression model and the CBFS, as summarized in Table 3. The performance of the regression model declines so rapidly with the decreasing number of the estimation sample that the average MSE of regression increases from 0.186 to 0.351 when the estimation sample size decreases from 50 to 20. The MSEs of the CBFS also show an increase of 0.034 to the decrease of estimation sample, because it would be better to select similar cases from a larger pool of dataset than from a small one. However, the sample size has a much greater impact on the performance of the regression model (F value of 1247) than on that of the CBFS. Therefore, when there are a few data points, it would be better to apply the CBFS rather than regression models.

Finally, increasing amount of error added to the dependent variable also greatly influences the forecasting accuracy, as shown in Table 3. The performance of the regression model rapidly declines with the increasing amount of error term. When less than 10% of error is involved, the regression model performs even better than the CBFS. However, the average performance of the regression model becomes inferior to that of the CBFS when more than 20% of error term is added to the dependent variable. Thus, the performance difference between regression models and the CBFS becomes more significant in forecasting with an increasing amount of error.

In summary, this simulation study shows that CBFSs are effective in forecasting, and they therefore can be directly applied to many business forecasting problems characterized in terms of many independent variables, a few data points, presence of error, and complex relationship among variables.

4. Conclusion and Future Research

Analogous cases have been used for prediction and forecasting in many business areas. In this paper, we develop a case-based forecasting system (CBFS) which identifies similar cases and applies their outcomes to forecasting. The system selects key attributes of a forecasting problem case, identifies similar cases from the database with respect to these attributes, and generates a forecast based on these cases identified.

We demonstrated the effectiveness of the CBFS in terms of its accuracy in predicting the outcome of the current problem based on the similar cases identified. We compared the forecasting accuracy of the CBFS with that of regression models developed by stepwise procedure under varied simulated problem conditions, and found that the CBFS outperforms regression models in most comparisons. The CBFS is quite robust against increasing amount of error, decreasing number of data points, increasing number of variables, and increasing complexity of relationship between independent and dependent variables. Thus, they can be directly applied to many business forecasting problems especially: (1) when there are many independent variables; (2) when there are only a few data points in the dataset; (3) when amount of error is expected in the dataset; (4) and when there are no appropriate forecasting tools available due to the complex relationship among variables.

The diversity of models and methods for the implementation of each subprocess of the system induces subsequent future research. The present CBFS employed specific models for relating similarity to distance and for determining how many cases to combine. A direct extension of this paper is an investigation into appropriate models for an effective CBFS. If they vary depending on the characteristics of forecasting problems, one should further explore which models are the most appropriate in which problem conditions.

There is much useful information indicating effectiveness of an CBFS in identifying similar cases and forecasting based on them. For example, the correlation between the target variable and the

composite of selected attributes indicates how effective the selected variables are in explaining the target variable, and the similarity distribution also informs us whether there are useful analogous cases or not. If the correlation between the target variable and the composite of selected attributes is small, or if there are no cases with a high similarity, it would be better not to completely rely on the system's retrieving and forecasting. Future research should also focus on the development of an CBFS which can extract and incorporate useful information from the knowledge base, can diagnose the effectiveness of the system's retrieving and forecasting processes, and can improve the process by incorporating this information into these processes.

Future research is required for applying the system to real business forecasting problems where no specific forecasting models and tools are effective, such as new product forecasting (cf. Lee, Wind, and Burke 1992a), advertising forecasting, forecasting sales and market shares, etc. It would also be interesting to investigate how the CBFS is effective when it is applied to the forecasting problems where many traditional forecasting models and tools are available, such as advertising-sales effect forecasting, time series forecasting, etc.

Reference

Ashton, A. H. (1986), "Combining the Judgements of Experts: How Many and Which Ones?," Organizational Behavior and Human Performance, 38, 405-414.

Burke, R. R. (1991), " Reasoning with Empirical Marketing Knowledge," International Journal of Research in Marketing , 8 (1), 75-90.

Carroll, J. D. and M. Wish (1974), Multidimensional Perceptual Models and Measurement Methods, Handbook of Perception, Vol. 2, Psychophysical Judgment and Measurement, Academic Press, Inc., 391-447.

Choffray, J. M. and G. L. Lilien (1986), "A Decision-Support System for Evaluating Sales Prospects and Launch Strategies for New Products," Industrial Marketing Management, 15, 75-85.

Clemen, R. T. (1989), "Combining Forecasts: A Review and Annotated Bibliography," International Journal of Forecasting, 5, 559-583.

Duda, O. R., and P. E. Hart (1973), Pattern Classification and Scene Analysis, Wiley-Interscience Publication, John Wiley & Sons, Inc.

Easingwood, C. J. (1989), "An Analogical Approach to The Long Term Forecasting of Major New Product Sales," International Journal of Forecasting, 5, 69-82.

Ghiselli, E. E. (1964), Theory of Psychological Measurement, McGraw-Hill Book Company.

Hogarth, R. M. (1978), "A Note on Aggregating Opinions," Organizational Behavior and Human Performance, 21, 40-46.

Lee, H., J. Wind and R. Burke (1992a), "A New Approach for Screening New Product and Service Concepts: Application to Financial Services," Working Paper, The Wharton School, University of Pennsylvania.

----------, ---------- and ---------- (1992b), "An Analogical Reasoning System for Managerial Learning and Problem Solving," Working Paper, The Wharton School, University of Pennsylvania.

Libby, R and R. K. Blashfield (1978), "Performance of a Composite as A Function of The Number of Judges," Organizational Behavior and Human Performance, 21, 121-129.

Mahajan, V. and J. Wind (1988), "New Product Forecasting Models: Direction for Research and Implementation," International Journal of Forecasting, 341-358.

Medin, D. L., and M. M. Schaffer (1978), "Context Theory of Classification learning," Psychological Review, 85, 207-238.

Mullick, S., G. Anderson, R. Leach, and W. Smith (1987), "Life-Cycle Forecasting," The Handbook of Forecasting, John Wiley & Sons, 321-335.

Nosofsky, R. M. (1988), "Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality," Journal of Experimental Psychology: Learning, Memory, and Cognition, 14 (4), 700-708.

Potvin J., G. Lapalme, and J. Rousseau (1989), "A Generalized K-Opt Exchange Procedure for The MTSP," INFOR, 27 (4), 475-481.

Sjoberg, L. (1980), "Similarity and Correlation," Similarity and Choice, Hans Huber publishers Bern, eds., Lantermann and Feger, 70-87.

Thomas, J. R. (1985), "Estimating Market Growth for New Products: An Analogical Diffusion Model Approach," Journal of Product Innovation Management, 2, 45-55.

Tou, T. J., and R. C. Gonzalez (1974), Pattern Recognition Principles, Addison-Wesley Publishing Company, Inc.

Wind, Y., A. Mahajan, and R. Cardozo (1981), New Product Forecasting: Models and Applications, Lexington, Mass.

Winkler, R. L. (1989), "Combining Forecasts: A Philosophical Basis and Some Current Issues," International Journal of Forecasting, 5, 605-609.