The Design of Controls in Supply Chain Management Sustainable Collaboration Using Decision Tree Algorithm

Se Hun Lim†

Sangji University, 660 Wosan-ong, Wonju-City, Kangwon-Province, South Korea

Summary
Prediction of Supply Chain Management (SCM) sustainable collaboration has attracted lots of interests in the SCM researchers, SCM managers and Chief Executive Officer (CEO), because it is very important domain affecting business performance. However, there isn't the guideline and control model for SCM sustainable collaboration. Thus SCM managers and CEO have difficulty in decision making for SCM sustainable collaboration pursuing and operation. In this study, we developed control model of SCM sustainable collaboration using Decision Tree Algorithms (DTA). We used logistic regression analysis (LRA) and multivariate determinate analysis (MDA) as a benchmark. And we has compared the performance of forecasting SCM sustainable collaboration through three types of models; LRA, MDA, DTA. Using DTA to forecast SCM sustainable collaboration is the most outstanding. This result provided useful information of SCM sustainable collaboration determining factors in the manufacturing and distributing companies.

Key words:
Decision Tree Algorithm (DTA), Logistic Regression Analysis (LRA), Multivariate Determinate Analysis (MDA), Supply Chain Management (SCM), Sustainable Collaboration, Business Forecasting

1. Introduction

Decision tree approach is a promising method for strategic planning and forecasting in business management and industrial applications. For example, C5.0 algorithm is widely used for forecasting bankruptcy, forecasting of customer churning, marketing strategy, customer service management, financial time analysis, industrial forecasting etc. In this research, C5.0 algorithm is applied to solve problems in forecasting Supply Chain Management (SCM) sustainable collaboration.

SCM sustainable collaboration is achieved until satisfactory SCM performance. Therefore, the sustainable collaboration of SCM implies its success. However, there are no guidelines to determine the need for SCM sustainable collaboration, nor are there any models for forecasting SCM.

Therefore, this research has developed a model for forecasting SCM sustainable collaboration using decision tree approach by adjusting the balance of measurements in the framework which was formed by Peter, Thomas [10], based on the Balanced Scorecard (BSC) presented by Kaplan and Norton [6]. Our results can provide pursuing SCM firms practical connotations.

This study is structured as follows. Section 2 introduces basic concepts of C5.0 and previous research applications in business management. In section 3, we described BSC variables to forecast SCM sustainable collaboration. Then explanations on the research and experiment structure are followed. Section 4 analyzes the empirical results. Finally, this article concludes and mentions limitations of the study.

2. An Application of Decision Tree Algorithm

Decision tree algorithm is an inductive learning method. It structures decision making rules using tree figure to solve problems of classification and prediction. We have two types of algorithm that support the decision tree algorithm. One is artificial intelligence technology based algorithms such as ID3 (Interative Dichotomizer 3), C4.5, and C5.0. The other one is based on statistics, i.e., CART (Classification and Regression Tree) and CHAID (Chi-square Automatic Interaction Detection) [1], [11].

We have several data mining studies employing C5.0 or Decision Tree in the area of finance for example, customer churning prediction of a credit card firm, bankruptcy prediction, and stock price prediction.

Braun and Chandler [2] suggest a decision making support model using ID3 to predict stock price fluctuation for the stock market analysts. Cronen et al. [4] adopt a recursive partitioning method that is an expansion of ID3 algorithm to predict firm bankruptcy and problems of classification of mortgage.

In the study, they prove superior predictability of the model. Lee, Jung and Shin [7] use C5.0 model to analyze customer churning classification in credit card market. They prove that C5.0 outperforms predictability of other

Manuscript received March 5, 2006.
Manuscript revised March 30 , 2006.

methods such as LRM and ANN. Johnson et al. [5] suggest decision tree - based symbolic rule induction system for categorizing text document automatically and verify its usefulness by far in solving practical problems. Monkol et al. [9] also employ decision tree - based algorithm (C4.5 and ID3) to predict fraud in on - line business transaction under e - business environment and prove its prediction accuracy.

On the other hand, Lee and Lee [8] improves prediction performance by combining ANN method and decision tree algorithm (C4.5) in the customer churning prediction of mobile telecommunication service users.

This article adopts a decision tree algorithm (C5.0) to develope a control model in SCM sustainable collaboration.

3. Research Method

3.1 Application Data

If you would like to itemize some parts of your manuscript, please make use of the specified style "itemize" from the drop-down menu of style categories In this study, we uses measurement tool which was developed by Peter, Thomas [10].

Table 1: BSC Description Statistics

	Feature Name	Mean	SD
[1]	Product and process innovation	4.361111	1.179722
	partnership management	4.333333	1.059007
	Information flows	4.231481	1.115787
	Threats and substitutes	4.425926	1.161724
[2]	Waste reduction	4.277778	1.092312
	Time compression	4.314815	1.116136
	Flexible response	4.305556	1.147597
	Unit cost reduction	4.342593	1.153313
[3]	View of product/service	4.314815	1.132759
	View of timeliness	4.259259	1.113808
	View of flexibility	4.351852	1.087867
	View of customer value	4.166667	1.045685
[4]	Profit margins	4.416667	1.086235
	Cash flow	4.268519	1.046637
	Revenue growth	4.240741	1.012726
	Return on assets	4.203704	1.039044

Each measurement items ware given values over a 7-point Likert-type scale. For developing control model in SCM sustainable collaboration purposes, the dependent variable, that is, sustainable collaboration, was measured with two types items, such as, '0' in non-sustainable collaboration and '1' in sustainable collaboration SCM.

This survey was administered from June through September, 2003, to SCM experts of distributing and manufacturing companies that are carrying out SCM. Out of 300 questionnairs distributed, 120 were collected and 108 of them were used in the analysis after discarding questionnairs with incomplete answers. The statistical information of the variables is shown in <Table 1> ([1] Learning Perspective, [2] Internal Process Perspective, [3] Customer Perspective, [4] Financial Perspective).

3.2 Research Method

This research compared the performance of forecasting SCM sustainable collaboration through three forecasting model of DTA, LRA and MDA model. The ratio for training data set and holdout data set is 80:20 for the test. These results consisted of 86 test data sets and 22 holdout data sets. Especially, in the C5.0 analysis, we used SPSS Clementine 8.1. The remaining default values were used as pruning level 75%, per minimum record of decision tree 2.

3.3 MDA

In this study, MDA model are used as a benchmark. In building MDA model for the purpose of predicting for SCM sustainable collaboration, we need to forecast a precise cut-off value showing a clear distinction between samples. MDA is a useful technique for forecasting of SCM sustainable collaboration. MDA function is represented as follows.

$$Z = W_1X_1 + \beta_2X_2 + \beta_3X_3 + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_iX_i$$

Where Z-scores refers a discriminant score, W represents a cut - off values. Z and Xi indicate dependent and independent variables respectively. Statistical analyses were done using SPSS 11.0.

3.4 LRA

We also employed LRA to compare the predictability of sustainable collaboration with that of MDA and C5.0 in SCM. LRA model or linear probability models are combination of multiple regressions and multiple discriminant analysis. The primary difference between LRA and multiple regressions are the use of a dichotomous dependent variable. Most critical is that the error term of a discrete variable follows the binomial

distribution instead of the normal distribution, thus invalidating all statistical testing performed in regression.

$$Y_j = \frac{1}{1+e^{-P}}$$

$$P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots\ldots + \beta_i X_i$$

When the dependent variable has value of 0 or 1 (dummy variable), a response function (Y estimation) shows S curve. This response function converges to 1 when the value of x increases. This function is sometimes referred as logistic function. If we define a vector of the observations of the independent variables as, and estimate their coefficients, we can calculate SCM sustainable collaboration the following logistic function. We analyze the data using SPSS 11.0 software and set 0.5 as a cutoff point.

4. Experimental Result and Discussion

4.1 Application Data

According to this empirical result, C5.0 gave the best prediction for SCM sustainable collaboration. The order of improving forecasting performance was as follows (C5.0 > LRA = MDA). <Table 2> describes the prediction accuracy of each mode for training data set and holdout data set. As <Table 2> shows, C5.0 achieved higher prediction accuracy than LRA and MDA by 1.02%, 1.02% for the holdout data respectively. In sum, C5.0 outperforms by far to compare with LRA and MDA.

Table 2: Margin specifications

Model	***LRA***	MDA	C5.0
Training Data	59.03	66.30	60.19
Holdout Data	81.80	81.80	82.82

4.2 Application Data

As we discussed C5.0 shows superior predictability in the SCM sustainable collaboration. In an application of decision tree rules, analyzing switched sustainable collaboration and non-sustainable collaboration. Our analysis using C5.0 results in about lots of SCM sustainable collaboration rules. <Table 3> shows exemplary rules (rule 1, 2, 3, 4) that decision making of sustainable collaboration. More detailed explanation on the rule 1, 2, 3, 4 is as follows.

Table 3: Decision Making Rule for Control of SCM

Sustainable Collaboration	Non-sustainable Collaboration
[Rule 1] 1 if V16 > 5 then 1	[Rule 3] 0 ifV16 <= 5 then 0
[Rule 2] 1 if V1 > 4 and V4 <= 5 and V8 > 4 and V11 > 3 and V15 <= 5 then 1	[Rule 4] 0 if V1 <= 4 and V8 <= 3 and V16 <= 4 then 0

In rule 1, the firms that have the following facts are likely to ongoing sustainable collaboration their partner companies next year (period): who choose (1) V16 > 5. In rule 1, the firms that have the following facts are likely to ongoing sustainable collaboration their partner companies next year (period): who choose (1) V1 > 4, (2) V3 <= 5, (3) V8 > 4, (4) V11 > 3, (5) V15 <= 5. In rule 3, the firms that have the following facts are likely to stop sustainable collaboration their partner companies next year (period): who choose (1) V15 <= 5. In rule 4, the firms that have the following facts are likely to stop sustainable collaboration their partner companies next year (period): who choose (1) V1 <= 4, (2) V8 <= 3, (3) V16 <= 4.

Based on the results of the SCM sustainable collaboration, the companies can set its decision making strategy effectively to minimize loss and can improve its operation performance.

This result showed importance characteristics of SCM sustainable collaboration to SCM enterprise manager and CEO. However, systematic work for expert experience and field research are needed to for successful SCM implementation.

5. Conclusion

In this research, we applied C5.0 to SCM sustainable collaboration forecasting. We used empirical data set for distributor and manufacturer pursuing SCM. The result showed the C5.0 achieved prediction accuracy comparable to that of MDA and LRA. This research result is very significant in confirming a more accurate decision-making

model for SCM. SCM managers can determine which control fits for various organization context best.

However, the research featured the following limitations.

First, Due to an insufficient number of samples, C5.0 tests were various experiments. Therefore, for future research it is necessary to collect more samples of corporations using SCM and to test.

Second, the variables for measuring SCM performance were used with the partially altered variables supplied by Peter, Thomas [10]. The feature of each field of corporation where SCM is used should be reflected in future research.

References

[1] Berry, M.J.A., and Linoff, G., *Data Mining Techniques : For Marketing, Sales, and Customer Support*, Wiley Computer Publishing (1997)

[2] Braun, Han I.,, and Chandler, J., "Predicting Stock Market Behavior Through Rule Induction : An Application of the Learning - From - Example Approach", *Decision Science*, 18 (1987) 415-429

[3] Breiman, L.J. Friedman, J., Olshen, R. A., and Stine, C. J., Classification and Regression trees, Beltmont, Wadsworth, (1984).

[4] Cronan, T., Glorfeld, L., and Perry, L., "Production System Development for Expert Systems using a Recursive partitioning Induction Approach : An Application to Mortgage, Commercial and Consumer Lending", *Decision Science*, 22 (1991) 812-845.

[5] Johnon, D.E., Oles, F.J., Zhang, T. and Goetz, T., "A Decision Tree Based Symbolic Rule Induction System for Text Categorization", *IBM System Journal*, 41.3 (2002) 428-437.

[6] Kaplan, R.S., and Norton, D.P., "The Balanced Scorecard Measures that Drive Performance", *Harvard Business Review*, 70.1 (1992) 71-79.

[7] Lee, G.C., Jung N.H., and Shin K.S., "Customer Churning Analysis by Using Data Mining in Credit Card Market", *Korea Journal of Intelligent Information System*, 8.2 (2003) 15-35 .

[8] Lee K.K. and Lee H.C., "A Study on the Combined Decision Tree (C4.5) and Neural Network Algorithms for Classification of Mobile Telecommunication Customer", *Korea Journal of Intelligence Information System*, 9.1 (2003) 139-155.

[9] Monkol Lek, Benjamin Anandarajah, Narciso Cerpa and Rodger Jamieson, "Data Mining Prototype for Detecting E-Commerce Fraud", *The 9th European Conference on Information Systems*, 9 (2002) 160-165.

[10] Peter, B.C. and Thomas, W.S., "Using the Balanced Scorecard to Measure Supply Chain Performance", *Journal of Business Logistics*, 21.1 (2000) 75-93.

[11] Quinlan, J. R. and Quinlan, J., *C4.5 Programming for Machine Learning*, Morgan Kaufman Publishers, (1997).

Se Hun Lim received his M.S. and Ph.D. degrees in MIS (Management Information Systems) from the Graduate School of Management at the Chung-Ang University in 2000 and 2003 and also his B.A. degree from the Chung-Ang University in 1998. He has been the Visiting Scholar of the Department of Information System and Decision Science at the University of North Texas. And he has been a member at the faculty (adjunct professor), of Department of Information Systems, Chung-Ang University. He is currently the Professor of the Department of Management Information System at the Sang-Ji University in Won-Ju City, Kang-Won Province, South Korea. His research interests include an application of Machine Learning Algorithms in Business Area, Data Mining and Computational Intelligence, e-business and SCM (supply chain management), RFID (Radio Frequency IDentification) etc. Professor Lim has published in many international and domestic journals.