Research Journal of Applied Sciences, Engineering and Technology 9(2): 91-97, 2015
DOI:10.19026/rjaset.9.1382
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: August 13, 2014 Accepted: September 14, 2014 Published: January 15, 2015

Research Article

Agent-based Modeling in Supply Chain Management using Improved C4-5

[1]C.P. Balasubramaniam and [2]V. Thigarasu
[1]Karpagam University, Coimbatore, Tamil Nadu 641021,
[2]Depertament of Computer Science, Gobi Arts and Science College, Gobichettipalayam, India

Abstract: The Supply chain can describe the activities that are involved in the chain, or the companies, or the different functions. In literature there are a lot of models describing the Supply chain from different perspectives. Currently supply chains performance measurement systems suffer from being too inward looking, ignoring external environmental factors that might affect the overall supply chain performance when setting new targets. The most efficient Supply chain is the one that has the lowest possible cost and at the same time meet the customer's expectations on service like delivery precision and lead-time. In this study decision based technique C4.5 is improved using correlation coefficient of Kendall for effective classification. The correlation coefficient of Kendall is adapted to improve the system. The C4.5 not only produce discrete attributes, but also continuous ones can be handled, handling incomplete training data with missing values and it is prune during the construction of trees to avoid over-fitting. The accuracy is calculated by sensitivity and specificity for the proposed and existing technique for the textile synthesis dataset. Obtained results will prove the efficiency of this proposed technique based on its accuracy.

Keywords: Decision trees, fuzzy classification, Kendall's rank correlation coefficients, supply chain

INTRODUCTION

In Recent Years, Supply Chain Management (SCM) has received an increased amount of interest both from researchers and in the industry. The SCM concept came up before the 1960s (Huan *et al*., 2004). More and more companies have to focus on their Supply chain in order to be successful in their business. Already in 1997 top managers had recognized the importance of having effective Supply chains to create competitive advantage according to Higginson and Alam (1997) and Cooper *et al.* (1997). A wide range of metrics for supply chain performance have been proposed using an equally diverse portfolio of methodologies (Estampe *et al*., 2013).

The margins for many companies are becoming smaller and smaller due to increasing demand from the customers on lower prices. Solvang (2001) writes that one of the biggest challenges for manufacturing Supply chains is to continuously improve their performance so that their competitiveness can be sustained in long term. To be able to survive on the market the companies have to cut cost in all areas and focus on SCM. Tummala *et al*. (2006) state to make changes to the Supply chain helps to lower cost and enables a company to more easily compete based on the price. Many concepts for Supply chain design and Supply chain modelling have been presented during the last couple of years with different focus according to Svensson (2003).

Usually a company's Supply chain is defined as "the network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hand of the ultimate customer" (Christopher, 1998).

Supply Chain Management and Supply Chain Cost are important parts of company. Company must have the following functions (Sotiris, 2000):

- Research and development
- Marketing and sales
- Supply
- Service

General administration and business controlling: Out of these five functions of a company, Supply is part of the Supply chain, but also parts of the other functions may be included in the scope of the Supply chain. The Supply chain structure differs between different types of products and services. Multi-agent systems may be particularly useful for modeling supply chain dynamics (Jain *et al*., 2009). The entities involved within a supply chain can be represented by agents able to perform actions and make autonomous decisions in order to meet their goals (Sycara, 1998; Wooldridge, 1999). Supply chain organization is, therefore, a distributed process where multiple agents apply their own retrieval

Corresponding Author: C.P. Balasubramaniam, Karpagam University, Coimbatore, Tamil Nadu 641021, India

and filtering capabilities. Multi Agent Systems (MASs) provide an appropriate infrastructure for supporting collaboration among geographically distributed supply chain decision-makers (Jain *et al*., 2009). As an example, in MASCOT (Multi-Agent Supply Chain coordination tool) system (Norman *et al*., 2003) a set of agents help users distributed across multiple companies to collaborate on the development and revision of supply chain solutions through an open and uniform communication and coordination interface.

Many works highlight the importance of a dynamic configuration of supply chains for market changes adaptation. In Jain *et al*. (2009) the author after a discussion about information related issues in a dynamic supply chain propose the definition of models based on the integration of agent technology and Petri networks to improve information flows and highlight potential risks for supply chain actors. In Selwyn (2005), a machine learning algorithm based on decision tree building allows for the choice of the best node at each stage of the supply network analyzing the combination of parameters such as price, lead-time, quantity, etc.

LITERATURE SURVEY

The premise of Supply Chain Management (SCM) is that the performance of a single company depends more and more on its ability to maintain effective and efficient relationships with its suppliers and customers (Chen and Paulraj, 2004; Croom *et al*., 2000). Therefore, managerial tasks are moving from an organizational scale to a supply chain scale (Lambert and Cooper, 2000) and thus encompass the inter-organizational integration and coordination of dispersed supply chain activities. Empirical research suggests that knowledge sharing and reuse between supply chain participants are important determinants of supply chain performance at both the strategic and operational level (Hult *et al*., 2006; Paulraj *et al*., 2008). The role of information systems to support this task is subject of much research (Gosain *et al*., 2004; Gunasekaran and Ngai, 2004; Rai *et al*., 2006).

Knowledge sharing and reuse between supply chain participants face many organizational obstacles such as confidentiality, trust and norms. However, fundamental prerequisites for knowledge sharing are means for exchanging, processing and interpreting the relevant domain knowledge by using one or more representations of this knowledge. Since such representations may be diverse and serve different objectives, formal ontology has been proposed to represent domain knowledge, enhance communication between participants and support interoperability of systems (Kishore *et al*., 2004). A formal ontology formally captures knowledge through concepts, relationships and axioms and can be regarded as the conceptual model of a knowledge base (Guarino, 1998). The application of ontology in SCM has led to a large number of ontologies for various SCM tasks, e.g., planning (Chandra and Tumanyan, 2007) as well as more generally representing arbitrary supply chains (Zdravkovic *et al*., 2011).

Fuzzy logic is a technique suitable for dealing with uncertainty and subjectivity, which becomes an interesting auxiliary approach to manage performance of supply chains. A descriptive quantitative approach was adopted as research method, based on the prediction model. Statistical analysis of the prediction model results confirmed the relevance of the causal relationships embedded in the model. The findings reinforce the proposition that the adoption of a prediction model based on fuzzy logic and on metrics of the SCOR model seems to be a feasible technique to help managers in the decision making process of managing performance of supply chains (Ganga and Carpinetti, 2011).

The modeling approach used in Anderson *et al*. (1989) stated that in measuring logistics performance, a comprehensive strategy of measurement is necessary for the successful planning, implementation and control of the different activities comprising the business logistics function. Stainer (1997) advocated that a set of performance measures is needed in order to determine the efficiency and/or the effectiveness of an existing system, or to compare competing alternative systems.

Murthy (1998) provided an overview of work in decision trees and a sample of their usefulness to newcomers as well as practitioners in the field of machine learning. Decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values.

PROPOSED METHODOLOGY

The proposed methodology of SCM consists of following algorithms.

Decision trees: Decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified and each branch represents a value that the node can assume (Sotiris and Panayotis, 2009). Instances are classified starting at the root node and sorted based on their feature values. Figure 1 is an example of a decision tree for the training set of Table 1.

Using the decision tree depicted in Fig. 1 as an example, the instance $\langle at1 = a1, at2 = b2, at3 = a3, at4 = b4\rangle$ would sort to the nodes: at1, at2 and finally at3, which would classify the instance as being

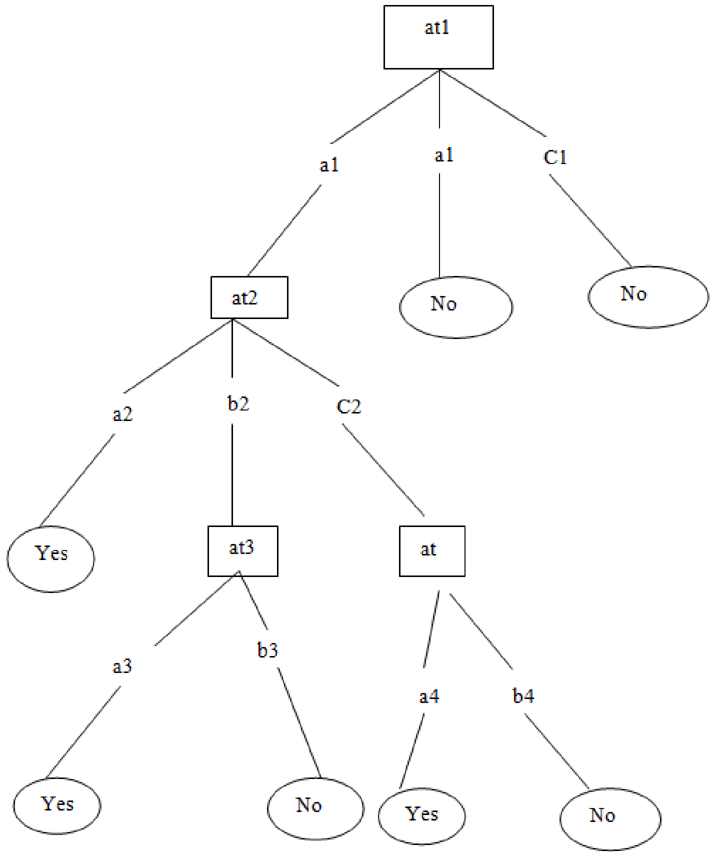

Fig. 1: A decision tree

Table 1: Training set

at1	at2	at3	at4	Class
a1	a2	a3	a4	Yes
a1	a2	a3	b4	Yes
a1	b2	a3	a4	Yes
a1	b2	b3	b4	No
a1	c2	a3	a4	Yes
a1	c2	a3	b4	No
b1	b2	b3	b4	No
c1	b2	b3	b4	No

positive (represented by the values "Yes"). The problem of constructing optimal binary decision trees is an NP complete problem and thus theoreticians have searched for efficient heuristics for constructing near-optimal decision trees.

C4.5 Algorithm: The most well-know algorithm in the literature for building decision trees is the C4.5 (Quinlan, 1993). C4.5 is an extension of Quinlan's earlier ID3 algorithm (Quinlan, 1979). C4.5 has a very good combination of error rate and speed.

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set $S = S_1, S_2, \ldots$of already classified samples. Each sample S_i consists of a p-dimensional vector, $(x_{1,i}, x_{2,i}, \ldots x_{p,i})$ where x_j represent attributes or features of the sample, as well as the class in which S_i falls.

At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sub lists.

This algorithm has a few base cases:

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

Counting Gain: here entropy is implemented it is defined as to measure or calculate the disorder of the data. It is defined as:

$$\text{Entropy}(\bar{y}) = -\sum_{j=1}^{n} \frac{|y_j|}{|\bar{y}|} \log \frac{|y_j|}{|\bar{y}|} \quad (1)$$

Iterating over all possible values of $|\bar{y}|$. The conditional Entropy is:

$$\text{Entropy}(j|\bar{y}) = \frac{|y_j|}{|y|} \log \frac{|y_j|}{|\bar{y}|} \quad (2)$$

Al last defined the gain:

$$\text{Gain}(\bar{y}, j) = \text{Entropy}(\bar{y} - \text{Entropy}(j|\bar{y})) \quad (3)$$

The goal is to maximize the gain, dividing by over all entropy due to split argument $\bar{y}$ by value j.

Kendall's rank correlation coefficients: Kendall correlation coefficient (Kendall, 1955) is also uses nonparametric method for correlation measure. It is also regarded as Spearman rank correlation coefficient. Spearman correlation is calculated from variables' rank rather Kendall correlation is associated with probability calculation. Kendal Correlation coefficient is denoted with the Greek letter τ (tau). Kendall-tau uses concordant or discordant values. The range of value of Kendall correlation coefficient is -1 to + 1. Let X and Y are the pair of measured and estimated inhibitory activity. Kendall tau coefficient is defined as in Eq. (4):

$$\tau = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \text{sgn}(x_i - x_j)\text{sgn}(y_i - y_j)}{n(n-1)} \quad (4)$$

where,

$$\text{sgn}(x_i - x_j) = \begin{cases} 1 \text{ if } (x_i - x_j) > 0 \\ 0 \text{ if } (x_i - x_j) = 0 \\ -1 \text{ if } (x_i - x_j) < 0 \end{cases}$$

$$\text{sgn}(y_i - y_j) = \begin{cases} 1 \text{ if } (y_i - y_j) > 0 \\ 0 \text{ if } (y_i - y_j) = 0 \\ -1 \text{ if } (y_i - y_j) < 0 \end{cases}$$

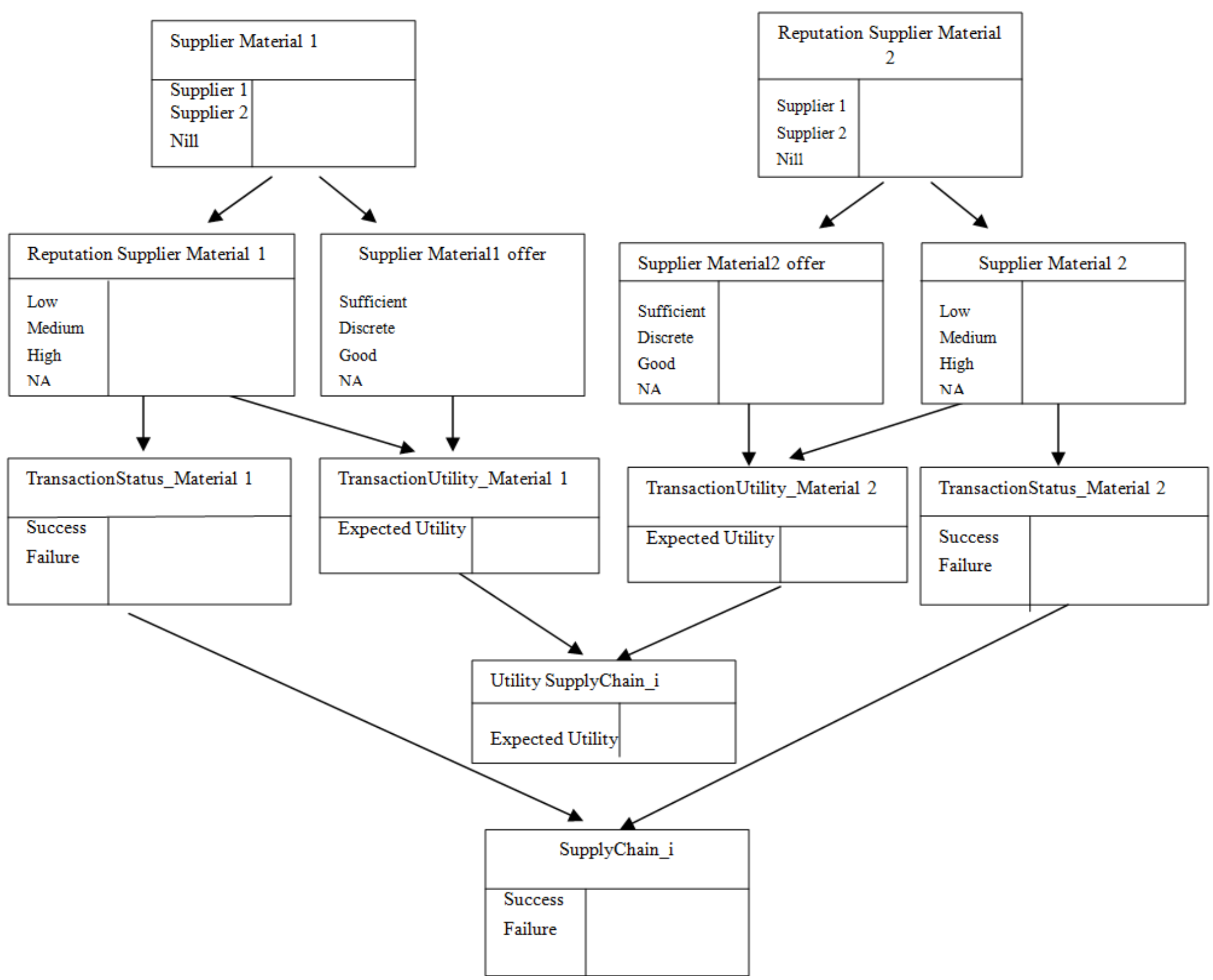

Fig. 2: Agent decision network for selecting the suppliers that maximizes the utility associated with the supply chain

This coefficient quantifies the discrepancy between the number of concordant and discordant pairs. Any two pairs of ranks $(x_{i,} x_j)$ and $(x_{i,} x_j)$ are said to be concordant when xi < xj and yi < yj, or when $x_{i,} > x_{j,}$ and $y_{i,} > y_{j,}$ or when $(x_{i-} x_j)(y_{i-} y_j) > 0$. Correspondingly, any two pairs of ranks $(x_{i,} y_i)$ and $(x_{j,} y_j)$ are said to be discordant when xi < xj and $y_{i,} > y_{j,}$, or when xi > xj and yi < yj, or when $(x_{i-} x_j)(y_{i-} y_j) < 0$. Similar to the two previous correlation coefficients, Kendall's tau ranges from -1 to +1, with the absolute value of τ indicating the strength of the monotonic relationship between the two variables (Fig. 2).

However, Kendall's tau can be 1 for even a wider range of scenarios than Spearman's correlation coefficient.

EXPERIMENTAL RESULTS

To evaluate the performance of the proposed work for SCM, synthetic dataset is used. That dataset contains labels, Supplier id, quantity, cost, item id, material, size, quantity, Product Cost and dealer cost values. These are given for the proposed improved C4.5 algorithm. The obtained results and its performance are measured and compared with existing technique for the input textile dataset.

The classification of this algorithm is viewed in tree structure where the decision tree is classified as (Fig. 3).

```
if x5<24
        then node 2
else if x5> = 24
        then node 3
else 4
if x1<3.5
        then node 4
else if x1> = 3.5
        then node 5
else 4
if x1<3.5
        then node 6
else if x1> = 3.5
        then node 7
else 6
        class = 4
        class = 5
if x2<2750
        then node 8
else if x2> = 2750
        then node 9
```

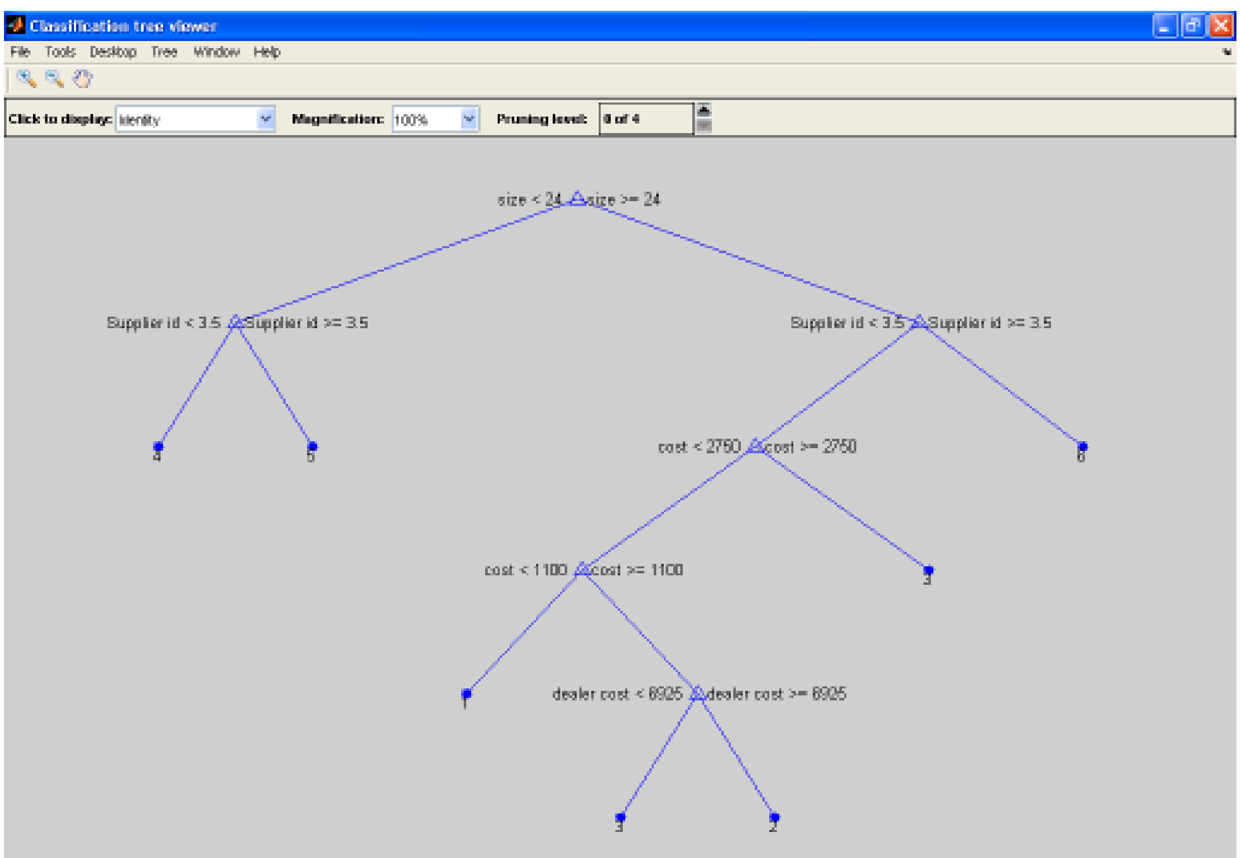

Fig. 3: Proposed technique viewed in tree representation

Table 2: Comparison table for accuracy

Performance Measures	Improved C4.5 for SCM (%)	Fuzzy Classification For SCM (%)
Accuracy	93.9	90.1

```
else 1
        class = 6
if x2<1100
        then node 10
else if x2> = 1100
        then node 11
else 1
        class = 3
        class = 1
if x9<6925 then
        node 12
else if x9> = 6925 then
        node 13
else 2
        class = 3
        class = 2
end
```

Table 2 gives the accuracy comparison for the proposed technique C4.5 and existing technique fuzzy based classification for SCM. The accuracy here is measured by sensitivity and specificity.

From the Table 2, the accuracy obtained by the proposed technique Improved C4.5 for SCM is 93.5% which is better when compared with existing technique of fuzzy based SOM whose accuracy is 90.1%.

The comparison graph for accuracy calculated by sensitivity and specificity is shown in graph. From the figure it proves that accuracy of proposed improved c4.5 with Kendall correlation coefficient gives maximum while comparing with existing fuzzy based technique for SCM.

CONCLUSION

The SCM environment is very important complex, highly dynamic and with many constraints. An improved C4.5 technique is proposed here for SCM using Kendall correlation coefficients. Here the performance of this technique is measured for textile dataset. The processing time and accuracy is quite good than the traditional C4.5 technique. And from the experimental analysis the accuracy from sensitivity and specificity produces good result than existing fuzzy classification. Further this can be extended for large dataset and for few more agents in SCM.

REFFERENCES

Anderson, P., H. Aronson and N.G. Storhagen, 1989. Measuring logistics performance. Eng. Cost. Prod. Econ., 17: 253-262.

Chandra, C. and A. Tumanyan, 2007. Organization and problem ontology for supply chain information support system. Data Knowl. Eng., 61(2): 263-280.

Chen, I.J. and A. Paulraj, 2004. Towards a theory of supply chain management: The constructs and measurements. J. Oper. Manag., 22(2): 119-150.

Christopher, M., 1998. Logistics and Supply Chain Management: Strategies for Reducing Costs and Improving Services. 2nd Edn., Pitman, London.

Cooper, M.C., D.M. Lambert and J.D. Pagh, 1997. Supply chain management: More than a new name for logistics. Int. J. Logist. Manage., 8(1): 1-14.

Croom, S., P. Romano and M. Giannakis, 2000. Supply chain management: An analytical framework for critical literature review. Eur. J. Purchas. Supply Manage., 6(1): 67-83.

Estampe, D., S. Lamouri, J.L. Paris and S. Brahim-Djelloul, 2013. A framework for analysing supply chain performance evaluation models. Int. J. Prod. Econ., 142(2): 247-258.

Ganga, G.M.D. and L.C.R. Carpinetti, 2011. A fuzzy logic approach to supply chain performance management. Int. J. Prod. Econ., 134(1): 177-187.

Gosain, S., A. Malhotra and O.A. El Sawy, 2004. Coordinating for flexibility in e-business supply chains. J. Manage. Inform. Syst., 21(3): 7-45.

Guarino, N., 1998. Formal ontology and information systems. Proceeding of 1st International Conference on Formal Ontology (FOIS, 1998), Trento, Italy.

Gunasekaran, A. and E.W.T. Ngai, 2004. Information systems in supply chain integration and management. Eur. J. Oper. Res., 159(2): 269-295.

Huan, S.H., S.K. Sheoran and G. Wang, 2004. A review and analysis of Supply chain operations reference (SCOR) model. Supply Chain Manag., 9(1): 23-29.

Hult, G.T.M., D.J. Ketchen, S.T. Cavusgil and R.J. Calantone, 2006. Knowledge as a strategic resource in supply chains. J. Oper. Manag., 24(5): 458-475.

Higginson, J.K. and A. Alam, 1997. Supply chain management techniques in medium-to-small manufacturing firms. Int. J. Logist. Manage., 8(2): 19-32.

Jain, V., S. Wadhwa and S.G. Deshmukh, 2009. Revisiting information systems to support a dynamic supply chain: Issues and perspectives. Prod. Plan. Control Manage. Oper., 20(1): 17-29.

Kendall, M.G., 1955. Rank Correlation Methods. Hafner Publishing Co., New York.

Kishore, R., R. Sharman and R. Ramesh, 2004. Computational ontologies and information systems I: Foundations. Communi. Associat. Inform. Syst., 14(Article 8): 158-183.

Lambert, D.M. and M.C. Cooper, 2000. Issues in supply chain management. Ind. Market. Manag., 29(1): 65-83.

Murthy, S.K., 1998. Automatic construction of decision trees from data: A multi-disciplinary survey. Data Min. Knowl. Disc., 2: 345-389.

Norman, M.S., D.W. Hildum and D. Kjenstad, 2003. Agent-based e-supply chain decision support. J. Org. Comp. Elect. Com., 13(3-4): 225-241.

Paulraj, A., A. Lado and I.J. Chen, 2008. Inter-organizational communication as a relational competency: Antecedents and performance outcomes in collaborative buyer- supplier relationships. J. Oper. Manag., 26(1): 45-64.

Quinlan, J.R., 1979. Discovering Rules by Induction from Large Collections of Examples. In: Michie, D. (Ed.), Expert Systems in the Microelectronic Age. Edinburg University Press, Edinburg, pp: 168-201.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco.

Rai, A., R. Patnayakuni and N. Seth, 2006. Firm performance impacts of digitally enabled supply chain integration capabilities. MIS Quart., 30(2): 225-246.

Selwyn, P., 2005. Machine learning for dynamic multi-product supply chain formation. Expert Syst. Appl., 29(4): 985-990.

Solvang, W.D., 2001. Architecture for supply chain analysis and methodology for quantitative measurement of supply chain flexibility. Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim.

Sotiris, K. and P. Panayotis, 2009. Predictive Data Mining: A Survey of Regression Methods. In: Khosrow-Pour, M. (Ed.), the Encyclopedia of Information Science and Technology. 2nd Edn., 6: 3105-3110.

Sotiris, Z., 2000. Supply chain management. Report Produced for the EC Funded Project INNOREGIO: Dissemination of Innovation and Knowledge Management Techniques, BPR Hellas, 2000.

Stainer, A., 1997. Logistics-a productivity and performance perspective. Supply Chain Manag., 2(2): 53-62.

Svensson, G., 2003. Holistic and cross-disciplinary deficiencies in the theory generation of Supply chain management. Supply Chain Manag., 8(4): 303-316.

Sycara, K.P., 1998. Multiagent systems. AI Mag., 19(2): 79-92.

Tummala, V.M.R., C.L.M. Phillips and M. Johnson, 2006. Assessing supply chain management success factors: A case study. Supply Chain Manag., 11(2): 179-192.

Wooldridge, M., 1999. Intelligent Agents. In: Gerhard, W. (Ed.), Multiagent Systems: A Modern Approach to Distribute Artificial Intelligence. The MIT Press, Cambridge, 1: 27-78.

Zdravkovic, M., H. Panetto, M. Trajanovic and A. Aubry, 2011. An approach for formalizing the supply chain operations. Lect. Notes Bus. Inf., 5(4): 401-421.