Supply Chain Enhancement Using Improved Chaid Algorithm for Classifying the Customer Groups

C.P.Balasubramaniam,

Ph.D Research Scholar ,

Karpagam Academy Of Higher Education,

Karpagam University,

Coimbatore,Tamilnadu

India

Dr.R.Gunasundari

Associate Professor,

Department of Information Technology,

Karpagam Academy Of Higher Education,

karpagam university

Coimbatore,Tamilnadu

India

Abstract: Lot of Companies is involving in the activities of the Supply Chain that is will used for different functions like company process management and relation between the suppliers and customers agents. The decision-tree based approach is used to make learn and recognize the logical methods of a tree structure. A state-of-the-art supply chain management [8] gives the coding rules as well as the Logical rules features needed by the system. Each attribute is classified and tested during the layout analysis and Logical features are collected and compared to the company's synthetic data set. The Agent-Based Modeling method is employed in the study is the Improved Chi -squared Automatic Interaction Detection" (I-CHAID method). All the user needs are tested around 50 dataset attribute contents belonging to the SCM with I-CHAID method; this will be representing the lower error rate for determining the logical labels are less than 5%. And also the efficiency like precision, recall, error rate will calculate and explains in detail functioning of I-CHAID with respect to the company's supply chain management.

Keywords: Decision-Tree Based Approach; Agent Based Modeling, Improved Chi-Squared Automatic Interaction Detection, Error Rate.

1. INTRODUCTION

Supply Chain Management is mainly based on the two core ideas. The first idea is that practically every product that reaches an end user represents the cumulative effort of multiple organizations. These organizations are mainly focused to collectively as the supply chain.

The second idea is that while the supply chains have existed for a long time, most of the organizations have only paid attention to what was happening within their "four walls." Very less businesses method, much less managed, the last movement chain of activities that ultimately delivered the products to the final customer. The result is not proper method and ineffective supply chain method.

Supply chain management, is an active management of supply chain activities to enlarge the customer value and achieve a sustainable competitive advantage. It represents a conscious effort by the supply chain firms to develop and run the supply chains in the most effective & efficient ways possible. Supply chain activities manage everything from the product development, sourcing, production, and logistics at the same time information of systems needed to coordinate these activities.

Figure1.Supply Chain Structure

The organizations that make up the supply chain are "linked" together through the physical flows and information flows. Physical flows develop the transformation, storage of goods, movement, and materials. These are the most visible piece of supply chain. Information deals allow the various method of supply chain partners to coordinate their long-term plans, and to control the day-to-day flow of goods and material up and down in the supply chain

Supply Chain Management (SCM) is the oversight of materials, information, and finances as they move in a process from the supplier, wholesaler, manufacturer, retailer to consumer.
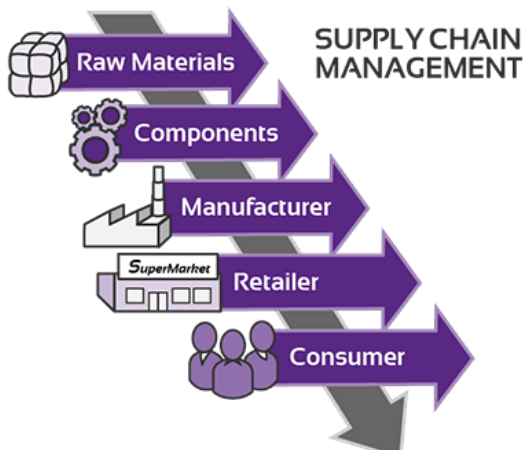
Figure2.Supply Chain Process

Supply chain management develops coordinating and integrating the flows on both within and among companies. The main goal of supply chain management system is to reduce the inventory. As a solution for the successful supply chain management, sophisticated software systems with the Web interfaces are competing with the Web-based Application Service Providers (ASP) who promises to provide a part or all of the SCM service for companies who rent their service.

2. LITERATURE REVIEW

Bolstorff and Rosenbaum (2003), understanding the SCM depends on the motivation and interest of those involved with this concept. A technology provider may associate the SCM with software, like Enterprise Resource Planning and Advanced Planning and Scheduling systems, third-party logistics providers align SCM with the distribution practices. Consulting companies may align it with their intellectual property, for example, along in the same line, it is possible to say that the same phenomenon is seen in the academia.

Halldorsson et al.(2007) discuss that SCM is unified theory and there is no one replace this theory. Depending on each situation, one can choose a theory as the dominant explanatory theory, and then complement it with one or several of the other theoretical perspectives". To establish a work of reference that allows us to mitigate the gap between the present SCM research and Practice and the theoretical explanations and to understand the SCM in practice such as economic perspective; socio-economic perspective; and strategic perspective. Also explaining the two research questions is something mentioned below like, how to structure a supply chain and how to achieve a goal of certain structure of supply chain. Finally their proposed system gives the final framework for looking at two different problem areas within the SCM such as third-party logistics and new product development.

Davis (1989) and Leblanc (1992) algorithms are based on the definition of a within-node homogeneity measure, unlike Segal's algorithm which tried to maximize between-node separation. In addition to that the decision tree is classified as the survival trees and regression tree. Survival tree [12] based analysis is a powerful non-parametric method of clustering the survival data [13] for prognostication to determine the importance and effect of various covariates.

Su and Fan (2004) extended the CART formally known as the classification tree and regression trees was designed by Brieman et al. (1984) for addressing the tree size selection and other issues related to the formation of the related attribute tree structure.

CART algorithm described as the three different procedural categories like growing a large tree, pruning the sequence of nested sub-trees, and finally selecting a best-sized tree. In addition to that there are two close approximations are available in the formation of analysis of correlated failure times when forming the CART tree. Marginal method is the marginal distribution approach of correlated failure times is formulated for classification method.

The approach is the infirmity model to the regression setting. The CART algorithm to multivariate survival [11] data by introducing a gamma distributed frailty to account for the dependence among the survival times [13] based on the likelihood ratio test as the splitting function.

Gordon and Olshen (1985) gave the first adaption of CART algorithm and it concise the recursive partitioning scheme of tree-structured [2] for classification, regression and probability class estimation are adapted to cover the censored survival analysis [3].

The assumptions only required are those which guarantee the identify ability of the conditional distributions of lifetime covariates. Thus, the techniques are applicable to more general situations than the famous semi-parametric model of Cox. This subject is used to the censorship of data. Wasserstein metrics is measure the distances between Kaplan-Meier [1958] curves and certain point masses.

A variable selection approach is useful to inference method, the Generalized Likelihood Ratio (GLR) test, is employed to address the hypothesis testing problems for existing the models.

Allen Zaklad et al. discussed that the Sustainable Supply chain improvement is the business process development, enabling technology, and social system transformation. They presented a model of supply chain intervention that will enable you to address the hidden side of supply chain operations in the context of business processes and technology.

Osman.A.H and Naomie developed the Semantic Plagiarism Detection Scheme mainly focused on Chi-squared Automatic Interaction Detection. It discussed based on the semantic text plagiarism detection technique based on the Chi-squared Automatic Interaction Detection.

It was very useful the analyses and compares the each text in Semantic Plagiarism Detection Scheme and each argument generated by the CHAID algorithm scheme [17] in order to select the important arguments were the another features are added of the method.

3. AGENT DECISION NETWORK FOR SELECTING THE SUPPLIER

Agent decision network is selecting the particular supplier that maximizes the usage of associated with the

supply chain. This system has been developed by the agents, and improve need to buy two different types of materials to manufacture the product.

Each material is associated with a list of possible supplier agents with it. The reputation node and the offer node are the maintained on the choice of the supplier where the offer node is modeled by a deterministic node characterize the cost of material offer proposed by the supplier.

This is determined by comparing the received offer to the market price and/or evaluating the quality of product and characteristics of the offered products. The supplier agent influences the commercial transaction method to acquire the specific material, represented by the node Transaction Status. The reputation and the offer most jointly determine the utility associated with the transaction between the customer agent and the selected supplier agent.

According to the probability distribution of reputation, the value of the expected utility is associated with the offers. To determine the utility of the entire sub-chain, the transaction utility nodes for all the materials are used together.

The Transaction probabilistic nodes consist of influence the probabilistic node Supply chain, raw materials, instead, which expresses the probability that a supply chain can be successfully developed.
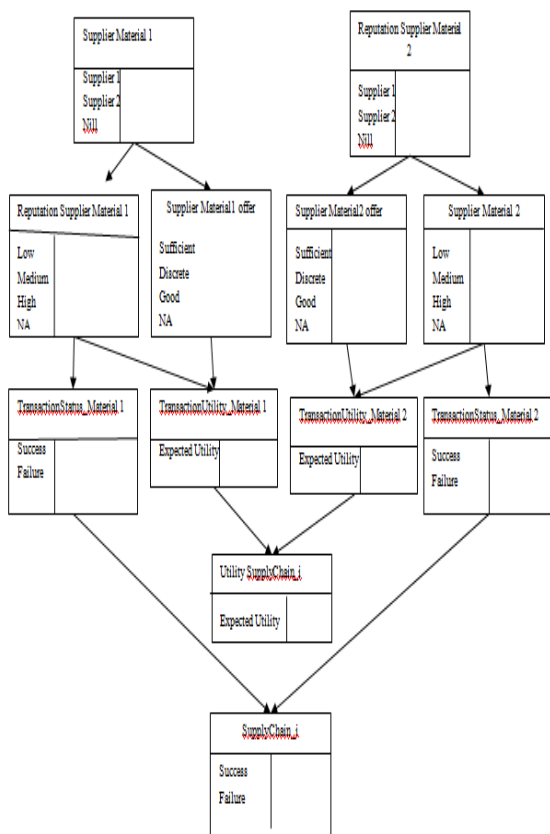

Figure 3.Agent Decision Network for Selecting the Supplier

4. IMPROVED CHAID ALGORITHM

Chi -squared Automatic Interaction Detection is the process of test, the mining data from the decision tree structured framework. The new skeleton are proposed in the CHAID named as Improved CHAID which has two processing schemas which are provided for the CHAID like insertion of logical rules and knowledge to support the decision tree.

In the Improved CHAID , there are two setups have been tested such as, each logical element have a own decision tree structured as well as single decision tree structured [2] for all the logical elements. Both trees are recognizing by the Improved CHAID with respect to the kernel of the system. The Improved CHAID are the trust with confidence on a decision tree.

The Improved CHAID algorithm is mainly focused on the data process whereby two, three or more tree elements are distinguished by the decision tree structure. Decision tree is one common method used in the data mining [16] to extract the predicted information. Morgan and Sonquist uses the regression trees in terms of decision tree for prediction and explanation process with the help of AID (Automatic Interaction Detection).

Generally Chi-squared Automatic Interaction Detection is discrimination and classification methods developed, based on the same representation paradigm by trees. Many methods are intended to increase the probability of solving some problem proposed in the past few era. Specially to improve the quality of the Quinlan's system, leading to the famous C4.5 method. This mobility emerged the concept of lattice graphs which was popularized by the induction graphs of the SIPINA method.

Rokotamalalaetal explains the concept of the decision tree construction for classification and discrimination problems based on the SIPINA software. The main problem is to predict the output value of an object from a set of variables, discrete or continuous. This prediction process is the problem solving in order to develop and finding a partitioning of the individuals elements on the decision tree.

The objective is to improve individual groups, the most homogeneous as possible from the point of view of the variable to be predicted. The main idea is to represent the empirical distribution of the attribute to be predicted by each and every node of the decision tree. Thus, the tree build favors the more discriminating attributes. Here, the difficulty is to choose among N attributes characterizing the structure elements that made it possible to have the best discrimination rate.

Depending upon the prediction of variable candidate and the variable, the characterizing has the two different conditions of things, generally the segmentation and statistical criteria available in the supply chain management. Both are used to entropy of Shannon and its alternatives. The segmentation define a contingency table crossing the variable to be predicted and the descriptor candidate.

All those actions happened for the process comprehension. Consider the notations to describe the numbers resulting from the crossing of the attribute class with V modalities and a descriptor with U methods.

Table 1. Number table during the crossing of two variables

$Q = P$	p_1	p_u	p_U	$\sum_{u=1}^{U} p$
q_1	w_{11}	w_{1u}	w_{1U}	
q_2	w_{21}	w_{2u}	w_{2U}	
q_v	w_{v1}	w_{vu}	w_{vU}	w_u
q_{V-1}	$w_{((V-1),u)}$	$w_{((V-1),u)}$	$w_{((V-1),U)}$	
q_V	w_{V1}	w_{Vu}	w_{VU}	
$\sum_{v=1}^{V} q$	w_v			w

To calculate the relevance of a variable in the segmentation, CHAID develops the independence deviation χ^2 defined by the following equation.

$$\chi^2 = \sum_{v=1}^{V} \sum_{u=1}^{U} \frac{\left(w_{vu} - \frac{w_v \times w_u}{w}\right)^2}{\frac{w_v \times w_u}{w}}$$

Equation (1)

The values of χ^2 are not bounded, they are in the range $[0, +\infty]$. The main disadvantage is the high emphasis of the descriptors having a high number of modalities. To decrease this negative impact, it is much more manageable to normalize by the number of freedom degrees. The formula T of $T_{schuprow}$ has values now in a range [0; 1]. This new concept of equation gives the Improved CHAID algorithm.

$$T_S = \frac{\chi^2}{\sqrt[w]{(V-1) \times (U-1)}}$$

Equation (2)

There are three lemma were used to find the significance of the decision tree which is based on the method of crossing of two variables.

Lemma 1:

If splitting of χ^2_{split} and $T_{S_{split}}$ which is given as the two different conditions for performing the decision tree

Condition 1: $\chi^2 > T_S$ decision tree were not performed.

Condition 2: if the threshold is increased of T_S to the $\chi^2 > T_S$ automatically the shortest decision tree was performed.

Lemma 2:

If merging of χ^2_{merge} and $T_{S_{merge}}$ which is given as the two different conditions for performing the decision tree as large and compact one. Before merging the different variable, it is need to verify the merging variable with the class distribution whether it has a statistical significance of threshold value and threshold level. This assumption gives the two conditions.

Condition 1: from lemma 1, when T_S values are decreased with the respect to the statistical significance the largest decision tree was formed.

Condition 2: simultaneously T_S values are increased with the respect to the statistical significance; the compact decision tree was formed.

Lemma 3:

Let us consider the level of process as "φ" and their independent significance tests as "τ" it gives the significance probability of ρ that will be got no significant differences in all these tests, which is described as following described condition.

the product of the individual probability $(1-\varphi)^\tau$, from this considerations of test, the study apply this in the following example consideration like level of 0.03, and τ is 10 however, it is got the individual probabilities as 0.74 from this the study now has a 26% chance that one of these 10 tests will turn of out & significant, despite each individual test performance only being at the 3% level. In order to guarantee that the overall significance test is still at the "φ" level, we have to adapt the significance level α′ of the individual test. This following results are relation between the overall and the individual significance level were compared with the following comparison results

$$(1-\varphi')^\tau = (1-\varphi)$$

Equation (3)

This equation can easily be solved for φ':

$$\varphi' = 1 - (1-\varphi)^{\frac{1}{\tau}}$$

Equation (4)

Which for small φ reduces to:

$$\varphi' = \frac{\varphi}{\tau}$$

Equation (5)

To get an overall significance level φ and perform τ individual tests, simply obtain the significance level for the individual tests by φ'.

5. EXPERIMENTAL RESULTS

The proposed experimental approaches are around five data set attributes were listed named as the Supplier Agent, Production Manager Agent, Dealer Agent, Client Agent, Inventory Agent that containing the 1000 data set and that is splinted in to two block each. In this era two approaches [19] are compared using the measures: recall, precision, and error rate with the proposed system.

The following table represents the three different approach are compared and shows the quality metrics like the recall, precision, insertion rate and error rate.

Table 1: Average Efficiency Quality Metrics

Exible and Generic Approach (Multi-Tagging Method)			Decision-Tree Based Approach (Data-Mining Method)			Agent-Based Modeling in Supply Chain Management (I-CHAID method)		
Efficiency quality metrics (Average)			Efficiency quality metrics (Average)			Efficiency quality metrics (Average)		
R ecall	Pre cision	Er ror rate	R ecall	Pre cision	Er ror rate	R ecall	Pre cision	Er ror rate
9 5.7%	93. 5%	5. 9%	9 3.0%	94. 0%	6. 4%	9 7.5%	92. 2%	4. 5%

The Agent-Based Modeling in Supply Chain Management (I-CHAID method) is more efficient to obtain better results Supply Chain Management. However the results of both methods are close to the proposed system but give the efficient results.
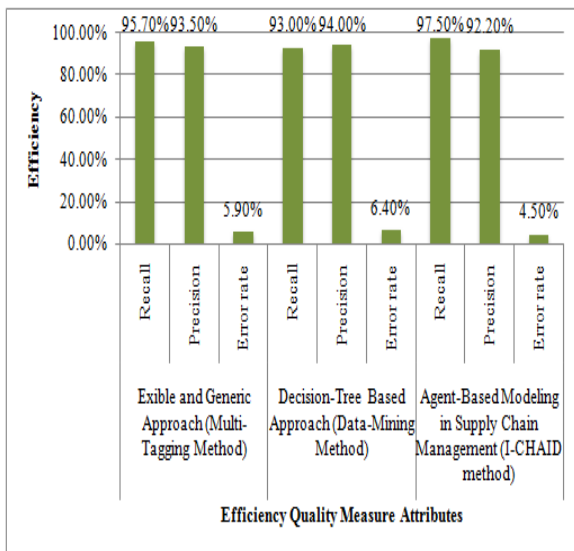

Figure 4. Efficiency Quality Measure

The Agent-Based Modeling in Supply Chain Management (I-CHAID method) is more accuracy to obtain better results in Supply Chain Management. However the results of both methods are close to the proposed system but the proposer system gives the accuracy.
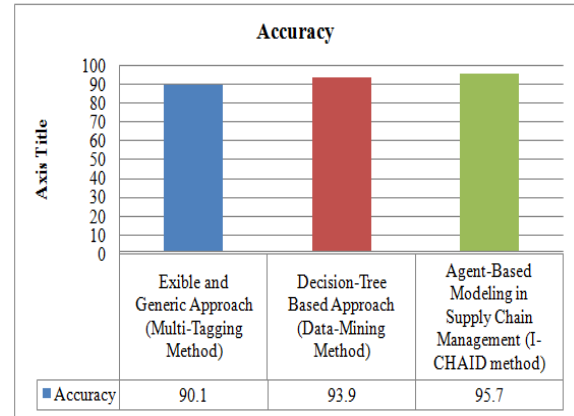

Figure 5. Various SCM Accuracy Comparisons

5. CONCLUSION

To determine the nature and performance of the proposed work for SCM with presence of CHAID improved and Bounded in extent of the Synthetic dataset is used in terms of Supplier, Production Manager, Dealer, Client and Inventory attribute set. Which contains the Client city, Client Name, Client Phone number, Client State, Cost, Dealer Name, Dealer id, Item Name, Product Name, Product Price, Product Quantity, Product id, Product name, Supplier Name, Supply Quantity. The above Synthetic dataset are managed and classified by the proposed improved CHAID framework. The obtained results of accuracy and its performance measures of recall, precision and error rate are measured and compared with the existing technique for the input supply chain [10] management dataset. The classification of this algorithm is viewed in tree structure where the decision tree is classified. Figure 4 gives the accuracy comparison for the proposed framework and existing technique like Multi-Tagging Method and *Data*-Mining Method [16].The accuracy here is measured by sensitivity and specificity. From the figure 3, the Efficiency quality metrics obtained by the proposed technique Improved CHAID for SCM in recall is 97.5% and precision is 92.2 % and error rate 4.50% which is better when compared with the existing technique of Exible and Generic Approach and Decision-Tree Based Approach. The comparison graph for Efficiency quality measure attributes are calculated by sensitivity and specificity is shown in the figure 3. From the figure 4 proves that the accuracy of proposed improved CHAID with SCM gives maximum while comparing with the existing technique for SCM.

6. REFERENCES

1. Gordon and Olshen, "Tree-structured survival analysis", Cancer Treatment Reports, Journal Article, Research Support, U.S. Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S. [1985, 69(10):1065-1069].
2. Xiaogang Su, Juanjuan Fan "Multivariate Survival Trees: A Maximum Likelihood ApproachBased on Frailty Models",Biometrics journal of the international Biometrics society, Volume 60, Issue 1, pages 93–99, March 2004.

3. A. Bela, T. Moinel, Y. Rangoni "Improved CHAID Algorithm for Document Structure Modeling"

4. Camarinha-Matos, LM. andAfsarmanesh, H, Collaborative networked organizations: a research agenda for emerging business models. Massachusetts, Kluwer Academic, 2004.

5. Lee, J.-H., and Kim C.-O. (2008). "Multi-agent systems applications in manufacturing systems and supply chain management: a review paper." International Journal of Production Research 46(1): 233-265.

6. Moyaux, T., Chaib-Draa, B., and D'Amours, S. (2007). "Information sharing as a coordination mechanism for reducing the bullwhip effect in a supply chain." Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews 37(3): 396-409.

7. Frayret, J.-M., D'Amours, and Montreuil, B. (2004a). "Coordination and control in distributed and agent-based manufacturing systems." Production Planning & Control 15(1): 1-13.

8. Forget, P., D'Amours, S., Frayret, J.-M., and Gaudreault, J. (2008b). Design of multi-behavior agents for supply chain planning: an application to the lumber industry. Supply Chains: Theory and Application. V. Kordic, I-TECH Education and Publishing: 551-568.

9. Bolstorff, R. R., and Rosenbaum, R. (2003). Supply chain excellence. New York, AMACOM.

10. Halldorsson, A., Kotzab, H., Mikkola, J.H., and Skjott-Larsen, T. (2007). "Complementary theories to supply chain management." Supply Chain Management: An International Journal 12(4): 284-296.

11. Davis, R. and Anderson, J. (1989): Exponential survival trees, Statistics in Medicine 8, pp 947-962.

12. Lebalanc, M.; Crowlry, L. (1992): Relative risk trees for censored survival data, Biometrics. v48. 411-425.

13. Su, X. G.; Fan, J. J. (2004): Multivariate survival trees: a maximum likelihood approach based on frailty models, Biometrics 60, pp. 93-99.

14. Kaplan, E.L.; Meier, Paul. (1958): Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53, 457-481.

15. Kass G. (1980), "an exploratory technique for investigating large quantities of categorical data, Applied Statistics" 29(2), 119-127, 1980.

16. Anita Prinzie, Dirk Van den Poel, "WITHDRAWN: Constrained optimization of data-mining problems to improve model performance: A direct-marketing application", http://www.researchgate.net/publication/222555006.

17. Gilbert Ritschard "CHAID and Earlier Supervised Tree Methods", Publications récentes du Départementd'économétrie, http//www.unige. ch/ses/metri /cahiers.

18. HongwenZheng , Yanxia Zhang "Feature selection for high dimensional data in astronomy" Accepted for publication in Advances of Space Research, arXiv:0709.0138v1 [astro-ph] 3 Sep 2007

19. Jose Manuel Serra, Laurent A. Baumes, "Zeolite synthesis modelling with support vector machines: A combinatorial approach", publication at: http:// www.researchgate.net /publication/6537094

20. C.P. Balasubramaniam, V. Thigarasu "Agent-based Modeling in Supply Chain Management using Improved C4-5", Research Journal of Applied Sciences, Engineering and Technology 9(2): 91-97, 2015, Maxwell Scientific Organization, 2015.