# Hybridized Model for Decision Support System with Data Mining and CBR

Luxmi Verma
*Department of Computer Science &Engineering*
*The NorthCap University, Gurugram, Haryana, India*

Sangeet Srivastava
*Department of Applied Sciences*
*The NorthCap University, Gurugram, Haryana, India*

Varun Sapra
*Department of Information Technology*
*Jagannath International Management School, India*

**Abstract- Today to flourish, every industry needs a support mechanism, this mechanism not only includes men and machine but also those models which provide them enough support to take a decision anticipating business scenarios. Decision support system (DSS) came into existence in 1970s and made a huge impact on the business processes and on the think tank of the industry. Many technical and organizational demands have made an impact on the revolutionizing of DSS approach. Earlier majority of models focused mainly on statistical analysis, goal programming and OLAP-based tools enabling multiple view on data. In this paper we introduced a model (KD2S2, Knowledge Driven-decision Support System) which emphasizes on the coalition between the research areas of data mining, CBR approach for decision support (DSSs). As data mining techniques extends the possibilities for decision support by discovering patterns and relationships, enabling an inductive approach to data analysis and CBR has emerged as a major research area due to its appeal as a methodology for building intelligent systems. This paper is the extension of our previous work where we proposed a conceptual framework for Decision support system based on Data Mining and CBR. Here we emphasize particularly on the CBR component of the model, illustrating an algorithm for matching user query with the existing cases in the CBR.**

**Keywords –Data mining, Case Based Reasoning, Knowledge Driven decision support system,**

## I. INTRODUCTION

In recent years data mining has become a very popular technique for extracting information from the database in different areas due to its flexibility of working on any kind of databases and also due to the surprising results [2]. An approach has been made to develop a decision support system which will take decision under complex environment. CBR is to solve problems by comparing a new case to previously experienced ones. With case-based reasoning, experiences are stored in memory. These experiences encode relevant features, courses of action that were taken, and results that ensued. When a new case arrives, it will find the situation that is most similar, and reuse, or revise it to match the new problem if the most similar problem does not match sufficiently. CBR has been successfully applied to a wide variety of areas including medical and engineering tasks. However, problems remain with the case based approach. Some of the most important issues that need to be addressed are:

• The integration raw facts and knowledge into a linking case base knowledge to enhance the potential of the candidate system and to improve the probability of finding the correct measure.

• Finding effective parameters for indexing and retrieving methods used for recovering the past cases. There are several main types of retrieval, knowledge based search, classification network search, nearest neighbor search, inductive approaches, etc.

• The automation of feeding the results from data mining tools to a CBR so that a corresponding case can be retrieved effectively. Cases are retrieved based on the knowledge stored in CBR, which further is a main issue of consideration.

## II. PROPOSED MODEL

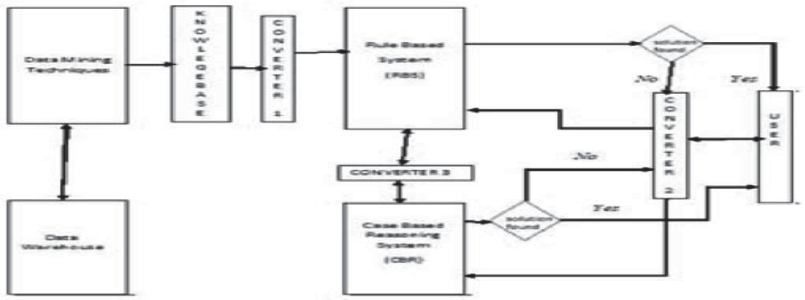


Figure 1. Framework for Knowledge Driven Decision Support System [1]

### A. *Procedure*

*Step 1* : Initially populate the case base with data mining result.
*Step 2 :* Take input from the user and convert into cases.
*Step 3 :* Search the case base for the relevant solution and present to the user.
*Step 4 :* If solution not in case base then search the knowledge base for the solution and update the case base.
End.

### B. *Components of the Model*

Knowledge Base - Knowledge base is an organized collection of facts about the system's domain. It consists of some encoding of the domain expertise for the system. This can be in the form of semantic nets, procedural representations, production rules, or frames. Knowledge base contains the knowledge necessary for understanding, formulating, and solving problems in the form of rules and expressed in the form. If the conditions are true then the actions are executed.

Data Mining - Data mining includes predictive data mining algorithms which result in models or interesting patterns that can be used for prediction and classification, and descriptive data mining algorithms for finding interesting patterns in the data, like associations, clusters and subgroups [8].The overall process of discovering and extracting valid, implicit and previously unknown knowledge from enormous databases is often referred to as data mining [2]. The popular data mining techniques are classification, clustering and association. Data mining engine is used to populate the case base. The data mining engine fetch the structured and unstructured data from the data base and data warehouse. Data mining techniques such as clustering, classification and association will be applied to the selected data results in generation of the knowledge in the forms of clusters, rules, etc.

Case-Based Reasoning-Case based reasoning (CBR) is a popular problem solving methodology which solves a new problem by remembering previous similar situations and reusing knowledge from the solutions to these situations [3] CBR enables retrieval of relevant data by comparing the user's current problem with previous situations (cases). Cases are situation-specific knowledge, stored in an experience base together with the necessary general knowledge (concept hierarchies, relationships, decision rules, associations, etc.) [2].

*Case Based Reasoning Sub-Component*

We have proposed a case based reasoning component which will have a case base (collection of relevant cases). A case is represented in foam (identification, attribute, value). Similar cases are identified in one location, i.e in certain types of sorting. A problem will be first transformed in to a case called problem case (in the format specified) and some sophisticated matching algorithm is used to retrieve the similar cases. Then re-use the similar case. The solution part of the retrieved case is modified, i.e., revised. That is the solution for problem case. It is presented to the user. At the same time, the problem case together with modified solution is added to the case base.

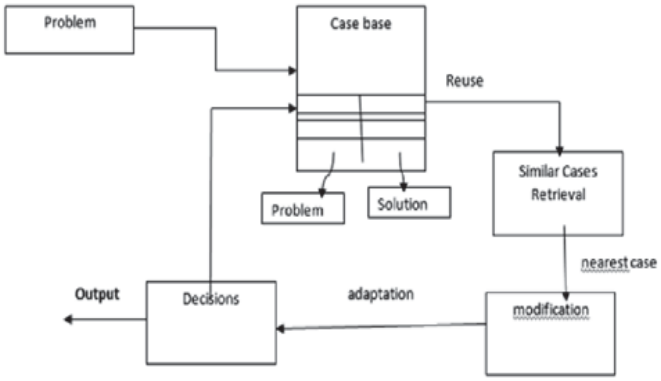


Figure 2. CBR Component

*Approaches to case retrieval from CBR*

The main concern of all the CBR system is the retrieval of similar cases rapidly and efficiently with the following objectives 1. The retrieve cases must be as few as possible 2. The retrieved cases must be as relative and similar to the current problem as possible [7]. CBR is an approach to problem solving that emphasizes the role of prior experience during future problem solving (i.e new problems are solved by reusing and if necessary adapting the solutions to similar problems that were solves in the past An important step in the CBR Cycle is the retrieval of previous cases that can be used to solve the target problem. Improving retrieval performance through more effective approaches to similarity assessment has been the focus of a considerable amount of research. There are various retrieval methods inductive, nearest neighbor, knowledge based method, Decision tree, Fuzzy based approach which have both the advantages and disadvantages [3-6].

Case 1
Problem: Sales is not as per last year
Model: Alto
Advertisement: NewsPaper
City: New Delhi
Year: 2013
State: Delhi
Solution : Diagnosis
1. Dealer incentives needs to be revised
2. Service station needs to be improved

Figure 3. Sample case sturcture

K-NN Approach- K-nearest search method has been extensively used in the case retrieval phase of CBR [9]. Traditional K-NN method adopts an exhaustive search scheme to scan the overall case base, and then select K prior cases which have the least dissimilarities with the new case from the case base. The performance of a CBR system is indeed influenced by the value of the parameter K when using the K-NN method. However, determining the value of K is never such a simple task, especially when users do not have sufficient domain knowledge. If K is too small to find a representative pattern, so that it may reduce the reasoning accuracy. Also, the reasoning accuracy may be reduced if K is so large since the selected prior cases would include some irrelevant or distorted cases. Up to now, several studies have developed their individual methods to determine the value of K for K-NN [10-13].

## III. PROPOSED APPROACH

Algorithm to retrieve similar cases from Case base to

*Step 1*: Consider a Target case T with T[i],T[i+1]…. T[n] features
*Step 2*: Consider an array S and fill the S with the first feature of the case base
*Step 3*: Compare T[0] with all the elements of S[i]…S[n]
For i = 1 TO N
Compare T[0] with S[i]..S[n]
*Step 4:* Store all the cases where T[0]==S[i] and generate the resultant set
*Step 5:* Repeat step 3 to 4 for rest of the elements of T with as, filling S with new features set values for every element of T.
*Step 6:* From the resultant sets count the number of occurrences for each case considering only those which are in resultant set and discard the rest.
*Step 7:* Fill the occurrences in an array and compare each occurrence with the minimal value and if the occurrence is greater than minimal value change the values of minimal with the occurrence value and record the position.
*Step 8:* The final value position will be the most similar case.

## IV. CONCLUSION

This algorithm finds the exact match rather then getting the approximation or the nearest neighbor. The exact match does not mean getting the result based on 100 % matching criteria but this algorithm matches the given criteria which may or may not be a single column value. It then matches the attributes within the column, thereby counting the occurrences of that attribute within the column, and then retrieving the value of the occurrences in each row so that the best possible match is found. In this work we have described the first phase of the three-phased model. This phase deals with the user query and retrieval of the best possible solution from CBR. The solution is appropriate for categorical attributes. Where in a particular domain if we want to search for model x with other supporting attributes, then we cannot replace model x with model y.

## REFERENCES


[1] L.Verma, S.Srivastava, S.Srinivsan, V.Sapra, " Integraiton of rule based reasoning system to support decision making", IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques 2014.

[2] Zhi-Wei, Shan-lin Yang, Yun Yang, and Feng-gang Li , "Case based reasoning framework based on data mining technique. Proceedings of the Third International Conference on Machine Learning and Cybernatics, Shanghai, 26-29 August 2004.

[3] Becerra-Fernandez, Irma, and David W. Aha, "Case-Based Problem Solving for Knowledge Management Systems." , FLAIRS Conference. In Proceedings of the welfth Annum Florida Artificial Intelligence Research Symposium Menlo Park 1999.

[4] Bergmann, Ralph, and Armin Stahl. "Similarity measures for object-oriented case representations." In European Workshop on Advances in Case-Based Reasoning, pp. 25-36. Springer Berlin Heidelberg, 1998.

[5] YanqingSun; ShuhuaYin; W angJi. Improved attribute reduction algorithm for case-based reasoning system based on rough set. Computer Engineering, 2010, 36(10):38-40.

[6] Chen, Youfen. "Research of the Case Retrieval Model Based on CBR." InIntelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on, pp. 1132-1136. IEEE, 2013.

[7] Hans-Dieter Burkhard Extending some Concepts of CBR—Foundationsof Case Retrieval N ets. In MarioLenz, BrigitteBartsh-Sporl Hans-Dieter 113

[8] Burkhardeds. Case-Based Reasoning Technology: From Foundations to Applications Springer 1998.

[9] Rashmi Kandwal, et al, " Health GIS and HIV/AIDS studies: Perspective and retrospective", Journal of Biomedical Informatics, 2009.

[10] Chieh-Yuan Tsai et al, "Developing a Significant Nearest Neighbor Search Method for Effective Case Retrieval in a CBR System ", International Association of Computer Science and Information Technology - Spring Conference-2009.

[11] I. Watson and F. Marir, "Case-based reasoning: A review.", Knowl. Eng. Rev., 1994, 9, pp. 355– 381.

[12] ZhansiJian, Liping Chen, Nianmeng Luo, "Similarity analysis in nearest-neighbor case retrieval", Computer Integrated Manufacturing Systems., June, 2007 , 13(6), pp. 1165-68.

[13] NitinBhatia , Survey of Nearest N eighbor Techniques. International Journal of Computer Science and Information Security. 2010, 8(2).