Volume 2, Issue 4, April 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Improved Case Based Reasoning(ICBR) Tool

Prof. Mythili Thirugnanam
SCSE,VIT University
tmythili@vit.ac.in

Bivash Kumar
Final year B.TECH,CSE,VIT University
bivashkumar41@yahoo.com

Awanish Kumar Gupta
Final year B.TECH,CSE,VIT University
awanishguptavit@yahoo.com

Abstract— **Case Based Reasoning approach is one of the most recent problem solving techniques; it is the process of solving new problem by referring the solutions of similar past problems. In addition to this, it is a prominent kind of analogy making in every day human problems. In past, the Knowledge Based Reasoning technique was used to solve the problem, but solution which is obtained from this approach is not highly appreciable and also it has some limitations such as knowledge elicitation, inability to access high volumes of information, difficult to implementation, high maintenance cost. Recently some of the pioneers contributed their research work towards addressing the above mentioned issues by introducing the Case Based Reasoning approach. The CBR technique is now being used in various fields such as structure extraction of industrial invoices [6], molecular biology [8], diagnosis of diseases [9], image processing [10], and etc... Recently there are many CBR tools developed to reduce the manual time and to bring the accuracy in retrieval of data. Some of the recent existing CBR tools namely Mycbr, FreeCBR, OpenCBR, ShellCBR and etc but it fails to adopt some of the quality attributes such as high complexity, lower degree of reusability, limited applicability etc. To address these issues a new CBR tool is to be developed by incorporating ease of use, accuracy in classification of data, friendlier GUI which is not provided by the existing tools. Therefore this paper presents an improved CBR (ICBR) tool which has enhanced performance and far away from the above mentioned limitations. The proposed Case Based Reasoning tool applies the mixture of experts for classification and similarity retrievals. The algorithm used for similarity retrievals and classification are namely nearest neighbour algorithm and Bayesian classification algorithm respectively. The proposed tool consists of three different phases- input phase, similarity determining phase and classification phase. In the input phase pre classified data will be loaded. So the user can create and edit as many numbers of classes and attributes he wants in a particular field. In the next phase the tool receives unclassified data as input from the user. In the last phase the tool retrieves the matching record or similar records based on the matching algorithm and the input data is also accurately classified using Bayesian classification algorithm.**

Keywords— **Case Based Reasoning, Bayesian Classification Algorithm, Nearest Neighbour Algorithm, Mixture of Attributes, Friendly GUI**

I. INTRODUCTION

Case Based Reasoning is the process of providing the solutions to real world problems based on the knowledge or information gathered (in a case base) from the solutions of the previous problems of similar nature. The CBR is mainly based on the reusability in which the case base is reused and matched against the problem with the help of similarity calculations. The main advantage of CBR approach is its closeness to the real time results and its more interactive Graphical user interface.

The CBR method is implemented in a four step process- retrieval of the most similar cases compared to the case base by identifying relevant problems and matching the case, reuse the retrieved case to solve new problem, revise and adapt the proposed solution and retaining the useful information from final solution in the case base. The CBR approach is specially used where the experts find it very difficult to eloquent their thought processes while solving problems. A good CBR tool should provide a range of retrieval mechanisms and allow mixture of these mechanisms when required and handle a large case base. The existing tools like mycbr and cbr shell java mainly perform the classification by setting the similarity of two objects in the initial phases or by using few numbers of objects attributes. Hence there is a need to develop a CBR tool that can classify the data by using mixture of attributes also called mixture of experts and logistic regression which improves the accuracy rate and achieves more realistic results. The proposed case based reasoning tool applies the mixture of experts for classification and similarity retrievals to get the better results utilizing the features of algorithms like Bayesian Classification Algorithm and Nearest Neighbor Algorithm respectively.

.

II. LITERATURE SURVEY

Peitao Cheng et. al., (2011),[1] proposed that Case Based Reasoning can be used to develop a plane geometry intelligent tutoring system (PGITS). The proposed system is very helpful in order to improve the learning and summarizing abilities of students. This system combines the rule based reasoning and

© 2012, IJARCSSE All Rights Reserved

case based reasoning approaches have increased the efficiency and accuracy of the system. In this approach when a student submits his problem, similar problems are searched in case library. If similar kind of problems is found then it is retrieved along with its solution otherwise the problem is solved by rule based reasoning. If the problem cannot be solved by the system it is sent to the experts. Every solved problem is stored in case library for further use. The main advantages of case are naturalness of representation, easy knowledge acquisition and the ability to express specialized knowledge. The disadvantage is that system uses large amount of space to store the old cases and hence retrieval is not quite fast. S. Margret Anouncia et al., (2010),[3] proposed that Case Based Reasoning can be applied for image interpretation that is easily understandable by users. The process requires the image information to be extracted by feature extraction method. After that these features are matched with the features of available objects of any domain and object is mapped correctly. To retain the exact features of actual image proper modeling of image content is required. This process is applicable in modeling for gray scale image interpretation emphasizing on welding defect classification which resulted in domain ontology of welding defect. Domain knowledge is created by formalizing the grey scale image information and its welding effects. The result obtained from this approach is more accurate than existing technique such as neural and statistical approach. But it is applicable for only enhanced images and further work is needed to be done for low quality images also. Hajar Mat Jani et. al., (2008),[4], proposed Case Based Reasoning approach for documenting frameworks. This approach helps the user to select from the several options representing features of object oriented framework and the system automatically comes up with the solution. There are four steps proposed to use case based reasoning approach - retrieve the previous cases from case base, reuse the previous solutions on match, revise the most similar case to solve new and retain the solution to new problem in case base. This approach has shaped correct results in very less time complexity and thus more effective to new framework users ultimately improving the effectiveness of the system. Hatem Hamza et. al., (2007), [6], proposed structure extraction of industrial documents invoices using Case Based Reasoning. Document Invoice Analysis is done in three main steps: problem elaboration, local problem solving, global problem solving. Many problem elaboration techniques have been proposed such as data extraction, P S(similar patterns) extraction using consecutive horizontal lines of similar patterns, KWS (key-words in vertices, and spatial relationships on edges) extraction by extracting keywords, document graph extraction. These techniques track similar patterns, keywords. Problem solving is done using P S extraction, KWS extraction using edit distance etc. Even though the proposed technique has given a recognition rate of 85%, still accuracy can be improved by using indexing method. Fulong chen et al.,[2007], [7], proposed Case Based Reasoning that can be used to classify SAR(synthetic aperture radar) images. The process of classification of SAR images mainly consists of four steps including SAR image processing, construction of case library, Case-Based classification and post classification processing. SAR image processing here includes speckle removing and geometric correction, while the case library is built using historical multi-temporal SAR images, historical land use map and other ancillary data. In case base classification the most dominant features is determined which is backscatter in this case. After removing some spots and classified reasons on the image it is classified and a new image is constructed that can show the property of a particular land. The proposed CBR approach gives assistance to identify nature of land at a particular time and accordingly a particular crop can be planted. Niloofar Arshad et al., (2005), [8], proposed Case Based Reasoning technique for class discovery in molecular biology, where the rules that define the domain knowledge are difficult to obtain and the number and the complexity of the rules affecting the problem are too large for formal knowledge representation. A method, MOE4CBR (mixture of experts for case-based reasoning), had been proposed that combines an ensemble of CBR classifiers with spectral clustering and logistic regression leading to the result that were very meaningful for a particular class label. The process includes clustering, feature selection, mixing the experts and finally classifying the input data. This method is applicable in life sciences domains such as data set from genes and protein expression profiling. But it suffers from curse of dimensionality that reduces its accuracy which can be improved by weighing the attributes. Nadipuram R. Prasad et.al., (2001) ,[9],proposed a method for combining fuzzy sets of case based reasoning and rule base reasoning for lung disease. Fuzzy set is a set of answers of questionnaires for determining the symptoms. A case base is constructed as a relational database and a similarity is measured between new case and cases stored in case base. Rule base reasoning represents the knowledge base on rules in the form of rules decided by the human brain. Combining the rule based and case base reasoning is done by taking the average of two and the value obtained is more realistic to that of separate methods but the method is bit complex. Valerie Ficet-Cauchard et. al.,(1998),[10], proposed Case Based Reasoning approach to develop an interactive system for development of image processing applications. The approach helps in the reusability of solutions or part of the solutions of the previous problems to solve the novel problem with the help of similarity calculations. The CBR approach also encourages the man/machine cooperation because of its closeness to human reasoning because of its user friendly GUI. The TMT(task method tool) model is used for knowledge acquisition in domain level and control level. The knowledge obtained is used to divide the problem in a tree of tasks and sub-tasks to use reusability by retrieving the previous cases. Morten Grimnes et al., (1996), [11] proposed a two layer Creek Case Based Reasoning architecture which is helpful in better

© 2012, IJARCSSE All Rights Reserved

understanding of medical images. The proposed method combines two cased based reasoners, one for segmentation of image and another for interpretation of image, correspond to the two layers of the architecture. The Case Based Reasoning is extremely appropriate for these two levels of analysis of image. This method decomposes the higher level of case base into low level sub-tasks which help in detailed understanding. The efficiency of the Creek architecture depends upon image segmentation robustness, learning capabilities, domain knowledge, context sensitive problem solving etc. For better efficiency, the degree of dependency is needed to be decreased.

The above survey indicates that all the existing systems have their own limitations. Some of the issues are, use of cosine formula for retrieval which is not effective, less prediction rate [7], use of hypothetical case base[8], dependency of the performance on the previous assumptions[10]. Some of the systems do not use a mixture of matching or classification algorithm to get better accuracy. In order to overcome these shortcomings there is a need to develop a new Case Based Reasoning tool that improves the accuracy and at the same time considers various issues such as lesser prediction rate, slower retrieval of cases into account.

S.No	Title	Proposed Domain	Pros	Cons
1.	Application of Case Based Reasoning in Plane Geometry Intelligent Tutoring System	Plane geometry intelligent tutoring system	Natural representation, Easy knowledge acquisition.	High complexity in storage of cases
2.	A knowledge model for gray scale image interpretation with emphasis on welding defect classification-ontology based approach	Classification of arc welding defects using radiographic images.	Better performance than statistical and neural approach.	Applicable only for enhanced images
3.	Implementing Case-Based Reasoning Approach to Framework Documentation	Documentation of object oriented framework	Helps in better learning of frameworks	Can't deal with new kind of frameworks.
4.	A case-based reasoning approach for invoice structure extraction	Document Invoice Analysis using CBR	Recognition rate is 85%.	Indexing method can be improved for better results.
5.	SAR Images Classification Using Case-based Reasoning Method	Classification of SAR images	Removal of fake cases	Low accuracy in classification of new SAR image.
6.	Data Mining for Case Based Reasoning in High-Dimensional Biological Domains	Integrating clustering techniques with CBR approach for classification in molecular biology.	Improved prediction rate in classification of molecular biology chemicals.	Very complex to implement.
7.	Approach to Combining Case Based Reasoning with Rule Based Reasoning for Lung Disease Diagnosis	Combining fuzzy set for rule based and case based reasoning	Results more realistic to that of rule based or case based when used separately.	Use of hypothetical case base to predict in lung disease diagnosis.
8.	An interactive Case Based reasoning system for the development image processing applications	Development of an interactive system for development of image processing applications.	It has more user friendly GUI	Low accuracy in image classification.
9.	A two layer case-based reasoning architecture for medical image understanding	Study of medical images	Helps in extracting meaningful information from images	Dependency of the performance on the previous assumptions

Table 1: The pros and cons of some of the works in the field of Case Based Reasoning are described.

Above comparative study shows that CBR approach is comparatively easy to use, more accurate and more practical in the above mentioned applications such as molecular biology, image processing. There is a need of a system with a user friendly GUI and accurate algorithms that can help users better similarity retrieval and classification.

III.

© 2012, IJARCSSE All Rights Reserved

IV. COMPARISON AMONG VARIOUS EXISTING TOOLS

	Mycbr	**Freecbr**	**Opencbr**	**Cbr shell**
Interface	Powerful GUI	Only standalone GUI	It has an interactive GUI.	More interactive user interface using java applet
Platform	Windows and Linux platforms.	Windows and Linux platforms.	Windows and Linux platforms.	Windows and Linux platforms.
Algorithm used for similarity retrievals	Similarity based retrieval functionality	Feature based distance measures.	Nearest Neighbor retrieval algorithm.	Nearest neighbor and threshold retrieval.
Advantages	It supports textual similarity modeling It supports structured object-oriented case representations Scalability is improved and results can be retrieved faster.	It can be Personalized according to user's preferences.	Multiple intelligent algorithm Flexible configuration improves scalability. Cross platform	Multiple diagnostic algorithms. It supports Cross platform use.
Limitations	Not applicable for new kind of data i.e. which is completely dissimilar	Set-up cost is high. Feature extraction, selection and weighting make system complex Less accuracy than feature free CBR.		Low degree of reusability Applicability for academic purposes only.

Table 2: Pros and cons of some of the existing tools in the field of Case Based Reasoning are described

V. PROPOSED FRAMEWORK

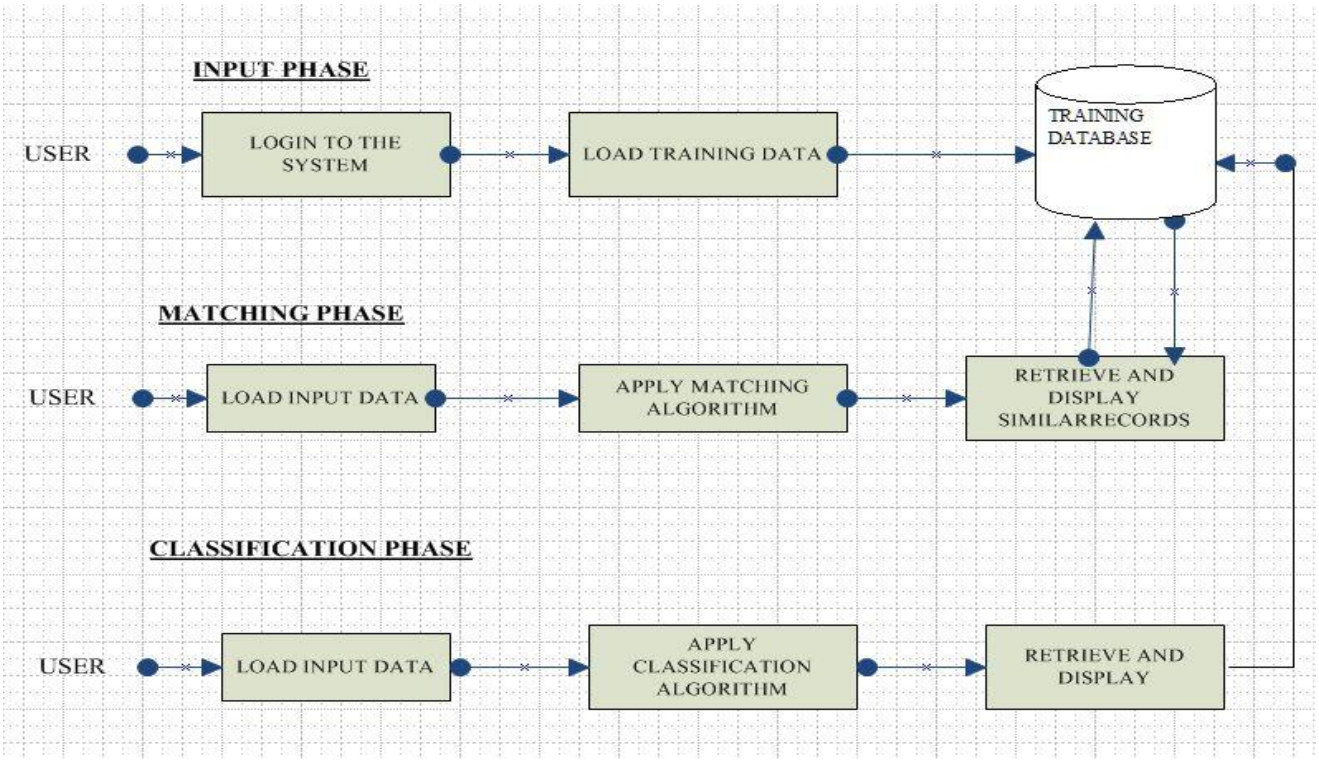

Fig 1: The framework of the proposed system

© 2012, IJARCSSE All Rights Reserved

VI. DETAILED DESCRIPTION

The developing system consists of three major phases namely input phase, similarity determining phase and classification phase. The detailed description of each phase is given below.

A. Input Phase

In this phase the user log on to the system. Next the training data is loaded into the system. The training data should be of CSV format. In the phase of loading the user specifies the class and its attributes. The user also defines the range of the values that an attribute can take and default values of each attributes. Next user manually enters the values of each attributes for a particular class. The data which is entered by user is stored into the disk in the database.

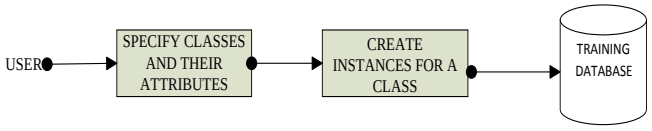

Fig: 2: Procedure for creating the classes and instances.

B. Similarity Determining Phase

In this phase, unclassified data will be taken as the input. Apply matching algorithm to identify and retrieve the similar records same as the input data. The developed ICBR tool supports mixture of attributes for similarity measures to retrieve the similar data. The ICBR tool will bring more accuracy in retrieval process and also improves the performance of the system. The Nearest Neighbor Retrieval algorithm is used for similarity measures. A typical evaluation function is used to compute nearest-neighbor matching using the equation given below.

$$similarity(Case_I, Case_R) = \frac{\sum_{i=1}^{n} w_i \times sim(f_i^I, f_i^R)}{\sum_{i=1}^{n} w_i}$$

Where wi is the importance weight of a feature, sim is the similarity function of attributes, and fiI and fiR are the values for attribute i in the input and retrieved cases respectively.

C. Classification Phase

Classification Phase, user has to provide the input data. Apply classification algorithm to determine the type of pattern/class of the given input data. The developed ICBR tool uses the training data for classification process and bayesian classification algorithm is used to classify the input data. Finally the classified data will be stored in the training data for further use. Advantages of using bayesian algorithm are computational time is very less when compare other classification techniques, also it gives a very good accuracy in case of uncertain data [11] and it uses a probabilistic method to classify the given data so the results are more practical and it can be used for multidimensional data. Bayesian algorithm is described here. For classification of the input data to an appropriate class bayesian algorithm is used. According to this algorithm the probability that a sample S belongs to a class C (P(S/C)) can be calculated as P(S/C) = (P(C/S) X P(S)) / P(C).

1. For each attribute the probability of a particular class is calculated. Suppose the attributes are Y1, Y2….Yn and the class is C1, C2. So for each attributes the probability P(Xn/Cn) is calculated as P (Xn/Cn) = P (Xn ∩ Cn) / P (Cn).
2. The final probability is calculated as
   P(X1, X2…Xn / C1) = P(X1/C1) X P(X2/C1) X…..P (Xn/C1).
   P(X1, X2…Xn / C2) = P(X1/C2) X P(X2/C2) X…..P (Xn/C2)
3. The class for which the probability value is higher is identified as the class for the input data.

VII. CONCLUSIONS

A new approach is proposed for developing a new ICBR tool which improves the usability of graphical user interface, brings more accuracy in retrieve of the similar data and classification of data. Based on the literature survey and comparative study of existing CBR tools, the proposed ICBR tool approaches a higher degree of reusability, moving towards the object oriented approach, using multiple similarity algorithms and a fine GUI which helps in closeness of human thinking to the CBR approach. Initial stage, the proposed new ICBR tool will be helpful in the diagnosis of various major diseases using symptoms of patients as cases. The proposed ICBR further can be enhanced in the area of decision support of various commercial and non-commercial firms, in industrial and architectural design based on reuse of previous components or parts of components and many more.

REFERENCES

[1] Peitao Cheng; Ke Zhao; Yatao Li; Wei Xu Application of Case Based Reasoning in Plane Geometry Intelligent Tutoring System proceedings on International Conference on Digital Object Identifier Publication Year: 2011 , Page(s): 4369 - 4373.

[2] Lahakone, A.U.; Senanayake S.M.N.A. A Hybrid Intelligent System Combining Self Organizing Maps and Case Based Reasoning for Evaluating Postural Control, 2010 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Publication Year: 2010 , Page(s): 38 - 43.

[3] S. Margret Anouncia, R. Saravanan A knowledge model for gray scale image interpretation with emphasis on welding defect classification- ontology based approach on Computers in Industry, Volume 61, Page no.742-749, Issue 8, October, 2010.

[4] Hajar Mat Jani , Lee Sai Peck Implementing Case-Based Reasoning Approach to Framework Documentation on IEEE journal Computer Science and information Engineering, page no. 590-594, issued on 31 March 2009.

[5] Jiangtao Ren, Sau Dan Lee, Xianlu Chen, Ben Kao, Cheng.R, Cheung.D, Naïve Bayes classification of

© 2012, IJARCSSE All Rights Reserved

uncertain data, Proceedings in International Conference on Data Mining, 28th Dec. 2009.

[6] Hatem Hamza, Yolande Belaid and Abdel Belaid, A case-based reasoning approach for invoice structure extraction on Document Analysis and Recognition, page no.327-331, issued on 23 September 2007.

[7] Fulong Chen , Chao Wang, Hong Zhang ,Bo Zhang, Fan Wu SAR Images Classification Using Case-based Reasoning Method Geosciences and remote sensing symposium, Vol. No. 40601058, page no. 2048-2051, July 2007.

[8] Niloofar Arshadi ,Igor Jurisica, Data Mining for Case Based Reasoning in High-Dimensional Biological Domains, IEEE transactions on knowledge and data engineering, vol. 17, page no.8,august 2005.

[9] Nguyen Hoang Phuong, Dang Huu Hung, Nadipuram R. prasad , Development of an Expert System for Lung diseases using fuzzy logic, Journal of Biomedical Soft Computing and human Sciences (BMFSA), Vo1.5, Page No.2, June 2001.

[10] Valerie Ficet-Cauchar, Christine Porquet, Marinnete Revenu, An interactive Case Based reasoning system for the development image processing application Journal of Springer Verlag Berlin heidelbelrg, volume 1488, page 437-447, 1998.

[11] Morten Grimnes and Agnar Aamodt A two layer case-based reasoning architecture for medical image understanding Journal of Springer Verlag Berlin heidelbelrg, volume 1168, page 164-178, 1996.

© 2012, IJARCSSE All Rights Reserved