World Applied Sciences Journal 29 (Computer Sciences, Engineering and Its Applications): 86-90, 2014
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2014.29.csea.2237

Data Mining a Boon: Predictive System for University Topper Women in Academia

R. Udayakumar, K.P. Kaliyamurthie, Khanaa and K.P. Thooyamani

School of Computing Science, Bharath University, Chennai-73, India

Abstract: This paper describes the PRADNYAVATI system framework, which is a predictive system for extracting the scholars in the University. This predictive system not only identifies the future scholars but it also helps to identify the performance-oriented professional inclination among students for the nation building. It also answers the traditional question, which requires extensive hands on analysis of the data, such as predicting the promising students, finding the success rate of female students, predicting failures in various streams. The past data inferences help in decision making for grooming as well as professional training to the current students. As a result, the placement activities will get help in selecting the promising candidates from all disciplines by smoothening and streamlining the recruitment process. The system is like a "time machine" which helps to predict the success parameters for future women empowerment. PRADNYAVATI uses the Data Mining Engine to find the inference, extract the rules from University database and forms the Rule base for further use. Rule Based reasoning and Case Based reasoning are combined to predict the women toppers.

Key words: Data mining · Predictive system · Rule Based reasoning · Case Based reasoning · University topper · PRADNYAVATI

INTRODUCTION

Data mining is a powerful technology, which has great potential that can be utilized to analyze and predict student's academic performance to help Academia using student data. This paper exhibits a framework for extracting the scholars in the University. Due to social pressures, usually in later life the women toppers sidetrack their careers and give importance to their families. A decade later most of the women toppers have gone into oblivion. It is a great loss for the nation because we are loosing the good qualities of women workforce like hard work, patience, decision making in critical situations.

The technology suggested automates the process of finding predictive information in a large university database. The traditional query and reporting tool will be used to describe and extract knowledge from the conventional database. A non-linear predictive model that learn through the training and rule induction will be used to discover the required result form the suggested system.

In subsequent sections we describe the concept of PRADNYAVATI system with its framework. The Data Mining Engine, which is responsible to extract the rules from database, is described in section 2.1. A discussion on Rule based reasoning and Case based reasoning is presented in section 2.3.1 and 2.3.2 respectively, followed by the concluding remarks.

System Concept: Predictive system can be well developed using Data Mining, Rule Based reasoning (RBR) and Case Based reasoning (CBR). The rules in the form of IF-THEN are formed for predicting promising students. Case based reasoning is more prominent for predicting the future performance of the students as the previous history and records are stored in the case base. Facts of the successful students are stored in detail to represent a case. System gives the predictions after analyzing and comparing of partial decisions of Rule based reasoning and Case based reasoning.

Corresponding Author: R. Udayakumar, School of Computing Science, Bharath University, Chennai-73, India.

Data Mining Engine: The primary task of the Data-Mining Engine is acquisition and representation of domain specific knowledge in the form of rules. This engine mines the database and finds the interesting patterns in the data and converts it into rules. It mines the knowledge such as student can achieve the target; student can work hard to achieve the target and no ability to achieve the target. This muti categorical response will help us to predict the success of the candidate. The system has its own grading, which uses education levels such as pre-primary, primary, secondary and graduation results. In gradation system we use percentage criteria and special parameters are also considered for finding the woman topper in university database. These parameters such as students interest in learning, grasping power, confidence level, goal of the student in life, physical and mental health, analytical thinking ability, seriousness in studies, extra curricular activities interest and family background etc. The detailed discussion is not covered in this paper.

Basic Ideas and Issues in RBR and CBR: RBR and CBR are regarded as two fundamental and complementary reasoning methods in the present state-of-the-art of knowledge-based systems. Rules often have unspoken qualifications and exceptions. Therefore one may need to go beyond rule-based knowledge, particularly to cases, which are a natural means of expressing those exceptions.

Reasoning: Reasoning is the process of thinking about something in order to make a decision [Cambridge Dictionary]. An expert facing a new problem is usually reminded of similar situations, recalls their results and perhaps the reasoning [1]. There are two main methods to reach a conclusion, top-down (or deductive) method and bottom-up (or inductive) methods [2].

RBR: In conventional RBR, both common sense knowledge and domain expertise are presented in forms of plausible rules (e.g. IF *<precondition(s)>* THEN *<conclusion(s)>).* This is based on a generalized relationship between problem description and conclusion. For example, an instance of a particular rule:

IF *(Seeta has secured A grade in pre-primary)*
AND *(Seeta has secured A grade in primary)*
AND *(Seeta has secured A grade in secondary)*
AND (*Seeta has secured A grade in graduation)*
THEN *(Seeta's Educational performance is consistent)*

Moreover, RBR requires an exact match on the precondition(s) to predict the conclusion(s). This is very restrictive, as real-world situations are often fuzzy and do not match exactly with rule preconditions. Thus, there are some extensions to the basic approach that can accommodate partial degrees of matching in rule preconditions.

The structure of RBR systems exhibits at least two components: the *knowledge base* (here, the *rule-base)* and the *inference engine.* The rule base contains the domain knowledge in the form of rules. The inference engine embodies the reasoning strategy for searching the rules, in a knowledge base, which enables to find an appropriate prediction for a particular student.

Apart RBR, current AI researchers have focused on how to deal cases and how to develop knowledge-based systems that use previously solved problems to perform CBR.

CBR: A case usually denotes a problem situation [3, 4]. It is a contextualized piece of knowledge, which comprises problem, solution and outcome [5]. Reasoning by re-using past cases is a powerful and frequently applied way to solve problems for humans [3, 4]. We, humans are robust problem-solver; we routinely solve hard problems despite limited and uncertain knowledge and our performance improves with experience [6].

In CBR, an *inductive reasoning* approach, is based on a *memory-centered cognitive model* in which past experiences can be remembered and adapted to guide problem solving [1]. Knowledge in CBR is knowledge representation (vocabulary) used, cases themselves, similarity metric used in identifying cases to be reused (that is retrieval) and the mechanism for adapting solutions (that is adaptation) [7].

It is a *methodology* for solving problems and is commonly described by the CBR-cycle that comprises four activities [5]:

- *Retrieve* similar cases to the problem description.
- *Reuse* a solution suggested by a similar case.
- *Revise* or adapt that solution to better fit the new problem if necessary.
- *Retain* the new solution once it has been confirmed or validated.

The basic principle of CBR systems is that of solving problems by adapting the solution of similar problems solved in the past. This is fundamentally different from classic rule-based systems, which require the formalization of all elementary knowledge in IF <*precondition(s)*> THEN <*conclusion(s)*> rules. The use of such knowledge alone involves mechanisms of deduction that are not always good simulations of expert reasoning. When solving a problem, to rely on the theory of particular domain, which may well, be expressed in rules. On the other hand, situations encountered in the past can be considered to solve and analyze.

Therefore, CBR is a decision-support method based on the idea of finding past cases most similar to the current problem in which decision must be made. Thus, CBR systems maintain a database of cases (here, case-base) related to the topics under consideration. When a new situation arises, the CBR system identifies and retrieves relevant cases, which already exist in database such facts, provide important clues or directing prediction to the current situation. Hence, the reasoning architecture of a CBR system has broadly consists of two major components: the *case-base* and an *inference-cycle.* Case representation scheme is a function of the case-size and the complexity of the case description [8].

When the number of cases in a case-base is very large, it is important to partition the case-base for efficient retrieval. Moreover, CBR raises a variety of research issues, which researchers [9] are addressing.

Knowledge Representation: If the given facts of a new case satisfy the preconditions of a rule, we can draw conclusions by using that rule. However, in actual cases, it is difficult to acquire rules for all possible eventualities and there is no reason why one should actively look for cases that do not overlap with the rules necessary to represent the previous case reports in some form that can be manipulated by programs [10]. Unlike the situation for rules, there is no standard representation for cases. In our work, cases are viewed as a collection of facts of successful female student.

Prediction

Rule-Based Prediction: System can produce comprehensive prediction using the rule base if any of the rules from the rule base triggers in the usual way, as in an expert system. When no rule is triggered by the facts of the new case, system can identify the rules that come closest to triggering. The identification depends on a score, which would be unity for a rule whose preconditions were all true for a given input [11]. Using the rules that come closest to triggering, system can generate some partial rule-based prediction for the new case. The partial prediction consists of the conclusion that would have followed if all the preconditions of the relevant rule(s) have been true, plus information focusing on the failed precondition (i.e., reasons why the conclusion cannot be accepted without reservations). The first step in generating partial prediction is to identify the suitable rules, which are nearly triggered as a consequence of the facts of the new case. A weight-based scoring mechanism is used to determine which rules are closest to triggering. The scoring formula for a particular rule Ri is:

$$\text{Score } R_i = \sum W_i * P_i / \sum W_i$$

if feature matches with precondition of rule Ri then $P_i = 1$ otherwise $P_i = 0$ W_i = relative weight for preconditions P_i.

Partial prediction is generated from the rules with the highest scores according to the equation. For each of these rules, a justification is provided describing the conclusion that would have been certain if it had triggered and which parts of the rule did or did not match. PRADNYAVATI provides no rule-based prediction when no rule from the rule-base is triggered and the scores of all rules are less than an appropriate threshold.

Case-Based Prediction: The case-base of PRADNYAVATI consists of two parts: case base, which serve as a repository for cases and a set of access procedures.

First Step: In the first step, PRADNYAVATI filters out the relevant cases from its case base then it selects all cases with respect to related features of the student being examined.

Second Step: The second step is for an assessment of similarity between the new case and the selected cases from the first step. The similarity is computed by matching of features of the cases. The similarity assessment is based on features, which defines the case in database.

Conceptual distance (D) is computed between selected cases. The measure of similarity taken over the cases in the case base is always below a threshold of Td.

Third Step: In the third step the system assesses similarity further by comparing the details PRADNYAVATI calculates the absolute distance between the cases by using a metric on the values of features that characterize them. Then up to five cases with the lowest distances from the current case, with respect to this metric, are selected for analysis and prediction purposes.

Analysis and Prediction: The system produces comprehensive prediction provided that at least one of the rules from the rule base has triggered. If none of the rules of the rule base is triggered, then the system can still suggest some partial or tentative prediction. In CBR, the system refers to the features of previously decided cases to identify the most similar cases to the situation that a user describes in input and produce the result.

Related Research: DENDRAL was the first expert system developed to interpret the mass spectrum of organic molecules [12]. The earliest and most successful rule-based expert system was MYCIN, which incorporated about 400 heuristic rules like IF-THEN to diagnose and treat infectious blood disease. After MYCIN many expert systems are developed [13]. Janet Kolodner, at Yale University, developed the first system of CBR, known as CYCRUS. It was a question-answering system with knowledge of the various travels and meetings of former US Secretary of State of Cyrus Vance [14, 15]. Almost every author of CBR cites Janet Kolodner's work on CBR. An overall view of CBR is provided by Aamodt & Plaza [4], Watson [5] Bradley [16]. Angi has presented survey on case adaptation [17]. Ashwin Ram, David Leake has also contributed in CBR. In short numerous works are available on CBR with varied applications.

Self-Reproducing Learning Mechanism is a type of learning technique which was developed from Intelligent Predictive Systems. It is very useful to be applied to Lotto Predictive System to increase the performance of prediction. Lotto Predictive System is also based on Human Super Nature Beliefs as part of its knowledge [18] Annie Ying developed an approach that also uses association rule mining on CVS version archives [19]. She especially evaluated the usefulness of the results, considering a recommendation most valuable or "surprising" if it could not be determined by program analysis and found several such recommendations in the MOZILLA and ECLIPSE projects [20].

CONCLUSION

This paper presents a framework for PRADNYAVATI system for mining the woman topper. To develop a frame work, we analyzed the case based, rule based, forecasting/predictive systems. In this, we are using gradation system along with the special parameters, to predict the success of the candidate. Some times it is difficult to acquire rules for the eventualities then for prediction we use the case based reasoning, where facts are stored in the database. The proposed framework may have several applications. It can be used to data mining course to clear the students perspective of data mining and predictive system research.

REFERENCE

1. Xu, L.D., 1995. Case based reasoning : A major paradigm of artificial intelligence, IEEE Potentials, 13(5): Dec 1994-Jan 1995, pp: 10-13.
2. Fatemeh Zahedi, 1993. Intelligent systems for Business: Expert System with Neural Network, Wadsworth publishing Company.
3. Agnar Aamodt, 1994. Knowledge representation system for integration of general and case-specific knowledge, Proceedings of Sixth IEEE International Conference on Tools with Artificial Intelligence, 1994, 6-9 Nov. 1994, pp: 836-839.
4. Agnar Aamodt and Enric Plaza, 1994. Case-Based Reasoning: Foundation Issues, Methodological Variations and System Approaches, AI Communications, 7(1): 39-59.
5. Ian Watson, 1999. Case-based reasoning is a methodology not a technology, Knowledge-Based Systems, 12(5-6): 303-308.
6. David B. Leake, (editor), 1996. Case-Based Reasoning: Experiences, Lessons and Future Directions" AAAI Press/MIT Press.
7. Cunningham, P. and A. Bonzano, 1999. Knowledge engineering issues in developing a case-based reasoning application, Knowledge-Based Systems, 12: 371-379.
8. Thooyamani, K.P., V. Khanaa and R.Udayakumar, 2013. Blue tooth broad casting server, Middle-East Journal of Scientific Research, ISSN:1990-9233, 15(12): 1707-1712.
9. Slade, S., 1991. Case-Based Reasoning: A Research Paradigm, AI Magazine, 12(1): 42-55.

10. Saravanan, T., V. Srinivasan and R. Udayakumar, 2013. A Approach for Visualization of Atherosclerosis in Coronary Artery, Middle-East Journal of Scientific Research, ISSN:1990-9233, 18(12): 1713-1717.
11. Bradley P. Allen, 1994. Case-based reasoning: business applications, Communications of the ACM, 37(3): 40-42.
11. Udayakumar, R., V. Khanaa and K.P. Kaliyamurthie, 2013. Optical Ring Architecture Performance Evaluation using ordinary receiver, Indian Journal of Science and Technology, ISSN: 0974-6846, 6(6): 4742-4747.
12. Lindsay, R.K., B.G. Buchanan, E.A. Feigenbaum and J. Lenderberg, XXXX. Applications of Artificial Intelligence in Organic chemistry: The DENDRAL Project, McGraw Hill
13. Joe Gallacher, 1989. Practical introduction to expert systems, Microprocessors and Microsystems, 13(1): 47-53.
14. Janet Kolodner, 1983. Maintaining organizations in a dynamic long-term memory, Cognitive Science, 7(4): 243-280.
15. Janet Kolodner, 1983. Reconstructing memory, a computer model, Cognitive science, 7(4): 281-328.
17. Angi Voss, 1996. How to solve complex problems with cases, Engineering Applications of Artificial Intelligence, 9(4): 377-384.
18. James Kuodo Huang, 2005. Self-Reproducing Learning, Data Mining And Intelligent Predictive Systems, IEEE, pp: 159-162.
19. A. T.T. Ying, 2003. Predicting source code changes by mining revision history, Master's thesis, University of British Columbia, Canada.
20. Thomas Zimmermann, Peter Weißgerber, Stephan Diehl and Andreas Zeller, 2004. Mining Version Histories to Guide Software Changes, ICSE.