ONTOLOGY-BASED LEARNER CATEGORIZATION THROUGH CASE BASED REASONING AND FUZZY LOGIC

Sohail Sarwar[1], Raul García-Castro[2], Zia Ul Qayyum[3], Muhammad Safyan[1] and Rana Faisal Munir[4]

[1] *Department of Computing Iqra University Islamabad,Pakistan*
[2] *Universidad Politecnica de Madrid (UPM), Spain*
[3] *University of Gujrat Gujrat, Pakistan*
[4] *University Politecnica de Catalunya, Barcelona, Spain*

ABSTRACT

Learner categorization has a pivotal role in making e-learning systems a success. However, learner characteristics exploited at abstract level of granularity by contemporary techniques cannot categorize the learners effectively. In this paper, an architecture of e-learning framework has been presented that exploits the machine learning based techniques for learner categorization taking into account the cognitive and inclinatory attributes of learners at finer level of granularity. Learner attributes are subjected to a pre-processing mechanism for taking into account the most important ones out of initial attribute set. Subsequently, couple of machine learning techniques namely Fuzzy Logic and Case Based Reasoning was employed on attributes selected for learner categorization. To best of our knowledge, these techniques have not been employed so far in learner categorization with quality of data and adaptivity while targeting semantic web.

KEYWORDS

E-learning, Learner Categorization, Case based Reasoning, Fuzzy Logic

1. INTRODUCTION

Internet has redefined every aspect of human life and hence the methods of educating the learners. Phenomenon of e-learning has greatly prevailed through ubiquity and universality in educating diverse communities of knowledge. Despite of all the benefits, e-Learning needs to prevail much in addressing content personalization for the learners (Sarwar, 2016). Learner is one of the key stake holders in an E-learning system along the Instructor and System Administrator. Instructor, an educationist with supervisory role, designs the learning contents, exercises/assignments and exams to educate and assess the learners.

Learner on the other hand, consumer of learning contents, undergoes the learning cycle of learning, assessments and corrections to master certain course(s). System Administrator, with a role of facilitator, harnesses the platform for instructor and learner in performing their respective roles. An e-learning system with its stakeholders is illustrated in Fig. 1 where focus of our work is pertinent to "Learner" (i.e. "Learner Categorization"). Learner categorization through learner profile aids in personalized recommendation of learning contests and subsequent adaptivity of these contents. Our current work focuses on learner categorization since typical techniques may not completely consider both academic (CGPA, Pre-Requisite score, Pre-Test score) taken implicitly and cognitive characteristics (learning style, aptitude and age) acquired explicitly for learner categorization at finer level of granularity. Few learner categorization techniques, after categorizing the learners, do not retain current information of categorization to take advantage of reusing information for future classifications. Others lack in having comprehensive set of axioms in categorizing the learners rightly. Lastly, few techniques claim to target the semantic web but explicit description of domain i.e. ontologies seem missing.
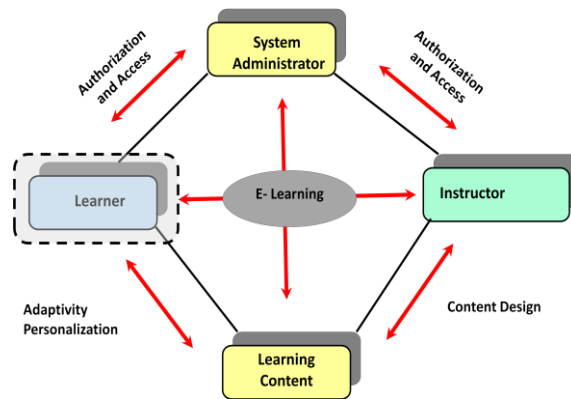
Figure 1. Stakeholders of-Learning System

Keeping above in view, phenomenon of machine learning techniques is employed for learner categorization that targets e-learning systems by modeling learner profile through ontology. It is dynamic enough to build learner's profile automatically with implicit parameters from real time data sources and explicit parameters acquired from the learner. Profile of learner is modeled by considering academic and cognitive aspects of learner using "LearnerOntology" coupled with "LearningContentOntology" to benefit from underlying technologies of semantic web. Once profile of learner is built, Case Base Reasoning (CBR) conjuncted with Majority vote classification (Agnar, 1994; Sankar, 2004) and Fuzzy Logic (FL) (Ying, 2004) is used to categorize the learners as illustrated in figure 2. Learners are categorized into one of the categories of 'Novice', 'Easy', 'Proficient' or 'Expert' based on their profiles. These learner categories have been introduced after consulting seasoned educationists, psychologists and literature (Agnar, 1994; Sankar, 2004; Thakaa, 2014; Ying, 2004).
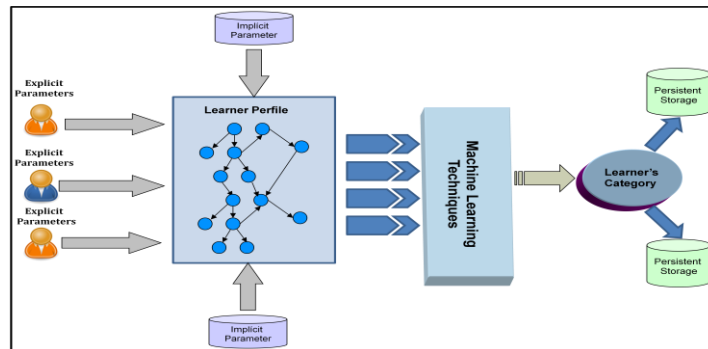

Figure 2. Learner Categorization using Machine Learning Techniques

The rest of the paper is organized as given in the follows: section 2 provides an overview of state of the art followed by section 3 describing the architecture. Section 4 elaborates implementation details followed by directions to for results and evaluations.

2. LITERATURE SURVEY

Educational Data Mining (Shute, 2010) termed as an emerging discipline is claimed to have a great room for developing methods and exploring unique types of data that come from educational settings. Using these methods has potential to facilitate better understanding of contents for students. Data mining techniques (Minaei, 2003) are employed for formative assessment of learners in order to provide a way for classifying the slow learners by identifying relation between academic achievements and his behavior in course of English language. Evaluation of learner is carried out through modes of listening, speaking and writing that helps in respective classification of learners.

Data mining techniques (Romero, 2007) have been used to predict failure ratio of students in two classes (Portuguese and Mathematics) while exploiting 25-29 predictive variables. Support Vector Machine, Decision Tree, Neural Network and Random Forest were employed on student dataset comprising of 800 students who appeared in final examination. Neural Nets and Decision Tree algorithms showed a predictive accuracy of 91% and 93% respectively for two-class dataset (pass/fail).

Any e-learning system has three mandatory components as suggested by experts of educational psychology; the content model (domain model), the learner model (user model) and the adaptive engine (Brusilovsky, 2010). Here adaptivity of learning contents based upon learner profile is discussed for recommending suitable contents.

3. OVERVIEW OF IMPLEMENTATION APPROACH

Modular architecture of proposed approach is presented in figure 3. There are three modules namely: Learner Ontology, Case base Reasoning (CBR) and Fuzzy Logic (FL).
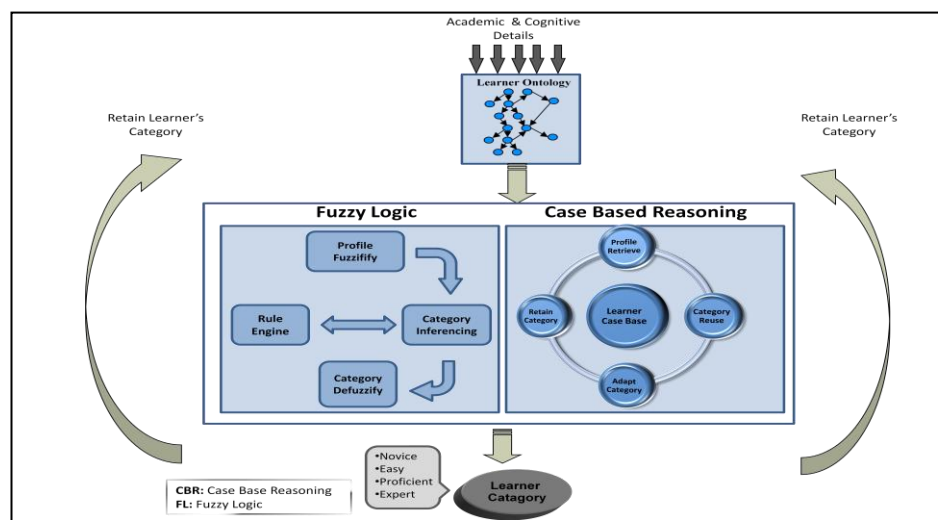

Figure 3. Proposed Architecture for Learner Categorization

3.1 Case Base Reasoning and Neural Networks

Case based reasoning targets to resolve the problems based on prior knowledge maintained in case base. Whenever new learners were enrolled in certain course, their profile was created by taking their personal details and ones pertinent to their aptitude and academics. Based upon this information, each learner was assigned a category reference to his cognitive strengths i.e. easy, novice, proficient and expert. This category was maintained along with rest of the profile details of learner in a repository. This repository serves as a "Case base" for our CBR model that not only plays a key role in categorizing new learners but is evolving over time. Phenomenon of how new learner is assigned a category using our CBR model is elaborated below:

Case Retrieval: provides query specific solution given the profile attributes of new learner (query case). Level of similarity is computed for the 'query case' vs 'cases in the case base'. This similarity index is computed using 'Tversky Ratio model (Sankar, 2004)' among query case and ones in case base. If cases retrieved from case base appear with exact similarity i.e. learner attributes in query case and cases in case base are same then new learner is assigned same category as that of similar learner in case base (termed as Reuse in CBR).

On the other hand, if retrieved cases are not exactly similar but similarity index falls between thresholds of 60 %-90 %), Case Adaptation is triggered (It is also called Revision). There may be another scenario where multiple cases are retrieved falling within stated range. Here decision of which case to adapt is made on the basis of 'Rank' assigned to each of the retrieved case through similarity index.

Case Revision aids in provision of possibly nearest solution to assign a category to certain learner, if exact match for new learner case is not found. Case adaptation is carried out through 'Majority Voting Classifier (MVC)'.

In MVC, occurrences of certain solutions are considered among the retrieved cases for classifying a certain learner. The learner category having a maximum number of occurrences is considered as the category of the new learner. In other words, the value of the nth element is considered for selecting the most probable candidate. For example, if the case retrieval process returns 10 cases (each case corresponding to 10 learners); 4 with category 'easy', 3 with category 'proficient', 2 with category 'novice' and 1 with category 'expert'; the category 'easy' is assigned to the new case (learner).

3.2 Fuzzy Logic

Fuzzy logic can be considered as knowledge-based systems incorporating human knowledge into their knowledge base through fuzzy rules and fuzzy membership functions (Ying, 2006) by manipulating the linguistic data of learner such as ("Novice", "Easy", "Proficient" and "Expert"). This module exploits the "Fuzzy Control Logic (FCL)" in order to categorize the learner.

Whenever a new learner comes in, input variables (feature attributes selected) corresponding to learner's profile are fed to the FL model in crisp form scaled over a numeric range. For example PreTestScore is an input variable with four ranges for Fuzzification through membership function i.e. poor (0-1.9), fair (2-4.9), good (5-7.9) and very good (8-10). These variables are fuzzified using the "Gaussian" membership function and represented in fig 3.

The Rule base of the fuzzy logic model aids in deciding the category of the learner. The knowledge required for the reasoning purpose is greatly dependent upon rules in the rule engine. Few of these rules (if-then-else) are given in the following:

> RULE 1 : IF PreTestScore IS poor OR CGPA IS fair OR LearningStyle is belowAverage THEN LearnerCategory IS novice;
>
> RULE 2 : IF CGPA IS average OR PreTestScore IS fair OR LearningStyle is average THEN LearnerCategory IS easy;

After rule engine yields certain value for the learner, it needs to be transformed into human-understandable format i.e. defuzzification. "Center Of Gravity" method is used to defuzzify the output of rule inference engine with other options of weighted average (Dipiti, 2001) and singleton methods.

Another important aspect of architecture is the representation of data pertinent to learner's profile. Since goal of presented system is to serve as component of semantic web based e-learning system, so learner's profile is maintained in ontology along with learning contents to benefit technologies of web 3.0. The learner's profile is modeled in the "Learner ontology". This ontology has been developed in a semi-automatic fashion where some of the concepts have been acquired implicitly from institute's repositories and others were incorporated manually in consultation with domain expert. The concepts in the learner ontology are envisaged to have reason among attributes learners, instructors and course contents through standard properties (associative, reflexive or transitive) or user-defined predicates. For example, if student 'A' has specialization of 'AI' and Instructor 'I' is teaching course of 'AI' the 'I' is likely to be supervisor of

'A'(assuming student and instructors have 1-1 relation).

4. EVALUATION AND DISCUSSION

In order to evaluate the proposed techniques, profiles of 400 students from different institutes and universities were used. The input for the evaluation of the given techniques consisted in four sets of new learners' profiles each having 20 profiles of learners. These profiles were subjected as input to all the ML models of CBR, and Fuzzy Logic for evaluating performance of ML techniques in terms of accurately categorizing the learners. The degree of accuracies exhibited by two machine learning techniques has been furnished in table 1.

Table 1. Percent Comparison of Accuracy in Learner Categorization

Technique	FL	CBR
Average (%)	49.67	67.35

In order to compare and analyze the accuracy of recommendations made by the CBR and Fuzzy Logic keeping in view the profiles of the learners. Domain experts have also suggested the categories given the profiles of the learners. Kappa coefficient (Sim, 2005) has been used to assert the relationship among recommendations by machine learning techniques and domain expert (DE).

An average of contents recommended by domain experts was taken as shown in table 2. This average was used alongside the contents recommended by CBR and FL for calculating the Kappa's coefficient. These results assert that CBR has better performance than FL due to its capacity of utilizing the profiles in the case base and dynamic nature of adaptive technique i.e. MVC. On the other side, FL has static rule base whose performance may be improvised with dynamic manipulation of if-else rules in fuzzy inference engine.

Table 2. Kappa Coefficient based Comparison of Accuracy in Learner Categorization

Set of Learner Profiles	% Recommendations		Accuracy Validation by DE		Kappa 's Coefficient
	FL	CBR	DE 1	DE 2	
Set 1	15	13	72%	81%	74%
Set 2	9	14	83%	77%	79%
Set 3	7	11	80%	85%	81%
Set 4	8	16	83%	68%	72%

5. CONCLUSION

Learner categorization targeted for e-learning systems is carried out through couple of machine learning ML techniques in this work. A comparative analysis for deciding the best one among Fuzzy Logic and Case Based Reasoning. CBR module uses similarity metrics in retrieving the relevant cases from the case base. The similarity metrics used with CBR seem trivial and static. So, different similarity metrics such as clustering or fuzzy logic would be employed to experiment unsupervised and supervised techniques for dynamic retrieval of relevant cases.

REFERENCES

Agnar A., Enric P., (1994), *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. AI Communications. IOS Press USA, pp. 39-59.

Brusilovsky, P., (1996), Methods and techniques of adaptive hypermedia, Journal of user modeling and user-adapted Interaction, Vol 6(3), pp. 87–129.

Dipiti S., Cheu L., (2001), Hybrid Fuzzy Logic-Genetic Algorithm technique for automated detection of traffic incidents on freeways, IEEE Intelligent Transportation Systems, pp 7194-7201.

Minaei B., Punch W., (2003), Using genetic algorithms for data mining optimization in an educational web-based system. *In Genetic and Evolutionary Computation, LNCS,* Vol 4(3), pp. 2252-2263.

Romero C., Ventura S., (2007), Educational Data Mining: A Survey from 1995 to 2005, *Expert Systems with Applications*, Vol. 33(1), pp. 135-146.

Sankar K., Simon C., Shiu K., (2004), *Foundations of Soft Case based Reasoning* by John Wiley & Sons, Inc.

Sarwar, S., Qayyum, Z.,Safyan, M., & Munir, F., (2016), Ontology based Adaptive, Semantic E-Learning Framework (OASEF), *Springer LNEE ICISA*, vol 376, pp. 1199-1212.

Shute V., & Towle B., (2010), Adaptive e-learning, *Educational Psychologist*, Vol. 38(2), pp. 105–114.

Thakaa Z., Mohammad A. and et al (2014), Classification Model of English Course e-Learning System for Slow Learners: Recent Advances in Information Science, *International Journal of Computer Science (IIJCS)*, Vol 4(7), pp. 474-486.

Van S., Ossevroot and et al, (2012), Influence of student characteristics on use of adaptive e-learning material, *International Journal of Computers & Education* Vol. 58(3), pp. 942–952.

Ying B., Dali W., (2006), *Fundamentals of Fuzzy Logic Control – Fuzzy Sets, Fuzzy Rules and Defuzzifications*, Advanced Fuzzy Logic Technologies in Industrial Applications.

Sim, J., Wright, C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. Physical Therapy. 85 (3): 257–268.