Classification Accuracy in Cognitive Load for Users Preference in Web Based Learning

L.Jayasimman
Assistant Professor
Dept. of Computer Application
J J College of Eng. and Tech. Trichy

E.George Dharma Prakash Raj
Assistant Professor
Computer Science and Eng.
Bharathidasan University
Trichy

ABSTRACT

With increasing popularity of web based learning, it is required to design the web layout to reduce cognitive load. Cognitive load theory is widely used to predict the effectiveness of the web based and multimedia learning. The cognitive load induced by instructional and multimedia modes are measured by indirect or subjective methods. Questionnaires are one common form of measuring cognitive load indirectly. In this paper, a questionnaire is prepared to identify the cognitive load of the student and his website preferences in a web learning environment. The cognitive attributes are used as the training input for the Naïve Bayes, Classification Regression Tree(CART), Random Forest and Random Tree for classification. Based on the response of the user, areas for improvement in layout of the web learning system are identified.

Keywords

User interface design, Cognitive approach, online learning, decision tree induction Introduction

1. INTRODUCTION

Web based learning content can be in the form of tutorials, online discussion group or virtual platform for hands on training [1]. Most of the Web based learning is enhanced using multimedia followed by self assessment tools [2]. A well designed user interface of web learning improves performance of the students. In general, features like information content, font, navigability, links influence the user's ease of use and satisfaction of the website. In web based learning systems, the cognitive load is the key for the effectiveness of the instructional and multimedia learning, thus the cognitive load and the students website preferences forms the basis of user's satisfaction. The graphical user interface and multimedia formats may hinder learning by imposing unnecessary cognitive demand. Therefore, proper design of website layout of a web learning environment is crucial.

Cognitive theory of learning advocates that learning is based on mental processes [3, 4, 5]. According to cognitive theory, the learning process comes out of experience, perception, memory and overtly verbal thinking [6]. Cognitive load theory fundamentals are used to explain the cognitive process of learning in web based instructions and multimedia methods [7]. In spite of decades of research, direct measurements of cognitive load in learning situations are not established. The cognitive load is assessed by indirect methods such as questionnaires, self-reports of mental effort [8].

The cognitive load is distinguished into three types: intrinsic cognitive load, extraneous cognitive load and germane cognitive load [9]. The intrinsic cognitive load is caused by the structure and complexity of the material. The extraneous cognitive load is caused by the format and manner in which the information is presented and germane cognitive load refers to the student's effort to process and understand the information. The intrinsic cognitive load cannot be influenced, whereas the extraneous and germane cognitive load can be influenced by the instructional design of the learning material. Research literatures have worked examples for low extraneous and optimal germane load [10, 11, 12].

In this paper, it is proposed to identify the relation between the cognitive load and the student's web layout preferences. A questionnaire is prepared to identify the cognitive load of the student and his website preferences in a web learning environment. The cognitive attributes are used as the training input for the Naïve Bayes, CART, Random Forest and Random Tree for classification. Based on the response of the user, areas for improvement in layout of the web learning system are identified.

2. RELATED WORKS

Feinberg, et al., [13] applied the principles of cognitive load theory to the design an instructional web site and discovered that cognitive load theory provided a sound baseline for the design of effective web-based instruction. To effectively enhance web-based instruction, the graphical user interface and multimedia formats must be developed in consideration of cognitive load principles. A graphical user interface and multimedia formats can increase extraneous cognitive load and have a negative impact on learning. But three learning techniques have direct application in the generation of web-based instruction using multimedia technology. These three include the split-attention effect, the redundancy effect and the modality effect. Studies through a multimedia website led to four findings which proved that cognitive load theory provides general design principles to reduce extraneous cognitive load caused by the web-based format of the instruction. It also proved that a) cognitive load theory provides a sound baseline for the design of effective instruction. b) That the principles of split-attention, redundancy effect and modality effect identified in cognitive load theory had clear applications in the design of web-based instruction. c) That cognitive load theory is consistent with general web design principles and provided an additional criterion for effective design of web-based instruction and finally d) That the graphical user interface and multimedia formats must be developed in consideration of cognitive load principles to effectively enhance web-based instruction.

Jan L. Plass [14] proposed a hybrid model that combines cognitive and software engineering approaches regarding the criteria for the design and evaluation of the user interface of foreign language multimedia software. Attempts to define generally applicable design and evaluation criteria for

multimedia software led to a number of different approaches and despite their comprehensive list of criteria these approaches are not specific enough to be usable for a particular subject matter area such as SLA. Evaluation criteria need to be based on domain specific learning processes and activities and on the cognitive processes that these activities involve. Using this approach, a taxonomy of SLA software features would be based on the pedagogy or principles of adult education (andragogy) and activities and instructional methods of language learning and would address how well individual software components can facilitate them. A *Cognitive Approach* appears to be the most appropriate basis both for the design and for the evaluation of user interfaces for SLA software since it incorporates both the user and the learning task into the design. The proposed approach involves a three step design which includes selection of instructional activity that supports cognitive processes of competence, selection of feature attributes and selection of designs features. It is still pragmatic to be practical. Based on this proposal, contextualized model of interface design, domain specific evaluation criteria are developed to describe how well the user interface is able to support the cognitive processes involved in the development of linguistic and pragmatic skills and competencies in SLA.

Martin Graff [15] suggested in this study that the two issues – whether information segmentation helps in facilitating users and whether it assists in improving their learning – need to be understood effectively through understanding an individual's cognitive style. Fifty participants were assigned to one of two web-based instructional systems featuring information on the subject of psychological ethics. The information in one of the web systems was segmented to a greater degree than the information in the other. Half the participants using each web system were given an overview of the system and half were not. After a given time using the system, participants were tested on the information from the web. The design of this study was a 2 X 2 between subjects design. The conditions were web structure in terms of page length and provision or absence of an overview. Participants were randomly assigned to one of these conditions and were given approximately ten minutes to read through the structure navigating through the pages in any manner they wished. They were then requested to answer recall questions and an essay question on information contained in the web structure. The findings suggest that cognitive style and segmentation affected performance, although the provision of the overview had little effect. The results are discussed in terms of a consideration of cognitive style in the design of web-based instructional systems.

Shee, et al., [16] proposed a multi-criteria methodology from the perspective of learner satisfaction to support evaluation-based activities at the pre- and post-adoption phases of the web-based e-learning system (WELS) life cycle. WELS has emerged as a new means of skill training and knowledge acquisition, encouraging academia and industry to invest in adopting this system. Traditionally, most evaluation related pre and post adoption tasks are carried out from viewpoints of technology. As users are a key group of stakeholders in influencing adoption of information systems, their attitudes toward this are important. Hence, this study based on the theory of multi-criteria decision making and the research products of user satisfaction from the fields of human–computer interaction and information systems. The evaluation of WELS alternatives was carried out through the learner interface, learning community, system content and personalization dimensions, the former two being based on

four criteria each, the third being based on three criteria and the fourth on two. It found that WELS learners regard learner interface as the most important dimension. As IS-related studies have pointed out the user interface is an area where a high level of interaction takes place, a well-designed, user-friendly learner interface becomes one of the critical factors in determining whether learners enjoy using the WELS. The proposed methodology supplies management in education and industry with a less complex and more appropriate and flexible way to effectively analyze currently deployed WELS.

Ping Zhang, et al., [17] provided a conceptual framework and foundation for systematically investigating features in a web environment that contributes to user satisfaction with a web interface. While browsing, a user constantly redefines his/her information problem; so the web environment plays a significant role in the browsing process. Specifically, a user browses for the following reasons: to gain an overview, to monitor a process, to shift/share cognitive load, to clarify ambiguous information, to develop a formal strategy, to discover/learn, and respond to environmental invitations These actions require specific web design features that address users' needs. The research uses a questionnaire developed to identify people with internal and external locus of control. Herzberg identified two types of employees - hygiene seekers and motivation seekers, leading to the conclusion that hygiene seekers are motivated by hygiene factors. Similarly, web users with high levels of empowerment defined as self-efficacy are likely to be more satisfied by interesting and challenging tasks than those with low levels of self-efficacy. Distinguishing between these users is important in explaining users' reactions to certain website designs.

3. EXPERIMENTAL SETUP

The cognitive behavior of 100 students belongs to affiliated colleges of Bharathidasan University was captured using questionnaires. They were initially subjected to go through a known subject and an unknown subject in a popular online learning website. The questionnaire is prepared to identify the cognitive load of the student and his website preferences in a web learning environment. Total of 15 questions was presented in the questionnaire. Typical questions in the questionnaire are as follows:

- How was the mental demand to understand the content of the website?

- As you were learning the content of the website, were your eyes diverted constantly by other content in the same web page?

- Were you annoyed by the way the content was organized in the web page?

- Were you frustrated that you could not understand all the content presented?

The questions were rated by the students in a scale of $1 – 5$, 1 being the lowest score and 5 being the highest. The rating is as follows: 5-Very high, 4-high, 3- Normal, 2-low, 1-very low. Class label indicating the score is assigned to all the 100 students. The cognitive attributes are used as the training input for the Naïve Bayes, CART, Random Forest and Random Tree for classification.

Naïve Bayes

Naive Bayes is one of the most efficient learning algorithms in data mining and has competitive performance in

classification. The effectiveness of Naïve Bayes stems from the conditional independence assumption. A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. Given a set of variables, $X=\{x_1,x_2, . . . x_n\}$, constructing the posterior probability from a set of possible outcomes $Y=\{y_1,y_2, . . .,y_n\}$. Using Bayes rule

$$P(Y_j|x_1,x_2. . .,x_n) = p(x_1,x_2,. . .,x_n|C_j)p(C_j)$$

Since Naive Bayes assumes that the conditional probabilities of the independent variables are independent that it can be decomposed to

$$p(X \mid Y_j) = \prod_{k=1}^{n} p(x_k \mid Y_j)$$

The posterior can be rewritten as

$$p(Y_j \mid X) = p(Y_j) \prod_{k=1}^{n} p(x_k \mid C_j)$$

Using Bayes' rule above, a new case X is labeled with a class level Y_j that achieves the highest posterior probability.

Classification and regression trees are decision trees used for classifying or predicting. The decision trees use layered approach for classification where each node in the tree represents a decision. The splits at the nodes correspond to all possible outcomes. The splits segregate the dataset into several subsets, which are identified by the leaves in the tree. A leaf gives the value of the categorical attribute of an instance. Most of the decision trees are based on if-then rules [18].

An instance to be classified is presented at the root of the tree, at each successive level the attributes in the instance are matched to mutually exclusive nodes. Finally, the leaf it reaches assigns the class to the instance. The sequence of decision or formation of nodes in the tree is determined during the training period.

CART

The CART algorithm was developed by Breiman, Friedman, Olshen, and Stone. CART is a binary decision tree, constructed by splitting a node into child nodes. This process is repeated beginning with the root node that consists of whole learning sample.

Random Tree

Random tree is a decision tree which uses K randomly selected attributes at each node. Random tree allows for class probabilities based on backfitting with no pruning [19]. The steps involved for growing random tree are:

1. A data set is created from the training dataset by sampling with replacement; the number of instance is same as the training dataset. So the new dataset contains duplicate examples.

2. The nodes and leaves are built using tree building algorithms by choosing a random number of attributes.

3. Each tree is grown fully without pruning.

Random Forest

Breiman [20] proposed an ensemble of classifiers called Random forest for classification and prediction. In random forest, multiple decision trees are built using bagging techniques to form subsets of randomly selected attributes from the original dataset. The splitting features are selected semi-randomly. A test instance is classified by combining the results of all the trees in the forest. The pseudo code for random forest algorithm is shown in Figure 1 [21]:

```
To generate c classifiers:
for i = 1 to c do
    Randomly sample the training data D with replacement to produce Di
    Create a root node, Ni containing Di
    Call BuildTree(Ni)
end for

BuildTree(N):
if N contains instances of only one class then
    return
else
    Randomly select x% of the possible splitting features in N
    Select the feature F with the highest information gain to split on
    Create f child nodes of N, N1,...,Nf, where F has f possible values (F1, ... ,Ff)
    for i = 1 to f do
        Set the contents of Ni to Di, where Di is all instances in N that match
        Fi
        Call BuildTree(Ni)
    end for
end if
```

Figure 1: Pseudo code for Random forest algorithm

The distribution of answers for some of the queries is given in Figure 2. This research tries to identify the cognitive load of the student based on the questionnaire and his website layout preference in a web learning environment.
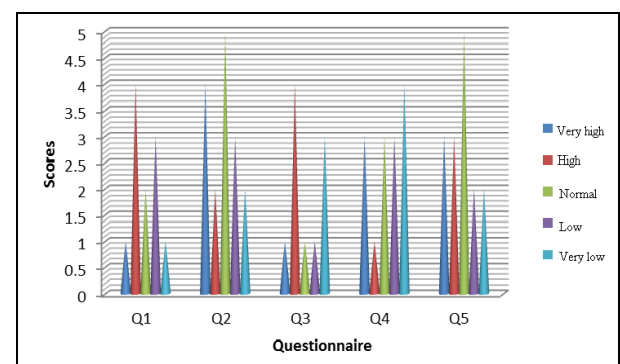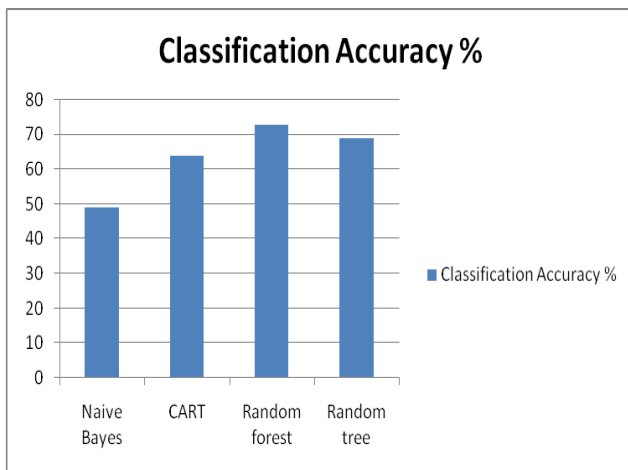

Figure 2: Frequency of answers for some of the questions.

The classification accuracy of the various classifiers for the cognitive input is tabulated in Table I and shown in Figure 3.

Table I: Classification Accuracy for various classifiers

| Technique Used	Classification Accuracy %
Naive Bayes	49
CART	64
Random forest	**75**
Random tree	69

Figure 3: Classification Accuracy for various classifiers

Figure 4 gives the root mean squared error (RMSE). RMSE is an error metric which measures the difference between the predicted values and the actual values. RMSE is a good measure of accuracy, it is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_i - x_i\right)^2}$$

Where y_i is the predicted value and x_i is the actual value. It is evident from Figure 4 that lower the RMSE, higher the accuracy obtained.
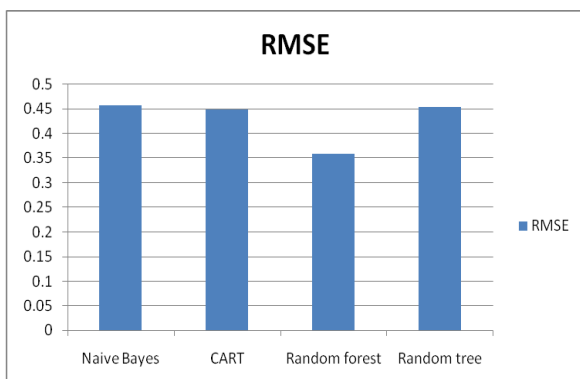
Figure 4: The Root Mean Squared Error

Other classification metrics used are precision, recall and f Measure. The precision, recall and f Measure are computed as follows:
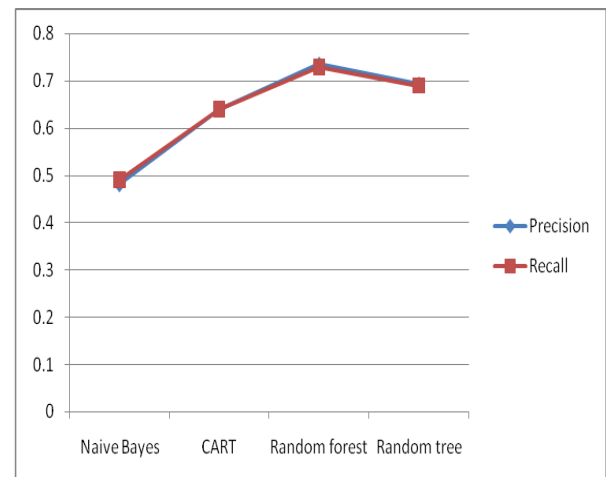
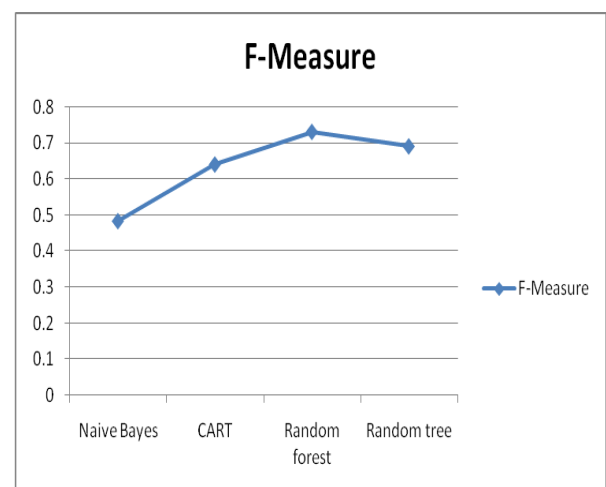$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$f\ Measure = \frac{2*recall*precision}{recall + precision}$$

Where TP is the true positives, FN is false negative and FP is false positives.

Figures 5 and 6 show the precision, recall and f Measure respectively.

Figure 5: Precision and Recall

Figure 6: F Measure for various classifiers

4. CONCLUSION

In this paper, a questionnaire is used to identify the cognitive load of the student and his website layout preference in a web learning environment. The cognitive behavior of the user is captured through questionnaire. The questionnaire helps identify the areas for improvement in layout of the web learning system which is used as the class label for the decision tree algorithm based on CART, Random forest, Random tree and the Naïve Bayes classifier. Random forest classifier predicts accurately with an accuracy of 75%.

5. REFERENCES

[1] Cook DA, Dupras DM. A practical guide to developing effective webbased learning. J Gen Intern Med 2004;19:698–707

[2] Clark RC, Mayer RE. E-learning and the science of instruction. San Francisco: Pfeiffer, 2003.

[3] Nathoo AN, Goldhoff P, Quattrochi JJ. Evaluation of an interactive case-based online network (ICON) in a problem based learning environment. Adv Health Sci Educ Theory Pract 2005;10:215–30

[4] Chen, S. Y. & Macredie, R. D. (2004). Cognitive Modeling of Student Learning in Web-Based Instructional Programs. *International Journal of Human–Computer Interaction*, 17(3), 375–402

[5] Chen, S. Y. (2002). A cognitive model for non-linearlearning in hypermedia programmes. *British Journal of* Educational Technology, 33(4), 449-460

[6] Chen, S.Y. , Magoulas, G. D. & Dimakopoulos, D.(2004). Cognitive styles and users' responses to structured information representation. *Journal of the American Society for Information Science and Technology*, 56(1), 70–83

[7] Mayer, R. E. (2001). Multimedia learning. New York: Cambridge University Press.

[8] Paas, F. G. W. C., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. Human Factors, 35, 737–743.

[9] Sweller, J., van Merriënboer, J. J.G ., &Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251–296.

[10] Kalyuga, S., Chandler, P., Tuovinen, J., &Sweller, J. (2001). When problem solving is superior to studying worked examples. Journal of Educational Psychology, 93, 579–588.

[11] Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 38, 43–52.

[12] Cooper, M., Tindall-Ford, S., Chandler, P., &Sweller, J. (2001). Learning by imagining. Journal of Experimental Psychology: Applied, 7, 68–82.

[13] Susan Feinberg and Margaret Murphy; Applying Cognitive Load Theory to the Design of Web-Based Instruction; Usability Testing and Evaluation Center, Illinois Institute of Technology, Illinois. US.

[14] Jan L Plass; Design and Evaluation of the User Interface of Foreign Language Multimeda software: A Cognitive Approach; Language Learning & Technology, Volume 2, Number 1, July 1998.

[15] Martin Graff; Learning from Web based instructional systems and cognitive style; British Journal of Educational Technology, Vol. 34, No. 4, 2003.

[16] Daniel Y. Shee and Yi-Shun Wang; Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications; Department of Information Management, National Changhua University of Education, Taiwan.

[17] Ping Zhang, Gisela M. von Dran, Ruth V. Small, Silvia Barcellos; Websites that Satisfy Users:A Theoretical Framework for Web User Interface Design and Evaluation; Proceedings of the 32nd Hawaii International Conference on System Sciences – 1999.

[18] Mitchell, Tom M. (1997), *Machine Learning*, McGraw-Hill.

[19] Frederick Livingston, "Implementation of Breiman's Random Forest Machine Learning Algorithm", ECE591Q Machine Learning Journal Paper. Fall 2005

[20] L. Breiman. Random Forests. *Machine Learning*, 45(1):5-32, 2001.

[21] G. Anderson. PhD thesis: Random Relational Rules. 2009.