International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012

Learning Content Management Using Machine Learning

Siddhartha Kumar Arjaria and Devshri Roy

***Abstract*—The popularity of web based learning has led to the development of many learning object repositories that store high quality learning materials. High quality learning materials are expensive to create. So it is very important to ensure reuse of learning materials. Reuse of learning materials are made possible by semantically tagging them with standard metadata. In all available learning object repositories the tagging is done manually. Manual annotation is time taking and labor intensive activity. In this paper, we have explored the feasibility of tagging learning materials automatically with IEEE LOM 9.0 metadata specification. Here, we present machine learning approach using k-nearest neighborhood and back propagation neural network to automatically identify the subject of learning materials. The classifier is tested & result shows about 84% & 93% accuracy for back propagation neural network & K-NN resp.**

***Index Terms*—BPN, classification, KNN, learning object.**

I. Introduction

The wide availability of learning content in the World Wide Web has given rise to new paradigms of learning and knowledge delivery. The use of learning and content management systems for delivering learning materials in web based learning is becoming an important subject of research for researchers [1].

The internet is a big source of good quality learning material. We can use this learning material for education purposes. But, this digital information is increasing rapidly and the student is not able to use these learning materials efficiently & effectively. Many learning object repositories (LOR) are available which stores learning materials which helps in delivering good quality learning materials relevant to the student's requirement. EdNA [http://www.edna.edu.au/edna/page1.html], Ariadne [http://www.ariadne-eu.org/] and Merlot [http://www.merlot.org/] are examples of some of existing LORs. For efficient retrieval of learning materials according to the requirement of a student, the learning materials are tagged with a set of metadata which describes educational artifacts such as topic of the document, type of the document etc. However, in order to facilitate the sharing and reuse of learning materials across different information repositories or learning management systems, the learning materials should be associated with some common metadata standard. Several metadata standards are used for description of learning objects like Dublin Core metadata initiative (DCMI, http://www.dublincore.org/),SCORM Metadata (http://www.adlnet.gov/scorm), Advance Distributed Learning Initiative (http://www.adlnet.org), IEEE Learning Object Metadata (http://ltsc.ieee.org) etc. [2].

Most of the available online learning object repositories have been developed manually. The authors, contributors and developers of the open repositories have the responsibility of manually attributing meta information to the learning objects. In the Health Education Assets Library (http://www.healcentral.org) and iLumina (http://www.ilumina-dlib.org/), the contributors are required to follow strict guidelines and fill up many forms to carefully ensure that the learning objects associated to the repository are according to their requirements. In LearnAlberta Online Curriculum Repository (http://www.learnalberta.ca/ login.aspx), the developer has to follow the specifications of resource development guideline such as learning object development guideline, metadata guidelines, instructional design guidelines etc.

Associating meta information to learning objects by humans is a labor intensive activity. Many contributors find the task of manual annotation and assigning of meta-tags uninteresting, and sometimes the tagging is not done satisfactorily. The development of a repository with manually annotated learning materials is expensive in terms of the time and effort required.

There is need of automatic extraction of metadata from learning materials for automating the development of LORs. Automatic extraction of metadata is especially important if we wish to harness the large number of documents available in the Web for building the LOR. We have worked to take forward the process of automatic metadata extraction from learning materials. We have worked for automatically extracting the element number 9 meta data element of IEEE LOM I. e. classification of learning objects based on their discipline (subject).

The work presented in this paper focuses on the automatic tagging of learning materials with a very important IEEE metadata element i. e. discipline (subject) of the learning material. We apply the machine learning approach to automatically classify the learning objects based on their discipline.

A domain ontological approach for document classification is suggested by Yi-hsing Chang [3] and achieved accuracy of classification of about 53%. Another approach given by [4] i. e. document Classification Algorithm Based on Kernel Logistic Regression achieved the highest classification accuracy up to 84%. In [5] the machine learning approach distance based classification is used and achieved highest accuracy of about 88%. This paper aims to increase the classification accuracy of learning object by using the neural network and K-NN approach for efficient

Manuscript received April 13, 2012; revised June 19, 2012.

Siddhartha Kumar Arjaria is with the Maulana AzadNational Institute of Technology Bhopal,462051 India (e-mail: arjarias@ gmail.com).

Devshri Roy is with the Maulana AzadNational Institute of Technology Bhopal,462051 India(e-mail: droy.iit@ gmail.com)

learning content management with least effort. Previous researches have also shown that NN can achieve accurate results, that are sometimes more accurate than those of the symbolic classifier [6].

II. Extraction of Ieee Lom 9.0 Metadata

There are three general methods by which information extraction can be accomplished: manually (writing custom code for each new type of document), which is labor intensive, wrapper learning (manually setting anchors within example documents to generate a set of rules for extraction),and fully automatic . Manual categorization of learning object is a time consuming and laborious task so there is need for automatic extraction. In this direction we proposed the automatic extraction of IEEE LOM 9.0 classification of learning object over discipline or subjects.

A. Feature Extraction

Web pages are collected from the World Wide Web. The input to feature extraction phase is learning object. The Learning Object is considered as sequence of words and symbols appear in different structured way. We remove the common words like 'the', 'to', 'and', 'a' etc.& repeated words with no meaning from Learning Object. A feature vector is created whose attributes are the words. Each word in the feature vector satisfies the criteria of minimum occurrence frequency i. e. the number of occurrences of the word in the document should be more than the minimum threshold frequency (both local & global to be 4 & 7 respectively). We count the occurrence of the words for each input document. After stemming & normalization process we get the frequency matrix for training and testing.

B. Classifier Overview

1) K-nearest neighbour (K-NN)

K-NN is one of the most popular classification rules, although it is an old technique. We are given c classes, ω_i , $i = 1, 2, \ldots, c$, and a point $x \in R_i$ and N training points, x_i, $i = 1, 2, \ldots, N$, in the l-dimensional space, with the corresponding class labels. Given a point, *x*, whose class label is unknown, the task is to classify *x* in one of the c classes. The rule consists of the following steps:

Among the N training points, search for the *k* neighbors closest to *x* using a distance measure (e.g. Euclidean, Mahalanobis). The parameter *k* is user-defined. Note that it should not be a multiple of *c*. That is, for two classes *k* should be an odd number

Out of the k-closest neighbors, identify the number k_i of the points that belong to class ω_i obviously, $\sum_{i=1}^{c} k_i = k$

Assign x to class ω_i for which $k_i \succ k_j \quad j \neq i$

In other words, *x* is assigned to the class in which the majority of the k-closest neighbors belong

2) Perception (MLP) neural networks

The MLP normally makes use of the gradient descent to compute the new value of the weights and biases. It is quickly able to adjust the network weights for good performance. The graph or space denoting the error of the system for every combination of weights and biases is called as the error space. The aim of any training algorithm is to find the global optima in this search space. BPA many times get trapped in local minima. This is due to the absence of any global guiding strategy or the attempt to cover the entire error space which is highly complex and dimensionality. The feed- forward neural network architecture used in this experiment consisted of one hidden layers along with input and output. The transfer function in hidden layer neurons and output layer neurons are sigmoid and purelin respectively. The performance function used was Mean Sum-squared Error (MSE).

where *i, j* and *k* indices referring to the neurons belonging to the output, hidden, and input layers respectively , p, m, and k are the number of neurons in input, hidden and output layer respectively.

x = Input vector

h = Weighted sum of input stimuli

v = Output vector of hidden layer

g = Weighted sum of v_j

y = Output vector of output layer

w_{ij} = Wt. connecting i^th unit of output layer and j^th unit of hidden layer

w_{jk} =Weight connecting j^th unit of hidden layer to k^th unit of input layer.

y = Actual output.

y_d = Desired output.

Actual outputs of the network have computed in forward path. Computations have done as follows

$$h_j = \sum_{k=1}^{p} w_{jk} x_k \tag{1}$$

$$v_j = \frac{1}{1 + e^{-h_j}} \tag{2}$$

$$g_j = \sum_{j=1}^{m} w_{ij} v_j \tag{3}$$

In the backward path the following computation will be done by neural network. Compute the error $e = y_d - y$

Compute $\delta_i = y.(1-y).(y_d - y)$

where, δ_i is used to distribute the error at the output unit back to the preceding layers. Update the weights connecting hidden layer to output layer using the following rule. $w_{ij}(t+1) = w_{ij} + \eta.\delta_i.h_j$

Compute $\delta_j = v_j.(1 - v_j).w_{jk}.\delta_j.x_k$

It is not necessary to propagate the error back to the input layer. δ_j is used to adapt the weights connecting the input layer to the hidden layer. Update the weights connecting input layer to hidden layer. $w_{ij}(t+1) = w_{ij}(t) + \eta.\delta_i.\delta_j.h_j$

III. Experimental Results

The learning objects are collected from the internet related to computer science subjects. The 80 learning objects related to four different subjects e.g. Computer Network, DBMS, and Operating System & Software Engg. with each one has 20 instances each are used for training.

TABLE I: Classifier Performance Table for BPN

INPUT TO BPN CLASSIFIER		CLASSIFIER OUTPUT (NO. OF DOCUMENTS)						
SUBJECTS	NO. OF DOC.	C.N.	DBMS.	O.S.	S.E	Precision	Recall	F1
C.N.	18	17	0	0	1	94.44	77.27	84.99
DBMS	17	5	9	2	1	52.94	90	66.66
O.S.	17	0	0	17	0	100	85	91.89
S.E.	15	0	1	1	13	86.66	86.66	86.66

TABLE II: Classifier Performance Table for K-NN

INPUT TO KNN CLASSIFIER		CLASSIFIER OUTPUT (NO. OF DOCUMENTS)						
SUBJECTS	NO. OF DOC.	C.N.	DBMS.	O.S.	S.E	Precision	Recall	F1
C.N.	18	17	0	0	1	94.44	89.47	91.88
DBMS	17	1	15	1	0	88.23	100	93.74
O.S.	17	1	0	16	0	94.11	88.88	91.42
S.E.	15	0	0	1	14	93.33	93.33	93.33

Similarly the 72 learning objects (18 from each subject) are used to test the classifiers. The term matrix for training will be for 74 learning objects and remaining 6 objects will be removed by TMG tool during tokenization and parsing. In the same way the test matrix will be reduced to 67 learning objects. 18 for Computer Network, 17 for DBMS, 17 for Operating System and 15 for Software Engineering

We are using 20-12-4 structure with tansig, logsig, logsig as an activation function with learning rate 0.35 and momentum of .45. It takes 4666 epochs to train the classifier. Fig. 1. shows the classification results. In table 1 we discussed the classification performance. We get about 84% classification accuracy of learning objects.

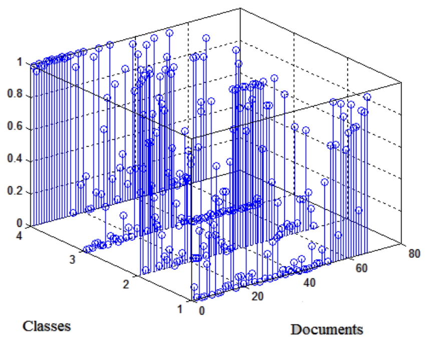

Fig. 1. Classification results of BPN

Then we are applying the KNN classifier with K=5.The Fig. 2. shows the classification results and table 2. Indicates the performance of KNN classifier and the overall accuracy of classifier is about 93%.

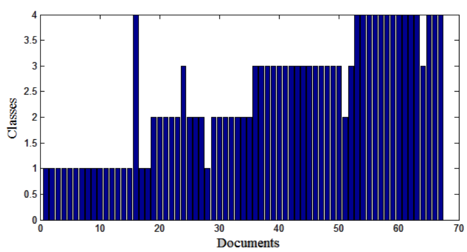

Fig. 2. Classification results of KNN

IV. Conclusion

In this paper, we have explored the feasibility of automatic extraction of IEEE LOM 9.0 i. e. subject of learning materials. This facilitates the creation of an e-Learning repository for storing these annotated learning materials, which can be used by different learning management systems. The idea is to make use of learning materials from the World Wide Web for developing metadata annotated learning object repository. Automatic annotation of learning materials is a difficult task. The KNN classifier and BPN classifier are designed to extract the subject of the learning materials. We found that the KNN classifier gives better result as compared to the BPN classifier.

References

[1] I. Süral, "Characteristics of a sustainable learning and computer management system (LCMS)," *Procedia-Social and Behavioural Science*, vol. 9, pp. 1145-1152, 2010.

[2] Standard for Learning Object Metadata, IEEE PISCATAWAY, NJ, 2002.

[3] Y. H. CHANG, "Automatically constructing domain ontology for document classification", *in Proc of 6th International Conference on Machine Learning and Cybernetics*, Hong Kong, pp. 19-22, August 2007.

[4] Z. Wang and X. Sun, "Document Classification Algorithm Based on Kernel Logistic Regression," in *Proc.of 2nd International Conference on Industrial and Information Systems*, pp. 76-79, July 2010.

[5] G. R. Ranganathan, Y. Biletskiy, and D. MacIsaac, "Machine learning for classifying learning objects," in *Proc. of CCECE'06*, Ottawa, pp. 280-283, May 2006.

[6] J. Clark, I. Koprinska, and J. Poon, "A neural network based approach to automated e-mail classification," in *Proc. o IEEE/WIC International Conference on Web Intelligence (WI 2003)*, Halifax, Canada, pp. 702–705, October 2003.

S. K. Arjaria has obtained the Bachelor of Engineering from Institute of technology, university of bilaspur(C.G.) in year 2001, and ahs obtained Master of technology from Govt. Engg. College, Raipur (declared N.I.T) in 2006. Currently he is pursuing Ph.D in computer science & engineering department from Maulan Azad National Insitute of Technology ,Bhopal(M.P.)

D. Roy has obtained the Bachelor of Engineering in Electronics from Maulana Azad National Institute of Technology, Bhopal in the year 1990, Master of Engineering in Computer Science Engineering from National Institute of Technology, Rourkela in the year 1998 and Ph D in the year 2007 from the Department of Computer Science Engineering, Indian Institute of Technology, Kharagpur, India.

She has worked with many prestigious institutes of India like Indian Institute of Technology, National Institute of Technology etc. in India. Currently she is working as an Associate Professor in Maulana Azad National Institute of Technology, Bhopal

Dr. Roy has received a research grant of worth Rupees 9.73 lakhs from Government of India to carry out a research project. Total number of papers published in referred Journals, International conferences and International workshops are 25. Current research interest includes Information Retrieval, natural language processing and application of Artificial Engineering techniques in Electronic and mobile Learning.