Effective Learning Experience using Text Classification

Pushpakala
Final CSE
Sri Sai Ram Institute of Technology

S.Sindhu priya
Final CSE
Sri Sai Ram Institute of Technology

Dr. A. Meena Kabilan
Asst. Professor
Sri Sairam Engineering College

Abstract - A classifier is used to classify certain words from the text document. Here the exact keywords should be matched with those of which present in the server. Hence, this restricts the user to type their answer in their own words. Therefore, adopting the proposed system would resolve this issue. The proposed technique includes a material recommendation system based on the student's performance. This would be predicted from their project work documents and online examination results. These online tests are conducted for each individual subjects and the questions would be designed by their faculties in a secure portal. The staffs would then feed the evaluation knowledge to the system for student's examination score generation. Based on the result, the students will be categorized and the study materials will then be rendered to a particular student depending upon the performance category he belongs.

Keywords: Text Mining, NLP technique, Word Net tools, Video transcoding, String matching algorithm.

I. INTRODUCTION

Discovering the relevant features from text documents (viz student answer sheets) and to evaluate the performance of students is nowadays playing a vital role in educational institution. So that appropriate learning materials could be rendered to them. Text mining also referred to as text data mining and roughly equivalent to text analytics refers to the process of deriving high-quality information from text.

This is done by devising of patterns and trends through means such as statistical pattern learning. Text analytics software's could be of significant use as they transpose words and phrases in unstructured data into numerical values which can then be linked with structured data stored in a database. They are further analyzed with traditional data mining techniques. To learn term features within relevant and unlabelled documents, two term-based models are used. In the first stage, a Rocchio classifier is used to extract a set of reliable irrelevant documents from the unlabeled set. In the second stage, these text documents are classified with the help of a SVM classifier. Alternatively, another two-stage model involving integration of thorough analysis (a term-based model) with pattern taxonomy mining can also be used and has evidently been proved as the best way to design a two-stage model for information filtering systems.

The proposed material recommendation system is based on student's performance. It is evaluated from project documents in their curriculum documents and from online examination result conducted every term. These online tests are conducted for individual subjects and a pool of questions is fed by their respective staffs in their portal. The staffs will also feed the evaluation knowledge to the system. Later after completion, students' scores are generated. The aggregate of project work and various online tests decide the student's overall performance.They are then categorized into three or more bands accordingly. The study materials will then be rendered to that particular student depending upon his performance band. Videos can also be played in the student portal for effective learning and to provide an interactive learning environment.

II. LITERATURE SURVEY

In paper [1] the author proposed an idea which includes an innovative and effective pattern discovery technique which included the processes of pattern deploying and pattern evolving, to improve the effectiveness of using and updating discovered patterns for finding relevant features and interesting information. Substantial experiments on RCV1 data collection and TREC topics demonstrated that the proposed solution achieves encouraging performance.

Zhao and his team in paper [2] proposed a framework called "Similarity Preserving Feature Selection" The authors showed, through theoretical analysis, that the proposed framework not only encompasses many widely used feature selection criteria, but also naturally overcomes their common weakness in handling feature redundancy. They begin with a conventional combinatorial optimization formulation for similarity preserving feature selection, then extended it with a sparse multiple-output regression formulation to improve its efficiency and effectiveness. They proposed set of three algorithms that were devised to efficiently solve the proposed formulations, each of which had its own advantages in terms of computational complexity and selection performance.

The author [3] implements the goal of supervised feature selection is to find a subset of input features that are responsible for predicting output values. The least absolute shrinkage and selection operator (Lasso) allows computationally efficient feature selection based on linear dependency between Input features and output values. They consider a feature-wise kernelized Lasso for capturing nonlinear input-output dependency. They first showed that with particular choices of kernel functions, non redundant features with strong statistical dependence on output values can be found in terms of kernel-based independence measures such as the Hilbert-Schmidt independence criterion. They then proposed that the globally optimal solution can be efficiently computed; which made the approach scalable to high-dimensional problems. The effectiveness of their proposed method is demonstrated through feature selection experiments for classification and regression with thousands of features.

III. MODULES DESCRIPTION

- Project and batch allocation
- Text mining for assessment
- Project Review and Student Assessment
- Material Recommendation and Interactive Student learning

3.1 Project and batch allocation

In this module coordinator upload project base paper on behalf of each and every student, and also allocate batches for all projects. Batches were created by coordinator by selecting number of student in batch and student ID's. The strength of each batch will be decided by the coordinator according to the project.

In upload process the base paper will be checked for duplication with previous batches in title level as well as in content level. In title level validation we use String Matching Algorithm. The content level checking is done by stripping down the pdf content is checked with previous batch abstract by using the content based algorithm. If the base paper is not duplicated server accepts the upload and updates the student record.

3.2 Text mining for assessment

Teacher prepares questions and answers for student assessment. Text mining process is done by natural language processing and word net tools. Natural Language Processing is a field of computer science, Artificial Intelligence and computational Linguistics concerned with the interaction between computer and human (natural) languages.

WordNet is a lexical database for the English Language. It groups English words into sets of synonyms called synsets provides short definitions and usage examples, and records a number of relations among these synonym sets.

PoS tagger (Part of Speech tagger) is implemented to extract the important keywords in the answer given by staff before assessment is done. The extracted Keywords are categorized as mandatory keywords, subordinate keywords, and technical keywords. WordNet tool is used to give the related synonyms to literal word in the subordinate terms. Now Teachers can feed the servers with the eligible terms in the categories to be present for student evaluation.

3.3 Project Review and Student Assessment

Student login with their credentials and then uploads the review materials in server. Reviewer gives the review marks for each student based on performance. Here we allotted three reviews, and give marks for student based on review performance, that is, zeroth review, first review and second review are conducted in online.

Student can write the assessment test and can submit the answers to the server. Student answers are evaluated later in server by extracting keywords using NLP technique and wordNet tool. The Machine will evaluate the answers by comparing it with the categorized terms given by the teachers. Depends upon the student answer they will give marks and prepare performance report. Review performance and assessment score are aggregated to find the overall performance.

3.4 Material Recommendation and Interactive Student learning

Teachers prepare the material for each subject and also give tags (good, best) for student material recommendation. Here we upload the materials like video, text, pdf. Video Transcoding is applied while video materials are uploaded for below average students. The Video Transcoding is done by the staff according to the student capacity. After finishing the assessment test, in student portal they get the materials based on overall performance calculated by server. If they have doubt while watching video content, students can interactively raise questions by simply clicking on the video frame. The video frames are previous indexed so that appropriate meta information's can be extracted for each frame. The student's questions and meta information from the current frame are send to server and can be reviewed by the staff. Once the staff login they will be notified with the questions and then staffs can reply to the question. The Student can now be able to view the answers given by the staffs.

IV. SYSTEM ARCHITECTURE

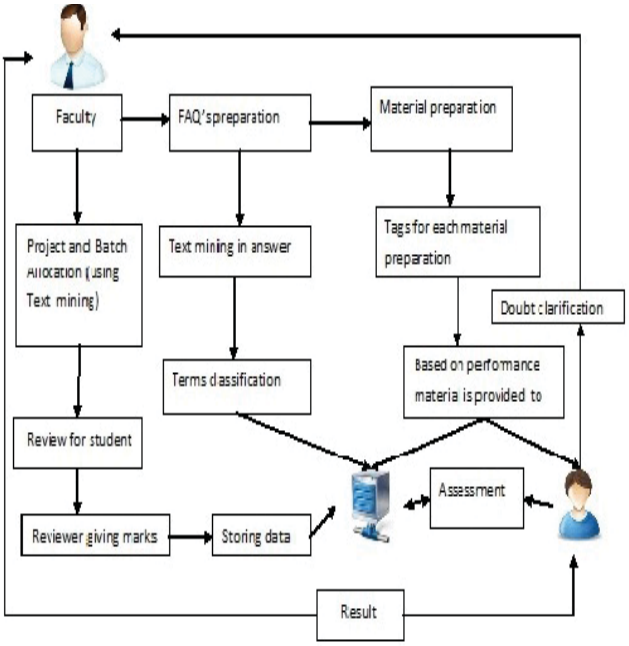

Fig. 1: Block diagram for effective learning

Fig 4.1 shows the description of how the data mining technique is applied at various stages to improve the learning experience of the students by conducting assessments and used to provide the necessary study materials.

V. CONCLUSION

In the proposed material recommendation system, student's performance in their projects and examinations are scrutinized and the students are awarded with performance bands based on their scores. Faculties handling those subjects would also be authorized to feed into question pool which could be use full while preparing question papers for the written tests. Finally, the study materials are provided to them according to their performance bands.

REFERENCES

[1] N.Zhong, Y . Li,and S-T . Wu,"Effective pattern discovery for text mining",in IEEE Trans .Knowl.Data Eng.,Data Eng.,vol. 24,no. 1,pp.30-44,Jan. 2012.

[2] Z.Zhao, L. Wang,H. Liu, and J. Ye,"On similarity preserving feature selection," in IEEE Trans. Knowl. Data Eng., vol. 25, no. 3,pp. 619-632,Mar. 2013.

[3] M.Yamada,W. Jitkrittum,L.Signal,E. P. Xing,and M. Sugiyama,"High-dimensional feature selection by feature-wise kernelized Lasso",Neural Comput., vol. 26, no. 1,pp. 185-207,2014.

[4] Object-Oriented Analysis and Design with Applications (3rd Edition) Grady Booch, Robert A. Maksimchuk, Michael W. Engel, and Bobbi J. Young April 30, 2007.

[5] Principles of Object-Oriented Analysis and Design James Martin and James J. Odell Jan 17, 2008.

[6] Software Engineering: A Practitioner's Approach Roger Pressman Jan 20, 2009.

[7] Software Engineering Shari Lawrence Pfleeger and Joanne M. Atlee Feb 27, 2009.

[8] Essentials of Data Base Management System - Alexis Leon and Mathews Leon - Vikas Publishing Limited, Chennai First Edition, 2009.

[9] Introduction to database Systems - Atul Kahat - Pearson Education – 2011.