International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume 2, Issue 3 (March'2016)

EFFECTIVE CLASSIFICATION OF TEXT AND IMPROVING LEARNING EXPERIENCE

I.Anisha Agnes,
UG-Scholars,
Velammal Institute of Technology,
Chennai,India.

K.Nivetha,
UG-Scholars,
Velammal Institute of Technology,
Chennai,India.

M.Rajeswari,
Assistant Professor,
Velammal Institute of Technology,
Chennai,India.

Abstract: One-to-many data linkage is an essential task in many domains, yet only a handful of prior publications have addressed this issue. Furthermore, while traditionally data linkage is performed among entities of the same type, it is extremely necessary to develop linkage techniques that link between matching entities of different types as well. In this paper we propose a new one-to-many data linkage method that links between entities of different natures. The proposed method is based on a one-class clustering tree (OCCT) which characterizes the entities that should be linked together. The tree is built such that it is easy to understand and transform into association rules, i.e., the inner nodes consist only of features describing the first set of entities, while the leaves of the tree represent features of their matching entities from the second dataset. We propose four splitting criteria and two different pruning methods which can be used for inducing the OCCT.The method was evaluated using datasets from three different domains. The results affirm the effectiveness of the proposed method and show that the OCCT yields better performance in terms of precision and recall (in most cases it is statistically significant) when compared to a C4.5 decision tree-based linkage

Keywords : Text mining, Text feature extraction, Text classification

I.INTRODUCTION

The main aim of this project is to discover relevant feature in text documents (student answer sheets) and to evaluate the performance of student so that appropriate learning materials can be rendered to them.Text mining also referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning.Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. With an iterative approach, an organization can successfully use text analytics to gain insight into content-specific values such as sentiment, emotion, intensity and relevance. Because text analytics technology is still considered to be an emerging technology, however, results and depth of analysis can vary wildly from vendor to vendor.

II.EXISTING SYSTEM

To learn term features within only relevant document and unlabelled documents, paper used two term-based models. In the first stage, it utilized a Rocchio classifier to extract a set of reliable irrelevant documents from the unlabeled set. In the second stage, it built a SVM classifier to classify text documents.

A two-stage model was also proposed in which proved that the integration of thorough analysis (a term-based model) and pattern taxonomy mining is the best way to design a two-stage model for information filtering systems.

III.PROPOSED SYSTEM

We proposed a material recommendation system based on student performance which will be predicted from the project work documents and the online examination results. The online tests are conducted for each individual subjects and the questions will be allocated for the same by the respective staff in their portal. The staffs will feed the evaluation knowledge to the system for student examination score generation. The aggregate of project work and online test decides the student's overall performance in three levels which is used to categorize them. The study materials will be rendered to the particular student depending upon the performance category he belongs. Videos can also be played in the student portal for effective learning which we innovate to interactive learning by simple click events on the doubtful content of the video.

IJCRCST © 2016 | All Rights Reserved

IV.MODULES

1. Project Allocation
2. Text mining for assessment
3. Project Review and Student Assessment
4. Material Recommendation and Interactive Student learning

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. In upload process the base paper will be checked for duplication with previous batches in title level as well as in content level. The content level checking is done by stripping down the pdf contents to text contents. If the base paper is not duplicated server accepts the upload and updates the student record.

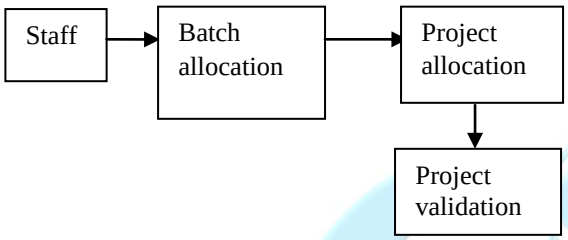

Figure 1: Batch allocation and validation

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. Pos tagger is implemented to extract the important keywords in the answer given by staff before assessment is done. The extracted Keywords are categorized 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.

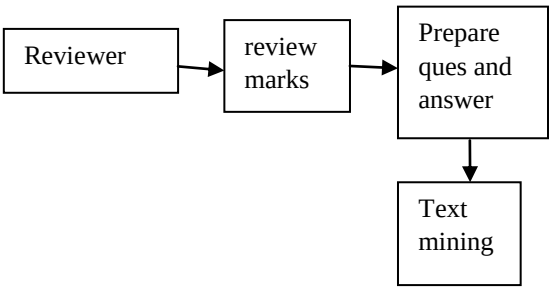

Figure 2: Performance evaluation by using NLP technique

3.Project Review and Student Assessment

Student login with his 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. 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.

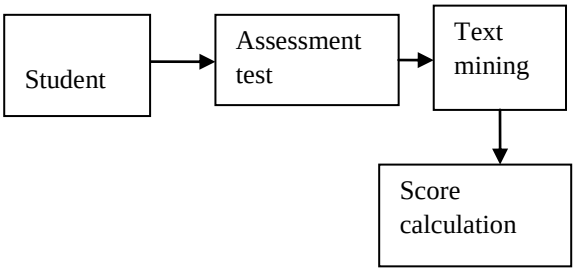

Figure 3: Review performance and assessment score

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. 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.

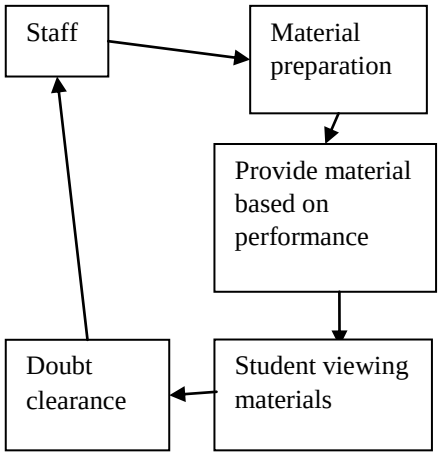

Figure 4: Material preparation and student learning

VI.ALGORITHM USED

- Nlp technique
- String matching algorithm
- content based algorithm
- Transcoding technique

VII.ENHANCEMENT

- Material Recommendation system
- Project title validation and content validation
- Transcoding in video material
- Interactive Student learning

VIII.CONCLUSION

Hence we developed a method to find and classify terms based on their teacher answer. Student test answer also classifies terms and evaluates answers based on teacher answer, and then teacher will provide material based on student performance.

IX.REFERENCES

[1] M. Aghdam, N. Ghasem-Aghaee, and M. Basiri, "Text feature selection using ant colony optimization," in Expert Syst. Appl., vol. 36, pp. 6843–6853, 2009.

[2] A. Algarni and Y. Li, "Mining specific features for acquiring user information needs," in Proc. Pacific Asia Knowl. Discovery Data Mining, 2013, pp. 532–543.

[3] A. Algarni, Y. Li, and Y. Xu, "Selected new training documents to update user profile," in Proc. Int. Conf. Inf. Knowl. Manage., 2010, pp. 799–808.

[4] N. Azam and J. Yao, "Comparison of term frequency and document frequency based feature selection metrics in text categorization," Expert Syst. Appl., vol. 39, no. 5, pp. 4760–4768, 2012.

[5] R. Bekkerman and M. Gavish, "High-precision phrase-based document classification on a modern scale," in Proc. 11th ACM SIGKDD Knowl. Discovery Data Mining, 2011, pp. 231–239.

[6] A. Blum and P. Langley, "Selection of relevant features and examples in machine learning," Artif. Intell., vol. 97, nos. 1/2, pp. 245–271, 1997.

[7] C. Buckley, G. Salton, and J. Allan, "The effect of adding relevance information in a relevance feedback environment," in Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 1994, pp. 292–300.

[8] G. Cao, J.-Y. Nie, J. Gao, and S. Robertson, "Selecting good expansion terms for pseudo-relevance feedback," in Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2008, pp. 243–250.