



Text Mining Techniques/Problems in Subjective Examinations

Alina Anjum

Department of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan

AlinaAnjum01@yahoo.com

Qamar Sultana

Department of Software Engineering
Sir Syed University of Engineering & Technology
Karachi, Pakistan

qqamarSultana@gmail.com

Abstract

In this Era of Fast-Moving Digital World, the world is still lacking the availability of Subjective E-Examinations System where a Bundle of Text can be submitted as a Subjective Answer & Subjective Evaluation may be performed. However, there are couple of Systems have been designed for Multiple Choice based Questionnaire Exams, although due to the Lack of Text Mining Techniques and Their problems, the world is still Lacking in Subjective Online Evaluations of Students either in English Text or Mathematical Solutions. There are various Natural Language Processing Techniques being used to Facilitate such issues but still we don't have a fool proof system for Online Score Calculations Based on Subjective Text. Our Survey will cater most of the recent Research works related to this aspect of Technology with their acquired solutions and the Deficiencies in those Solutions as Why we are still unable to have such a System which we can implement in Various Government and Private institutions for Online Examinations.

Keywords - Subjective Answers, Natural Language Processing, Subjective Evaluations, Text Mining, E-Examinations

I. INTRODUCTION

In this smart technological era the digital data acquiring techniques manages a huge amount of data. Now-a-days data is composed of structured, unstructured and semi structured data and stored in data ware houses. And it creates a big problem for organization to extract the meaningful data from that huge and massive amount of data, for solving these issues related to extraction of meaningful data text mining techniques are introduced. Like other organizations educational sector is also facing some issues. The goal of our research is to extract the meaning full data from different educational contributors, and what problems they are facing to gather the meaningful data from different kind of knowledge portals. There are multiple techniques used for it like information extraction, categorization, clustering and summarization. We will be discussing various Technologies being used to resolve text mining issues. The main goal of Higher Education Commissions is to achieve meaningful information from the data where they can asses the students via digital systems. The two main areas of Higher Education Commission are

Educational Data Mining (EDM) and Learning Analytics (LA) for making the ground of using Text Mining and the outcomes of the applications of Text Mining. wide range of Web based platforms are there to facilitate online education, such as web-based Adaptive and Intelligent Educational Systems (AIES), web based Intelligent Tutoring System (WITS), Learning Management System (LMS), Massive Open Online Courses (MOOCs) [2]. The data for mining can be received from various resources including the internet sources as well. Our main focus is to discuss the issues in subjective answers in Online Examinations. The aim is to measure the performance of Student through online assessment process. Monitoring of Performance comprised of assessments and evaluation processes, which plays an important role in providing valuable information which help students, instructor, administrators, and policy makers in higher education institutions to make decisions [1]. E-Learning which is know as Online Web Based Education is increasingly introducing huge amount of data in terms of various user interactions and provides various learning platforms, we can say it a Blessing for Online Education of Students. But E-Learning can equally be viewed as an exponentially growing nightmare, in which unstructured information chokes the educational system without providing any articulate knowledge to its actors. We would be discussing various techniques which have been implemented for Text Mining issues and the Problems being faced in those techniques.

II. TEXT MINING & IT'S TECHNIQUES

The manual framework for assessment of Subjective Answers for specialized subjects includes a great deal of time and exertion of the evaluator [6]. If the Human evaluates an Online Subjective answer, the evaluation and assessment may be based on varied in various conditions, it may also be based on various emotions. Due to such reason, there being multiple techniques which have been introduced to resolved subjective evaluation issues. The improvement of e-learning frameworks can lead to different dangers identified with investigation of multivariable learning information [7].

In every Educational condition, one of the principle goals is to guarantee that learning forms permit to get understudies and their learning ways. This is the place Data Mining Educational or Educational Data Mining (EDM) for its abbreviation in

English, give a major an incentive to instructive establishments and to all elements that help various procedures in learning exercises [10].

Over the previous decade, numerous methodologies have been suggested that abuse data mining procedures to tackle arithmetic word issues. The existing work can be arranged based on number juggling and logarithmic word issues that comprehend single and different variables in an equation [12]. But still there is a vast need for subjective text mining. Following are some of the techniques discussed in various past research works to resolve those issues.

A. *Classification & Clustering*

This is one of the Assessment Technique discussed in [1]. This technique uses the approach of putting the set of students in different classes as who is taking interest in classes and who is going to be the drop out. This technique creates the class for each category of student thus it enhances the efficiency of e-learning process. Then the various similar classes are put into groups which is known as Clustering through clustered groups, teachers can come to know the various deficiencies in students through the logs generated from LMS. The professor can take the decisions based on the created classes and clusters but these methods can not be used for student marks evaluation and full assessment.

B. *Fuzzy logic methods*

Data mining methods discussed in [5] used in intelligent system. The methods derive such system where the analytics can be created about the interaction of student in the system. For Example, if a question is placed by the professor, it would count the interaction of students as who has interacted with the posting of professor. But again, this does not fulfil the requirement of full proof assessment system.

C. *Natural language processing*

A technical approach being used in software engineering also discussed in [2]. Natural language handling (NLP) is a field in software engineering used to control characteristic language content or speech. It can process and break down a lot of normal language information by utilizing calculations for semantic and syntactic investigation. Educational Sectors had embraced NLP procedures throughout the years; be that as it may, the developing of Big Data, versatile innovations, internet-based life, and MOOCs brought about the making of numerous new research openings and challenges. The Educational Sector's urgent need for NLP is to extract information from Subjective Essays and answers from subjective Questions. Due to emergence of Big Data and Un structured semantics in the digital world, NLP is also facing certain issues in text processing including the Lack of vocabulary available, understanding of text through grammar by introducing the abilities of Parts of Speech, connecting different words from vocabulary for the creation of text, variations in structured and unstructured data. Apart from the

above issues, there are certain cases where a single word means different in different contexts for Example the word "Live", it can be used to denote dwelling somewhere and it is also used for Live transmissions, these things depends upon the situation and environment where the word is being used. NLP is still unable to resolve such issues that's why education sector cannot completely acquire this methodology. The meanings of the words can also be varied based on the feelings of a Human who is writing the answer and as well as it may be based on the circumstances.

More significance ought to be given to E-learning Assessment in Data Science ventures, since Assessment-related information have end up being exceptionally valuable to remove significant information and the other way around, that is, Data Science procedures have demonstrated to give significant data to all the more likely get Assessment forms [11].

D. *Preprocessing and Processing tools*

One of the tools used for Data Mining during text preprocessing as discussed in [3,4]. The data miner personnel upload some text which needs to be processed. In preprocessing phase, the text is matched with certain predefined rules of fixed text. If the uploaded text matches with the rules it overcomes the restrictions otherwise certain validations are applied. In the processing phase evaluations are performed on preprocessed text which was matched with predefined rules. This does not fulfil the requirement of education sector because they cannot define the set of answers in form of certain rules as the student can answer a question in various ways with different vocabulary.

E. *Systematic Review Method*

The method uses the guidelines proposed for SLR by Kitchenham et al. (2009) [8]. it uses certain phases for analysis including Planning (identifying keywords), Conduction Review and Assessing the Quality. The quality assessment is based on the identification of keywords but this does not fulfil the mining the full sentence.

F. *Assessment Algorithm based on Keyword Expansion*

The algorithm proposed in [9] works upon the following modules comprised of Pre-Processing, Keyword Expansion, matching & grading Module. The author has defined the process of designing the paper of online assessment. The methods provide the way of assessment based on keywords expansion where the semantic meaning of the word cannot be understood by the system and the marks calculation is based on keyword match count.

G. *K-Nearest Neighbors Algorithm (KNN)*

The algorithm discussed in [13] is in the top 10 algorithms being used for Data Mining Process which uses

text matching techniques. It matches the text with its neighbor text which they denote with K and K may be any number i.e. 3,4,5. The algorithm mainly focuses on the similarity of the text with earlier predefined text. In educational sector if we implement this algorithm, the system would be dependent upon text matching criteria although the student can write the answer in various different ways of text writing.

H. Text Mining using Corpora-based Sentiment Dictionary

A dictionary of words is created which the algorithm checks and matches with certain text as discussed in [14]. In the first stage of implementing this technique, a dictionary of sentiment is created and in the second step the set of words is matched with the sentiment dictionary which is manually created in order to Mine the text. The technique is based on semantic words matching thus does not fulfill the requirement of educational sector for mining subjective text.

III. RECOMMENDATIONS

From the above survey, we can see that most of the earlier research works are based on semantic matching techniques based on keywords although the researches in text mining based on full sentence with successful results are still lacking in number. To resolve the issues in Educational sector, standards must be fixed at international level for each standard of class and the course content must be categorized. For each category of course in each class, a system can be developed where the teacher can start putting questions and the student can submit the answer thus at initial level of implementation of this system, teacher can mark the answer as acceptable or not, a data warehouse would be built of various accepted answers against a single question i.e. each question can have multiple set of acceptable answers. This system can be implemented at international level in order to increase the size of data ware house. Once the data ware house is built, the above techniques can be used for matching the text from the acceptable answers thus the scoring can be performed by the system. This recommendation would mostly be based on data mining techniques from a Huge data ware house.

IV. CONCLUSION

In this paper, we discussed various techniques which are being used in text mining and the issues as why they are not yet implemented in the education sector successfully until now. We have also provided some recommendation for implementing an online successful examination system based on the concept of maintaining historical data in data ware house.

ACKNOWLEDGMENT

This Research was supported by the faculty members of Sir Syed University of Engineering and Technology. We are thankful for their guidelines and making us to able to put efforts on this research.

REFERENCES

- [1] Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49.
- [2] Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6), e1332.
- [3] Giannakopoulos, T., Foufoulas, Y., Dimitropoulos, H., & Manola, N. (2019, January). Interactive text analysis and information extraction. In *Italian Research Conference on Digital Libraries* (pp. 340-350). Springer, Cham.
- [4] Saa, A. A., Al-Emran, M., & Shaalan, K. (2019, March). Mining student information system records to predict students' academic performance. In *International conference on advanced machine learning technologies and applications* (pp. 229-239). Springer, Cham.
- [5] Dharmarajan, K., Balasree, K., & Abirami, K. Students E-Learning Performance Improvement and Predicting the Students Learning Interest Using Data Mining.
- [6] Patil, P., Patil, S., Miniyaar, V., & Bandal, A. (2018). Subjective Answer Evaluation Using Machine Learning. *International Journal of Pure and Applied Mathematics*, 118(24).
- [7] Gushchina, O., & Ochepovsky, A. (2019). Data Mining for the E-Learning Risk Management. *Turkish Online Journal of Distance Education*, 20(3), 181-196.
- [8] Saa, A. A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. *Technology, Knowledge and Learning*, 24(4), 567-598.
- [9] Alrehily, A. D., Siddiqui, M. A., & Buhari, S. M. INTELLIGENT ELECTRONIC ASSESSMENT FOR SUBJECTIVE EXAMS.
- [10] Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, (33), 235-266.
- [11] Lara, J. A., Aljawarneh, S., & Pamplona, S. (2020). Special issue on the current trends in E-learning Assessment. *Journal of Computing in Higher Education*, 32(1), 1-8.
- [12] Rehman, T., Khan, S., Hwang, G. J., & Abbas, M. A. (2019). Automatically solving two-variable linear algebraic word problems using text mining. *Expert Systems*, 36(2), e12358.
- [13] Sriwanna, K. (2018, February). Text classification for subjective scoring using K-nearest neighbors. In *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)* (pp. 139-142). IEEE.
- [14] Rizun, N., & Waloszek, W. (2018, September). Methodology for Text Classification using Manually Created Corpora-based Sentiment Dictionary. In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018)-Volume (Vol. 1, pp. 212-220)*.
- [15] Yuksel, M. E., & Fidan, H. (2019). A Decision Support System Using Text Mining Based Grey Relational Method for the Evaluation of Written Exams. *Symmetry*, 11(11), 1426.
- [16] Hung, H. C., Liu, I. F., Liang, C. T., & Su, Y. S. (2020). Applying Educational Data Mining to Explore Students' Learning Patterns in the Flipped Learning Approach for Coding Education. *Symmetry*, 12(2), 213.