Research Article

AUTOMATED ESSAY SCORING USING FEATURE EXTRACTION METHOD

Selvani Deepthi Kavila[1]*., Lakshmi S V S S[2] and Rajesh B[3]

[1,2]Department of CSE, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam-531162, India
[3]Department of CSE, GIT, Gitam Deemed to be University, Visakhapatnam, India

ARTICLE INFO

ABSTRACT

Evaluation is the major metric in any institution to judge the student and their learning skills. Essay writing is one of the major ways in which students are tested. Now, a days this evaluation is done manually. This requires a lot of human effort and time. In this paper, evaluation technique is automated. Since machine cannot understand the evaluation measures we give them in terms of keywords which act as features. Here initially the student written essay is compared with keywords mentioned by admin. If the similarity is less than 20%, then the essay is not up to the standard. Otherwise, it is evaluated using segmentation, stop words removal, calculating word frequency, forming numerical features, sentence validation, spell check. By using features extraction methods a general essay have been evaluated and after evaluating this general type of essay the final result is produced. The result obtained by proposed model is compared with manual judgment and it showed that the proposed model is much better than the manual judgment.

INTRODUCTION

Essays are considered by many researchers as the most useful tool to assess learning outcomes, implying the ability to recall, organize and integrate ideas, the ability to express oneself in writing and the ability to supply merely than identify interpretation and application of data..Many researchers claim that the subjective nature of essay assessment leads to variation in grades awarded by different human assessors, which is perceived by students as a great source of unfairness. This issue may be faced through the adoption of automated assessment tools for essays. A system for automated assessment would at least be consistent in the way it scores essays, and enormous cost and time savings could be achieved if the system can be shown to grade essays within the range of those awarded by human assessors. The assessing of an essay automatically by the system is done using the domain called "Text mining". Text Mining is a new field that tries to extract meaningful information from natural language text (Lokesh Kumar-2001).

Text mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High quality in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, entity relation modeling (i.e., learning relations between named entities).

Corresponding author: **Selvani Deepthi Kavila**
Department of CSE, Anil Neerukonda Institute of
Technology and Sciences, Visakhapatnam-531162, India

Text analysis involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information, extraction, data mining techniques including association analysis, visualization, and predictive analytics. The goal is, essential to turn text into data for analysis via application of Natural Language processing (NLP) and analytical methods. The phrase "text mining" is generally used to denote any system that analyses large quantities of natural language text and detects lexical or linguistic usage patterns (Wikipedia).Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. preprocessing prepares raw data for further processing. The steps for preprocessing are a) Sentence segmentation is the process of dividing written text into meaningful units, such as words, sentences, or topics. In English and some other languages, using punctuation, particularly the full stop / period character is a reasonable approximation. However even in English this problem is not trivial due to the use of the full stop character for abbreviations, which may or may not also terminate a sentence. When processing plain text, tables of abbreviations that contain periods can help prevent incorrect assignment of sentence boundaries. As with word segmentation, not all written languages contain punctuation characters which are useful for approximating sentence boundaries (Duyu Tang, 2014).b) Stop words are words which are filtered out before or after processing of natural language data (text).. Some tools specifically avoid removing these stop words to support phrase search. Any group of words can be chosen as the stop words

for a given purpose. One of the pioneers in information retrieval, is credited with coining the phrase and using the concept.

Rest of the paper is described as follows: section 2 describes the literature survey based on feature selection methods, section 3 described the proposed system architecture and methodology, section 4 shows the performance analysis and results, section 5 shows the conclusion and future work.

LITERATURE SURVEY

Background work related to feature extraction

Jill Burstein and Daniel Marc (2000) has cited their research for to assign an essay score based on the quality of writing characteristics in an essay. It points out the importance of the application's modularity with regard to experiments that evaluate the integration of new features and the re-use of modules for evaluations that contribute to the adaption of the system towards the generation of feedback. Beata Beigman Klebanov, Nitin Madnani, Jill Burstein, Swapna Somasundaran (2014) has cited their research for selection of information from external sources. The aim of this model was to improve the automated scoring of such essays and the evaluation will proceed as every essay E is responding to a test that contains a lecture L and a reading R. In this model first it takes an overlap model and based on the formulas it takes number of tokens and estimate the probability. Later, in position model it checks the position of x in lecture. After performing this, it assigns the scores to all essays. Finally the results highlight that the effectiveness of an importance model depends on the genre of the source text. It doesn't give the accurate result to the essay scoring. Semire Dikli (2006) has cited their research in automated essay scoring. AES systems provide the student with a score as well as feedback within seconds. Compare to the human scoring it gives accurate result but four types of AES systems are there, by all these systems AES provide the total score, and feedback. Firstly, PEG (project essay grader) contains only training stage and scoring stage. It just trains the system and does not provide any feedback because it concentrates on structures only and ignores the semantic analysis failing to detect the content related features of an essay the system does not provide instructional feedback to the students. Intelligent Essay Assessor (IEA) analyses and scores an essay using a semantic text analysis method called Latent Semantic Analysis. It measures similarity among words and texts. In order to evaluate the overall quality of an essay, LSA needs to be trained on domain representative texts. Intelligent Essay Assessor (IEA) focus on content related features rather than the form related ones, it provides no feedback. Siddhartha Ghosh, and Sameen S Fatima (2010) has cited their research for automated essay scoring that the existing systems are not giving the accurate grade and score and it implemented the present AES system, Project Essay Grade (PEG), Intelligent Essay Assessor (IEA), Electronic Essay Rater (ERater), are the models which are present in AES system and it uses various methods in scoring. Once the essay input is given, it gives the grade and as well as a proper feedback to improve. Fei Dong and Yue Zhang (2016) have cited their research for scoring systems are challenging since it relies not only on grammars, but also on semantics, discourse and pragmatics. It will check the grammar and structure of the essays and later gives the score. Sometimes its time consuming since features need to be

carefully handcrafted and selected to fit the appropriate model. A score-specific word embedding (SSWE) for word representations and a two-layer bidirectional long-short term memory network (LSTM) to learn essay representations. By combining SSWE, LSTM outperforms traditional SVM model. Using LSTM alone does not give significantly more accuracy compared to SVM. Bayesian Linear Ridge Regression and Support Vector Regression are chosen as state-of-art baselines. It uses word embedding with an embedding matrix. Randy Elliot Bennett and Anat Ben-simon (2005) has cited their research for AES system that involves four stages in scoring the essay, for each essay separately independently assigned initial weights. Weights are assigned on a 0-100 scale with the sum of dimensions are equal to 100. Later they will check the essay with other essays and take the mean weight of the essays. In second stage it gives the individual responses to the essay writers. Third stage includes that it checks the essay using NLP techniques, and by viewing the stage1 results it analyze the essay. Danielle S. McNamara, Scott A. Crossley, Rod D. Roscoe, Laura K. Allen, Jianmin Dai (2015) has cited their research in AES system that essay scoring is an enormously complex task that involves a choices, and preferences on the part of the grader. The focus of this is to describe a new method of AES that we have designed using hierarchical classification and report on its reliability. Across AES systems, a method is followed. First, a set of target essays are divided into a training set and a test set. A computational algorithm is there in training set, using this features it automatically calculated from the text. John K. Lewis (2013) has cited their research in AES system. All AES systems need to be trained through the creation of a knowledge base .Most systems grade essays based on style and content. The AES system that evaluates essay by PEG and erater techniques. It gives the score to individual student without comparison. AES systems admit that they are limited to identifying words and phrases that are characteristic of a strong answer, no AES can understand context or the deeper meaning of language typical of quality writing. Fridolin Wild, Christina Stahl, Gerald Stermsek, Gustaf Neumann (2005) has cited their research in AES system to develop by using LSA. This approach eliminating noise in word application. Using LSA to assess written essays enables grade ranges similar to those awarded by human graders. The process of auto-scoring can be divided into five sub-steps: text preprocessing, weighting, calculation of the SVD, correlation measurement and correlation method. Mark D. Shermis, Jill Burstein, Derrick Higgins, Klaus Zechner (2008) has cited their research in AES system that project grade essay was used in this that produced impressive results. The essay length is important up to a point, but beyond a certain threshold it carries little additional weight. Where that point is becomes a function of the average essay length. There are three general approaches, two of which involve the collection of empirical data and one that is currently more of a theoretical option. In the first approach, two samples of essays are collected, one for model building and the other for model evaluation/confirmation. Each of which has been rated by multiple raters. Robert Williams and Heinz Dreher(2014) has cited their research in AES system that in order to automate the grading of essays some method of capturing the meaning of the words, sentences, and paragraphs must be found. Markit provide exact performance in providing the result and equally good as other systems but problem with is it compares the essay with another essay and provides

similar score to all students and there is no change in providing feedback.

Proposed Technique

The proposed technique used in this paper, for developing automated scoring system in a field of text mining. Automated Essay Scoring systems implemented by using different techniques and methods to provide accurate grade and score. The Figure1 shows the architectural diagram representing the overall system framework. It consists of five phases.

In the first phase, the admin will add topics for each class and keywords for each topic. Then student will submit an essay related to that topic which is given as an input to the system. System now evaluates this essay and gives appropriate score according to the predefined rules.
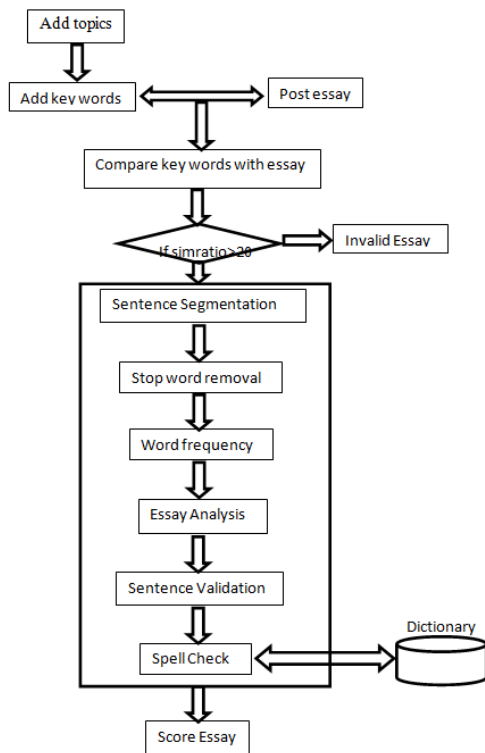

Figure 1 System Architecture

In the second phase, text preprocessing is done. In preprocessing stage stop word removal is done. Stop words removal is a process in which all the stop words like a, an, the, about, all punctuations, etc. are removed.

In third phase, the extracted keywords are compared with the keywords given by the admin to determine the relevance of the essay to the given topic. Then determine similarity ratio based on set theory i.e. obtain the intersection and union of words from the essay and from keywords given by admin. If the simratio is greater than 20 only then essay is evaluated, else, student is asked to rewrite the essay.

In fourth phase, the actual evaluation process is done using feature selection methods. The features to be considered for the evaluation process are word frequency, topic relevance, essay analysis, sentence validation, spell checking.

In fifth phase, a final score is given based on score from all the five features. These scores are then stored in database according to their classes.

The proposed algorithm is presented below in Table 1.

Table 1 Steps for Feature Extraction Method

Algorithm

Input: General Essay related to a topic
Output: Score for given essay based on different features for each student
Step-1: Admin will post topics for each class and add keywords for each topic.
Step-2: Student then selects class and write an essay related to that topic.
Step-3: Keywords are then compared with the words in the essay to get similarity ratio. It is calculated as follows:
Simratio = (words (essay) ∩ words (keywords)÷ words (essay) ∪ words (keywords)) * 100
Step-4: If, simratio ≥ 20 then go to step 5, else, go to Step 12.
Step-5: To split the sentences search for '.' and consider it as one sentence. Total numbers of sentences are also counted.
Step-6: To remove stop words from the text, each word from the input text is compared with the stop word list. If the word is a stop word, it is replaced by an ' ' (empty string) , else the process continues with the next word until all words of the text are compared.
Step-7: To obtain the word frequency, the occurrence of each word is counted. When the same word is repeated its count is increased. These occurrences are used to obtain weight of the essay.
Step-8: To get analysis of the essay consider word count after stop word removal, total number of sentences, average word length. The average word length is obtained by calculating sum of length of all words divided by total number of words.
Average word length = Length of all words ÷ Number of words
Step-9: To validate a sentence each word is tagged with its parts of speech using a POS tagger. Now get the link between one word to other. This sentence forms are now compared with certain rules. If they satisfy the rules they are considered to be valid sentences, else, not valid. The probability is calculated as:
Probability = (No. of valid sentences ÷ Total No. of sentences)*100
Step-10: To check spellings of all the words we use an en-US dictionary. Each word from essay after stop word removal are compared with the words in dictionary.txt. The count of misspelled words is used to find the probability.
Probability = (No. of misspelled words ÷ Total No. of words) * 100
Step-11: A final score is given based on scores from step-6 to step-10.
Step-12: End.

PERFORMANCE ANALYSIS AND RESULTS

Essay scoring system is developed for generating the scores of each student for the primary and secondary education students using feature extraction methods. The documents related to general topics like corruption, demonetization and so on are collected from present or past scenario and it is given to both the primary and secondary education students. Standard measures such as accuracy and error are used for evaluation.

Accuracy: Accuracy can be defined as the amount of uncertainty in a measurement with respect to an absolute standard. Accuracy specifications usually contain the effect of error due to gain and offset parameters.

Accuracy Percentage = (100%) – (Error Percentage)

Error: All measurements are subject to error, which contributes to the uncertainty of the results.

Error Percentage = ((|Experimental-Manual|) / Manual) * 100

A manual judgment is generated for all the general topics. A sample judgment scores as shown in Table 2 for the first 10 students that are studied.

Table 2 Scores of each student generated manually and by the proposed system

Student Reg. No.	Manual Scores	System Scores
1	31	28
2	39	35
3	35	33
4	40	37
5	41	37
6	43	38
7	40	37
8	40	36
9	38	35
10	40	35

The score of Error and accuracy percentage of each student are calculated as shown in Table 2

Table 2 Error and accuracy Percentage of scores

Student Reg. No.	Error Percentage	Accuracy Percentage
1	9.7	90.3
2	10.3	89.7
3	5.8	94.2
4	7.5	92.5
5	9.8	90.2
6	11.7	88.3
7	7.5	92.5
8	10.0	90.0
9	7.9	92.1
10	12.5	87.5

The graph is drawn to show the scores of each student generated manually and system efficiency of the proposed system.
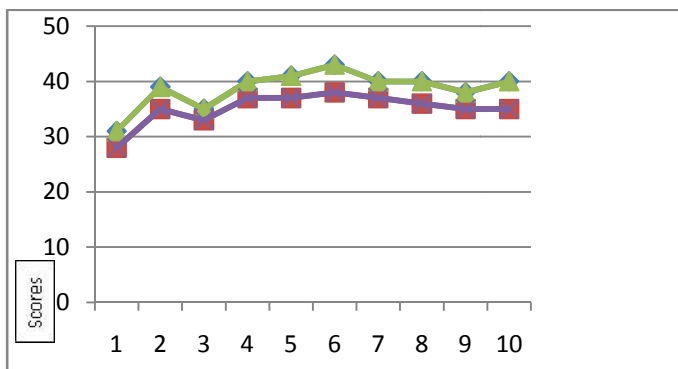

Figure 2 Performance comparisons for system and manual generated scores

Figure 2 shows the performance of the valuation of 10 students for system and manual generated scores. The X-Axis represents the student register number and the Y-Axis represents the scores of each student. It can be interpreted from the graph that there is very less difference between both the scores. Hence, our proposed system is valid.

CONCLUSION AND FUTURE SCOPE

The feature extraction methods play a major role in the text mining process. In this paper various feature extraction methods have been studied and the importance of various feature extraction methods is investigated using various sources. In the existing system, essays are corrected manually which is a tedious work. To check the performance of the proposed system, accuracy and error are used as evaluation measures. The performance of the proposed system is also evaluated with the comparison of the manual judgment. By observing the scores of each student generated by the system is better than the manual judgment. Hence it is proved that the efficiency of the proposed system is 90%.

Though our proposed system, automated essay scoring tool overcomes many problems that occur with manual evaluation such as man power, partiality etc., this system has certain limitations such as some sentences that are grammatically correct but might miss determiners where they should have been. In grammatical correction, some more rules can be defined so that sentence can be evaluated effectively, synonyms checking can be done, punctuation checking can be added, suggestion box can be added in student's user interface and semantic similarity checking of essays can be done. The speed and effectiveness in evaluating the essays can be increased. Once essay evaluation gets successful above some threshold, this automated evaluation can be applied to evaluation of other kind of questions also.

References

1. B. Wielings, "On system scalability", Research paper, volume3, pages1-4, 2003.
2. Beata Beigman Klebanov, Nitin Madanani, Jill Burstein, Swapna Somasundaran,"content importance models for scoring writing from sources" in association for computational linguistics, volume 4, pages 247-252, June 2014
3. Danielle s. Mcnamara, Scott a. Crossley, Rod d. Roscoe, Laura k. Allen, Jianmin Dai in a "hierarchical classification approach to automated essay scoring", volume 23,page number 35-59,Jan 2015
4. Duyu Tang, "A joint segmentation and classification" in conference on empirical analysis in conference on empirical methods in natural language processing, volume5, pages 477-487, 2004.
5. D.Radev, Abu Jabara, "Assessing Agreement on classification tasks", The Kappa statistics, volume 4, Pages 3-5, 2011.
6. Fei and Zhang," automatic features for essay scoring- an empirical study",in conference on empirical methods in natural language processing,volume 1,pages 1072-1077, 2016
7. Fridolin Wild,Christina Stahl, Gerald Stermsek and Gustaf Neumann in "parameters driving effectiveness of automated essay scoring with LSA", vol 5,pages 7-19,2005
8. Herzberg, "automated essay scoring in the classroom", Research paper, volume5, pages 45-57, 2011.
9. Jill Burstein and Daniel Marc, "benefits of modularity in an automated essay scoring "in creative Commons attribution teaching and research. Volume 5 ,pages 22-29,year 2000
10. John k. Lewis, "ethical implementation of an automated essay scoring system" in faculty and staff-articles & papers,vol 7, pages 47 -61, Jan 2013
11. Lokesh Kumar, "Contrasting automated and human scoring of essays", research paper, volume 2, pages 1-5, 2001.
12. Lawrence M. Ruden, "An Evaluation intelli-metric essay scoring system using responses" to GMAT, in *journal of management administration council*, volume2, pages 1-12, 2005.
13. M.Sydes, "Readability, the limitations of an approach through formulas", research paper, volume2, pages1-5, 1997.

14. Randy Elliot Bennett and Anat Ben Simon, "toward theoretically meaningful automated essay scoring" in national center for education statistics, vol 1,pages 60-77, 2005

15. Shyane Miel, "Project essay grade", Research paper, volume3, pages 1to 5, February 2014.

16. Semire Dikli,"automated essay scoring "in Turkish online *journal of distance education,* volume 7, pages 49-62, Jan 2006

17. Siddhartha Gosh, and Sameen s Fatima,"design of an automated essay grading system" in Indian context in *international journal of computer applications,* volume 1, pages 60-65, 2010

18. Yigal Attali, "Automated essay scoring with E- rater" in *the journal technology learning and assessment,* volume 4, pages 5-30, 2006

How to cite this article:

Selvani Deepthi Kavila *et al* (2018) ' Automated Essay Scoring Using Feature Extraction Method', *International Journal of Current Advanced Research,* 07(4), pp. 12161-12165. DOI: http://dx.doi.org/10.24327/ijcar.2018.12165.2132

****