

Automated Grading System using Natural Language Processing

¹Amit Rokade, ²Bhushan Patil, ³Sana Rajani, ⁴Surabhi Revandkar, ⁵Rajashree Shedge
^{1,2,3,4,5}Department of Computer Engineering, ^{1,2,3,4,5}Ramrao Adik Institute of Technology

Mumbai, India

{amipro.n, bhushan3819, sana.rajani21, surabhirevandkar29, rajashree.shedge,}@gmail.com

Abstract--

Most of the articles which cover automated grading consider keyword matching to be a crucial aspect while grading answers. Even though these are important, it is human to forget several uncommon terms and instead replace them with words that have a similar meaning. In this paper, a solution to grading of papers of theory based subjects is obtained where in Automatic Paper Grading will be performed using Natural Language Processing. Machine learning techniques like Semantic Analysis will be adopted. As a single answer can be presented in a number of ways by different students, matching keywords is inefficient. That is why, using ontology, extraction of words and their synonyms related to the domain is done which makes the evaluation process holistic as presence of keywords, synonyms, the right word combination and coverage of concepts can now be checked. The above mentioned techniques will be implemented with Ontology and will be tested on common input data consisting of technical answers. The results will be analyzed and an unbiased, high accuracy automated grading system for a theory based subject will be obtained with very little error rate which is comparable to a differential human-to-human error rate. The algorithm is designed based on the responses collected during the survey conducted amongst teachers regarding their parameters when correcting papers manually.

Keywords-- natural language processing, machine learning, ontology, semantic analysis.

I. INTRODUCTION

Natural language processing (NLP) [1] is a machine's ability of understanding the spoken human language. NLP is the process of computer analysis of input provided in a human language(natural language) and conversion of this input into a useful form of representation. The development of NLP applications is challenging because computers need the humans to "speak" to them using a precise and unambiguous programming language. NLP is a field of computational linguistics, artificial intelligence and computer science and is concerned with the interactions between computers and human language. Among all the uses of NLP, Information Extraction[2], Fighting Spam[3], Machine Translation, Summarization, Processing and Grading Answers are a few applications that are the most common. Using NLP along with semantic analysis, ontology and several other Machine Learning (ML) techniques, we aim to develop a

system which corrects theory answers in an efficient and unbiased way.

Our main aim is to provide a system for grading of online answers on the basis of content and style of writing by not only focusing on the usage of keywords but by also determining the semantic relationship between words and relevance checking thereby allotting scores based on the knowledge of the student and his/her concept clarity and not just keyword hits.

Our paper is organized as follows: The chapter II contains a study of the related work. The chapter III gives our problem definition along with the aim of our software. Chapter IV deals with the proposal of our system. Finally, chapter V gives the conclusion.

II. RELATED WORK

The implementation of essay grading softwares can be traced back to when Ellis Page marked the beginning of the automation process of grading when he developed Project Essay Grade (PEG) [4] which is the industry's most researched AI system has been used largely to help students grade essays over the past few years.

Next came Intelligent Essay Assessor (IEA) [5]. IEA was developed with the purpose of information retrieval. The idea is to identify which of several calibration documents are identical to the new input document based on the least frequent index value.

The Educational Testing Service's Electronic Essay Rater (e-rater) [5] uses syntactic variety, discourse structure and content analysis which in turn makes it a combination of PEG and IEA. To measure syntactic variety, e-rater counts the number of infinitives, subordinates, clauses and notes the occurrences of modal verbs to calculate ratios of these syntactic features per sentence and per essay. e-rater uses 60 different features for structural analysis.

An Intelligent Grading System for Descriptive Examination Papers Based on Probabilistic Latent Semantic Analysis proposed a system which scores descriptive or theory exam papers, with the help of Probabilistic Latent Semantic Analysis (PLSA). For scoring papers, semantic similarity between a student's answer paper and a model answer paper is used.

Aaron Bloomfield [7] published a paper in which he described the evolution of a system designed to allow digital grading of traditional paper-based exams or homework

assignments. In particular his system now had the ability to monitor an enormous range of data, from individual question scores, to variances for the same questions across different graders, to question and exam statistics from one semester to the next. The system boasted a number of benefits over paper based grading.

Laurie and Maiga [9] developed an Auto-assessor with the goal of automatically scoring students' short answers based on the semantic meaning and analysis. This system consists of a architecture where components are created with the aim to reduce the original sentences to their mathematical form. These mathematical forms are the intermediate stages to the preprocessing of correct answer and student answer.

Manvi Mahana, Mishel Johns, Ashwin Apte presented a paper on Automated Essay Grading Using Machine Learning [10] which uses a data set >10000 essays, obtained from kaggle.com. The essays are ramified into 8 different sets based on context and features such as no of words per essay, total sentences, total parts of speech etc from the training set essays. They used a regression analysis: simple linear regression model to gain knowledge from these features and generate parameters for testing and validation. They made use of forward feature selection algorithm to obtain a combination of features that outputs the ideal score prediction.

Ade-Ibijola, Wakama and Amadi developed an Expert System named Automated Essay Scoring (AES) for scoring free text answers [11] based on Information Extraction. This system comprises of three primary sections as: Knowledge Base, Inference Engine and Working Memory. Inference engine uses NLP technique to perform pattern matching with the help of inference rules to the data stored in knowledge base. The NLP module is made up of: a Lexical Analyzer, a Filter and a Synonyms Handler module.

Shihui Song, Jason Zhao developed the Essay Scoring software [12] wherein they created an automatic essay grading system to score around 13,000 essays from Kaggle.com, which is a platform for predictive modelling and analytics competition.

Ms. Shweta Patil presented a paper in International Journal of Engineering Research & Technology (IJERT) in which she discusses various techniques underpinned by computer assisted assessment system as well as current approaches of Computer Assisted Assessment (CAA) [13] and utilizes it as a framework for designing her new framework. She states that there are 3 main kinds of techniques for grading responses viz. Statistical, Information Extraction and Full Natural Language Processing.

M. Syamala Devi and Himani Mittal [14] published a paper in the International Journal on Natural Language Computing (IJNLC) in which she brought together several ML techniques to try and capture the latent relationship between words. The techniques included Latent Semantic Analysis (LSA), Generalized Latent Semantic Analysis (GLSA), Maximum Entropy (MaxEnt) and BiLingual Evaluation Understudy (BLEU).

Table 2.1 Comparative Analysis of Literature Survey

Sr.No	Year of Paper	Name of the proposed system	Authors	Description	Disadvantages
1	2004	An Intelligent Grading System for Descriptive Answers	Yu-Seop Kim, Jung-Seok Oh, Jae-Young Lee, Jeong-Ho Chang	Developed an intelligent grading system, which scores descriptive examination papers automatically, based on Probabilistic Latent Semantic Analysis (PLSA).	They used semantic similarity between a student's paper and a model paper and are used for building linguistic semantic knowledge rather than estimating contextual semantic similarity.
2	2010	Evolution of a digital paper exam grading system	Aaron Bloomfield	Described the evolution of a system designed to allow digital grading of traditional paper-based exams or homework assignments	This system only deals with enormous range of data, from individual question scores, to variances for the same questions across different graders, to question and exam statistics from one semester to the next.
3	2010	Design of an Automated Essay Grading System in Indian Context	Siddhartha Ghosh and Dr. Sameen Fatima	The aim of the system is to overcome the problems of influence of local language in English essays while correcting and by giving correct feedback to writers.	This system only proposes a new framework to overcome the problems of influence of local Indian languages in English essays.
4	2011	Automated essay grading using machine learning	Laurie and Maiga	Developed software with an aim to automatically score student short answers based on the semantic meaning of those answers	Each word from correct answer in canonical form is compared with the word from student response which is in canonical form and finally scores are awarded for student response. Only keyword matching is checked.
5	2012	Automated Essay Grading Using Machine Learning	Manvi Mahana, Mishel Johns, Ashwin Apte	Aims to build an automated essay scoring system using a data set of ≈13000 essays.	It is not constructive for context specific essays and features that are grammar and usage specific are not satisfied. Synonyms of the keywords are not considered while automating.
6	2012	Automated Essay Scoring (AES) an	Ade-Ibijola, Wakama	AES is based on Information Extraction (IE).	In order to compare the performance of

		Expert System (ES)	and Amadi	This ES is composed of three primary modules as: Knowledge Base, Inference Engine and Working Memory	the systems some sort of unified measure should be defined. Lack of standard data collection is identified.
7	2013	Automated Essay Scoring software	Shihui Song, Jason Zhao	They evaluated Linear Regression, Regression Tree, Linear Discriminant Analysis, and Support Vector Machines on our features	This software did not explore structural and syntactical style.
8	2014	Evaluating Student Descriptive Answers using NLP	Ms. Shweta Patil	Full Natural language processing(NLP) is used which involves parsing of text and find the semantic meaning of student answer and finally comparing it with moderators answer and assignment of the final scores.	Statistical Technique is only based on keyword matching, hence considered as poor method. It cannot tackle the problems such as synonyms in student answers, nor does it takes into account the order of words, nor can it deal with lexical variability.
9	2016	Machine Learning techniques with Ontology	M. Syamala Devi and Himani Mittal	Their system included techniques like Latent Semantic Analysis, Generalized Latent Semantic Analysis, Maximum Entropy and BLEU. These techniques were applied to Subjective Evaluation.	This work can be enhanced by using extended Ontology including the concepts of all computer science subjects.

III. PROBLEM DEFINITION

The existing grading systems consider ‘keywords’ as a major parameter while scoring answer responses, as a result of which a student who uses the most keywords will score the most. But, there are students who do not use ‘keywords’ instead use synonyms and remember the logic behind them. Scoring should be more generalized and knowledge based so that students don’t focus on rote learning and try to understand the logic behind it. Knowledge oriented learning will in turn motivate students to actually learn instead of just rote learning. Since the existing systems don’t focus on the equivalent-approach the scoring process is not exactly wholesome even though it uses semantic analysis and checks for relevance and grammar. While manual grading of papers assures that the criteria of the knowledge testing is fulfilled,

the process is quite time consuming. Other than this, each moderator has a different point of view when checking answers as some may prefer a pictorial representation of data while some may give more weightage for explaining concepts with the help of examples.

Performing evaluations through computers using intelligent techniques ensures uniformity in marking as the same inference mechanism is used for every student. In the automated systems, subjective answers are evaluated on the basis of content and style of writing. If standard keywords are found in student’s answer then the answer is presumed to be correct however, as explained above, we understand that, marking answers by just counting the hits of keywords is deficient. A more practical approach is required which evaluates the answers on the basis of not just the presence of the keywords but makes use of the semantic relationship between words and concepts. Thus, aim of our software is to make the evaluation process as user friendly as possible by making it knowledge oriented rather than “word” oriented. In this system, a solution to grading of theory papers is proposed which makes use of Natural language processing (NLP), Semantic Analysis(SA) and Ontology.

IV. PROPOSED METHODOLOGY

For a system based on keyword matching which also allows synonyms, we introduce a system which takes the students answer as input and processes it by analyzing it with the correct answer from the database. The Framework for the System Architecture is as shown in figure 3.1

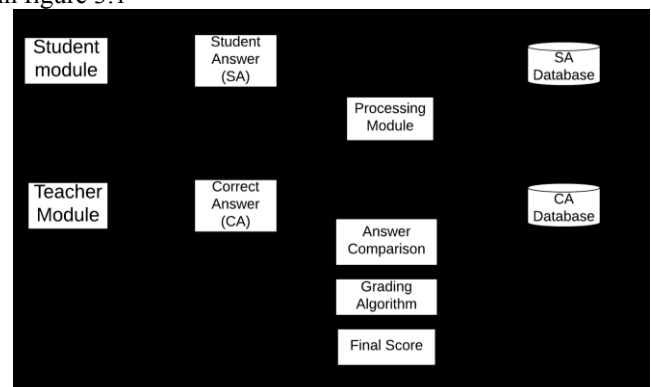


Fig 3.1 Framework of the System Architecture for Automated Grading System

- The student will login and submit the response to a particular question
- This response will be sent to the processing module where the words are mapped to the dictionary designed and will be further saved in the Student Answer database
- A correct answer database will be provided in order to compare the student’s answer Both, the Student Answer db and the Correct Answer database will be used together during answer comparison during which the variance in the used words will be checked (the variance marking will be

pre determined)

- The compared answer will now be graded using the grading algorithm (the algorithm will be designed after conducting a survey with the teachers)
- The total final score of the paper will be calculated.

The working of the various modules in figure 3.1 is explained as follows:

A. Student Module and Student Answer (SA): This module comprises of a student webpage developed using CSS and Python. Here the student is given questions on the respective theory paper. The student writes his/her answers in a non-grammar checking text box. As the test is successfully submitted ,the student entered text is then extracted from the web page in the backend and is sent to the Processing Module.

B. Processing Module and SA (Student Answer) Database: Processing Module is a Python program that takes input the raw and unstructured data (student answer) and converts it into concept map.This conversion is done by the following:

- Tokenization : Raw data is split into different words called tokens and each word in the data that contributes to little to no meaning is eliminated to reduce overall file size
- Ontology :Mapping of keywords based on facts presented in the passage is done at this stage .

C. SA Database and CA (Correct Answer) Database: It is the location where the student and model answers are stored in the form of concept maps. This can be stored in a server or a physical storage device.The format of these 2 databases is similar to each other as it makes the system easier to evaluate them

D. Score Calculation :

- Answer Comparison : The answer is compared with the help of keywords, synonyms and underlying thesaurus of the word in consideration and is then graded.
- Grading Algorithm and Final Score: The Grading Algorithm used here focuses not only on scoring a student based on similarity between SA and CA but also grading the paper based on the representation of students understanding of the subject which will be predefined. Grading is done in a sequential way. If the Student Answer perfectly matches the scope and content of Model Answer then full marks are provided.The first step here is matching the keywords , we try to find out relationship between 2 similar words and grade these accordingly(for eg. relation between “database” and “database” gets X by factor of 1) and then integrate it at the final stage to calculate the final score.

While scoring the following concepts

are used:

- Frequency(hit -miss) & percentile: The number of hits per paper are counted and highest marks are given accordingly with respect to percentile ranges.
- Parameters: Students answers are graded considering the following parameters and using the relative scoring algorithm. The parameters to grade are as follows:
-Key points
-Relevant required concept to the asked question
-Example solved of the concept
-Sequence of key points
-Correct Grammar

Thus, our system will maximize the efficiency by using the concept of knowledge based evaluation rather than considering only important keywords.

V. CONCLUSION

Our Automated Grading System focuses on a knowledge oriented approach rather than just keyword matching hit ratio. We plan on using ontology to map domains related to a given keyword. LSA and dictionary mapping will ensures that relevant answers get marks. Grammar and syntax is also checked however it will not affect the overall score of the response provided the concept is properly explained.

Using this development and enhancing it further more, a system for correcting mathematical answers, algorithms, flowcharts etc can be developed.

REFERENCES

- [1] Daniel Jurafsky and James H. Martin, Speech and Language Processing, 2e, Pearson Education, 2009
- [2] Application of Natural Language Processing - Information Extraction available at:
<https://www.lifewire.com/applications-of-natural-language-processing-technology-2495544>
- [3] Application of Natural Language Processing - Fighting Spam available at:
http://support.netdoor.com/mailgate/TMWD_IBF_Whitepaper_032604.pdf
- [4] Project Grade Essay(PEG) available at:
<http://www.measurementinc.com/products-services/automated-essay-scoring>
- [5] Intelligent Essay Assessor (IEA) and E-rater available at: <http://pareonline.net/htm/v7n26.htm>

[6] Kim YS., Oh JS., Lee JY. and Chang JH., "An Intelligent Grading System for Descriptive Examination Papers Based on Probabilistic Latent Semantic Analysis" Springer, Berlin, Heidelberg Journal, vol 3339, pp 1141-1146 available at:
https://link.springer.com/chapter/10.1007/978-3-540-30549-1_114

[7] Aaron Bloomfield,"Evolution of a digital paper exam grading system", Frontiers Education Conference, pp 27-30, 2010 available at:
<http://ieeexplore.ieee.org/document/5673292/>

[8]Siddhartha Ghosh and Dr. Sammen S Fatima,"Design of an Automated Essay Grading (AEG) system in Indian Context." International Journal of Computer Application (0975-8887), Vol.1, pp.112010

[9]Laurie and Maiga,"Automated Essay Grading Using Machine Learning" ,Technology for Education Conference (T4E), 2011 available at:
<http://ieeexplore.ieee.org/document/6004365/>

[10] Manvi Mahana, Mishel Johns and Ashwin Apte,"Automated essay grading using machine learning", Stanford University , dec-2012 available at :
<http://cs229.stanford.edu/proj2012/MahanaJohnsApte-AutomatedEssayGradingUsingMachineLearning.pdf>

[11]Ade-Ibijola, Wakama and Amadi,"Automated Essay Scoring (AES) an Expert System (ES)",International Journal of Computer Applications, vol.51, no.10,2012 available at:
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.258.7570&rep=rep1&type=pdf>

[12] Shihui Song and Jason Zhao,"Automated Essay Scoring Using Machine Learning" Stanford University-2013 available at: <http://cs229.stanford.edu/proj2013/SongZhao-AutomatedEssayScoring.pdf>

[13] Ms. Shweta M. Patil and Ms. Sonal Patil,"Evaluating Student Descriptive Answers using Natural Language Processing" Internation Journal of Engineering Research and Technology (IJERT), 2014 available at:
<http://www.ijert.org/view-pdf/8850/evaluating-student-descriptive-answers-using-natural-language-processing>

[14] M. Syamala Devi and Himani Mittal,"Machine Learning techniques with Ontology", International Journal on Natural Language Computing (IJNLC), vol5, no.2, 2016 available at :
<https://arxiv.org/ftp/arxiv/papers/1605/1605.02442.pdf>