International Journal of Research Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

A Novel Approach for Automated Essay Scoring using Vector Space Models and Natural Language Processing Techniques

[1]Srujana Inturi, [2]A Sangeetha

[1]Asst Professor, Department of CSE, CBIT, Hyderabad

[2]Asst Professor, Department of CSE, CBIT, Hyderabad

ABSTRACT— *In the field of education, essay questions is considered as the most appropriate question types for assessment compared to closed questions to evaluate the knowledge of the students. However, evaluation of answers of essay type questions consumes a long time, effort and includes unavoidable human errors. Currently there are many current Automated Essay Scoring (AES) systems however those structures are not successful sufficient to offer an correct rating and most of them aren't freely available. The recognition of this research is to present a unique approach for AES the usage of Vector Space Models (VSMs) and Natural Language Processing strategies. It employs version answer primarily based evaluation for the scoring technique. In order to deal with variations inside the college students' essay solutions, NLP techniques (lemmatization, tokenization, managing of spelling mistakes, relation of gadgets, higher and decrease case of phrases, short term decision) were used. Proposed approach does not want any pre-schooling prior to every essay questions compared to maximum of the prevailing systems.*

1. INTRODUCTION

The Intelligent Essay Assessor is business software that grades essays as it should be as professional human graders. The need for laptop-assisted evaluation of gaining knowledge of outcomes is connected to two inter-associated elements in today's training and schooling markets. First, instructors want to automate the evaluation and assessment method particularly in mass publications. Secondly, a student, especially while following a web course, may also need to assess the diploma of his or her personal studying process prior to an exam. Evaluation is a extensive concept which covers each formal and informal feedback, executed either explicitly or implicitly. In this paper, we use the time period assessment in reference to formal evaluation (i.e. measuring the studying outcomes with a numerical grade). Many of the automatic assessment structures do now not completely make use of the capacity of available technology. Too frequently, teachers and students ought to be glad with automated more than one preference questions. Essay assignments, as compared with multiple preference and choice duties, have many blessings. Written responses require college students to generate solutions which show better order thinking abilities such as synthesis and analysis.

While assessing the exceptional of college students' essays, people are vulnerable to numerous forms of mistakes, or "rater results". Such errors include the halo effect and stereotyping. The halo effect refers to a situation where the assessor's selection is prompted with the aid of the earlier affect he/she has of the pupil as opposed to at the real performance inside the take a look at this is being assessed. In stereotyping, judgments approximately the fine of the work are primarily based at the impression of the assessor approximately the group (e.g. Ethnic or gender) that the man or woman whose paintings is being assessed belongs to. Automatic grading of essays is substantially extra demanding and highly-priced than grading a couple of choice or different selection duties. Automated assessment systems must preserve the advantages of written responses, be capable of carry out as appropriately as human raters, growth the section of assessment and decrease grading-associated expenses. In addition to decreased costs, computerized assessment of essays can assist attain better accuracy and objectivity. Since computers can grade essays more unexpectedly than human beings, essay writers can get immediately feedback. An automated evaluation device is not laid low with errors because of lack of consistency, fatigue or bias. Research to automate the grading of essays has been going on because the 1960's. Several fashions have been evolved. The quality known are Project Essay Grade (PEG), e-rater and methods based totally on Latent Semantic Analysis. Two thoughts not unusual to most of these methods is that they expect human ratings to be the best estimate of the true excellent of essays and people computers can't independently determine the best. These structures need to use essays scored by means of people for developing assignment-specific scoring fashions. The use of computers in any such relatively disturbing challenge as evaluation of essays has raised numerous questions and has even generated pretty significant competition; the earliest tactics, in particular PEG, had been primarily based solely on the floor characteristics of the essay such as the duration in words and the numbers of commas. Despite superb consequences wide-range reputation turned into no longer carried out in the training community. One of the main worries became that scoring essays with such easy, oblique measures could have a leveling affect thereby getting rid of creativity.

The automated detection and correction of spelling mistakes in prose has received a considerable amount of attention (an annotated bibliography is given by using Peterson). However, users spend a large amount of time typing commands to the consumer interfaces of programs, and they make typographical errors similar to those made even as getting into prose. Although designing and imposing a nicely-engineered and tolerant person interface calls for good sized effort, it's far viable that some simple techniques, such as correcting the spelling of keywords, may be applied at low price. In discussing ideal attributes of precise consumer interfaces, assert that spelling correction in ordinary interactive packages is straightforward, for the reason that typically an error is made in a context in which most effective a limited of keywords (fewer than a hundred) are applicable. If spelling correction is actually "truthful" then there's no cause why nearly all user interfaces must not provide one of these facility, even to the quantity of retrofitting a corrector into existing applications.

2. RELATED WORK

D. Kerr, H. Mousavi and M. R. Iseli report tested the overall performance of SemScape, a rule-primarily based Natural Language Processing (NLP) gadget for routinely grading student essays for content. As against maximum different textual content mining systems, SemScape's scoring process is completely semantic and does not consist of any indicators of length or writing style. SemScape turned into capable of appropriately extract propositions from scholar writing, with a median precision of seventy nine% and an average doesn't forget of sixty five%. The 35% of propositions that had been now not efficiently extracted were break up between Wrong (20%) and Missing (15%) propositions. Wrong propositions were in large part due to wrong pronoun resolution and Missing propositions were in large part because of unusual or misguided sentence systems for which there had been no applicable TextGraph rules. We plan to feature extra TextGraph guidelines to address uncommon sentence structures and reduce the share of Missing propositions. Rules may also be brought to the pronoun resolution process to lessen the proportion of Wrong propositions with the aid of, as an instance, not permitting the pronoun "it" to clear up to a person.

Bailey. S, & Meurers. D discussed an structure which is predicated on shallow processing techniques and achieves an accuracy drawing close ninety% for content errors detection on a learner corpus they gathered from novices finishing the sporting events assigned in a actual-existence English as a Second Language (ESL) magnificence. Even for the small facts sets to be had in the area of language mastering, it seems that gadget gaining knowledge of may be powerful for combining the proof from various shallow matching features. The proper performance confirms the viability of the usage of shallow NLP techniques for which means errors detection. By developing and checking out this model, they wish to make contributions to bridging the distance between what's sensible and possible from a processing attitude and what's suitable from the angle of present day theories of language training.

In natural language processing there are numerous possibilities to pick velocity over accuracy. For example, whilst tagging a sentence you possibly can use a Hidden Markov Model tagger or an easy trigram tagger. In these times we made the choice to change accuracy for velocity. When implementing the smarts of Raphael Mudge gadget, he has opted to use simpler algorithms and cognizance on obtaining extra records and growing the first-class of information our gadget learns from. As others have pointed out, with enough records the complicated algorithms with their tricks give up to have a bonus over the less complicated methods. His real-word error detector is an instance of simplicity over complexity. With his easy trigram language model, we had been able to accurate almost a quarter of the errors within the dyslexic creator corpus. Raphael Mudge ought to improve the performance of our real-phrase error corrector actually by way of including more confusion sets. Raphael Mudge outline "do what works" as favoring combined strategies for locating and correcting errors. He used each statistical and rule-based method to hit upon real word mistakes and accurate grammar mistakes. Here he proven a manufacturing software program carrier

machine used for proofreading documents. While designing this gadget for production he referred to several regions of improvement. He defined how we carried out a comprehensive proofreading solution the usage of an easy language model and a few neural networks. He also showed that there are benefits to a software program provider from using large language models.

3. FRAMEWORK

In this paper we proposed a novel approach for automated essay scoring. This proposed system utilized Vector Space Models (VSM) and Natural Language Processing (NLP) techniques.

A. Vector Space Model

The Vector Space Model (VSM) or time period vector version is an algebraic version used for Information Filtering, Information Retrieval, indexing and relevancy rankings. It represents herbal language documents in a proper way by way of using vectors in a multi-dimensional area which has simplest nice axis intercepts. It becomes used for the primary time by the SMART Information Retrieval machine which becomes developed at Cornell University within the Nineteen Sixties. The VSM formal operational process may be divided into 3 stages. The first is file indexing. Here content material bearing phrases are extracted. The 2nd offers with weighting of indexed terms. Finally, third is answerable for calculating similarities between the input question and indexed files.

Document Indexing

Document indexing carries document preprocessing which in truth would possibly consist of stopword elimination and, or stemming. Non-linguistic strategies for indexing have also been implemented. Probabilistic indexing is based on the idea that there's some statistical difference inside the distribution of content bearing words, and stopwords. Another indexing approach is probably indexing approach which uses serial clustering of phrases in text.

Term Weighting

The term weighting for the vector area version is dealt with by means of information. There are 3 principal elements of time period weighting: time period frequency element, time period collection frequency factor and report vector period normalization component. The cease time period weight is probably made out of all or a subset of noted elements.

Obtaining Similarities

The report similarity is decided by using associative coefficients based totally on the inner made of a report vector and a question vector (queries are dealt with as ordinary documents), wherein phrase overlap suggests similarity. The internal product is generally normalized. In maximum cases the cosine coefficient, which measures the perspective among file vectors is used as the similarity measure.

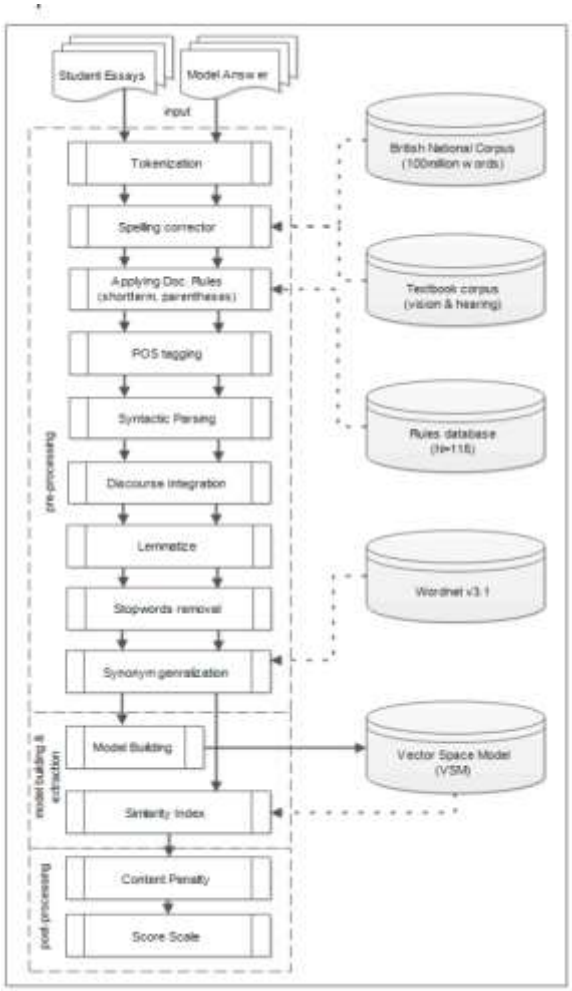

Fig1. Overview of the Proposed Framework

B. NLP Techniques

Syntactic Parsing

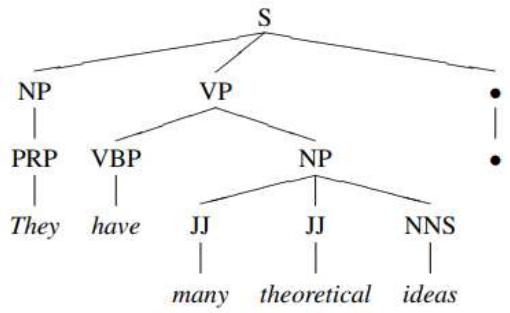

Fig2. Textual parsing

Basically, the Syntax provides policies to put together phrases to shape components of sentence and to put together those additives to shape sentences.

The parse tree is rooted with S, denoting Sentence; the sentence is composed of a noun phrase (NP) followed by a verb phrase (V P) and period. The leaves of the tree are the words in the sentence, and the preterminals (the direct parents of the leaves) are part-of-speech tags.

Lemmatization

For grammatical motives many students are the use of diverse form of the same word such as go, going, is going (morphological editions). But these phrases are associated with a same semantic base. Lemmatization can reduce those inflectional forms. In lemmatization 'go, going, is going' will become 'go'.

Stop Words Removal

Stop phrases are regularly occurring insignificant words and their presence in a textual content makes it hard to examine with other texts. Therefore, phrases inclusive of "the, for, and, in…and so on." are eliminated. Stop words list is customized from the NLTK inbuilt forestall phrases corpus.

4. EXPERIMENTAL RESULTS

In this experiment, we need to collect the data which is essay kind of data. Here, first we load the

dictionaries later, upload the sample as well as student essays.

Question 1: What is a pointer?

Table 4.1 Human and System score for Question 1

Student ID	Human Score	System Score
1	0	0
2	0	15
3	40	38.9
4	75	82.3
5	85	84.6
6	95	80.2
7	85	88.9
8	95	90.3
9	90	89
10	98	95.5

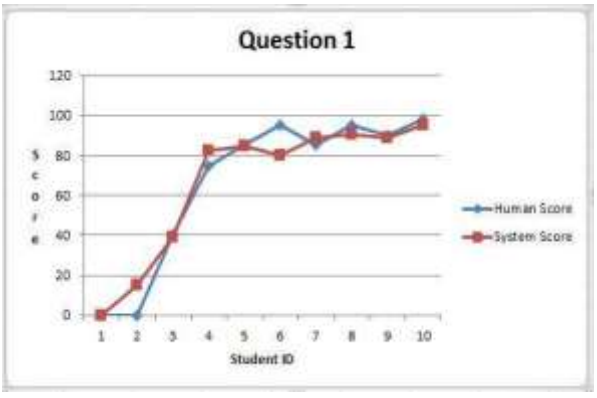

Fig 4.1 Comparison of Score for Data Set 1 Line Graph

Fig4.1 shows a graph which shows the comparison of the system score against human score for Data Set 1. It shows that the system score is almost similar to human score.

Question 2: What is Java?

Table 4.2 Human and System Scores for Question 2

Student ID	Human Score	System Score
1	85	88.8
2	90	99.5
3	95	89.6
4	5	11
5	95	90.1
6	80	90.6
7	90	94.3
8	95	96.7
9	80	88.2
10	90	93.7

Table 4.2 gives the human and system scores for the answers given by the students

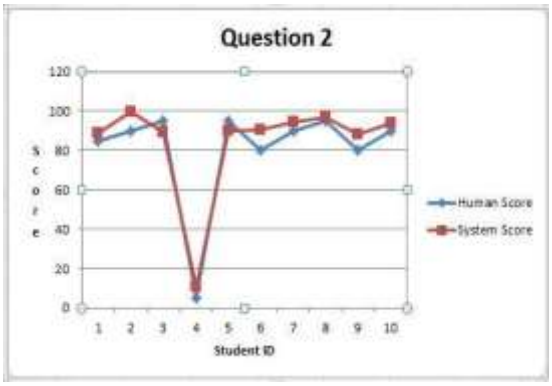

Fig 4.2 Comparison of Score for Data Set 2 Line Graph

Fig 4.2 shows a graph which shows the comparison of the system score against human score for Data Set 2. It shows that the system score is almost similar to human score.

5. CONCLUSION

In this paper, authors offered a novel approach for automating short essay grading using Latent Semantic Analysis and Natural Language Processing. On most of other Latent Semantic Analysis (LSA) based totally systems, its miles required to feed the device with the pre-marked education essays. Those training information set is taken into consideration because the semantic space and every scholar answer essay is considered as a similarity question with a view to calculate the score. In this proposed approach, Student solutions together they considered as a semantic area and similarity became calculated in opposition to the deviation from the version answer. Thus, its miles considered as a dynamic semantic area. This dynamic semantic area changed into computed the use of Vector Space Models and it became dynamically constructed upon each essay query set utilizing pupil solutions. Consequently, similarity indexing became initiated to examine each pupil solution against the model answer query.

REFERENCES

[1] D. Kerr, H. Mousavi and M. R. Iseli, "Automatically Scoring Short Essays for Content: CRESST Report 836," National Center for Research on Evaluation, Los Angeles, CA, 2013.

[2] Bailey. S, & Meurers, D. (2008), Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein, & R. De Felice (Eds.), Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Applications (pp. 107-115). Stroudsburg, PA: The Association for Computational Linguistics.

[3] Raphael Mudge, "The Design of a Proofreading Software Service" Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing, pages 24–32, Los Angeles, California, June 2010. @ 2010 Association for Computational Linguistics

[4] P. Norvig, "How to Write a Spelling Corrector," n.d.. [Online]. Available: http://norvig.com/spell-correct.html. [Accessed 17 Oct 2014].

[5] BNC, "The British National Corpus version 3 (BNC XML Edition) Distributed by Oxford University Computing Services on behalf of the BNC Consortium," URL: http://www.natcorp.ox.ac.uk/, 2007.

[6] Focus on Life Science, "California, Grade 7 (Student Edition)," Columbus, OH: Glencoe/McGraw-Hill, 2007.

[7] S. Bird, E. Klein and E. Loper, Natural Language Processing with Python, 2nd ed., ,1005 Gravenstein Highway North, Sebastopol, CA 95472: O'Reilly Media, Inc, 2009.

[8] C. D. Manning, Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, J. Steven, McClosky and David, "The Stanford CoreNLP Natural Language Processing Toolkit.," 2014.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

[9] B. Santorini, "Part-of-speech tagging guidelines for the Penn Treebank Project,Technical report MS-CIS-90-47,," Department of Computer and Information Science, University of Pennsylvania., Pennsylvania, 1990.

[10] G. A. Miller, "WordNet: A Lexical Database for English. Vol. , No. 11:," Communications of the ACM, vol. 38, no. 11, pp. 39-41, 1995.

[11] C. Fellbaum, "WordNet: An Electronic Lexical Database," MIT Press, Cambridge, MA, 1998.