SUBJECTIVE ANSWER EVALUATION USING MACHINE LEARNING*

Anshi P
Dept. of CSE
KMCT College of Engineering for Women
Calicut,India
anshipalyil@gmail.com

Dilsana
Dept. of CSE
KMCT College of Engineering for Women
Calicut,India
dilshanavv8@gmail.com

Fidha PP
Dept. of CSE
KMCT College of Engineering for Women
Calicut,India
fidhapp03@gmail.com

Sahala EK
Dept. of CSE
KMCT College of Engineering for Women
Calicut,India
eksahala@gmail.com

Anu KS
Head of Dept
KMCT College of Engineering for Women
Calicut,India
anuks@gmail.com

***Abstract*—Manually evaluating subjective papers is a difficult and time-consuming undertaking. One of the biggest obstacles to employing artificial intelligence (AI) to analyze subjective articles is inadequate comprehension and acceptance of the data. There have been several attempts to use computer science to grade students' responses. To do this, the majority of the job, however, makes use of conventional counts or certain terms. Additionally, vetted data sets are also lacking. In order to evaluate descriptive responses automatically, this paper suggests a novel method that makes use of a variety of machine learning, natural language processing, and tools, including Wordnet, Word2vec, word mover's distance (WMD), cosine similarity, multinomial naive bayes (MNB), and term frequency-inverse document frequency (TF-IDF). Answers are assessed using keywords and solution phrases, and The ability to anticipate answer grades is taught to a machine learning model. Overall, WMD outperforms cosine similarity, according to the results. The machine learning model could also be employed independently with sufficient training. The results of the experiment indicate an 88MNB model accuracy. By employing MNB, the error rate is further decreased by 1.3**

***Index Terms*—component, formatting, style, styling, insert**

I. INTRODUCTION

The manual method of subjective response evaluation in scientific fields involves a large time and resource commitment from the evaluator. Computers, however can only currently be used to assess multiple-choice tests. When it comes to the theoretical review of responses, a teacher is required to look across the answer sheet. As a result, the teacher must put more effort into grading answer sheets than into educating the students. Subjective answers can be evaluated using a variety of criteria, including the inquiry's substance and writing style. An key duty is evaluating personal responses. The quality of a valuation made by a human can vary based on their emotional state. Using intelligent ways to evaluate students using computers ensures uniformity in marking since all students use the same inference procedure. The subjective examinations serve as a basis of all university and board level exams. Attendance at subjective exams is high. The descriptive response will allow the moderator to determine how much knowledge the student has acquired throughout the course of his academic career. Due to the Covid-19 pandemic, everyone now works from home, requiring automation in the process. There is a desire for a program that can quickly assess responses and give reliable scores. Ad- ditionally, this program will help a lot of people. In assessing colleges, universities, and coaching facilities the question and effective answer. Computer evaluation of these issues is difficult, primarily because natural language is confusing. Several preparation steps must be taken before dealing with the data. Pro- cedures, such as data cleansing, must be followed. tokenization, too. Then, you can evaluate the textual information. Using several techniques, such as document similarity, concept graphs, latent semantic structures, and ontologies. You can evaluate the final result depending on language, similarities and structure . Due to its defining feature the two sides observe characteristics, context, and subjective tests both educators and students are more challenging and command- ing. A subjective response must be given in order for the each word must be care- fully

Identify applicable funding agency here. If none, delete this.

read by the checker. The final result is significantly impacted by the mental condition, level of Objectivity and weariness. Therefore, it takes much longer to allow a system perform this time-consuming and at times vital task of assessing subjective responses. and economical with resources . Machine evaluation of ob- jective answers are comparatively straightforward and useful. the program that may easily feed feedback from students one-word answers to queries. However, managing it is more harder to give subjective answers. In order to save time and money, a framework is created in this paper as a solution to this problem. They evaluate the answers in picture format based on a variety ofbkeywords, length, and distinctive character, in addition and others. We have created an algorithm that will use the algorithm to determine how many marks to give the student. We have evaluated this algorithm to make it more precise and accurate.

II. LITERATURE SURVEY

A. SUBJECTIVE EVALUATION:A COMPARISON OF SEVERAL STATISTICAL TECHNIQUES

Evaluation of subjective examinations using computerized tools has been a topic of research for more than four decades.Several statistical and mathematical tech- niques have been proposed by various researchers.The research work compares sev- eral methods proposed earlier like Latent Semantic Analysis (LSA), Generalized Latent Semantic Analysis (GLSA), Bilingual Evaluation Understudy (BLEU), and Maximum Entropy (MaxEnt) on common input data.The techniques are im- plemented using Java programming language, MatLab, and other open source tools.The database used for testing is collected by conducting tests of students of graduate level in the field of computer science.

B. A SURVEY ON THE TECHNIQUES,APPLICATIONS,AND PERFORMANCE OF SHORT TEXT SEMANTIC SIMI- LARITY

STSS is crucial for question-answering, info retrieval, and sentiment analysis.Traditional ML methods struggle with short text due to manual rules, ontologies, and graphs.Recent STSS advances lack documentation in existing literature.The systematic literature review (SLR) aims to:Address semantic similarity challenges.Identify suitable deep learning techniques.Classify contextual language models.Find relevant short text datasets.Highlight research challenges and future improvements.The SLR's goal is to aid researchers in enhancing short text semantic analysis.

C. SUBJECTIVE ANSWER EVALUATION USING MACHINE LEARNING

Subjective paper evaluation is described as a complicated and tiresome task when done manually.A lack of curated data sets for evaluating subjective papers has been observed.The paper proposes a new approach that incorporates various machine learning and natural language processing techniques, as well as tools such as Wordnet, Word2vec, WMD, cosine similarity, MNB, and TF-IDF.The proposed method involves using solution statements and keywords for answer evaluation, and a machine learning model is trained to predict grades.The results of the study suggest that WMD outperforms cosine similarity in terms of overall performance.

D. FACTORS AFFECTING SENTENCE SIMILARITY AND PARAPHRASING IDENTIFICATION

Sentence similarity determines if two sentences are similar in structure and mean- ing.Factors like sentence representation, similarity measure, and word weighting affect sentence similarity detection.The study evaluates the impact of three factors on similarity detection and paraphrasing identification.Different word embedding models, clustering algorithms, and weighting methods were tried.The clustering algorithms were applied to an Arabic paraphrasing benchmark with 1010 labeled sentence pairs.

E. SUBJECTIVE ANSWERS EVALUATION USING MACHINE LEARNING AND NATUARAL LANGUAGE PROCESSING

Using machine learning and NLP.It focus on application of techniques such as stop word removal, Latent Semantic Analysis,word2vec,bag of words ,text stemming, fuzzy approach, and document plagarism detection.

F. TEXT SIMILARITY ANALYSIS FOR EVALUATION OF DESCRIPTIVE ANSWERS

Application of Machine Learning and Data Science in education.Challenge of manually evaluating descriptive answers.Scoring parameters such as answer size, language, keywords, similarity index, and copying check.Developed by Siamese Manhattan LSTM .

G. DOCUMENT PROCESSING: METHOD OF SEMAN- TIC TEXT SIMILARITY ANALYSIS

Text similarity analysis is vital in NLP Traditional methods like BOW.Techniques like Word2Vec improved semantic analysis.

H. SENTENCE SEGMENTATION AND WORD TOK- ENIZATION

First pre-processing steps for most NLP task Text obtained can noisy and do not necessarily follow the orthographic sentence and word boundary rules .Tokeniza- tion split paragraphs and sentences into smaller units that can be more easily assigned meaning Segmentation can accurately identify sentence boundaries

I. CONCEPTUAL GRAPHS BASED APPROACH FOR SUBJECTIVE ANSWERS EVALUATION

A student's answer is evaluated by comparing it with a model answer of the question Model answers cannot exactly match with the students' answers due to variability in writing .Conceptual graphs for both student as well as model answer and compute similarity between these graphs using techniques of graph similarity measures. Based on the similarity, marks are assigned to an answer.

J. PAIRWISE DOCUMENT SIMILARITY MEASURE BASED ON PRESENT TERM SET

Most of the similarity measures judge the similarity between two documents based on the term weights and the information content that two documents share in common . Introduces a novel text document similarity measure based on the term weights and the number of terms appeared in at least one of the two documents.

K. RE-EVALUATING WORD MOVER'S DISTANCE

The word mover's distance (WMD) is a fundamental technique for measuring the similarity of two documents WMD outperforms classical baselines such as bag-of-words (BOW) and TF-IDF by significant margins in various datasets .The optimal transport (OT) distance is an effective tool for comparing probabilistic distributions .

L. AUTOMARKING: ASSESSMENT OF OPEN QUESTIONS

It will focuses on automating the assessment of open-style questions in Learning Management Systems (LMSs) using Natural Language Processing (NLP). Recent advancements in NLP, Information Extraction, and Information Retrieval offer a more linguistic approach.The system will utilize a component-based architec- ture.System uses pre-processing, synonym matching, and a grading algorithm. Two texts that use similar words would be considered semantically similar using LSA.

M. A GENERAL FRAMEWORK FOR SUBJECTIVE INFORMATION EXTRACTION FROM UNSTRUCTURED ENGLISH TEXT

The paper's main objective is to present an information extraction (IE) strategy for handling subjective information from unstructured text.Testing done in eval- uating company news and its impact on stock prices. NLP improves querying in information systems, including database queries.Four phases - part-of-speech tagging, syntactic parsing, relation generation, criteria evaluation.

N. INFORMATION RETRIEVAL WITH CONCEPTUAL GRAPH MATCHING

Conceptual graphs are used for representing text contents in information re- trieval .A method for measuring similarity between texts represented as conceptual graphs is presented .The paper discusses previous work on text comparison, introduces conceptual graphs, and explains the transformation of text into graphs .Comparing texts in these representations is crucial for various text processing applications.

O. MEASUREMENT 0F TEXT SIMILARITY

It is the basis of natural language processing tasks. It assess and evaluate various methods and techniques for measuring the similarity between texts .It play major role in information retrieval ,automatic question answering ,machine translation ,dialogue systems ,and document matching .Text similarity measurement method is described by two aspects: text distance and text representation.

P. EMPIRICAL EVALUATION AND STUDY OF TEXT STEMMING ALGORITHMS

It is preprocessing step for Natural Language Processing (NLP) to standardize word forms.Especially for languages with ambiguous structures like Arabic and Urdu.Categories : intrinsic and extrinsic.Stemming errors :Under-stemming, over- stemming, and mis-stemming.Stemming algorithms : Rule based and Statistical.

REFERENCES

[1] P. Patil, S. Patil, V. Miniyar, A. Bandal, "Subjective Answer Evaluation Using Machine Learning ,"International Journal of Pure and Applied Mathematics , vol. 118, No. 24, 2018.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, "A survey on the techniques, applications, and performance of short text semantic similarity," Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.

[3] H.Mittal, M. S. Devi, "Subjective evaluation: A comparison of several statistical techniques," Appl. Artif. Intell., vol. 32, no. 1, pp. 85–95, Jan 2018.

[4] M. Alian and A. Awajan, "Factors affecting sentence similarity andparaphrasing identification," Int. J. Speech Technol., vol. 23, no. 4,pp. 851–859, Dec. 2020.

[5] M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska and S. S. Band, "Subjective Answers Evaluation Using Machine Learning and Natural Language Processing," in IEEE Access, vol. 9, pp. 158972-158983, 2021,doi: 10.1109/AC- CESS.2021.3130902.

[6] V. Bahel and A. Thomas, "Text similarity analysis for evaluation of descriptive answers," 2021, arXiv:2105.02935.

[7] A . W. Qurashi, V. Holmes, and A. P. Johnson, "Document processing: Methods for semantic text similarity analysis," in Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA), Aug. 2020, pp

[8] A. W. Qurashi, V. Holmes, and A. P. Johnson, "Document processing:Methods for semantic text similarity analysis," in Proc. Int. Conf. Innov. Intell. Syst. Appl. (INISTA), Aug. 2020, pp.

[9] K. Sirts and K. Peekman, "Evaluating sentence segmentation and word Tokenization systems on Estonian web texts," in Proc. 9th Int. Conf. Baltic (HLT) (Frontiers in Artificial Intelligence and Applications) vol.328, U. Andrius, V. Jurgita, K. Jolantai, and K. Danguole, Eds. Kaunas, Lithuania: IOS Press, pp. 174–181, Sep. 2020.

[10] G. Jain and D. K. Lobiyal, "Conceptual graphs based approach for subjective answers evaluation," Int. J. Conceptual Struct. Smart Appl., vol. 5, no. 2, pp. 1–21, Jul. 2017.

[11] M.OghbaieandM.M.Zanjireh,"Pairwise documents imilarity measure based on present term set," J. Big Data, vol. 5, no. 1, pp. 1–23, Dec. 2018

[12] R. Sato, M. Yamada, and H. Kashima, "Re-evaluating word mover's distance," 2021, arXiv:2105.14403.

[13] L. A. Cutrone and M. Chang, "Automarking: Automatic assessment of open questions," in Proc. 10th IEEE Int. Conf. Adv. Learn. Technol. ,Sousse, Tunisia, Jul. 2010, pp. 143–147.

[14] H. Mangassarian and H. Artail, "A general framework for subjective information extraction from unstructured English text," Data Knowl. Eng., vol. 62, no. 2, pp. 352–367, Aug. 2007.

[15] M. Montes-y-Go´mez, A. Lo´pez-Lo´pez, and A. Gelbukh, "Information retrieval with conceptual graph matching," Proc. Int. Conf. Database Expert Syst. Appl., vol. 1873, Jan. 2000, pp. 312–321.

[16] A. Jabbar, S. Iqbal, M. I. Tamimy, S. Hussain, and A. Akhunzada, "Empirical evaluation and study of text stemming algorithms," Artif. Intell. Rev., vol. 53, no. 8, pp. 5559–5588, Dec. 2020.