

Study of Subjective Answer Evaluation using Natural Language Processing and Machine Learning

Sapana Sachin Baheti

Lecturer, Department of Computer Technology.
Amrutvahini Polytechnic, Sangamner, Ahmednagar, Maharashtra, India
sapanabaheti@gmail.com

Abstract: *Now a days, as we are moving towards automation there is a need for an automatic subjective answer evaluation system. Manual evaluation is a time and energy-consuming task. Currently, the online answer evaluation is available for mcq based question, hence evaluation of the theory answer is hectic for the checker. Evaluator manually checks the answer and allot the marks. Sometimes quality of evaluation may change according to mood of evaluator. The current system takes more manpower and time to evaluate the answer. It will be very helpful for educational institutions if the process of evaluation of descriptive answers is automated to capably assess student's exam answer sheets. The system will evaluate the answer based upon the number of words and number of letters from extracted text from the pre-processed data.*

Keywords: OCR, NLP, Contextual similarities, Semantic Analysis Grammatical Correction, TFID, Cosine-Similarity, Contradiction, Antonyms, Synonyms, Machine Learning

I. INTRODUCTION

Manual answer evaluation is a very tedious task. The manual checking is very time consuming process and also requires lots of manpower. Also, the paper checker is not able to give marks equally. So, this system will evaluate answer based on some keyword and also manpower will be saved. Only one has to scan the paper then, based on the keyword in the answer the system will provide the marks to the question according to the dataset present. Also, By this system, the evaluation error of the marks to the particular question will be reduced. So, this system will evaluate answer based on some keyword and also manpower will be saved.

Optical character Recognition, conversion of images of text into a character. We can segment the each characters in one word. Using classification algorithm, to recognize the text .The process of Recognition a text using classification algorithm. Result is compared with the text that should appear in that specific part of the image. Optical character recognition or optical character reader (OCR) is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example: from a television broadcast). Widely used as a form of data entry from printed paper data records whether passport documents, invoices, bank Statements, computerized receipts, business cards, mail, printouts of static data, computerized receipts, business cards, mail, printouts of static data, or any suitable documentation- it is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted)text-to-speech key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and vision. Early versions needed to be trained with images of each character, and worked on one font at a time. Advanced systems capable of producing a high degree of recognition accuracy for most fonts are now common, and with support for a variety of digital image file format inputs. Some systems are capable of reproducing formatted output that closely approximates the original page including images, columns, and other non-textual components. Optical character recognition (OCR) technology is a business solution for automating data extraction from printed or written text from a scanned document or image file and then converting the text into a machine-readable form to be used for data

processing like editing or searching. Optical character Recognition (OCR) is a conversion of scanned or printed text images, handwritten text into editable text for further processing. This technology allows machine to recognize the text automatically. It is like combination of eye and mind of human body. An eye can view the text from the images but actually the brain processes as well as interprets that extracted text read by eye. In development of computerized OCR system, few problems can occur. First: there is very little visible difference between some letters and digits for computers to understand. For example it might be difficult for the computer to differentiate between digit “0” and letter “o”. Second: It might be very difficult to extract text, which is embedded in very dark background or printed on other words or graphics.

II. LITERATURE SURVEY

[1] A model to evaluate subjective Answer papers using Semi Automated Evaluation technique is created. For that first they create Question base which contains question type, sub type, question and marks. Then Answer base is created with model answer. Evaluated answer is mapped using hash index which referred as question number. The student answer is evaluated by considering semantic meaning and length of the sentence.

[2] The proposed system is designed to evaluate answers for five students providing five different answers. The standard answer is stored in the database with the keywords, meaning and the description of that answer. Then each answer is evaluated by matching the keywords as well as its synonyms with the standard answer. It will also check the grammar and spellings of the words. After the evaluation, the answer is graded depending on the correctness of it.

[3] E-assessment system developed to checks the answer sheet of the student and provides marks to the same. The system consists of an algorithm that compares the student’s answer against three reference answers given by three different faculties and the answer with most close results and with highest precision is taken into consideration and marks are allocated accordingly. Algorithm based on TFIDF,

Grammar check, WMD, cosine and Jaccard Similarity. Both the answers need not be exactly the same or word to word. This approach can be a quick and easy way for the examiners by reducing their workload.

[4] The proposed system receives solution sets from admin and student’s answers. Then stop words are removed from them in order to generate keywords. After keyword generation it checks for similarity and by calculating similarity it also checks the relation of the keywords along with sentences with the sentences in the dataset which finds the exact similarity and correctness of the sentence with the datasets. If the sentences match with the datasets it generates similarity score as per the overlapping percentage. It also checks the synonyms and similar words before relating the keywords in order to increase the accuracy of the overlap. Data duplication technique is used to compare the previous answers submitted by students and on the basis of uniqueness of answers, grades are generated.

[5] The proposed system is developed an algorithm which will evaluate theoretical answers and give marks according to the keyword matching which will reduce manual work and saves time with faster result evaluation. A person should collect the answer copy from the student and scan it. The machine will take the image as input and will evaluate the answer based on the length of the answer and important keywords covered which are specified by the teacher with each answer which is to be evaluated.

[6] Developed an algorithm which will evaluate theoretical answers and give marks according to the keyword matching which will reduce manual work and saves time with faster result evaluation. A person should collect the answer copy from the student and scan it. The machine will take the image as input and will evaluate the answer based on the length of the answer and important keywords covered Which are specified by the teacher with each answer which is to be evaluated.

[7] The handwritten text image, given as input to the handwriting conversion module is extracted and converted to machine encoded text using Optical Character Recognition (OCR) Algorithm. The machine encoded text is given as input to the evaluation module. Evaluate module evaluates the answer based on grammatical meaning of the sentences; number of keywords matched and gives marks as output.

III. NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a branch of AI that helps computers to understand, interpret and manipulate human languages like English or Hindi to analyze and derive its meaning. NLP helps developers to organize and

structure knowledge to perform tasks like translation, summarization, named entity recognition, relationship extraction, speech recognition, topic segmentation.

1) Stemming

Stemming is the process of reducing inflection (prefix, suffix) in words to their root forms such as mapping a group of words to the same stem even if the stem itself is not a valid word in the Language. It reduces a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming will remove suffix and prefix from the word.

2) Lemmatization

To extract the proper lemma of word, it is necessary to look at the morphological analysis of each word. Rather than stemming, lemmatization provides lemma which has a dictionary meaning. The key to Lemmatization is linguistics.

3) Stopwords

Stopwords are the most common words in any natural language. For the purpose of analyzing text data and building NLP models, these stopwords might not add much value to the meaning of the document.

IV. WEIGHTING MODULE FOR EVALUATION OF ANSWER

4.1 Answer length

In the existing system, we observed that the student gets a full mark on writing only keywords. And so, answer length is also an important factor. As student may write all keywords and grammatically correct short sentences. Such answer would get full marks for keywords and grammar but less marks for short length.

4.2 Keyword Matching

Keywords is another important factor which we consider to check whether student cover all important concept (key₁ point) in his answer or not. If the keyword matches the maximum marks will be allotted as per the keyword considering its correctness. But if few the keywords differ with the model answer keywords then the marks will be deducted accordingly.

4.3 Contextual Similarity

We have designed an algorithms which give similarity between two answers and if two answers are not contextual similar then it will give 0 else it give percentage of similarity. We studied different python libraries which give high similarity ratio even though the answers are not contextually similar.

4.4 Semantic Similarity

When two answers are similar in context, we use a specialized API to check if they share the same meaning. This tool compares the deeper sense of the text, not just the words used. It helps us understand if the ideas in both answers align well. Although this method improves evaluation, it's important to know it may have limitations like dealing with unclear language. Still, it's a useful tool for getting a better understanding of students' responses.

4.5 Contradiction

The Existing systems were unable to find contradicting sentences, because the system removed important stopwords (not, never, doesn't...etc.) while preprocessing and also not considering key antonyms present in the answer. We use POS tagger to find contradicting sentences from student answer against model answer.

4.6 Grammar Check

To check and detect grammatical mistakes and spelling errors Grammar, Textgear API is used in our system. When the text is sent to Grammar Textgear's API, it returns a number of grammatical error present in text

4.7 Cosine Similarity

Cosine similarity is a standard point of reference to measure how similar the documents are irrespective of their size. The similarity is represented as the dot product of two text vectors.

4.8 Antonyms and Synonyms

NLTK WordNet is a lexical database of English language. Nouns, verbs, adjectives and adverbs that are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. The lemmas() function of the synset. It returns synonyms as well as antonyms of that particular synset. Synonyms are considered to find semantic similarity International Research Journal of Engineering and Technology between two different answer having same contextual

meaning. Antonyms are considered to find conflicting or contradicting sentences from student answer and model answer.

V. MACHINE LEARNING ALGORITHMS

5.1 NAIVE BAYES

Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. It is a good and mostly used classification Algorithm. However it provides less accuracy for this model. Naive Bayes uniquely assumes that features are independent of each other within each class. Naive Bayes believes that each feature (like a word in text) is chosen independently of others when deciding on a class (like spam or not spam). This makes the math easier and helps with tasks like sorting emails. It's like assuming each part of an email (like words) doesn't depend on the others when deciding if it's spam or not.

5.2 Decision Tree Classification

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

VI. DIFFERENCE BETWEEN MACHINE LEARNING AND NLP

1. Machine learning is a subset of AI that allows a machine to learn from past data without explicitly programming it.

NLP is also a subset of AI, but it requires machine learning to be used effectively.

2. Machine learning aims to find patterns in data and then make predictions based on those patterns, which are often complex, to answer business questions, detect and analyze trends, and help solve problems.

The goal of NLP is for computers to comprehend texts/languages in the same way that humans do. Once this is accomplished, computer systems will be able to comprehend, infer, summarize, translate, and generate accurate, natural human text, and language.

3. Machine learning's main applications include online recommender systems, Google search algorithms, Facebook auto friend tagging suggestions, etc.

The main applications of NLP are speech recognition, sentiment analysis, translation, chatbots, market intelligence, automatic grammar checking, etc

4. The machine learning algorithm must be monitored, and regularly maintained to keep it running.

Because the system is designed for a single and specific task, it cannot adapt to new domains and problems due to limited functions, and it isn't completely dependable.

5. Machine learning requires a large amount of data to adequately capture the relationships that may exist between input features as well as between input features, and output features.

NLP requires machine learning to provide accurate responses, and automate some of these processes.

6. Machine learning employs two techniques:

Supervised learning: This involves training a model on known input and output data to predict future

outputs. Unsupervised learning: This involves discovering hidden patterns or intrinsic structures in input data.

NLP employs two techniques:

Syntactic analysis: Involves analyzing a string of symbols in natural language, computer languages, or data structures that follow formal grammar rules.

Semantic analysis: Checks whether the generated parse tree follows the programming language's rules.

7. Machine learning is primarily concerned with accuracy and pattern recognition.

NLP is concerned with computer-human language interactions, specifically how to program computers to process, and analyze large amounts of natural language data.

VII. CONCLUSION

In this paper, I have studied an algorithm which will evaluate theoretical answers and give marks according to the keyword matching which will reduce manual work and saves time with faster result evaluation. A person should collect the answer copy from the student and scan it. The machine will take the image as input and will evaluate the answer based on the length of the answer and important keywords covered which are specified by the teacher with each answer which is to be evaluated.

Subjective Answer Checker System Using NLP and Machine Learning would be helpful for online universities, schools and colleges level exam.

Subjective Answer Evaluation software assign mark to subjective question based on Answer length, keyword matching, Grammar check, cosine similarity and Contextual similarity against Model answer provided by faculty and student answer. We also developed algorithm to find contradicting statements from student answer with model answer. Even answer written by student does not match word to word with model answer provided by teacher, but still our system capable to evaluate answer based on context

REFERENCES

- [1] Chhanda Roy, Chitrita Chaudhuri, "Case Based Modeling of Answer Points to Expedite Semi-Automated Evaluation of Subjective Papers", in Proc. Int. Conf. IEEE 8th International Advance Computing Conference (IACC), 2018, pp. 85-9.
- [2] Aditi Tulaskar, Aishwarya Thengal, Kamlesh Koyande, "Subjective Answer Evaluation System", International Journal of Engineering Science and Computing, April 2017 Volume 7 Issue No.4.
- [3] Saloni Kadam, Priyanka Tarachandani, Prajakta Vetaln and Charusheela Nehete, "AI Based E-Assessment System", EasyChair Preprint, March 18, 2020.
- [4] Vishal Bhonsle, Priya Sapkal, Dipesh Mukadam, Prof. Vinit Raut, "An Adaptive Approach for Subjective Answer Evaluation" VIVA-Tech International Journal for Research and Innovation Volume 1, Issue 2 (2019).
- [5] Prince Sinha, Sharad Bharadia, Ayush Kaul, Dr. Sheetal Rathi, "Answer Evaluation Using Machine Learning" Conference-McGraw-Hill Publications March 2018
- [6] Rosy Salomi Victoria D, Viola Grace Vinitha P, Sathya R, "Intelligent Short Answer Assessment using Machine Learning" International Journal of Engineering and Advanced Technology (IJEAT), Volume-9 Issue-4, April 2020.
- [7] Sakshi Berad, Pratiksha Jaybhaye, Sakshi Jawale, "AI Answer Verifier", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue 01; Jan 2019
- [8] V. Lakshmi and Dr V. Ramesh, "Evaluating students descriptive answer using natural language processing and artificial neural networks", in International Journal of Creative Research Thoughts (IJCRT), Volume 5, Issue 4, December 2017, pp. 3168- 3173.