Automatic Assessment of Descriptive Answers for Online Examination using Semantic Analysis

Vaishnavi Adhav[1], Itisha Adhav[2],Tanuja Bhujbal[3], Sibiya S Babu4, Dr. S. N. Zaware[5]
DEPARTMENT OF COMPUTER ENGINEERING AISSMS Institute of Information Technology
Kennedy Road, Near R.T.O., Pune 411 001, Maharashtra (INDIA).

Abstract- Question Answering is a specialized form of information retrieval. Given a collection of documents, a Question Answering system attempts to retrieve correct answers to questions posed in natural language. Open-domain question answering requires question answering systems to be able to answer questions about any conceivable topic. Such systems cannot, therefore, rely on hand crafted domain specific knowledge to find and extract the correct answers. Closed-domain question answering deals with questions under a specific domain and can be seen as an easier task because NLP systems can exploit domain-specific knowledge frequently formalized in ontology's. Alternatively, The proposed system provides the semantic based QA evaluation using Artificial Neural Network (ANN). Our system also comes under closed domain QA where we are supposed for accessing online based subjective examination.

Keywords- Online Subjective Examination, Sentiment Analysis, Machine Learning, Paraphrase, Evaluation process.

I. INTRODUCTION

Question answering is a specified form of information retrieval. Our work comes under closed domain question answering. We are working on assessment of answer for online subjective examination. Examination and evaluation are part of every course module so are even in online examination, objective based examination are already available but subjective examination are in need of time as subjective assessment is considered as best way of evaluation of ones subject understanding & knowledge. In our research work we have discussed two issues related to the answer method i.e. length & paraphrasing. And have obtained a pattern extraction by creating a sequence for a given answer. Our system has a centralized file system which includes the reference material as well as the model answer for questions. These are used for matching and evaluating a candidate's answer. For every correct answer a confidence factor of being positive is assigned when the required selective pattern of candidates answer matches with the model answer**.**

II. LITARATURE SURVEY

Maram et al. [1] introduces an Automatic evaluation of an essay (AEE) system which is written in Arabic. The system presents a hybrid approach which integrates the LSA and rhetorical structure theory (RST) algorithm. LSA method supports the semantic analysis of the essay, and the RST to evaluate the writing method and the cohesion of the essay. The LSA method finds the similarity ratio among two texts even if they do not include similar words. The system processes input essay into two phases is a training phase and testing phase. The training phase is made up of three parts: calculating the average of words per essay, calculating the most ten visible words on a given topic and applying LSA algorithm. The testing phase passes through a number of processes:1) calculating LSA distance.2) calculating the number of a vernacular.3) calculating a number of repeated sentences.4) calculating the length of the essay.5) calculating number of spelling mistakes.6) applying RST algorithm.7) checking cohesion of essay related to the topic.

Anirudh et al. [2] propose an automated evaluation system for descriptive English answers that contains multiple sentences. The system evaluates the student's answer with an answer-key for questions of professional courses. It depends on a group of algorithms for natural language processing which are Wu and Palmer, Longest Common Substring (LCS),the similarity score of LCS combined with a similarity score of Wu-Palmer technique using the similarity matrix method. LSA uses SVD on the similarity matrix that formed of both sentences. SVD produces two vectors representing two sentences. The similarity between two sentences is computed using cosine similarity. Pure PMI-IR combines all similarity scores of word pairs among sentences in one value using the similarity matrix method. The multi-class Logistic regression technique combines results of all five techniques to produce a score for the answer.

DimitriosAlikaniotis et. al. [3] Automatic Text Scoring Using Neural Networks System introduce a model that forms word representations by learning the extent to which specific words contribute to the text's score. Using Long-Short Term Memory networks to represent the meaning of texts, we demonstrate that a fully automated framework is able to achieve excellent results over similar approaches. In an attempt to make our results more interpretable, and inspired by recent advances in visualizing neural networks, we introduce a novel method for identifying the regions of the text that the model hasfound more discriminative.

According to system [4] proposes an approach of evaluation of online descriptive type students' answers using Hyperspace

Analog to Language (HAL) procedure and Self-Organizing Map (SOM) method. To evaluate students' answer, the student writes the answer and sent as input to HAL. HAL constructs a high dimensional semantic matrix from a collection of an n-word vocabulary. Method for construct matrix through motivation a window of length "1" by the corpus through one-word increment. HAL ignores sentence boundaries, punctuation and converts each word to numeric vectors expressing information on its meanings for words. Inside window computes the distance between two words is "d", then computes "(l-d+1)" which denotes the weight of an association among two words.

kumaran and Sankar [5] propose a technique of an automated system for assessing the short answers using ontology mapping. Three stages of assessing the short answers are RDF sentence builder, ontology construction, and ontology mapping. In the first stage, the system constructs the RDF sentence for every sentence in student answer and model answer after reading the model answer and student answer as input in plaintext form. The system parses each sentence and builds the grammatical relationships each sentence. It uses Stanford typed dependency parser to represent dependency relationships. In the second stage, the RDF sentences are as input to ontology constructor to construct an ontology for them. The authors use sequential and coordinate links to construct RDF graph for the RDF sentences.

Raheel and Christopher [6] propose a system that provides a novel approach for automated marking of short answer questions. To compute the grade for the student's answer, authors introduce the architecture for the system that is composed of three phases to address the student's answer. Three phases are 1) spell checking and correction that is implemented by an Open Source spell checker like JOrtho.2) parsing the student's answer using the Stanford Parser. This statistical parser can be creating parses with high accuracy. The parser offers the following results which are the part of speech tagged text and design dependency grammatical relations among singular words. 3)The Third phase of the processing answer is a comparison between the tagged text with syntactical structures specified by authors in Question and Answer Language.

The Automatic marking system for a student's answer examination of the short essay was introduced by Mohd et al. [7]. The system applied to sentences were written using the Malay language that requires technique to process it. The technique mentioned in which is the syntactic annotation and the dependency group to represent the Grammatical Relations(GR) from Malay sentences.

A new automated assessment algorithm for assessing the Chinese subjective answers was proposed by Runhua et al. [8]. The algorithm called Automated Word and Sentence Scoring (AWSS) assesses the student answers for the level of word and sentence. From fundamental problems of the Chinese, Natural Language Processing is the word segmentation, but this problem solved by the Institute of Computing Technology, the Chinese Lexical Analysis System (ICTCLAS). It assesses the student's answer to the standard answer in two phases as follows:1) compute similarities between two words depend on How-Net.

According to system [8] to compute the similarity of sentences depending on dependency structure among words of a sentence. This phase parses the sentence by the language technology platform (LTP) to find out the dependency structure of the sentence. The method of computing dependency structure is finding a valid pair which is a noun, verb or subjective linking to the head of the sentence. Then, computing the sentence similarity based on dependency structure as functions mentioned by this system.

Xia et al. [9] design automatic scoring algorithm for a subjective question. They use the idea of a one-way approach degree depending on the closeness theory of fuzzy mathematics. The authors are calculating the closeness of two fuzzy sets which are set "A" denoted by the standard answer string and set "B" denoted by the student answer string. A fuzzy set is an ordered collection from a single character that decomposed from a string. To compute a one-way approach degree between two fuzzy sets "A" and "B", "B" contain n characters and one-way approach degree denoted by δ (B, A)=m/n whereas m denotes by the effective sum number of the set B in each element in the set A. δ (B, A) introduce B close to A unidirectional closeness. The introductory algorithm provides the aim of the system.

Zhenming et al. [10] propose a novel web-based online objective examination system for computer science education. This system conducts the examination and auto-marking of objective questions and operating questions. The system transmits answers and questions into the bit stream after encoding to ensure security and intrusion. It is the password protected system and camera are used to monitor the activities of students. The auto-grading system can automatically grade the answers, that are collected from the examination system.

III. OBJECTIVES OF SYSTEM

The Objective of the proposed application is as follows:

- To identify the correct answer possibility for all question set.
- Identify the patterns of answer and find the similarity for classification
- Implement the NLP approach for identifying the words probability.
- Finally use the classification approach for label classification.

IV. PROPOSED METHODOLOGY

The proposed system will provide data preprocessing, data normalization, stop word removal, porter stemmer algorithm,

TF-IDF calculation etc. has used for features selection.This module will classify all answers base on similarity weights. Confidence in the correctness of an answer can be increased in a number of ways. Finally it will shows the system accuracy with clustering result with accuracy as well as false positive ratio.

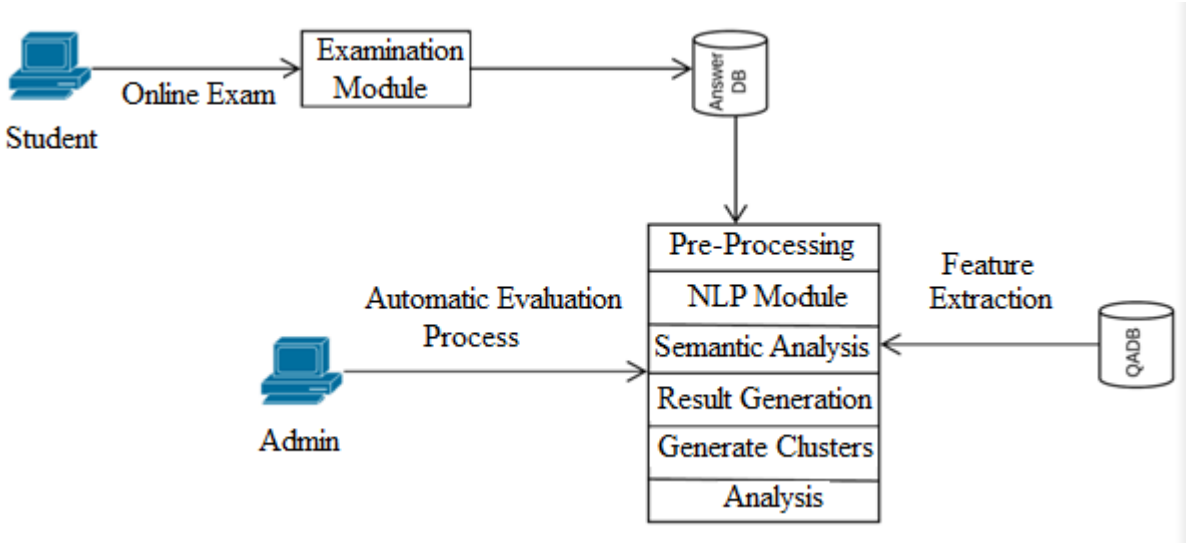

Fig.1: Proposed System Architecture

Module 1: Question Answering
In this module, admin first upload the QA dataset which is technically verified according to the specific subject. The each question have specific marks as well as desired answer in answer set. The connectivity of each question and answer has managed by different indicators.

Module 2: User authentication and validation
This module admin creates the multiple users and set the authorization process, once user has validate by authentication system, he can give the exam and add the answer for various questions.

Module 3: Subjective Examination
Basically this module set the exam with time and scheduling, according to that each user provide the answer for desired questions.

Module 4: Features extraction using NLP
This module will provide data preprocessing, data normalization, stop word removal, porter stemmer algorithm, and features extraction etc has used for features selection. Once an answer has been identified, the shallow parsing performed is leveraged to extract only the relevant word or phrase in answer to the question. The use of a part-of-speech tagger can help to enable recognition of answer candidates within identified model answer. Answer candidates can be ranked based on measures of distance between keywords, numbers of keywords matched and other similar heuristic metrics.

Module 5: Semantic Analysis
This module will classify all answers base on similarity weights. Confidence in the correctness of an answer can be increased in a number of ways. One way is to use a lexical resource like WordNet (Synonyms) to verify that a candidate response was of the correct answer type, and it will evaluate the marks automatically according to the current weights (it will be 0.01 to 0.99).

Module 6: Classification Results
Finally it will shows the system accuracy with clustering result with accuracy as well as false positive ratio.

V. SYSTEM ANALYSIS

Algorithm1 : Stop word Removal Approach
Input: Stop words list L[], String Data D for remove the stop words.
Output: Verified data D with removal all stop words.
Step 1: Initialize the data string S[].

Step 2: initialize a=0,k=0

Step 3: for each(read a to L)

If(a.equals(L[i]))

Then Remove S[k]

End for

Step 4: add S to D.

Step 5: End Procedure

Algorithm2 Stemming Algorithm.

Input : Word w

Output : w with removing past participles as well.

Step 1: Initialize w

Step 2: Intialize all steps of Porter stemmer

Step 3: for each (Char ch from w)

If(ch.count==w.length()) && (ch.equals(e))

Remove ch from(w)

Step 4:if(ch.endswith(ed))

Remove 'ed' from(w)

Step 5: k=w.length()

If(k (char) to k-3 .equals(tion))

Replace w with te.

Step 6: end procedure

Algorithm3 TF-IDF
Input : Each word from vector as Term T, All vectors V[i…n]
Output : TF-IDF weight for each T
Step 1 : Vector = {c1, c2, c3….cn}
Step 2 : Aspects available in each comment
Step 3 : D = {cmt1, cmt2, cmt3, cmtn}
and comments available in each document
Calculate the Tf score as
Step 4 :tf (t,d) = (t,d)
t=specific term
d= specific document in a term is to be found.
Step 5 :idf = t → sum(d)

Step 6: Return tf *idf

VI. RESULTS AND DISCUSSION

For the system performance evaluation, calculate the matrices for accuracy. The system is executed on java 3-tier architecture framework with INTEL 2.7 GHz i3 processor and 4 GB RAM with supervised learning approach. We have tested around 100 samples with of QA dataset using proposed NLP and ANN. The below figure. 2 shows the accuracy of proposed system of experiment 1, it shows the accuracy as well as error rate of system.

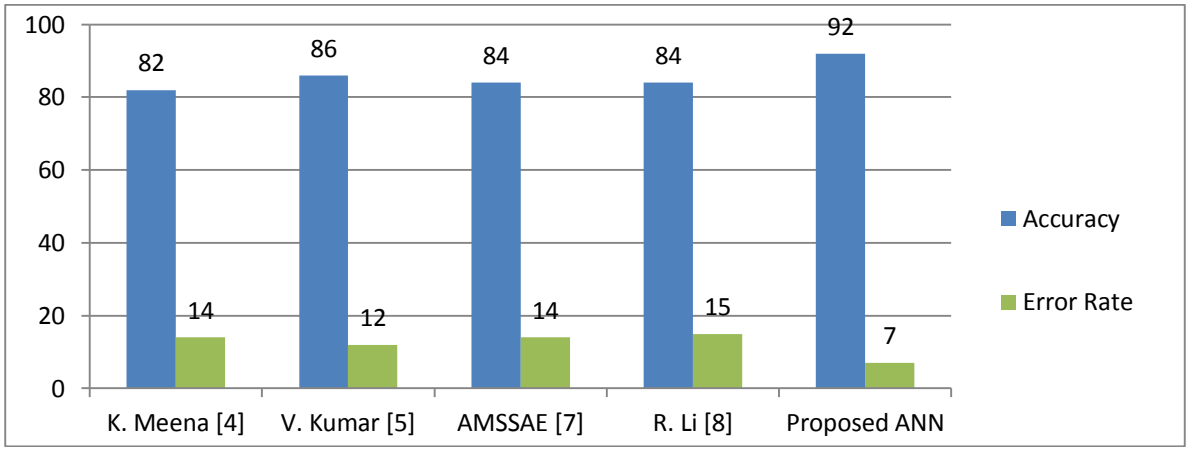

Fig.2: Proposed overall system evaluation.

The figure 2 shows the system comparative performance with someexisting systems. Proposed system provides the accuracy better than [4,5,7,8]

VII. CONCLUSION

The proposed system used for semantic evaluation for subjective answer using Artificial Neural Network (ANN).System consist two modules admin and students, admin first upload the standard QA dataset and student also load the answers dataset. System uses NLP approach for during the preprocessing phase and ANN for generate the similarity weight of system. According the gained weight system will assign the proportional marks to specific answer and finally generate the class cluster for individual student. Apply sentiment analysis with polarity dataset is the future work of system.

VIII. REFERENCES

[1]. M.F. Al-Jouie, A.M. Azmi, "Automated Evaluation of School Children Essays in Arabic," 3rd International Conference on Arabic Computational Linguistics, 2017, vol.117, pp.19-22.

[2]. A.Kashi, S.Shastri and A. R.Deshpande, "A Score Recommendation System Towards Automating Assessment In Professional Courses," 2016 IEEE Eighth International Conference on Technology forEducation,2016,pp.140-143.

[3]. Alikaniotis D, Yannakoudakis H, Rei M. Automatic text scoring using neural networks. arXiv preprint arXiv:1606.04289. 2016 Jun 14.

[4]. K.Meena and R.Lawrance, "Evaluation of the Descriptive type answers using Hyperspace Analog to Language and Self-organizing Map", Proc. IEEE International Conference on Computational Intelligence and Computing Research, 2014, pp.558-562.

[5]. Vkumaran and A Sankar," Towards an automated system for short-answer assessment using ontology mapping," International Arab Journal of e-Technology, Vol. 4, No. 1, January 2015.

[6]. R. Siddiqi and C. J. Harrison, "A systematic approach to the automated marking of short-answer questions," Proceedings of the 12th IEEE International Multi topic Conference (IEEE INMIC 2008),Karachi, Pakistan, pp. 329-332, 2008.

[7]. M. J. A. Aziz, F. D. Ahmad, A. A. A. Ghani, and R. Mahmod, "Automated Marking System for Short Answer examination (AMSSAE)," in Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on, 2009, pp. 47-51.

[8]. R. Li, Y.Zhu and Z.Wu,"A new algorithm to the automated assessment of the Chinese subjective answer," IEEE International Conference on Information Technology and Applications, pp.228 – 231,Chengdu, China, 16-17 Nov. 2013.

[9]. X.Yaowen, L.Zhiping, L.Saidong and T.Guohua," ," 2nd InternationalSymposium on Computer, Communication, Control and Automation, Vols. 347-350, pp. 2647-2650, 2013.

[10]. Y. Zhenming, Z. Liang, and Z. Guohua, "A novel web-based online examination system for computer science education," in 2013 IEEE Frontiers in Education Conference (FIE), 2003, vol. 3, pp. S3F7–10.