Evaluation of Answers using Computational and AI Models

Project Team Members: Navya Sibi, Vedant Phuse, Devdatta Bidkar, Snehal Patil

Supervisor Name: Dr. Kishore Bhurchandi
Department of Electronics and Communication Engineering
Visvesvaraya National Institute of Technology, Nagpur-INDIA

Problem Formulation

We were given a set of problem statements like:

1. Fitness Device Biomedical Signal Processing
2. Automatic Subjective Answer Evaluation
3. Image Inpainting and Outpainting

Out of these, we decided to choose the Automatic Subjective Answer Evaluation problem statement, which is essentially Evaluation of answers using Computational and AI Models

Motivation

1. Subjective paper evaluation demands the checker check every word in the descriptive answer for many students
2. This becomes a monotonous and time-consuming process
3. Automatizing of evaluation process makes the process impartial and increases its efficiency

Introduction

The uneven distribution of evaluators and examinees in the society, impartial and favored marking, monotonous and time-consuming evaluation process has created a need for a system that would solve these problems and make the process efficient. Moreover, a subjective answer demands the checker check every word of the answer for scoring actively, and the checker's mental health, fatigue, and objectivity play a massive role in the overall result. Thus, we are developing a system using the concepts of Natural Language processing that will programmatically evaluate long answers.
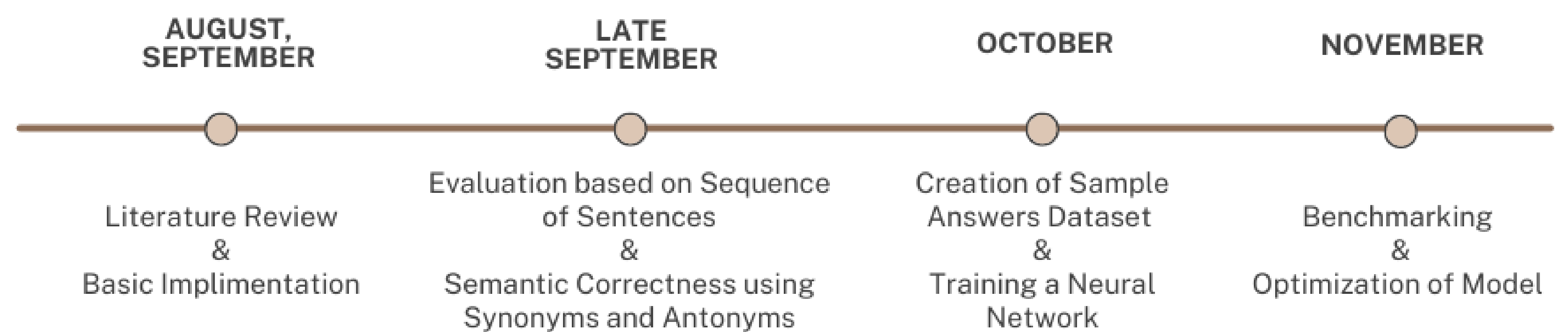
References

1. M. F. Bashir, H. Arshad, A. R. Javed, N. Kryvinska, and S. S. Band, "Subjective answers evaluation using machine learning and natural language processing," *IEEE Access*, vol. 9, pp. 158 972–158 983, 2021.
2. C. N. Tulu, O. Ozkaya, and U.Orhan, "Automatic short answer grading with semspace sense vectors and malstm," *IEEE Access*, vol. 9, pp.19 270–19 280, 2021.

Literature Survey

1. Subjective Answers Evaluation Using Machine Learning and NLP

   - In this paper, the answers are evaluated using the solution and provided keywords using various Similarity-based techniques.
   - Two score prediction algorithms are proposed, which produce up to 88% accurate scores. Various similarity and dissimilarity thresholds are studied, and various other measures such as the keyword's presence and percentage mapping of sentences are utilized to overcome the abnormal cases of semantically loose answers.

2. Automatic Short Answer Grading With SemSpace Sense Vectors and MaLSTM

   - This paper makes the use of SemSpace and Manhattan LSTM (MaLSTM), based on multi-layer LSTM Networks for Automatic Short Answer Grading (ASAG).
   - The model was first trained on the open source Mohler dataset
   - A new dataset, CU-NLP was developed for this study, where initial data was converted into SemSpace Sense Vectors after which it was trained on a model built on the MaLSTM Network.
   - Two identical LSTMs are trained for Student and Model answers, and each calculates vectorial similarity to find the Manhattan distance which is used to find the similarity value.

TimeLine

AUGUST, SEPTEMBER	LATE SEPTEMBER	OCTOBER	NOVEMBER
Literature Review & Basic Implimentation	Evaluation based on Sequence of Sentences & Semantic Correctness using Synonyms and Antonyms	Creation of Sample Answers Dataset & Training a Neural Network	Benchmarking & Optimization of Model

Methodology

A crude logic of text-preprocessing and basic answer evaluation was implemented using Python which included:
1. Loading the sample answers from an excel file and loading it into a dataframe.
2. Preprocess the student answer chunk of text: remove stopwords and punctuation and lemmatize the text sentence-wise.
3. For each word in the preprocessed text, we apply the evaluation algorithm which involves checking for
– presence of keywords
– sequence of keywords
– sequence of sentences
– presence of synonyms/antonyms and their effect
4. Allot marks according to a marking scheme decided.
5. Create a dataset that has the set of answers. Train a neural network that can evaluate the answers based on the 2 inputs i.e model answers and student answers.
6. Testing the model and increasing the efficiency.

Results

- The crude logic of answer evaluation aided in alloting marks to a text file in the basic way
- The marks alloted were on the basis of presence of keywords and their order
- The total marks for an answer is the sum of the individual marks of each sentence

Conclusion

1. Generation of keywords, Removal of stop words, Evaluation of answers by comparison of keywords and the relevance of the order is done
2. From literature review, we observe that LSTMs are the preferred ways to solve NLP based AI Problems. The quality of the dataset plays a key role in deciding the performance of the model.