AUTOMATED ESSAY GRADING SYSTEM USING DEEP LEARNING

Mr. P. V. Hari Prasad[1], G. Himaja[2], Ch.Abhigna[3], K.Saroja[4], K. Nikhil Venkatesh[5]

[1]Associate Professor, Dept. Of Computer Science and Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India

[2, 3, 4, 5] Bachelor of Technology, Dept. Of Computer Science and Engineering, Dhanekula Institute of Engineering and Technology, Andhra Pradesh, India

---***---

ABSTRACT: *In the view of educational institutions, assignments or essays play an important role in assessing the ability of students to understand and recall the topics explained to the students. The manual evaluation of these papers takes a lot of effort and time of the evaluators hence resulting in a time consuming process. The solution to grade a large number of papers effectively within a stipulated time is to let the machine do the grading. The automated grading system will not only reduce the time of evaluation but comparing it with human scores will also make the score more realistic. The project aim is to develop a system which grades an essay or a paper without any manual involvement. When an essay is loaded into the proposed grading system, the system accepts the essay given as the input and grades it using deep learning techniques and its layers such as LSTM and dense layers. This proposed system is useful while grading papers in smart schools and educational institutions.*

Keywords: Grading, evaluation, score, deep learning, layers, educational institutions

1. INTRODUCTION

Essays play a vital role in trying out or assessing instructional achievement, integration of ideas and capacity to take into account of a student. If a student has better understanding of a concept, then he or she can apply the concept in a better way compared to those who don't have a clearer understanding of the concept. Grading students' assignments creates awareness about his or her learning ability, understanding of concepts and their application to the teachers. If the time taken for evaluation of these essays is somewhat reduced, then teachers can give more attention to preparing more notes for the next classes, gather up more conceptual examples for the better understanding of the students.

Automated essay grading or scoring has been the topic within the discipline of laptop science for the reason that 1960's. The earliest beginning of automated essay grading system has been found inside the works of Ellis Batten Page. He has positioned forth the possibility of scoring essays using computers and also posted a paper in 1968. Many advanced their own versions of computerized essay grading systems. Peter Foltz and Thomas Landauer developed a system the use of a scoring engine referred to as Intelligent Essay Assessor. It turned into used for scoring essays for undergraduate courses at some point of 1970's and continues to be getting used as an assessor for diverse state and country wide exams.

Our project aims at developing a model using deep learning techniques which automatically grades an essay upon submission. We grade our essay on a scale of 1-10 and the result would be displayed on the screen using a message box.

2. RELATED WORK

Earlier systems of essay grading used a dataset of 13000 essays obtained from Kaggle.com. These essays were divided into 8 sets based on the context. These systems considered features such as word count, number of long words, sentence count, and parts of speech counts and so on. In the earlier systems, the data was split and validated using 5 fold cross validation to train the model. The model was built from the linear regression algorithm to which the folds resulted from the cross validation were given as input to train and test the model. There were other essay grading systems which used algorithms such as Support vector machine (SVM), Na�e Bayes classi fier, and Random forest and other machine learning algorithms.

3. DATASET

For this project, we have used essays written by students from grade 7 to grade 10. We have acquired this dataset from the William and Flora Hewlett Foundation which was provided for Automated Student Assessment Prize Competition on Kaggle. This dataset consists 8 sets of essays which are written in ASCII text. All the essays from these 8 sets are generated from a single prompt. The dataset consists of around 12000 essays. The dataset we get is a TSV file i.e., a tab separated value file consisting of essay Id, essay set, essay and scores of two raters in 2 domains and the resolved score between the two raters in each domain. We have used 10% of testing and the other 90% for training our model. Each essay in the given dataset is approximately of 150 to 550 words in length.

4. PROCESS

Our approach to tackle this problem involves the following steps:

4.1. Data Gathering

We have acquired our data from the the William and Flora Hewlett Foundation from the Kaggle.com (as mentioned in the above section: Dataset). We have collected a total of around 12000 essays in which each essay is in the ASCII text format. The approximate length of each essay is around 150 to 550 words and is ideal for this project. We made use of only 8000 essays form the given dataset.

4.2 Data preprocessing:

To preprocess the data, we have imported and used a package belonging to Python programming language called NLTK (Natural Language Toolkit). In data preprocessing, we first remove all numbers, whitespaces and default stopwords (will, being, so, few, as, yours, had, have, and, not). Stopwords are words which do not play a part in the meaning of a sentence. So, it makes sense to remove them as they do not possess much value to the meaning. Then we split the “cleaned essay” into tokens.

From this, we *extract features* like word count, character count, average word length, misspelled words, prevalence of the submitted essay and POS tagging. To get the misspelled word count, we have compared our data with a text file called big.txt which consists of large collection of words. As deep learning or machine learning models cannot understand text data when given as input, we have to convert out text into a format which the model can understand and take in to process it, which is a numerical format or vector format. To produce feature vectors, we have used model architecture from word2vec called Continuous Bag Of Words (CBOW) which takes in text corpus as input and pops out feature vectors as its output. The cause and usefulness of Word2vec is to group the vectors of comparable words together in vector space. That is, it detects similarities mathematically. Word2vec model creates vectors which might be allotted numerical representations of word functions, features consisting of the context of character words.

4.3. Training Model

To train the model we, 5 fold cross validation is applied on the dataset. The model used here is a deep learning model called Sequential model. The reason to choose Sequential model is that it is a simple model which is just a linear arrangement of layers chosen. We can add our layers in the order we want to perform our computations. The layers we have implemented are 2 LSTM (Long short-term memory) layers and a single dense layer.

LSTM stands for Long-Short Term memory layer which is artificial recurrent neural network architecture. By stacking or using a 2 layered LSTM model, we have multiple hidden memory cells. So our networks become deeper thus allowing our network to perform better as the success of the learning sometimes depends on the depth.

A dense layer is a simple regular layer of neurons in a neural network. Each neuron takes the input from all the neurons in the previous layer, thus fully connected. We have also used Dropout technique with a value of 0.5 thus enabling it to drop a fraction of neurons to minimize overfitting as much as possible.

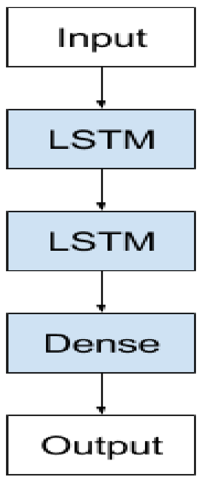

Fig-2: Model architecture

In the output layer we have used relu activation function (Rectified Linear Unit or ramp function) as no normalizing of training labels is required. The output layer then produces the output of the essay as a discrete value or a single integer.

5. RESULT

We have created a user interface using the Tkinter package from Python which takes an essay as input. The essay taken as input will be pre-processed i.e. numbers, symbols and stopwords are removed; the cleaned essay will be converted into feature vectors. The feature vectors will be passed as input to the neural network consisting of the above mentioned layers and the score or grade of the essay according to the features considered (word count, character count, average word length, misspelled words, and prevalence) will be displayed on the screen via a small message box.

Consider that the input essay given is,

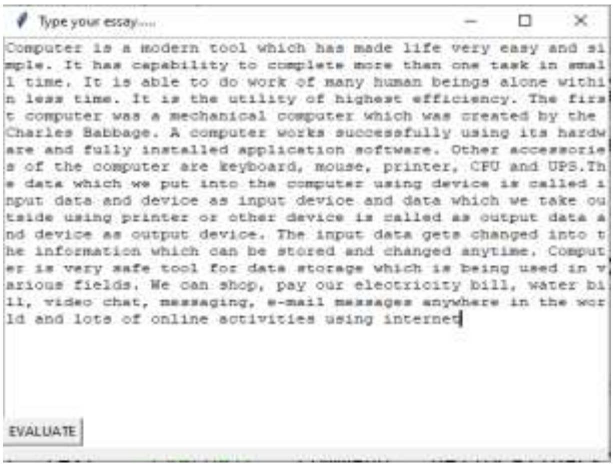

Fig-3: Input given

The output displayed will be as:

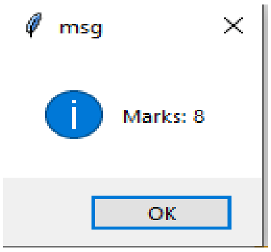

Fig-4: Score for the given essay

6. CONCLUSION

Our model gives out good predictions on the basis of the features considered such as word count, sentence count, prevalence, parts of speech count. The performance on context and sentiment rich essays can be made better by better training our model with larger and more complex datasets and advanced NLP features. The average weighted kappa we have achieved by using this process is above 0.5 which is normally ideal.

7. FUTURE WORK

Although, we have predicted the scores of the essays submitted, even according to the prevalence, there is further scope for this project. It can be made better by training the model with larger and multiple datasets and also on more complex features to achieve better performance and accuracy. By performing this project, we have come to know the potential of neural networks in processing natural language problems or issues which becomes useful in real world.

ACKNOWLEDGEMENT

We would like to express our gratitude to our mentor, Mr. P.V. Hari Prasad, our project co-ordinator Mr. K. Sandeep, our head of the department Dr. S. Suresh for their unwavering support and guidance through every stage. We also express our gratitude to the faculty of our college for giving us unlimited support and motivation.

REFERENCES

[1] Manvi Mahana, Mishel Johns, Ashwin Apte. Automated Essay Grading System using Machine Learning, CS229 Machine Learning-Autumn 2012.

[2] Y.Harika, I.Sri Latha, V.Lohith Sai, P.Sai Krishna , M.Suneetha. Automated Essay Grading System using Feature Selection. p-ISSN: 2395-0072. Volume: 04 Issue: 03 | March -2017

[3] Abhishek Suresh, Manuj Jha. Automated Essay Grading using Natural Language Processing and Support Vector Machine. IJCAT - International Journal of Computing and Technology, Volume 5, Issue 2, February 2018

[4] Hassan, Samer and Mihalcea, Rada. “Semantic relatedness using Salient Semantic Analysis” 2012 http://www.cse.unt.edu/ rada/papers/hassan.aaai11.pdf (Accessed: 25 April 2012)

[5] Kaggle. ”Develop an automated scoring algorithm for student-written essays.” (2012). https://www.kaggle.com/c/asap-aes

[6] Drolia, S., et al., Automated Essay Rater using Natural Language Processing. International Journal of Computer Applications, 2017. 163(10).