**ENLP Project Report (Group - 4)**

**Project Definition:**

**Short Answer Grading with Semantic Similarity and Alignment based features.**

Description: Automatic Short Answer Grading(ASAG) is the problem of assigning holistic grades to the student answers. ASAG problem can be modeled as a regression problem as well as classification problem. In classification based formulation, the task can be modeled as n-way classification problem.

**Dataset Description**

* **Wikipedia Dump**

- The Wikipedia dump – 500Mb contains the relevant pages from the Wikipedia corpus for

the questions under consideration.

- Currently the targeted questions are from Data-Structures Domain

* **Expert Graded Question Answer Dump between 0 to 5**

- It contains 89 questions from Data-Structures Domain, along with reference answers.

- Total 2000 student responses to these questions.

- Every answer is graded between 0 to 5.

**Flow of Work**

**1. Wang2Vec Model**

- It considers 2 major modifications to the original models in Word2Vec that improves word

embeddings obtained for syntactically motivated tasks.

- It makes the network aware of the relative positioning of the context words.

- Word2Vec uses

1. Continuous Bag of Words Model

2. Skip-gram Model.

- Wang2Vec uses

1. Cwindow

2. Structured Skip gram Model.

- The Wikipedia corpus is supplied to Wand2Vec model, to get the word embeddings

for the words of the corpus keeping semantic analysis in focus.

This model is trained for window size of 5, using structured skip gram method, which

returns the word embedding vectors of size 50 indexed for each word of Wiki corpus

**2. Feature Vector Construction and Output Class**

- The dataset of student given answers is divided into 2 parts –

* + - 1. Training Set – 70%
    - II. Testing Set – 30%

**Following Operations are performed on both the sets:**

* Feature I - Using model, constructed in step I, The vector of Reference answer - Vr and Student Given answer - Vs is computed, Using these vectors the cosine similarity is computed as Feature1
* Feature II – The Edit distance between Vr and Vs.
* Feature III – The Wordnet Distance between Vr and Vs.
* Feature IV – The TERP – Term Edit Rate between Vr and Vs.
* Output Class – The grade (Score) between 0 – 5, given by experts.

Output Class is actual score given for the question by expert – it ranges from 0 to 5 at distance of 0.5, so it defines total 11 classes for classification

**3. Training Neural Network**

Using Keras python library, we have trained neural network with 3 hidden layers, for 100 epochs, over trained dataset and estimated the accuracy on testing dataset.

Result Achieved : Accuracy - 52.3%