CS 491 NLP Project Report 1

Group No.: **G18**

Group Members:

Sr No.	Name	Enrollment No.	Email
1	Akshat Khanna	U101115FCS046	akshat.khanna@st.niituniversity.in
2	Chinmaya Bansal	U101115FCS077	chinmaya.k.bansal@st.niituniversity.in
3	Mayank Singh	U101115FCS200	mayank.singh@st.niituniversity.in
4	Sangamesh Kotalwar	U101115FCS210	sangameshn.kotalwar@st.niituniversity.in

Reference Paper Title: **Summarizing Lengthy Questions**

Authors: Tatsuya Ishigaki, Hiroya Takamura and Manabu Okumura

Summary Of Paper

In this research paper, authors have proposed the task of question summarization. They first analyzed question-summary pairs extracted from a Community Question Answering (CQA) site [here *Yahoo! Answers Comprehensive Questions and Answers version 1.0*], and found that a proportion of questions cannot be summarized by extractive approaches but requires abstractive approaches. Summarizing a question, which can often be lengthy, helps respondents understand the question. Approaches used in generic summarization tasks are often classified into two different types: extractive and abstractive. Extractive approaches select and order units, which are usually sentences or words, from the input text. Abstractive approaches, rather than selecting units, generate a summary using words not found in the input text. By using the data, they trained extractive and abstractive summarization models, and compared them based on ROUGE scores and manual evaluations. The experimental results show an abstractive method using an encoder-decoder model with a copying mechanism achieves better scores for both ROUGE-2 F-measure and the evaluations by human judges.

Work done so far

1. Reading of paper and going through references to understand the concept more.
2. Since the paper was based on the dataset *Yahoo! Answers Comprehensive Questions and Answers version 1.0,* which was not publicly available, so we applied for the dataset as research purpose. Now we have got access to the dataset so we will be proceeding with the analysis.

Plan of work and responsibilities:

October

1. Obtain dataset
2. Clean dataset
3. Filter out data according to our requirements and assumptions considered in the reference paper.
4. Going through reference papers given in the research paper we are following.

Responsibilities:

Akshat - Cleaning and filtering dataset 1, referring other related paper

Chinmaya - Cleaning and filtering dataset 1, referring other related paper

Mayank - Cleaning and filtering dataset 2, referring other related paper

Sangamesh - Cleaning and filtering dataset 2, referring other related paper

November

1. Analysis using extractive approach
   -> Rule based approach
   -> Machine learning based approach
2. Analysis using abstractive approach
   -> Vanilla encoder - decoder
   -> Encoder - decoder with attention
   -> Encoder - decoder with copying mechanism

Responsibilities:

Akshat - Rule based approach, vanilla encoder - decoder, encoder - decoder with copying mechanism

Chinmaya - Rule based approach, vanilla encoder - decoder, encoder - decoder with attention

Mayank - Machine learning based approach, encoder - decoder with attention

Sangamesh - encoder - decoder with copying mechanism, machine learning based approach.

December

1. Experiments
2. Evaluation
   -> Evaluation with ROUGE
   -> Manual Evaluation
3. Qualitative Analysis
4. Final report submission

Responsibilities:

Akshat - Experiment and evaluation with ROUGE

Chinmaya - Evaluation with ROUGE and manual evaluation

Mayank - Manual evaluation and qualitative analysis.

Sangamesh - Qualitative analysis, cross checking evaluations and final report