Project workplan
Text summarizer for chat messages in a group chat

Andrea Auletta   Davide Baggio   Marco Brigo
Marco Bernardi   Sebastiano Sanson

May 7, 2024

Contents

1 Short summary

Our Natural Language Processing (NLP) project aims to automate the summarization of group chat messages, providing users with concise summaries of conversations without the need to read through all messages. Imagine a scenario where you're part of bustling team chat or a lively online community. Messages fly back and forth very quickly about various topics and discussions. Obviously chat logs accumulate rapidly and it might happen that you miss some important points, requiring significant time and effort to read through the entirety of the convesation history. Our aim is to save users time and cognifitive effort when facing situations like this. By leveraging state-of-the-art natural language processing techniques our system will analyze the content of group chat messages, summarizing the most relevant information, the key points, main topics, and notable highlights of the discussion, presenting users with a condensed version of the conversation that is easy to digest and comprehend.

The idea is to try several existing approaches that can be useful to reach an optimal goal. Then we will try to create our own version by using the techniques that we found to be the most effective.

2 Tasks

Number	Task	Mandatory	Participants
1	Study of the state-of-the-art of the topic	Yes	Auletta, Baggio, Bernardi, Brigo, Sanson
2	Dataset collection	Yes	Auletta, Baggio, Bernardi
3	Data preprocessing	Yes	Brigo, Sanson
4	Summarization of text messages	Yes	Auletta, Baggio, Bernardi, Brigo, Sanson
5	Summarization of images	No	Auletta, Brigo, Sanson
6	Summarization of audio	No	Baggio, Bernardi
7	Test and evaluation of the model	Yes	Auletta, Baggio, Bernardi, Brigo, Sanson
8	Documentation	Yes	Auletta, Baggio, Bernardi, Brigo, Sanson

3   Deliverables

Task	Deliverable	Description	Measurable objectives
1	Report of the state-of-the-art	Each member of the group will be assigned a certain number of papers that they will have to read and summarize. At the end of this process, a final report condensing all the information will be produced, showcasing the state-of-the-art in the field.	We will report about at least 15 papers in our state of the art analysis (3 for each of us)
2	Dataset	A dataset of chat messages will be collected	Collect a minimum number of chats, the number will not be small but will be chosen depending on the model that will be developed
3	Preprocessed and analyzed dataset	The dataset will be preprocessed and analyzed to be used in the model. A notebook file will be produced	
4	Model (text messages)	An implementation of the model that can summarize text messages will be produced	
5	Model (images)	The model will be able to understand images and will summirize their content	
6	Model (audio)	The model will be able to understand audio and will summirize their content	
7	Evaluation report	The model will be tested and evaluated, and a report will be produced	
8	Documentation	Will be produced all the documents which are useful to the comprehension of the project	

4 Metrics