Group No.: G14

Group Members:

No.	Name	Email @st.niituniversity.in
1.	Ajinkya Bedekar	ajinkya.bedekar@st.niituniversity.in
2.	Dhruva Agarwal	dhruva.agarwal@st.niituniversity.in

Reference Paper Title:

MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations

Authors:

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, Rada Mihalcea

Summary of the paper in your own words:

Emotion recognition in conversations (ERC) is a challenging task that has recently gained popularity due to its potential applications. Until now, however, there has been no large scale multimodal multi-party emotional conversational database containing more than two speakers per dialogue. To address this gap, the authors proposed the Multimodal EmotionLines Dataset (MELD), an extension and enhancement of EmotionLines. MELD contains approximately 13,000 utterances from 1,433 dialogues from the TV-series Friends. Each utterance is annotated with emotion and sentiment label, and encompasses audio, visual, and textual modalities. The authors proposed several strong multimodal baselines and show the importance of contextual and multimodal information for emotion recognition in conversation.

MELD has multiple use-cases. It can be used to train emotion classifiers to be further used as emotional receptors in generative dialogue systems. These systems can be used to generate empathetic responses. It can also be used for the emotion and the personality modeling of users in conversations.

By being multimodal, MELD can also be used for training multimodal dialogue systems. Although by itself it is not large enough to train an end-to-end dialogue system, the procedures used to create MELD can be adopted to generate a large-scale corpus from any multimodal source such as popular sitcoms. The authors define multimodal dialogue system as a platform where the system has access to the speaker's voice and facial expressions which it exploits to generate responses. Multimodal dialogue systems can be very useful for real time personal assistants such as Siri, Google Assistant where the users can use both voice and text and facial expressions to communicate.

In this work, the authors introduced MELD, a multimodal multi-party conversational emotion recognition dataset. They described the process of building this dataset, and provided results obtained with strong baselines methods applied on this dataset. MELD contains raw videos, audio segments, and transcripts for the multimodal processing. Additionally, they also provide the features used in our baseline experiments. They believe this dataset will also be useful as a training corpus for both conversational emotion recognition and multimodal empathetic response generation. Building upon this dataset, future research can explore the design of efficient multimodal fusion algorithms, novel ERC frameworks, as well as the extraction of new features from the audio, visual, and textual modalities.

Work done so far:

By now, we have completed the following steps:

- Read the research paper completely
- Understood each and every part of the research paper
- Downloaded the dataset
- Checked the dataset to see what all things are available in it
- Started to develop an algorithm for implementing Emotion Recognition in Conversations.

Plan of work for the rest of the semester including who will do what:

Plan of work for the rest of the semester:

- Complete developing the algorithm: By 15th October 2019
- Implement the algorithm using Python Code: By 31st October 2019
- Train the model with dataset available with research paper: By 15th November 2019
- Testing the trained model: By 20th November 2019

Who will do what:

- Ajinkya Bedekar: Develop and implement the algorithm for emotion recognition in conversations
- Dhruva Agarawal: Test and train the model on dataset available with research paper