

Chapter 1

Literature Survey

1.1 Related Work

Previous works have extensively studied sentiment and emotion analysis in language, while the relationship between emotion and sarcasm has been largely unaddressed. Most of the existing research has focused on the detection of sarcasm [Joshi et al., 2017, 2018]. Research studying the impact of sarcasm on sentiment analysis [Maynard and Greenwood, 2014] showed that sarcasm often has a negative sentiment, but the associated emotion(s) is important to frame the response and follow-up communication. For tweet analysis, NLP researchers have tried to detect sarcasm and perform sentiment analysis together, while some try to improve sentiment analysis performance using sarcasm detection [Bouazizi and Ohtsuki, 2015].

Castro et al. [2019] is the first multimodal data set annotated for the sarcasm detection task. This data contains a balanced set of 345 sarcastic and 345 non-sarcastic video utterances. Castro et al. [2019] is a subset of Multimodal Emotion Lines Dataset (MELD) data set [Poria et al., 2018] which is a multimodal extension of EmotionLines data set [Chen et al., 2018]. MELD contains about 13,000 utterances from the TV-series Friends, labeled with one of the seven emotions (anger, disgust, sadness, joy, neutral, surprise, and fear) and sentiment. Chen et al. [2018] is a textual data set comprising 29,245 utterances from the series Friends and private Facebook messenger dialogues.

In Chauhan et al. [2020], authors annotated the MUsTARD dataset with emotions and sentiment, and showed that in a multi-task setting, the primary task for sarcasm detection yielded better results with the help of secondary tasks of emotion and sentiment analysis. Since our study purely focuses on the understanding the speaker’s emotion while using sarcasm, we used their annotated basic emotions, as well as annotate the dataset with arousal and valence to understand the degree of emotion. Research on sarcasm detection has led to the creation of some datasets. i-Sarcasm is a tweet dataset with intended and perceived sarcasm annotations for sarcasm detection [Oprea and Magdy, 2020]. The

pragmatic categories of sarcasm, metaphor, and irony are related to each other [Musolff, 2017].

1.2 Emotion Recognition Approches

1.2.1 Textual Emotion Recognition

Keyword Spotting based [Joshi et al., 2016a] approach is dependent on the existence of specific emotional terms in the input that conveys emotion. This method makes use of Emotion Lexicons or dictionaries such as WordNet Affect, NRC Emotion Lexicon, DepecheMood, and others. These are the emotion keywords to look for in a statement to link it with a certain emotion. Lexical Affinity [Joshi et al., 2016a] based technique improves on the Keyword Spotting method in that, in addition to looking for emotional elements, random words are given some probabilistic affinity in this method. These are Rule-based approaches, there are several deep learning based approached also present

1.2.2 Speech Emotion Recognition

Now we will discuss some of the initial approaches for using the audio modality for emotion classification.

MDRE based Approach

Initial model proposed by [Yoon et al., 2018], authors have proposed a novel **multimodal dual recurrent encoder model** that simultaneously utilizes text data, as well as audio signals, to permit a better understanding of speech data. Their model encodes the information from audio and text sequences using dual RNNs and then combines the information from these sources using a feed-forward neural model to predict the emotion class. Their extensive experiments show that their proposed model outperforms other state-of-the-art methods in classifying the four emotion categories, and accuracies ranging from 68.8% to 71.8% are obtained when the model is applied to the IEMOCAP dataset. They have Textual features and audio they have only used MFCC and Prosodic features.

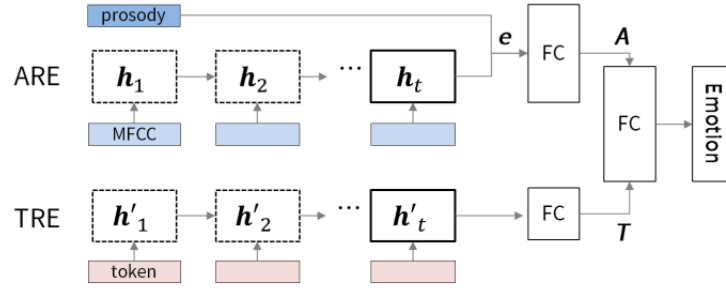


Figure 1.1: Multimodal dual recurrent encoder. The upper part shows the ARE, which encodes audio signals, and the lower part shows the TRE, which encodes textual information.

CNN-LSTM based approach

Another approach for speech emotion recognition is proposed by [Etienne et al. \[2018\]](#). Authors have built a neural network for recognizing emotions in speech, using the IEMOCAP dataset. Unlike the prior results, in order to measure the model performance, we performed 10- fold cross-validation, which is more appropriate for the IEMOCAP dataset. To address the issues of scarcity and class imbalance they have employed data augmentation by means of VTLP and minor class oversampling. Following the modern trends in speech analysis, they have used a mixed CNN-LSTM architecture, exploiting the capacity of convolutional layers to extract high-level representations from raw inputs. In this paper they have investigated the effect of batch normalization, an indispensable tool in most image recognition tasks. In order to preserve the signal structure as much as possible, they performed the normalization layer-wise as well as batch-wise.

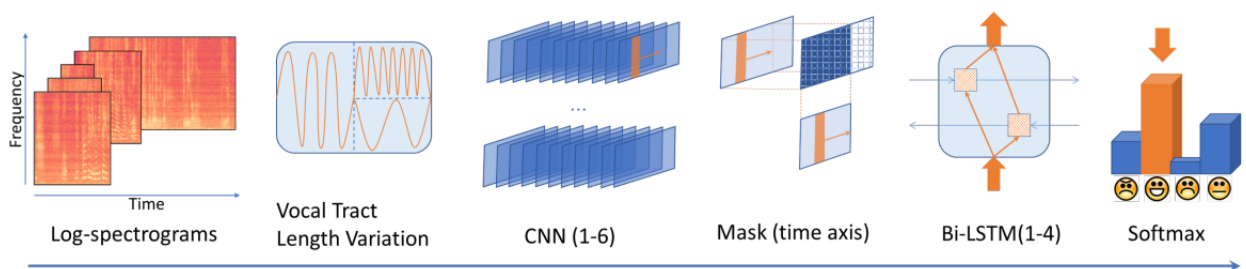


Figure 1.2: CNN-LSTM based architecture

1.3 Sarcasm Detection Approches

Automatic Sarcasm Detection has previously been defined as a classification task. There has been a lot of work done in this area. Existing methods can be divided into three categories:

Rule Based Approach: Riloff et al. [2013] published a classic work that presents a contrast model that employs a bootstrapping-based technique to automatically learn lists of positive sentiment words and negative situation phrases from sarcastic tweets. Congruity-based methods that employ incongruity as a feature have also been proposed in the past. This includes semantic inconsistency, emotion inconsistency, language model inconsistency, and contextual inconsistency.

Statistical Approach: In addition to text-based characteristics such as n-grams, punctuation, and intensifiers, incongruity in text has been utilised as a feature. Then, for classification, a statistical model such as SVM is utilized. Riloff et al. [2013] also used an SVM classifier to predict sarcasm. The article combines SVM and the rule-based method and gets an F-score of 51%.

Deep Learning Based Approach: The usage of deep-networks such as CNN and LSTM, etc, is part of many methods. Joshi et al. [2016b] proposes a sarcasm detection deep learning technique using LSTM/CNN and F.R.I.E.N.D.S. conversational transcripts of a popular tv program. The model contains two sub-networks: a network for utterances and a network for sequences. The bidirectional RNN with GRU cells are being used in another model suggested by Kolchinski and Potts [2018]. The model integrates the inclination of author to be sarcastic in two methods to increase their performance: using a Bayesian prior and by using author embeddings.

Multimodal Sarcasm Detection

The first multimodal approach for sarcasm detection is proposed by Castro et al. [2019]. They have given the multimodal learning method for the identification of sarcasm. MUS-tARD, a new data collection comprising of sarcastic and not-sarcastic video gathered from various sources, was introduced for study on this topic. They have also created models using SVM on three modes, including text, voice and visual inputs. The findings of baseline studies have confirmed their theory that sarcasm detection requires multimodality. In several evaluations, they found that the multimodal variants were shown to significantly outperform their unimodal counterparts, reducing the relative error rate of up to 12.9%. Chauhan et al. [2020] has developed a deep learning-based multi-task model that solves all three issues at the same time, namely sentiment analysis, emotion analysis, and sarcasm detection. With the aid of emotion analysis and sentiment analysis, the two secondary tasks in their scenario, their suggested multi-tasking framework provides improved performance for the primary task, sarcasm identification, and registered an F1 score of 72.57 in the sarcasm detection challenge.

1.4 Summary

In this chapter, we have discussed about the previous approach for emotion recognition and sarcasm detection. In next chapter, we will discuss about the basics of deep learning.