Report

Introduction

Paper Title: ***Understanding Emotions in Text Using Deep Learning and Big Data***

When you read, "Why don't you ever text me", does it convey an angry emotion or sad emotion?

Understanding Emotions in Textual Conversations is a hard problem in absence of voice modulations and facial expressions.

Description

In this task, you are given a textual dialogue i.e. a user utterance along with two turns of context, you have to classify the emotion of user utterance as one of the emotion classes: happy, sad, angry or others.

Data

The training data set contains 15K records for emotion classes i.e., happy, sad, and angry combined. It also contains 15K records not belonging to any of the aforementioned emotion classes. For Test Data, 2703 records of unlabelled data is provided to be used to evaluate models we create. Final testing would be done on as of yet unreleased data.

Examples from the dataset -

turn1	turn2	turn3	label
Don't worry I'm girl	hmm how do I know if you are	What's ur name?	others
When did I?	saw many times i think -_-	No. I never saw you	angry
Money money and lots of money😍😍	I need to get it tailored but I'm in love with it 😍	😁😁	happy
Bcoz u dont know wat is to miss someone	but sometimes one can't express the same	😥	sad

Methodology

The approach used was a standard one for these sort of short text classification tasks - pre-trained word vectors in conjunction with LSTMs .

We concatenated turn1, turn2 and turn3 and then created embedding matrix using Glove embeddings.

For LSTM and CNN models, we passed that matrix as Embedding layer.

For other ML models, we transformed the tweets into numeric representation using embedding matrix.

Evaluation

For evaluation, we considered f1 score as the metric.After performing some data preprocessing and some visualizations, we applied five models on the data and got the following results.

We can see that (Glove + LSTM) outperformed all other models.

	Happy			Sad			Angry			Macro	Micro
Model	Precision	Recall	f1	Precision	Recall	f1	Precision	Recall	f1	f1	f1
Glove + Gaussian NB	0.12	0.39	0.19	0.08	0.62	0.14	0.16	0.42	0.23	0.27	0.46
Glove +SVM	0.13	0.42	0.20	0.18	0.49	0.26	0.23	0.67	0.34	0.38	0.65
Glove + Random Forest	0.46	0.16	0.24	0.59	0.18	0.28	0.55	0.47	0.50	0.49	0.84
Glove+ LSTM	**0.57**	**0.65**	**0.61**	**0.49**	**0.74**	**0.59**	**0.59**	**0.77**	**0.67**	**0.70**	**0.88**
Glove +CNN	0.44	0.53	0.48	0.41	0.72	0.52	0.54	0.82	0.65	0.64	0.86