Assignment No 2

ML-based Emotion Classification

Group Members:

Faizan E Mustafa (3438705)

Milena Voskanyan(3522556)

Motivation

- Does contextualized models improve emotion prediction ?

- Does No. of samples for an emotion affect model scores ?

- How the results change when form of text is changed?.
  - Tweets vs Dialogues.

Model and Corpus

- **ML Model:** Linear Regression with Bag of Words features.
  - Unigram features to ensure complete context independence.

- **DL Model**: LSTM
  - Bi-directional 2 layer to ensure context dependence.

- **Corpus Requirements:**
  - Should have same labels and classification type e.g Ekman and multi-class.
  - So we can compare emotion prediction for different form of text.

- Following multi-class Ekman labeled datasets were chosen.
  - TEC
  - Dialy Dialogue. (no-emo label excluded)

Experimental Design

- Equal representation of classes in Train / Test set.
  - Stratified 3 fold cross-validation used.
  - 10 percent Dev set extracted from train set for LSTM.

- Metrics:
  - F1 , Precision and Recall reported for each emotion class.

Results

- Result for an emotion is positively correlated with No. of samples.
  - Geen: Greater than 8000 samples have best score.
  - Red: Less than 1000 samples have worst score.

- ML model outperforms DL model.
  - DL has better F1 score in case of Surprise and Fear.(Dialogues)
  - Precision and Recall: ML better but no specific pattern.

- Corpus not fully comparable as no of samples are different.

		TEC		Dialogues	
		ML	DL	ML	DL
Joy	P	0.61	0.66	0.84	0.83
	R	0.78	0.66	0.95	0.89
	F1	0.68	0.66	0.89	0.86
Sadness	P	0.45	0.37	0.66	0.44
	R	0.43	0.48	0.47	0.42
	F1	0.44	0.42	0.55	0.42
Surprise	P	0.53	0.45	0.58	0.59
	R	0.47	0.47	0.45	0.45
	F1	0.5	0.46	0.5	0.5
Fear	P	0.57	0.53	0.27	0.2
	R	0.43	0.38	0.03	0.07
	F1	0.49	0.44	0.05	0.11
Anger	P	0.45	0.25	0.36	0.24
	R	0.28	0.15	0.19	0.19
	F1	0.35	0.18	0.25	0.21
Disgust	P	0.4	0.16	0.37	0.21
	R	0.16	0.13	0.12	0.12
	F1	0.23	0.14	0.19	0.16

Analysis

- Both models confuse gold emotions with joy frequently.
  - Except for two instances of sadness.
  - Large No. of samples of joy could be the reason.

- Gold label joy is confused with remaining frequent emotions.
  - Sadness, surprise and anger.

		Gold					
		Joy	Sadness	Surprise	Fear	Anger	Disgust
ML	TEC	Sadness	Joy	Joy	Joy	Joy	Joy
	Dialogue	Surprise	Joy	Joy	Joy	Joy	Joy
DL	TEC	Sadness	Joy	Joy	Joy	Sadness	Sadness
	Dialogue	Anger	Joy	Joy	Joy	Joy	Joy

Analysis

Text	True	ML	DL
Next Saturday ? I ' m sorry . I ' Ve promised to go to a Chinese Opera with my daughter .	Joy	Sadness	Joy
I love when your having a great day and then one thing ruins your whole day	Anger	Joy	Anger

Possible improvements:

- Use comparable corpus to compare scores of different form of text.
  - Ideally have same No. of sample per emotion.
  - Data augmentation could be applied.

- Attention based model could handle context better.

Thank you for listening.