Assignment Discussion #3 and Project Discussion

Deepak Singh Baghel 203050005
Ankush Agarwal 203050007
Nilesh Kshirsagar 203059004

12th October, 2021

Assignment Discussion #3

WSD continuation, NEI

Problem Definition: NEI

- Given a sentence/document, mark each token as 1/0 as per whether the token is a Named Entity or not

- If the named entity consists of multiple words just continue with 1s until a non-NE appears

- E.g. *The_0 State_1 Bank_1 of_1 India_1 is_0 the_0 largest_0 bank_0 in_0 the_0 country_0 . _0*

Feature Engineering

- Features:
  1. POS tag of current word
  2. POS tags of 4 context words
  3. 6 Context Words
  4. Is word capitalized
  5. Length of the word
  6. Is it first word of the sentence
  7. Is previous word a named entity

Justification of Feature Set

- Named entities mostly appear with some specific pos tags such as noun and adjective
- Context words and their POS Tags are useful to identify multi word named entities
- First character of NE is capitalized
- Since start word of the sentence is also capitalized, to avoid confusion with NE, another feature is added indicating if current token is first word of sentence
- Many named entities are multi word, such as name of person or organization, hence previous word tag is a useful feature

DATA

- CoNLL 2003

- Used load_dataset('conllpp') method from huggingface to load the dataset
- Used tokens from dataset directly

Performance

- Precision : 0.953

- Recall   : 0.951

- F1-score : 0.952

- Accuracy : 95.1%

Confusion Matrix

Predicted → ⬜ Actual (rows)	0	1
0	36725	1452
1	818	7440

Result Interpretation

- POS tag is not completely accurate and since model uses POS tag as one of the features, more accurate POS tag would improve model accuracy
- Undersampling and oversampling did not improve accuracy

Project

What is the "Problem"

- Problem Statement

  To identify the emotion of speakers in a conversation based on audio and text modalities

  - Input : Audio file and corresponding transcript

  - Output : Emotion(Sad, happy, neutral etc.)

Why is the problem important

- Emotion analysis is useful in identifying how users feel about a product based on reviews
- It is used in healthcare domain to identify and monitor certain conditions such as stress, anxiety and depression
- It is also useful in development of virtual assistants

What is hard about the problem

- Incorporating more number of emotions and achieving better accuracy

What has been done on this problem so far

Basic

- Yoon S, Byun S, Jung K. Multimodal speech emotion recognition using audio and text. In 2018 IEEE Spoken Language Technology Workshop (SLT) 2018 Dec 18 (pp. 112-118). IEEE.

State-of-the-Art

- Siriwardhana S, Reis A, Weerasekera R, Nanayakkara S. Jointly Fine-Tuning" BERT-like" Self Supervised Models to Improve Multimodal Speech Emotion Recognition. arXiv preprint arXiv:2008.06682. 2020 Aug 15.

Your tackling of the problem

- Preprocessing (data imbalance, lemmatization, removal of special symbols).

- Feature Extraction for audio features such as speech, harmonics, speech energy, pause and central movements and text features such as TF-IDF

- Create vectors of these features and fuse them for modalities

- Classification using Ensemble model(RF, SGD, MLP, MNB, LR) into angry, happy, sad, fear, neutral, surprise

- Sahu G. Multimodal speech emotion recognition and ambiguity resolution. arXiv preprint arXiv:1904.06022. 2019 Apr 12.

- Dataset : IEMOCAP

- Tools : Librosa, scikit-learn, xgboost, pytorch, Colab GPU

- Performance Metrics : Accuracy, F1-score

Thank You