



NLP final Presentation

組員: 0716027 柯秉志、0716081 葉晨、0716083 羅子涵



Outline

- Introduction
- Method
- Experiment
- Conclusion

Introduction

Prompt

I didn't finish my NLP midterm this morning. It is so hard for me.

Utterance

NLP midterm is too hard, I think I cannot pass the course.

Don't worry, the professor will give you more chances.

predict

32 emotions





Introduction

Prompt	Utterance	Label (Emotion)
I have someone always asking me questions about how to use mobile.	Hi_comma_ I have a co work always asking me questions about how to use her mobile. It is almost always the same questions.	17 (anxious)
	It must be so frustrating to have to answer the same thing over and over. Is she older?	
	...	



Method

- only use Prompt
1. **Simple Baseline:** tf-idf
 2. **Medium Baseline:** word2vec / GloVe + CNN / RNN-based model
 3. **Hard Baseline:** BERT



Method

→ tf-idf

$$w_{x,y} = \text{tf}_{x,y} \times \log \left(\frac{N}{\text{df}_x} \right)$$

TF-IDF

Term x within document y

$\text{tf}_{x,y}$ = frequency of x in y

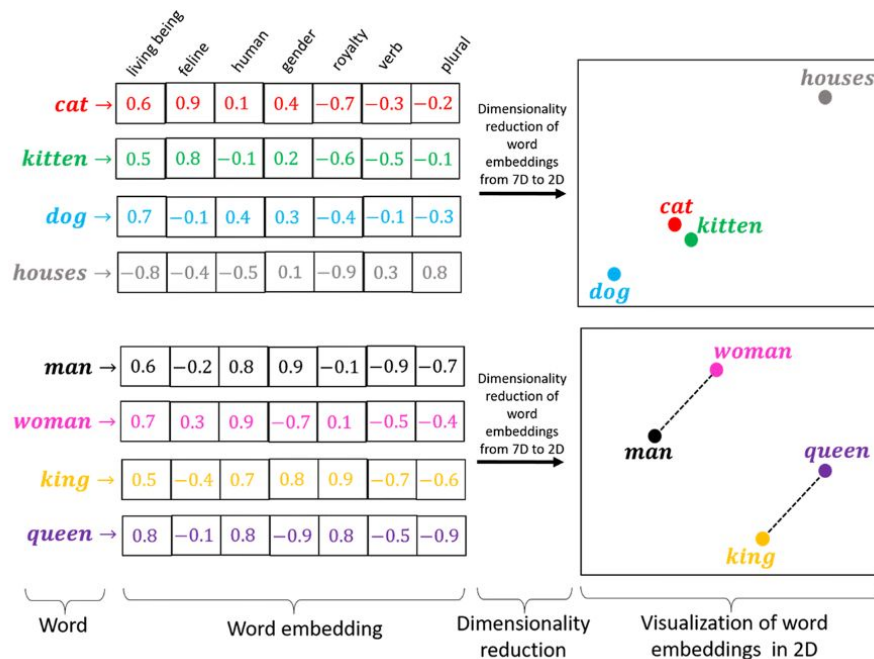
df_x = number of documents containing x

N = total number of documents

Method

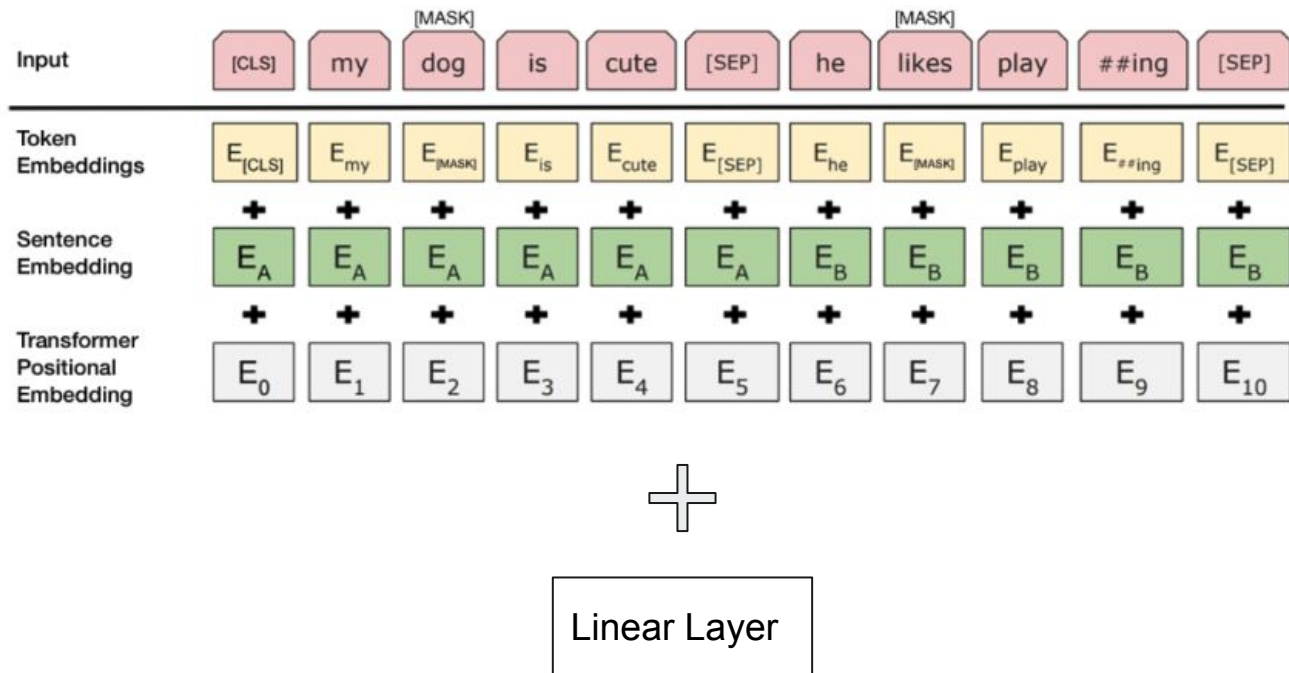
→ word2vec : predictive model

→ GloVe : count-based model



Method

→ Bert





Experiment - model testing

Model		Accuracy
TFIDF	Random Forest	0.36662
TFIDF	CNN	≈0.39
GloVe	CNN	≈0.46
Word2Vec	Random Forest	0.34876
Word2Vec	LSTM	0.39743
Word2Vec	GRU	0.409
Bert		0.5323



Experiment - Bert

- **Input data type**
 - prompt
 - prompt + utterance
 - all utterances
 - part of utterances
- **Stop words**
 - yes
 - with white list
 - without white list
 - no
- **Voting**
 - yes
 - no



Experiment - Bert

- **First attempt: accu 0.53230**
 - only prompt
 - no stop words
 - no voting
- **Best attempt: accu 0.56356**
 - prompt with all utterances
 - no stop words
 - with voting



Conclusion / Improvement

- Bert performs better than other three methods (TFIDF, GloVe, w2v).
- Texts in bert **need no preprocessing**. (stemming, stop word...)
 - I remember going → I **rememb** go → i re ##me ##mb go
- Length of sentence doesn't matter.
- Bert → Large base Bert / RoBERTa
- Try other method to group utterance
- Upgrade hardware