pdfmark=/PUT,Raw={ThisPage}«/Trans « /S /Replace » »

pdfmark=/OUT,Count=2,Dest=slide.3,View=/XYZ null null null,Title=Problem Statement pdfmark=/OUT,Count=0,Dest=slide.4,View=/XYZ null null null,Title=Problem Definition pdfmark=/OUT,Count=0,Dest=slide.5,View=null null null,Title=Data Set pdfmark=/OUT,Count=3,Dest=slide.6,View=/XYZ null null null,Title=Text Processing pdfmark=/OUT,Count=0,Dest=slide.7,View=/XYZ null null null,Title=Preprocessing pdfmark=/OUT,Count=0,Dest=slide.8,View=/XYZ null null null,Title=Word2vec pdfmark=/OUT,Count=0,Dest=slide.9,View=/XYZ null null null,Title=word2vec pdfmark=/OUT,Count=0,Dest=slide.10,View=/XYZ null null null,Title=Trainning Modele pdfmark=/OUT,Count=0,Dest=slide.15,View=/XYZ null null null,Title=Conclusion pdfmark=/OUT,Count=0,Dest=null null null,Title=Thanks for watching

Back

logos/tulip-logo.eps

pdfmark=/PUT,Raw={ThisPage} /Trans «/S /Replace» /Type /FLIP01 FINAL PRESENTATION

# FLIP01 FINAL PRESENTATION

Guanzhang Huang
Xi'an Shiyou University

February 27, 2020

logos/tulip.eps

**Overview**

logos/tulip.eps

## Problem Statement

## Problem Definition

Using 8 years daily news headlines to predict stock market movement

## Data Set

- data information

Table 1: The head of the data

|   | Date | label | Top1 | ... |
|---|------|-------|------|-----|
| 0 | 2008-08-08 | 0 | b"Georgia 'downs two Russian warplanes' as cou... | ... |
| 1 | 2008-08-11 | 1 | b'Why wont America and Nato help us? If they w... | ... |
| 2 | 2008-08-12 | 0 | b'Remember that adorable 9-year-old who sang a... | ... |
| 3 | 2008-08-13 | 0 | b' U.S. refuses Israel weapons to attack Iran:... | ... |
| 4 | 2008-08-14 | 1 | b'All the experts admit that we should legalis... | ... |
| 5 | ... | ... | ... | ... |

**Text Processing**

mark.eps

logos/tulip.eps

## Preprocessing

- Get rid of the HTML tag
- Remove the punctuation
- Cut into the word /token
- stopwords
- Reorganize into new sentences

stop words for example:
"'d", "'ll", "'m", "'re", "'s", "'t", "'ve" "ZT", "ZZ", "a" ,"a's" ...

pdfmark=/PUT,Raw=(This Page)/Trans <</S /Replace>> pdfmark

mark.eps

logos/tulip.eps

**Word2vec**

embedding_size

learning_rate

nagetive_sample

pdfmark=/PUT,Raw={ThisPage</Trans<</S/Replace>>>/Repla... **word2vec**

mark.eps

logos/tulip.eps

model[ok]

128D:array([-0.29960674,0.03145241...])

# Framming Modele

## Modeling

- LSTM
- SVM
- CNN

## LSTM

Train on 1611 samples, validate on 378 samplesEpoch 1/31611/1611
[==============================] - 50s - loss: 0.6942 - acc: 0.5208 - val_loss: 0.6935 - val_acc:
0.5079Epoch 2/31611/1611 [==============================] - 51s - loss: 0.6722 - acc: 0.5971 -
val_loss: 0.6904 - val_acc: 0.5106Epoch 3/31611/1611 [==============================] - 50s -
loss: 0.5941 - acc: 0.7492 - val_loss: 0.6831 - val_acc: 0.5661378/378
[==============================] - 3sTest score: 0.683092702633Test accuracy:
0.566137566138prediction accuracy: 0.566137566138

**CNN**

Train on 1611 samples, validate on 378 samplesEpoch 1/11611/1611 [==============================] - 14s - loss: 0.6900 - acc: 0.5317 - val_loss: 0.6974 - val_acc: 0.5079352/378 [==========================>...] - ETA: 0sTest score: 0.69742725829Test accuracy: 0.507936509829prediction accuracy: 0.507936507937

# The score of models

| | model | score |
|---|---|---|
| | | Table 2: The score of models |
| 1 | LSTM | 0.566137566138 |
| 2 | CNN | 0.507936507937 |
| 3 | svm | 0.595238095238 |

pdfmark=/PUT,Raw=(ThisPage)/Trans /S /(Replace) HANDOUT PRESENTATION

logos/tulip.eps

## Conclusion

## Conclusion

At present, we can only simply use the function call of the algorithm in NLP, and the parameter setting inside the function is still in the initial stage

I need to master a variety of natural language processing methods

The next step is to improve the quality of the algorithm based on the internal principles of the function

pdfmark=/PUT,Raw=(ThisPage)/Trans...

ark.eps

logos/tulip.eps

**Thanks for watching**