Zhaomin Xiao

http://pitt.edu/~zhx36
zhx36@pitt.edu | 412.628.4394

EDUCATION

UNIVERSITY OF PITTSBURGH
MS in Information Science
Expected April 2019 | PITTSBURGH, PA
Cum. GPA: 3.59

SHENZHEN UNIVERSITY
BS in Mathematics and Applied Mathematics
July 2017 | SHENZHEN, CHINA
Cum. GPA: 3.23 TOP 20%

LINKS

Github:// **zhaomin1995**

COURSEWORKS

GRADUATE
Human Language Technology
Machine Learning
Data Mining
Algorithm Design
Introduction to Neural Network
Iterative Method

UNDERGRADUATE
Probability and Statistics
C Programming
Computing Method
Ordinary Differential Equation
Stochastic Processes
Discrete Mathematics

SKILLS

PROGRAMMING
Java • C • Matlab • R
Python • MySQL • LaTeX

FRAMEWORK
Keras • Scikit-learn • NLTK

REFERENCES

Diane Litman
Department of Computer Science
dlitman@pitt.edu
Catalin Trenchea
Department of Mathematics
trenchea@pitt.edu
Jon Walker
Department of Information Science
jdw8@pitt.edu

PROJECT EXPERIENCE

BUILD NATIVE LANGUAGE IDENTIFICATION SYSTEM
Oct 2018 - Dec 2018 | School of Computing and Information

- Extract features like char n-gram, POS n-gram and punctuation mark counts.
- Logistic regression with different penalty is used to do feature selection and classification.
- Search for the best hyperparameter using GridSearch with 5-fold cross-validation.
- Precision, recall and F1-score on test set are 69%.

CLASSIFY THE HATE SPEECH AND OFFENSIVE LANGUAGE
Sep 2018 – Dec 2018 | School of Computing and Information

- Clean the data collected by making use of Twitter API.
- Use word-based TF-IDF and POS-based TF-IDF to achieve a very good baseline.
- Select various features using logistic regression.
- SVM, logistic regression and deep learning techniques are used to classify tweets.
- Obtain 89% accuracy which is very close to the start-of-the-art.

CLASSIFY TOXIC COMMENTS WITH BIDIRECTIONAL LSTM
May 2018 – Aug 2018 | School of Computing and Information

- Use Keras and bidirectional LSTM to classify comments based on toxic information contained in comments.
- Various dropout rates are used to avoid overfitting.
- Use Maxpooling layers to extract features.
- Many optimizers such as SGD, RMSprop and Nadam, are used to improve accuracy and accelerate the training procedure.
- Use GPUs to speedup the training procedure.

PREDICT SNOW ACCUMULATION OF MONTANA
March 2018 – June 2018 | School of Computing and Information

- Attend weekly group meeting with team members and Norway Team.
- Model the snow accumulation using dynamic Bayesian network, Bayesian search and Naïve Bayes.
- Tune parameters via cross-validation and use ROC curve to compare models.

RESEARCH EXPERIENCE

DERIVE RANDOM WALK IN DIFFERENT DIMENSIONS
March 2015 – July 2015 | Shenzhen University

- Determine the recurrence and transience of random walk in 1-dimension using Stirling's formula.
- Generalize the case in 1-dimension to the case in multi-dimension using the same method.
- Analyze the case of unsymmetrical probability of different directions in various dimensions.

AWARDS

2014 Second Class Academic Scholarship in Shenzhen University