

# Wu Xueyang

xueyangwu@yeah.net | +86-15121115836

## EDUCATION

### SHANGHAI JIAO TONG UNIVERSITY

IEEE HONOR CLASS STUDENT,  
MAJOR IN CS  
GPA: 87.14 | 2013-Present

## COURSEWORK

Data Structures  
Algorithms  
Computer Architecture  
Computer Network  
Machine Learning

## SKILLS

### PROGRAMMING

C++ • Python • JAVA

### NEURAL NETWORK TOOLS

Keras • Theano • Torch

## AWARDS

- 2015 Academic Excellence Scholarship (Top 10%)
- 2015 MSRA 2015 Penta-Hackthon: Best UI Design Award (Rank 2/31)
- 2014 CUMCM Second Prize in Shanghai (Top 15%)
- 2014 Merit Student of SJTU (Top 10%)
- 2014 Academic Excellence Scholarship

## INTERESTS

- Machine Learning
- Natural Language Processing
- Artificial Intelligence
- Data Mining
- Information Retrieval

## EXPERIENCE

**RESEARCH INTERN** June 2015 - Present | Speech Lab, SJTU

**LEADER OF THE DEPARTMENT OF ACADEMY AND INNOVATION** June 2014 - June 2015 | Student Union of SEIEE of SJTU

**VOLUNTEER** Sep 2013 - Present | Kinds of Activities

## RESEARCH AND PROJECTS

### PUNCTUATION PREDICTION FOR CHINESE SPEECH TRANSCRIPTS

- July 2015 - Present
- Punctuation prediction is necessary for downstream processing on speech transcripts. My work not only focuses on punctuation prediction but also semantic segmentation.
- Sequence labeling methods are applied to predicting punctuation, treating every punctuation (including "no punctuation") as a label to each word in the text.
- I have tried CRF, CRF with various features and RNN-LSTM, improve the performance from around 40% to 80%.
- Multiview and multitask structure are also tried to adopt to this task.
- Currently, I am trying to combine punctuation prediction with parsing.

### AUTOMOBILE COMMENTS ANALYSIS

**SYSTEM** Mar 2016 – May 2016 | Machine Learning Course Project

- This system consists of three part: topic categorization, key-phrase extraction and sentiment analysis.
- We crawled comments from forums talking about automobiles, some of which includes labels so that we could utilize these labeled data for supervised learning.
- Different CNN were applied to topic categorization and sentiment analysis with accuracy 88% accuracy and 92%, respectively.
- I also researched on word embedding for contrasting meaning and akin meaning but it was useless to improve the performance of sentiment analysis.

**COMPUTER GO** Oct 2015 - Jan 2016 | Rank 2/13

- The searching space of GO is extremely large and not like other board games, GO doesn't have a well-defined evaluation function for heuristic searching.
- We used Monte Carlo Tree Search with RAVE-AMAF to find the best next-move. Additionally, statistical method was also applied to imitating expert play in the opening moves by building a Joseki tree.

### SEARCH ENGINE OF BOOK WITH IMAGE RECOGNITION AND NATURAL LANGUAGE UNDERSTANDING

Dec 2014 - Jan 2015 | Best Course Project

- The search engine were built with Lucene, based on 1.8 million books' information that we crawled from website.
- We used SIFT algorithm to detect and describe local features in images.
- Different from other engines that can only search books according to book title, we allow fuzzy search in terms of author, type, content and comment. What's more, I designed a system based on parser to understand users' natural language input.

### INNOVATIVE AND ENTREPRENEURIAL PROJECT OF SJTU : INTELLIGENT LECTURE ORGANIZING APPLICATION

Jan 2015 - May 2016

- This project aims at automatically generating notes for lecture and helping users get their notes organized.
- Lecture notes are outlined by terminology extraction, which I designed to detect domain specific keywords from text using TF-IDF, and other syntax features.