

SHIYI SHEN

shiyis@cs.brandeis.edu | 412 - 482 - 7499 | <https://github.com/shiyishen3447> | <https://nlpwshiyi.me>

EDUCATION

Brandeis University

Waltham, MA

MS in Computational Linguistics | **GPA: 3.67/4.0**

May '22

Coursework: Fundamentals of NLP I&II; Stats NLP; Informational Retrieval; Fundamentals of Artificial Intelligence; Natural Language Annotation for Machine Learning; Syntactic Theory; Semantics: Truth, Form, and Meaning; Data Structures and the Fundamentals of Computing; Advanced Programming in Java

University of Pittsburgh

Pittsburgh, PA

BA in Linguistics, Anthropology, and Japanese Studies | **GPA: 3.60/4.0**

July '18

Coursework: Intro to Computational Linguistics; Syntactic Theory; Linguistic Variation and Change; Applied Linguistics; Pragmatics of Japanese; Language and Media of South Asia; Intro to Translation in Japanese;

SKILLS AND ACHIEVEMENTS

Programming languages: Java; Python | **Languages:** Mandarin Chinese(native); Japanese(advanced)

Platforms and Tools: Tensorflow; Pytorch; Open AI; Jupyter Notebook; Github; Emacs; Vim; Textmate; Visual Studio; NLTK; Pynini; Scikit learn; Pandas; Anaconda; Homebrew

Achievements: *Awarded with 50% funding for graduate studies in computational linguistics*

Sep '18

Full tuition remission with SEALS scholarship for intensive language training

May '14

PROFESSIONAL EXPERIENCE

Beth Israel Deaconess Medical Center | Harvard University

Boston, MA

Bio-NLP Researcher

· **bio-markers extraction:** finished a project on bio text-mining and bio-markers extraction leveraging RNA/DNA data of patients with pancreatic ductal adenocarcinoma (PDCA) and understood the basics and contribution of cancer prognosis and treatment predictions.

Lobos Management

Pittsburgh, PA

Bilingual Office Assistant

Oct 2017 - Dec 2017

· **data entry and maintenance:** maintained important credentials and real estate data in the filing system. Created a coding system for the benefit of organizing billing information. Handled bilingual emails and tasks.

PROJECTS

Sentiment Analysis with Movie Reviews on IMDb

Utilized Google pre-trained word2vec model to preprocess data and K-means to locate the center words and SVM to do classification, after cross-validation, and the F1 Score on test data reaches 0.65

Logistic Regression Movie Review Classifier

Tuned parameters for LR model with cross validation for classifying movie reviews, roc auc attains 0.773

NER with Conditional Random Fields (in Mandarin Chinese)

Improved conditional random fields model for NER in Mandarin Chinese; obtained F1 Score 0.96

HMM POS Tagger with Finite State Transducers

Built a HMM POS tagger backed by FSTs for Brown corpus; prediction result obtained F1 of 0.92