Le Fang

+86 15623929315
fredfang1203@gmail.com

SUMMARY

Computational linguistics researcher with three years' natural language processing (NLP) experience and six years' programming experience, broadly interested in NLP, machine learning, and artificial intelligence.

EDUCATION

Northwestern University, Evanston, IL **GPA:** 3.83
Master of Science, Computer Science, Dec 2022
Project: *Step in Story Sifting*

Pennsylvania State University, University Park, PA **GPA:** 3.72
Bachelor of Science (cum laude honors), Information Sciences and Technology, Dec 2018
Honors Thesis: *The Role of Information Density in Bilingual Code-Switching*

RESEARCH EXPERIENCE

Story Sifting Project, Northwestern University Jan 2022 – Jun 2022
Independent Researcher – supervised by Dr. Ian Horswill

- Found evidence that a new domain-specific language STEP is more flexible and simpler than its contemporary counterparts for story sifting (a creative approach to emergent narrative).
- Illustrated how to identify compelling patterns of micro-stories and automatically generate narrative with STEP.

Applied Cognitive Science Lab, Pennsylvania State University Jan 2019 – Dec 2019
Research Assistant – supervised by Prof. David Reitter and Dr. Jesús Calvillo

- Collected a new dataset of Chinese-English text with 1476 pairs of code-switched and non-code-switched sentences.
  - Verified correctness and fluency of translations by using a survey in Amazon Mechanical Turk.
  - Aligned each code-switched sentence with a non-code-switched sentence.
- Investigated whether cognitive effort affected code-switching among bilinguals.
  - Extracted known psycholinguistic features as control variables.
  - Calculated variables of interest: word surprisal and word entropy.
  - Predicted code-switching by fitting a binary logistic regression model.
- Found that code-switching in written language is reliably affected by sentence length, word length, word frequency, and mostly word surprisal.
- Released replication package, including codes for data processing and model selection, on GitHub for future code-switching research.

Honors Thesis Project, Pennsylvania State University Jan 2018 – Dec 2018
Independent Researcher– supervised by Prof. David Reitter, Dr. Jesús Calvillo, Dr. Jeremy Cole

- Summarized well-known factors that affect code-switching from previous literature.
- Built a Chinese-English text corpus that contains 4740 code-switched sentences.
  - Utilized Beautiful Soup 4 to scrape Chinese-English corpus from online forums.
  - Implemented Stanford Chinese Word Segmenter to divide sentences with Chinese Penn Treebank.
  - Built Chinese and English relative frequency dictionaries with Google 1-gram corpora to identify code-switched sentences from the raw corpus.
  - Hired five Chinese-English bilinguals to translate code-switched sentences into Chinese sentences.
  - Preprocessed and combined the translated sentences with original code-switched sentences to build a code-switching corpus.
- Found that code-switching is more likely to occur when the words to be produced in the initial language have high information density (measured by relative frequency and high surprisal of the words).

ACADEMIC EXPERIENCE

Natural Language Processing – Programmer Mar 2022 – Jun 2022

- Developed a person name identifier to find hosts from Twitter tweets about Golden Globe Awards.
- Wrote a parser to get cooking tools and ingredients in recipes; transformed recipes into different styles.

Machine Learning – Programmer Jan 2022 – Mar 2022

- Utilized decision tree models to predict binary output given sample datasets.
- Implemented K-Nearest Neighbors models with Manhattan and Euclidean distance measures.
- Compared K-means and Gaussian Mixture Models to classify data into multiple groups.
- Trained polynomial regression model with discrete datasets of various distributions.
- Explored basic ideas of neural networks, e.g., perceptrons and convolutional networks.
- Practiced reinforcement learning with multi-armed bandit and Q-learning problems.
- Learned about other ML models, e.g., support vector machine, active learning, gradient descent, local search, genetic algorithms, and recurrent neural networks.

Artificial Intelligence – Programmer Sep 2021 – Dec 2021

- Compared breadth-first search, depth-first search, and A* search in finding the best way to travel between any pairs of landmarks given a map.
- Implemented Minimax with and without Alpha-Beta Pruning for a Hawaiian Checker Game.
- Studied ideas of knowledge representation and reasoning, decision trees, and probabilities.
- Explored concepts in responsible AI: explainability, interpretability, transparency, accountability, and bias.

Inclusive Making – Practitioner Jan 2021 – Mar 2021

- Surveyed issues on inclusivity and accessibility of technology by interviewing minority communities.

TECHNICAL SKILLS

Programming: Python, Java, C#, R, HTML, PHP, Microsoft SQL, Lisp
Software: Stanford NLP software, SRI Language Modeling Toolkit, Tableau, GitHub, RStudio, Unity
General: NLP, object-oriented programming, web scraping, database

PUBLICATIONS

Published:

- Jesús Calvillo, Le Fang, Jeremy Cole, David Reitter. 2020. *Surprisal Predicts Code-Switching in Chinese-English Bilingual Text.* Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).