Yurun Song

Compute Science (PhD.)

- Irvine, CA, US
- Email: songyr888@gmail.com

Skills

Machine Learning	**5+ yrs.**
Natural Language Processing	**4+ yrs.**
Big data/ Data science	**4+ yrs.**
Python / PyTorch	**4+ yrs.**
LaTex / Markdown	**2+ yrs.**
Medical AI	**2+ yrs.**
Software Engineering	**3+ yrs.**
Java / Matlab	**3+ yrs.**
C++	**2+ yrs.**
English	**IELTS**
Japanese	**Beginner**

Biography

I am a PhD student from University of California, and I am passionate about exploring the possibilities of NLP in the future.
As a major area of computer science, NLP has always been a passion of mine. I have been organizing and participating in many individual and team projects to improve my research ability in the field of NLP, as well as my leadership.
In terms of technology, I have good project experience in PyTorch, Keras and have certain research experience in natural language processing, such as text generation, automatic code generation, Machine Translation evaluation, etc. At the same time, I like to focus on the practical application of the model. Because I think the only technologies worth pursuing are those that can be applied in real life.
Furthermore, I have conducted personal research on various medical AI applications, including the Deep Learning Detection of Arrhythmia, EfficientNet with RandAugment for Diabetic Retinopathy Detection, and others. Recent years, my focus has been on multimodaity, code generation and large language model pretraining area.
During the past year as a PhD student, and in the four years prior while working on NLP. I have come to fully appreciate the significance of specializing in data science and NLP, both for individual and group projects. I am currently seeking a full-time position within the artificial intelligence industry, where I can apply my expertise to practical tasks and add value to the organization I join.

Experience

Student Researcher **Dec/2022 - Now**

BlackBerry

Research related to utilize LLM for the purpose of generating code, managing and executing queries in the 5G Core network.

Researcher **Dec/2021 - Jun/2022**

Multi-Intents-Detection

Aktify Inc (https://aktify.com)

Design a Multi-intents Detection Model for Aktify Company. Using their conversation data and classify customs' intents behind. The project is about to submited to EMNLP conference in 2022.

Technical Leader **Dec/2020 - Nov/2021**

Automatic Code Generation (Prof. Ian. G. Harris)
University of California, Irvine

Supervised by Prof. Ian. G. Harris, I and my team members are developing a new AI model to translate Natural Language to Python Code. And we submitted our work to EACLconferences.

Research Team Member **Feb/2021 - May/2021**

Dr. Kishore Papineni
Google Research Intern 2021

Guided by Dr. Kishore Papineni, I and my partner research on a new topic about document level MT evaluation metric. We are about to focus on the combination of semantic and cross-lingual document metric.

Education

Sep/2022 -
Computer Science (PhD.)
University of California, Irvine

Sep/2019 - Nov/2020
Advanced Computing (MSc.)
Imperial College London, UK
Machine learning for Imaging
Natural Language Processing
Computer Vision

Sep/2016 - May/2019
Computer Science (BSc.)
King's College London, UK
Software Engineering Group Project • Distributed system • Artificial Intelligence

Interests

- Painting
- Science Fiction
- Cooking
- Travel

Contact

- Irvine, CA, US
- 949 617 8642
- songyr888@gmail.com
- github.com/Rain9876

First Author Jun/2020 - Nov/2020
Summer Internship (Prof. Lucia Specia)
Imperial College London

Mentored by Prof. Lucia Specia, I worked with my partner on her new proposal in the 2020 summer. We focus on developing a compound evaluation metric in Machine Translation field. We submitted the Paper to 2021 NAACL-HIT Conference and get the acceptance.

Projects

- **SentSim: Crosslingual Semantic Evaluation of Machine Translation (2020)**
  Design a compound evaluation metric to assess both mono-lingual and cross-lingual translations from semantic aspects. It achieves the SOTA in many evaluation tasks and is superior to previous metrics.
- **Text Generator in Clinical NLP (2020)**
  An individual project focus on automatically generating discharge instructions for the clinician based on the patient's hospitalization course. T5-Base model was implemented and reached SOTA in the project.
- **EfficientNet with RandAugment for Diabetic Retinopathy Detection (2020)**
  An individual project focus on applying EfficientNet architecture to classify Diabetic Retinopathy Images for patients. The performance of my model was superior to previous work in this project.
- **Deep Learning for detecting irregular heart beat (2019)**
  An individual project classifies electrocardiography (ECG) signals and detects arrhythmia symptoms using neural networks. The model and approaches of this project achieved great performance close to SOTA.

Presentations

SentSim: Cross-lingual Semantic Evaluation of Machine Translation (May 2021)
2021 Annual Conference of North American Chapter of the Association for Computational Linguistics.

Text Generator in Clinical NLP (Sep 2020)
Imperial College London

EfficientNet with RandAugment for Diabetic Retinopathy Detection (2020)
Imperial College London

Deep Learning for detecting irregular heart beat (2019)
King's College London

Publications

PCMID: Multi-Intent Detection through Supervised Prototypical Contrastive Learning (2023)
Yurun Song*, Junchen Zhao*, Ian G. Harris, Spencer Koehler, Amirali Abdullah, EMNLP Findings 2023

GAP-Gen: Guided Automatic Python Code Generation (2022)
Junchen Zhao*,Yurun Song*, Ian. G. Harris EACL SRW 2023

SentSim: Crosslingual Semantic Evaluation of Machine Translation (2021)
Yurun Song*, Junchen Zhao*, Lucia Specia, **NAACL-HIT 2021** Conference.
URL: https://www.aclweb.org/anthology/2021.naacl-main.252.pdf