Songtao Wang

Ms.c in Computer Engineering

National University of Singapore

📞

✉ e1011051@u.nus.edu

⌂ https://github.com/songtao-x

EDUCATION

•**Harbin Institute of Technology, China** *Sep,2018-Jul,2022*

Bachelor Degree in Engineering, Automation GPA: 3.74/4

•**National University of Singapore** *Aug,2022-Feb,2024*

Ms.c in Computer Engineering CAP: 4.5/5

RESEARCH INTEREST

My research interest span a wide range of areas in AI including statistical ML with Bayesian Inference, natural language processing, reinforcement learning, and large language model.

PUBLICATION

•**Benchmarking Large Language Models on Communicative Medical Coaching: a Novel system and Dataset**

Hengguan Huang, **Songtao Wang***, Hongfu Liu, Hao Wang, Ye Wang, ACL Findings 2024*

- **Description**: In this paper, we managed to create a multilingual role-play simulation with LLM agents. We built the benchmark to achieve the assessment of LLM coaching efficacy in a simulated patient-doctor-coach environment supported by human annotation, forging a novel intersection among education, medical and AI. Additionally, we introduce the Generalized Chain-of-Thought (GCoT) approach which improves upon CoT by embedding generalizable variables within reasoning paths. These variables are elements shared across various data samples' reasoning steps, facilitating the creation of structured feedback and seamless external knowledge integration.

- **Contribution**: I played a pivotal role in creating the simulated patient-doctor-coach environment for LLM communication, designing procedures of generating the benchmark dataset, and conducting comprehensive experiments on various prompting methods. My significant contribution also includes coming up with the idea of inferring latent variables by LLM itself and following two-steps prompting framework.

•**Verbalized Probabilistic Graphical Modeling with Large Language Models**

*Hengguan Huang, Xing Shen, **Songtao Wang**, Dianbo Liu, Hao Wang, arXiv:2406.05516, under review, June 2024*

- **Description**: Our vPGM method circumvents the traditional necessity for manual graphical modeling and extensive data collection by leveraging LLM to understand and inference about latent variables and their interdependencies according to Bayesian principles. In addition, we proposed Bayesian confidence calibration method to refine the reliability of predictions. Experiments on compositional reasoning tasks demonstrate superior performance in confidence elicitation and text generation quality, underscoring the potential of integrating Bayesian principles with LLMs and refining reliability of LLMs

- **Contribution**: I managed to design the pipeline for guiding LLMs to discover the vPGM structure, identifying latent variables and their interrelations. Following this, according to Bayesian Inference formulas, I realized them into specialized prompts that enabled the approximation of each latent variable's distribution, effectively allowing the LLM to sample from the graphical model constructed. Furthermore, I also conducted experiments on performing vPGM on my previous work medical coaching dataset and achieved significant improvement.

•**Incomplete multi-view clustering with multiple imputation and ensemble clustering**

*Guoqing Chao, **Songtao Wang**, Shiming Yang, Chunshan Li, Dianhui Chu, Applied Intelligence, 2022.* Jan, 2022

- **Description**: We proposed a new multi-view clustering algorithm that could handle missing values in the machine learning with multiple imputation and weighted ensemble clustering method based on Bayesian estimation which has been proved to reach state-of-the-art performance.

- **Contribution**: I managed to perform multiple imputation to handle incomplete datasets to make it available to implement clustering algorithms. Furthermore, given that many complete datasets acquired from last step with much of uncertainty, I implemented ensemble clustering trying to gather them all. And I designed a weighted algorithm based on Bayesian estimation to combine all the results from ensemble clustering whose weights are based on the accuracy on its branch of algorithm. Finally, it will contribute to performing last step hierarchical clustering.

THESIS

•**NLP based algorithms for evaluating student portfolios**

Master thesis   Apr, 2023

– **Description**: Developed an evaluation system for NUS's Pioneer House students using machine learning algorithms and hierarchical clustering, with an emphasis on unsupervised learning approaches such as VAE and HDBSCAN.

– **Contribution**:My work included critical feature analysis using correlation, PCA, SVM, and more. I developed an LSTM-based framework for predicting student enrollment probabilities, integrating feature importance into the recommendation process. Most importantly, I mainly focused on unsupervised learning. Considering that the portfolio is complicated high-dimension data, I utilize VAE to extract latent variable provided for HDBSCAN which was demonstrated to have a better performance.

•**Decision making simulation of autonomous driving based on deep reinforcement learning**

Bachelor Thesis   Aug, 2022

– **Description**: This is my bachelor graduation thesis. It was designed to implement to deep reinforcement learning algorithm on simulating decision making procedures of racing cars in TORCS system.

– **Contribution**: I was responsible for programming the entire network, implementing the DDPG algorithm which is one of the extension of Actor-Critic framework for critical decision-making regarding steering and acceleration, complemented by a self-designed adapted Q-function.

WORK EXPERIENCE

•**National University of Singapore, School of Computing, Sound Music Computing Lab**

Research Assistant, supervised by Professor Wang Ye   Apr,2023-Jan,2024

– During this period, I started to get familiar with large language model(LLM) and the beneath model GPT. I was committed to the program 'ChatCoach' which finally led to the paper "Benchmarking Large Language Models on Communicative Medical Coaching: a Novel system and Dataset".

–**Harbin Institute of Technology**

Research Intern, supervised by Professor Guoqing Chao   Jan, 2021-Jan,2022

∗ I stared to learn the machine learning, specifically in spectral clustering.

SELECTED AWARD

∗**China Undergraduate Mathematical Contest in Modeling**

Third-class Price   China, 2021

∗**Mathematics competition of Chinese College Students**

First-class Price   China, 2019

∗**Chinese high school Olympiad Mathematical Competition**

First-class Price   China, 2016

RELATED MODULES

Large Language Model: In this course, I explored the possibility of integrating uncertainty estimation into PPO framework.

Pattern Recognition: I learned the basis of several machine learning algorithms including Bayesian predictive distribution, PCA, LDA, LR, SVM and I'm required to programmed them from scratch. Also, I managed to perform CNN for a facial recognition task.

Machine Learning: In this module, I learned the mathematical principle about machine learning and reinforcement learning. I was also required to finish the DQN, DDQN from scratch and its application on game

Deep Learning: I learned principle and application of RNN, LSTM, Transformer, GPT and so on. It also showed me the most popular deep learning generation algorithms including VAE, GAN, stable Diffusion and NeRF. Thanks to the module, I got acquaintance about the mathematical formulas about them. During this period, we accomplished the semantic *segmentation tasks with LinkNet*, and the *Style Transfer tasks with CycleGAN*.

ADDITIONAL INFORMATION

Natural Languages: Chinese(native), English, GRE:320

Technical Skill: Python, C/C++, R, matlab

Tools and Framework: Pytorch, Huggingface, ROS

Interest: Go/Weiqi(4th grade), electronic piano(8th grade), basketball, football