“Pat” Navapat Nananukul

MS.c. in Computer Science @ University of Southern California

github.io

nananuku@isi.edu

Research Interests

Natural Language Processing and Social Science, Knowledge Graphs, Large Language Models

Education

University of Southern California, USA
Master of Science (M.S.) in Computer Science — Aug. 2021 - Aug. 2023

Chulalongkorn University, Thailand
Bachelor of Science (B.S.) in Computer Science & Mathematics — May 2013 - May 2017

Research Experience

Student Researcher at **Information Science Insititute (ISI), AICS lab,** University of Southern California — USA
mentored by Prof. Mayank Kejriwal (USC) on Knowledge Graph and Large Language Model — June. 2023 - Present

I. Cost-Aware Blocks Prioritization for Entity Resolution in LLMs (in progress)
(1) Proposed and implemented blocks prioritization algorithms to integrate state-of-the-art commercial LLMs with entity resolution task. (2) Utilzed feedback from LLMs to improve the cost/performance ratio of entity resolution task. (3) Analyzed and compare the performance and cost to perform entity resolution using LLMs of the algorithms with state-of-the-art entity resolution algorthms.

II. Cost-Efficient Prompt Engineering for Unsupervised Entity Resoltion (under review)
(1) Designed various automated prompt techniques, including attributes selection, persona, few-shot prompting, data formatting for unsupervised entity resolution. (2) Anaylzed the performance of engineered prompts and provide comprehensive results on cost/performance tradeoff measures with statistical test.

III. Large Language Models Hallucination Ontology/Knowledge Graph (under review)
(1) Collected and consolidated hallucinated prompts from multiple resources. (2) Designed and created extendable hallucination ontologies to systematically store hallucinations types and relation. (3) Proposed potential applications such as, benchmarking hallucinations in LLMs and improve fine-tuning LLMs process.

Research Intern, University of Southern California — USA
mentored by Prof. Mohammad Rostami on Domain Adaptation and Computer Vision — Mar.2023 - May 2023

I. Unsupervised Federated Domain Adaptation for Segmentation of MRI Data Using Multi-Site Source Domains (under review)
(1) Implemented a domain adaptation algorithm for MRI image segmentation task using 3D-UNET and SEGJDOT architecture. (2) Implemented ensemble algorithm to improve the segmentation performance from multiple source domains. (3) Analyze the results and perform theoretical analysis to show that our domain adaptation and ensemble beat existing approach on MICCAI 2016 multi-site dataset.

Course Research Project

Early Parkinson’s Disease Detection using CNN-LSTM Model for Time-series Keystroke Data (2023)
- Proposed a CNN-LSTM model that beats baseline models such as SqueezeNet, MobileNet, and AlexNet in prediction Parkinson’s disease.
- Addressed imbalance data issue by performing SMOTE to the original keystroke dataset.

Text Generation Model for Short Story Generation: An Experimental Study (2022)
- Experimented with text generation models and algorithms such as Markov chains, LSTM, GPT-2, GPT-3, BART.
- Analyzed text generation results, and short story quality with potential improvement for story generation task.

Knowledge graph for European Soccer with News Clarification Feature (2022) [Slides]
- Performed an end-to-end knowledge graph building including crawling, entity resolution, ontology design, visualization, and UI building for knowledge graph application.
- Built a european soccer player knowledge graph to serve our web-application that is able to suggest and clarify soccer jargons in soccer news.

Work Experience

New york city, USA

Senior Data Analyst/Scientist at Agoda (USA) Sep. 2018 - July 2021

Key member of the data analytics and report management team. The role involves increasing data accuracy, driving process improvement, analyzed the impact of strategic initiative through designing a data-driven A/B testing and experimentation. My technical contributions include:

- **Automation:** Built internal data preprocessing tools, including EDA, ETL processes, and data visualization for management team
- **Machine Learning:** built statistical and machine learning models to predict sales performance, market management target, and potential opportunities.
- **Experiment:** Evaluated the impact of strategic initiatives through experiementation and A/B testing. Building success metrics and data models that forecast the market.
- **Visualization:** Enhanced the efficacy of sales and market management teams by developing tableau dashboards and alerts that allowed team members to identify market opportunities and issues.
- **Training:** Oversee the design and development of technical training for new hires in addition to training external teams on data management, SQL skills.

Bangkok, Thailand

Business Intelligence Developer at Agoda June 2017 - Sep. 2018

- **BI Product:** Provided end-to-end BI product and data engineer solutions including ETL processes, analysis, OLAP and visualizaion using Tableau.
- **Visualization:** Experienced in large-scale Tableau dashboard with more than 300K traffic from more than 1000 unique users.

Talks/Presentations

Artificial General Intelligence Micro-Seminar for Undergraduate Students (2023)

- Support Professor Mayank Kejriwal on a demo about common-sense reasoning on LLMs such as ChatGPT, BARD.

Programming Skills

- **Coding:** C/C++, R, Python, SQL, MATLAB, R, Git
- **Deep Learning Toolkits:** Tensorflow, PyTorch
- **Data Warehouse:** Hadoop, Big Query, MapReduce, Apache Spark
- **Data Analysis Tools:** Tableau, Google Data Studio, Dbeaver, Metabase, Jira, AWS, Postman, Agile

Leadership and Awards

- **Head of Student Financial** of Computer Science Department - Chalulongkorn University (CS class of 2017)
- **Deutscher Akademisher Austauschdienst (DAAD) Scholarship** winner (ranked #1 in CS field) to be Thailand's representative exchange student intern as a research intern at Germany.