Research Statement

Ching Nam Hang

My research interests lie at the intersection of big data analytics and artificial intelligence (AI), with an emphasis on their implementation in healthcare and education. Specific research projects include the theoretical exploration and practical applications of machine learning for pandemic response, the management of misinformation during crises, and the augmentation of self-learning in large-scale educational systems. My work is devoted to turning data-driven insights into practical, real-world applications, notably improving our response to global health crises and nurturing autonomous learning. My professional journey within the vast realms of big data analytics and AI consistently enriches my knowledge and problem-solving skills, thereby laying a solid foundation for the development of innovative solutions with substantial and wide-ranging impacts.

Ph.D. Research

Machine Learning for Pandemic Response

In the face of the COVID-19 pandemic, an array of digital contact tracing (DCT) strategies has been proposed to curb virus transmission. Existing literature, however, largely focuses on DCT application design and deployment, with less emphasis on machine learning optimization. In [1], we conduct a comprehensive review of machine learning- enhanced DCT strategies, formulating relevant research questions and outlining potential solutions using existing machine learning methodologies. We introduce a new taxonomy categorizing DCT strategies into forward, backward, and proactive contact tracing. We examine the integration of privacy-preserving strategies, such as federated learning, with big data and machine learning to enhance the privacy and security of DCT systems. Additionally, we demonstrate how convolutional neural networks, graph neural networks (GNN), and transformer-based learning optimize DCT strategies. As a solution to limited data availability in the early phase of pandemics, we show how generative adversarial networks (GAN) can be leveraged to confront such challenges in DCT.

Machine Learning for Infodemic Risk Management

The COVID-19 pandemic has led to an infodemic of unprecedented scale with the rapid spread of misleading or false information via online social networks. Network analysis plays an indispensable role in fact-checking efforts, modeling and learning infodemic risks through large-scale graph computations and statistical processes. In [2], we present MEGA (*M*achine Learning-*E*nhanced *G*raph *A*nalytics), a novel framework enabling automated machine learning for mega-scale graph datasets. Within MEGA, we employ a blend of feature engineering and graph neural networks to enhance learning performance and enable parallel computation on extensive graphs. Infodemic risk analysis serves as a unique application of MEGA, encompassing spambot detection through triangle motif counting and influential spreader identification via distance center computation. Performance evaluations using the COVID-19 Infodemic Twitter dataset underscore MEGA's superior computational efficiency and classification accuracy. Further examining this issue, in [3], we introduce TrumorGPT, a maximum truth-seeking AI. TrumorGPT synergizes large language models (LLMs) like RoBERTa and dynamic influence tracking mechanisms, such as trumor centrality, to facilitate automated fact-checking on social media platforms, demonstrating efficacy on large-scale datasets.

Artificial Intelligence for Large-Scale Learning

The COVID-19 pandemic has significantly reshaped learning environments and modalities, elevating the importance of self-learning in STEM education. In [4], we offer a comprehensive review of transformer-based LLMs for AI-assisted programming, showcasing the potential of LLMs in facilitating programming learning. In [5], we propose an AI chatbot system designed to optimize flipped learning. This system employs deep learning techniques to blend peer instruction with just-in-time teaching. Additionally, we formulate optimization problems to ideally balance the allocation of pre-class and in-class learning time. In further pursuit of these aims, in [6-7], we develop AI-powered software tools to foster interactive self-learning in STEM education.

Future Research Directions

While big data analytics and AI are ubiquitously and extensively used in various forms in modern society, research on large-scale machine learning is widely acknowledged to be a challenging field. What is the cheapest technology to automate large-scale machine learning? I plan to undertake the following long-term directions in my future research to meet these challenges:

- Leveraging large language models for richer data interpretation: Large language models such as GPT-4 have shown exceptional prowess in understanding and generating human-like text. However, capitalizing on their potential to comprehend the nuanced meaning within vast and diverse data domains remains a considerable challenge. I intend to investigate novel techniques such as, preference-based reinforcement learning, to unlock the full potential of large language models for richer data interpretation across various applications.
- Advancing generative AI for data augmentation and synthetic data generation: Generative AI, especially GANs, have shown tremendous potential in creating synthetic data that mimic real-world distributions. This is particularly important in scenarios with limited data availability or privacy concerns. To advance this research frontier, it is important to create more robust and versatile generative models that can support a wide array of machine learning applications.
- Graph-based machine learning: GNNs represent a powerful tool for processing data structured as graphs. This is particularly relevant in a world increasingly connected, where relations and networks dictate how information flows. The implications range from social networks to biological networks, recommendation systems, and the interpretation of molecule structures. It is crucial to push the boundaries of this field by exploring scalable and efficient algorithms for graph-based learning, as well as studying how to combine these approaches with large language models for enhanced understanding of complex, interconnected data.

Through these endeavors, I hope to contribute significantly to the evolving landscape of machine learning, specifically in the areas of large language models, generative AI, and graph-based learning. Beyond the scope of these areas, I'm also keen on exploring the integration of machine learning with other disciplines, such as healthcare and education, to maximize the societal impact of AI advancements. While having specific milestones is crucial to steer research efforts, I remain open-minded and alert to new research opportunities. As the realms of big data and AI continue to expand, I am committed to ongoing learning, adapting to new challenges, and leading innovative research at the frontiers of technology.

References

[1] **C. N. Hang**, Y. -Z. Tsai, P. -D. Yu, J. Chen and C. W. Tan, Privacy-Enhancing Digital Contact Tracing with Machine Learning for Pandemic Response: A Comprehensive Review, ***Big Data and Cognitive Computing, special issue on Digital Health and Data Analytics in Public Health***, accepted for publication, 2023.

[2] **C. N. Hang**, P. -D. Yu, S. Chen, C. W. Tan and G. Chen, MEGA: Machine Learning-Enhanced Graph Analytics for Infodemic Risk Management, ***IEEE Journal of Biomedical and Health Informatics***, accept pending minor revision, 2023.

[3] **C. N. Hang**, P. -D. Yu and C. W. Tan, TrumorGPT: Maximum Truth-Seeking AI for Effective Fact-Checking through Dynamic Influence Tracking, to be submitted to ***IEEE INFOCOM***, 2023.

[4] M. F. Wong, S. Guo, **C. N. Hang**, S. W. Ho and C. W. Tan, Natural Language Generation and Understanding of Big Code for AI-Assisted Programming: A Review, ***Entropy, special issue on Statistical Machine Learning with High-Dimensional Data and Image Analysis***, accepted for publication, 2023.

[5] C. W. Tan, **C. N. Hang** and L. Ling, Optimal Flipped Learning: Blending Peer Instruction with Just-in-Time Teaching using AI Chatbots, submitted to ***IEEE Transactions on Learning Technologies***, 2023.

[6] J. Li, C. W. Tan, **C. N. Hang** and X. Qi, A Chatbot-Server Framework for Scalable Machine Learning Education through Crowdsourced Data, ***ACM Conference on Learning at Scale***, 2022.

[7] C. W. Tan, L. Ling, P. -D. Yu, **C. N. Hang** and M. F. Wong, Mathematics Gamification in Mobile App Software for Personalized Learning at Scale, ***IEEE Integrated STEM Education Conference***, 2020.