Research Statement

Dai Bing Tian
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-9603; Email: btdai@smu.edu.sg
20 Dec 2024

Background

My primary research interests are data mining and machine learning. I always feel it is interesting to discover insights from data and to explain insights from data and models. My primary research area for machine learning includes natural language processing with applications in medical data and educational data. In addition to data mining and machine learning, I am also interested in (1) building systems to assist an entire machine learning pipeline from data collection to model refinement and then to providing human-interpretable feedback; and (2) combining machine learning with decisions making and optimization problems.

Research Areas

Context Aware Research in Knowledge Graph and Natural Language Processing

It is very challenging to build an AI system that thinks and speaks like a human. Current AI systems often lack understanding of the context, and that is exactly why when we human are interacting with bots, we feel the bots are always talking or asking out of context. Therefore, it is important to have context aware AI systems. The emergence of ChatGPT has shown us a possibility to combine globally trained large language models (LLM) with locally available context information. My next-step research is to tune LLM with local context, so that answers can be answered more precisely according to the context under which questions are posed. If we could have a context aware AI system, all domain specific question and answer tasks can be automatically taken care of, and this saves huge cost for the business.

It is however not straightforward to build such context-aware AI systems. We either need to build the knowledge graph with the help of LLM, or directly fine-tune the pre-trained embedding of questions with the context information, and then provide answers which are generated with the encoded context information.

Pedagogical Research

One interesting question in pedagogical research is how to make education more effective. In this aspect, I think it is necessary to apply data mining and machine learning techniques to study students' behavior data, and to derive more effective strategies to improve teaching effectiveness.

To be more accurate in quantifying students' learning, I have also proposed to combine with context to model how students are learning in a pedagogical knowledge

graph. This should not be limited to just one course; it can be extended to a programme or to a conventional stage of education. Therefore, I have been diligently seeking research grants that can support my research in pedagogy. I have considered both education grants for primary and secondary education, as well as education grants for adult learning. My aim in educational research is (1) to alleviate teachers from mundane tasks, allowing them to dedicate more time to foster creativity; and (2) to nurture students' self-directed learning and match the evolving trends of changes in education.

Optimization Research

The third research question I am working on is on decision making and optimization research. In particular, how do we optimize vehicles and routing with coalited objectives or objectives at different levels of priority. These are common questions to logistic companies, and I wish to apply machine learning techniques to enhance such decision-making process.

Gen AI Validation

Gen AI has gone beyond human's ability in handling images and language processing. This leaves a big question to us, how would human know what Gen AI generates is "correct". More research is needed to find out a systematical framework in validating AI's output. This direction is what I plan to put more effort into in the next few years.

Selected Publications and Outputs

Fuel-saving route planning with data-driven and learning-based approaches: A systematic solution for harbor tugs, by Shengming WANG; Xiaocai ZHANG; Jing LI; Xiaoyang WEI; Hoong Chuin LAU; Bing Tian DAI; Zhe XIAO; QIN, Zheng (2024). *The Thirty-Third International Joint Conference on Artificial Intelligence* (IJCAI-24): 7483-7490. https://doi.org/10.24963/ijcai.2024/828

MWPToolkit: An Open-Source Framework for Deep Learning-Based Math Word Problem Solvers, by Yihuai Lan, Lei Wang, Qiyuan Zhang, Yunshi Lan, Bing Tian Dai, Yan Wang, Dongxiang Zhang, Ee-Peng Lim (2022). *The Thirty-Sixth AAAI Conference on Artificial Intelligence* (AAAI 2022): 13188-13190.

The Gap of Semantic Parsing: A Survey on Automatic Math Word Problem Solvers, by ZHANG, Dongxiang; WANG, Lei; ZHANG, Luming; DAI, Bing Tian; SHEN, Heng Tao. (2020). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42 (9), 2287-2305. https://doi.org/10.1109/TPAMI.2019.2914054

Integrated telegram and web-based forum with automatic assessment of questions and answers for collaborative learning, by CHEONG, Michelle L.F.; CHEN, Jean Y. C.; DAI, Bing Tian. (2018). *Proceedings of IEEE International Conference on Teaching, Assessment and Learning for Engineering TALE 2018: Wollongong, Australia, 2018 December 4-7,* Piscataway, NJ: IEEE. https://doi.org/10.1109/TALE.2018.8615137