Xin Du

Research Statement

My research is motivated by the challenge of discovering interesting patterns from the data [18]. Specifically, the objective is to find subpopulations which can be summarized by descriptive variables, like ‘age $\geq 25 \land$ Smoker $=$ yes’, which referred to as ‘subgroups’ [13]. These subgroups indicate that the interactive dependencies between target variables are exceptionally different from those ones in the whole data. To do so, multiple statistical measure and modeling methods are proposed to summarize those patterns. And specific search algorithms are implemented to enable the discovery of interesting subgroups within sufficient time and computing budget. My research is focused on providing algorithms and models for finding such interesting local patterns on diverse datasets and application scenarios. Feature importance has long been considered as an important factor for the machine learning models. Exlorational data analysis provides an efficient way to understand and discover such influential patterns. This links the data mining methods to the trustworthy machine learning problems, which suggests that a trustworthy model should be transparent, robust and explainable. For instance, explainable methods like Lime [20] attempts to construct a local linear model with human understandable features to approximate the global non-linear model. Adversarial attack methods [17] employ gradient descent based algorithms to find examples that are similar to the in-distribution samples from the data but could alternate the outcome of the model [12]. From the aspects of learning theory, this refers to the over-confident problem when making prediction with the out-of-distribution data. This problem indicates that a single aggregated statistical measure is not sufficient to provide trustworthy evaluation for the performance of machine learning models in complex scenarios. In order to solve the emergent problems strives from these scenarios, my research leverages the exceptional model mining with focus on learning problems, which consists of the following sections:

Exceptional Model Mining

Exceptional Model Mining (EMM) is a local pattern mining approach that cares about how differently a model would perform in subpopulations, as compared to the same model but fitted on the whole data [9]. The main interests of EMM is to discover exceptional interactive patterns in the subset of data with descriptive statistics by comparing the patterns related to models’ performance on whole data and subgroup data. Model is a tool to summarize the patterns in the data. My research in EMM mainly focuses on the multi-modal interaction in the data. One of my work [6] investigates the structural heterogeneity in network data and its effects on the learned network representation models. My research claims that the structural heterogeneity in networks can bias the network representation models across subgroups, hence biasing the decision making in downstream tasks like node classification or link prediction. This work also evaluates the statistical significance of the discovered exceptionality score of subgroups, which shows that without the network connection patterns inferred from a trained black-box model could be very different across certain attributes associated with the data. In other words, attributes carry sufficient information that is distinguishable for the graph structures.

One of my another work in EMM [5] investigates the Spatio-temporal behavior on collective social media: behavior in this setting can be exceptional in three distinct ways: in terms of spatial locations, time, and texts. We develop a Bayesian Non-Parametric Model (BNPM) to automatically identify spatio-temporal behavioral patterns on the subgroup level, explicitly modeling the three exceptional behavior types. The proposed graphical model with Gibbs sampling method allows us to learn the summary statistical pattern behind the multi-modal data distributions. This allows us to provide an effective way to measure the exceptionality of a behavioral pattern indicated by collective social behavior.

Causal Inference

Learning to infer the causal relations and causal effects have long been an important topic for machine learning communities with broad applications like healthcare and sociology [14, 3]. Answering causal questions make it possible to shift from what to why, and has been raised as advancing studies in recent machine learning research community [19]. For instance, one could estimate the impact of a new drug on specific individuals to assist clinical planning and improve the survival rate. Causal relation could prevent us from learning spurious correlations between variables which would lead to false discoveries. One of my work [8] studies the problem of estimating the Conditional Average Treatment Effect (CATE) from observational data. Heterogeneity is an important property that exists across the data distributions, with regard to descriptive variables. For instance, one clinical treatment could be quite effective on specific groups of people but ineffective on other groups. Techniques for causal inference range from matching based optimization methods [22] to regression based statistical methods [4, 21]. My research strives from the matching based method with optimization algorithms from operational research [23], and employs the recent advances of domain adaptation methods which suggest that a well structured representation learning model could improve the matching and counterfactual reasoning ability of neural networks significantly on multiple causal inference benchmarks [15].

Trustworthy Machine Learning Systems

Descriptive variables could indicate certain properties in subgroups of data. These groups could be used to make predictions for a specific target variable. Hence, those conjunction and combination of descriptive variables could be used as predictive local decision rules for classification. Local decision rules could provide both high predictive performance and explainability. However, the stability of those explanations given by the rules is still underexplored. One of my work [7] proposes to investigate the robustness of decision rule ensembles across different environments. This work introduces causal graph and group variance as the regularization term in the optimization process for searching robust explainable rules. Main concerns behind this work are that given a learned representation about the data, how much information can we know about the trustworthiness of that representation? Following this direction, another work of mine proposes to evaluate the trustworthiness of image classification models by random data augmentation, e.g. the vision transformer based image models like ViT and DeiT. In high-stake applications like clinical health care, it is necessary to understand the failure modes of the model under potential environments.

Cognitive Computing and Robustness

Using historical data to predict the potential outcome of a policy is of high importance in Cognitive Computing due to the high-cost of randomized controlled trials. Deep learning models are well-known for the strong capability of extracting high-level features based on large amount of training data.

However, they are also known to be vulnerable to copy or amplify the biases in training data [16]. This problem could bring more risks when the training and application environments mismatch, e.g. the demographic information. This may cause the model to learn spurious features in the data instead of the true features that indicate the semantic label. In application scenarios like domain specific fields, the mismatch between target and pre-training environemnts could bring more problems. Current solution for this problem is to leverage huge amount of data to pre-train a general model, which is referred to as foundation model [1], or large language model represented by General Purpose Transformers (GPTs) [2]. These models are used as the backend to suppose the multiple applications tasks like generation, reasoning and planning. My research is inspired by this pre-training and fine-tuning framework and is designed to solve the problems when data are not easily acquired for the general pre-training. For instance, in medical diagnosis domain, spcialist models are required to perform diagnostic and prognostic tasks to assist the decision making process. However, due to the sensitity of patients data and the different measurements for bio-factors, the interactive patterns indicated by the data model could be different. This work aims to explore the automated learning problem in such environments to learn a transferable model across different tasks, e.g. perform causal inference to estimate the survive rates. In such a problem setting, data scarcity and heterogeneity are commonly existed. Sometimes the computational resources are not shared between individuals or organizations. This brings federated learning to the front-end for efficiently training a model that can be used by different stakeholders with guaranteed performance [10]. New challenges are raised for the optimization of AutoML systems.

In socio-economic scenario, this problem is euqually important regarding to the data sensitivity. Potential solution is to leverage the meta-learning techniques [11] to discover a meta-distribution of data that improves the generalisibility of the model, which strives from the data centric learning paradigm. When applied to different task endpoints and stakeholders, and also citizen centric scenarios, quality and modularity of data greatly influence the excellence of models with limited capacities. Comparing to the traditional data centric learning paradigm with standard data processing pipeline, the selection trade-off between data sources are significant to the learning process. AutoML systems face a problem to build a transferable data preprocessing pipeline and learn a meta-data distribution to improve the capability of the models. My future research would be focused on this direction to explore the new learning paradigm for the solution of such problems.

References

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. J. Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *ArXiv*, abs/2005.14165, 2020.

Sabrina Casucci, Yuan Zhou, Biplab Sudhin Bhattacharya, Lei Sun, Alexander

Nikolaev, and Li Lin. Causal analysis of the impact of homecare services on patient discharge disposition. *Home Health Care Services Quarterly*, 38:162 – 181, 2019.

Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4:266–298, 2008.

Xin Du, Yulong Pei, Wouter Duivesteijn, and Mykola Pechenizkiy. Exceptional spatio-temporal behavior mining through bayesian non-parametric modeling. *Data Mining and Knowledge Discovery*, 34(5):1267–1290, 2020.

Xin Du, Yulong Pei, Wouter Duivesteijn, and Mykola Pechenizkiy. Fairness in network representation by latent structural heterogeneity in observational data. In *34th AAAI conference on Artificial Intelligence (AAAI2020)*, 2020.

Xin Du, Subramanian Ramamoorthy, Wouter Duivesteijn, Jin Tian, and Mykola Pechenizkiy. Beyond discriminant patterns: On the robustness of decision rule ensembles. *arXiv preprint arXiv:2109.10432*, 2021.

Xin Du, Lei Sun, Wouter Duivesteijn, Alexander Nikolaev, and Mykola Pechenizkiy. Adversarial balancing-based representation learning for causal effect inference with observational data. *Data Mining and Knowledge Discovery*, pages 1–26, 2021.

Wouter Duivesteijn, Ad J Feelders, and Arno Knobbe. Exceptional model mining. *Data Mining and Knowledge Discovery*, 30(1):47–98, 2016.

Alireza Fallah, Aryan Mokhtari, and Asuman E. Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In *Neural Information Processing Systems*, 2020.

Chelsea Finn, P. Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 2017.

Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

Francisco Herrera, Cristóbal José Carmona, Pedro González, and María José Del Jesus. An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems*, 29(3):495–525, 2011.

Guido Imbens and Donald B. Rubin. Causal inference for statistics, social, and biomedical sciences: An introduction. 2015.

Fredrik D. Johansson, Uri Shalit, and David A. Sontag. Learning representations for counterfactual inference. In *International Conference on Machine Learning*, 2016.

David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Fairness through causal awareness: Learning causal latent-variable models for biased data. In *Proc. FAT**, pages 349–358. ACM, 2019.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Katharina Morik, Jean-François Boulicaut, and Arno Siebes. Local pattern detection: International seminar dagstuhl castle, germany, april 12-16, 2004, revised selected papers (lecture notes in computer science/lecture notes in artificial intelligence), 2005.

Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, 62(3):54–60, 2019.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

Paul R. Rosenbaum. Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17:286–327, 2002.

Elizabeth A. Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25 1:1–21, 2010.

Lei Sun and Alexander G Nikolaev. Mutual information based matching for causal inference with observational data. *The Journal of Machine Learning Research*, 17(1):6990–7020, 2016.