Yo

ABSTRACT

Most of the current machine learning approaches to IR—including search and recommendation tasks—are mostly designed based on the basic idea of matching, which work from the perceptual and similarity learning perspective. This include both the learning of features from data such as representation learning, and the learning of similarity matching functions from data such as neural function learning. Though many models have been widely used in practical ranking systems such as search and recommendation, their design philosophy limits the models to the correlative signals in data. However, advancing from correlative learning to causal learning in search and recommendation is an important problem, because causal modeling can help us to think outside of the observational data for representation learning and ranking. More specially, causal learning can bring benefits to the IR community on various dimensions, including but not limited to Explainable IR models, Unbiased IR models, Fairness-aware IR models, Robust IR models and Causal Reasoning IR models. This workshop focuses on the research and application of causal modeling in search, recommendation and a broader scope of IR tasks. The workshop will gather both researchers and practitioners in the field for discussions, idea communications, and research promotions. It will also generate insightful debates about the recent regulations on AI Ethics, to a broader community including but not limited to IR, machine learning, AI, Data Science, and beyond. Workshop homepage is available online.1

KEYWORDS

Causality; Causal Learning; Counterfactual Learning; Search; Recommendation; Information Retrieval

1 MOTIVATION AND APPROPRIATENESS

State-of-the-art search and recommendation models extensively rely on complex machine learning and latent representation models such as matrix factorization and deep neural networks, and they work with various types of information sources such as ratings, text, images, audio or video signals. Though many complicated models are developed for these tasks, most of the models are designed based on the basic idea of learning correlative signals from data, such as representation learning for feature extraction, and neural function learning for similarity matching. However, advancing from correlative learning to causal learning in search and recommendation is an important problem to explore, because causal modeling can help us to think outside of the observational data for representation learning and ranking.

Causal learning can bring benefits to the IR community on various dimensions: (1) Explainable IR models—through counterfactual reasoning over data it helps to develop new models for explainable search and recommendation; (2) Unbiased IR models—through interventional or counterfactual learning it helps to remove bias from data or model for unbiased search and recommendation; (3) Fairness-aware IR models—counterfactual fairness helps search and recommendation models to avoid sensitive feature profiling for counterfactually fair ranking; (4) Robust IR models—causal learning based on both factual and counterfactual data can help to train more robust search and recommendation models that are resistant to shilling attacks, spams and echo chambers; and (5) Cognitive reasoning models—causal reasoning, together with logical reasoning, is one of the most important ingredients to advance intelligent systems from perceptual AI to cognitive AI, enabling cognitive-aware models in various IR systems such as conversational search and recommendation that actively interact with human users.

The motivation of the workshop is to promote the research and application of causal modeling for various IR tasks such as search and recommendation, under the background of AI Ethics in a broader sense. As noted above, causal modeling can not only help to improve IR systems on traditional perspectives such as ranking accuracy, but also on various ethics-related perspectives such as explainability, fairness and robustness.

In a broader sense, researchers in the broader AI communities have also realized the importance of advancing intelligent systems from correlative learning to causal learning, which aim to address a wide range of causal learning problems in deep learning, computer vision and natural language processing tasks. Recently, a series of AI regulations have also entered into force, such as the EU General Data Protection Regulation (GDPR) and The California Consumer Privacy Act, which emphasize the ethical perspectives of intelligent systems such as transparency, fairness and reliability. Causal learning—with the potential to enhance model transparency, fairness and robustness in various ways—is becoming more and more important to our ultimate goal of ethical AI. As an important branch of AI research, this further highlights the importance and urgency for our IR community to discuss and address the causal learning problems of various search and recommendation systems.

In this workshop, we hope to not only present state-of-the-art research on causal models for search and recommendation, but also generate insightful debates about how causal modeling can enhance
ethical decision support systems, to a broad scope of community including but not limited to IR, machine learning, AI, Data Science, and beyond. To achieve this goal, we will not only invite speakers/panelists from technical sectors, but also humanity sectors such as legal and political science researchers.

2 THEMES AND TOPICS OF INTEREST
The purpose of the workshop is to gather researchers and practitioners of the community to communicate the latest ideas and research achievements on causal modeling for search and recommendation, discuss about the advantages and disadvantages of existing approaches, and share the ideas of future directions. Based on this workshop, we would like to not only present the latest research achievements, but also connect researchers in the community that are interested in the causal learning topic to promote this direction in the following years. The main themes and topics of the workshop include but are not limited to:

- **New Models for Causal IR**
  - Interventional learning/reasoning models
  - Counterfactual learning/reasoning models
  - Causal mining models
  - Sequential causal modeling
  - Novel causal priors for IR
  - Novel causal structures for IR
- **Theoretical Guarantees for Causal Models**
  - Theory for treatment-effect estimation
  - Theory for causal structural models
  - Theory for unmeasured confounder
  - Identifiability of causal models
  - Theory on causal graph discovery
- **Causal models for Explainable IR**
  - Causal explainable search
  - Causal explainable recommendation
  - Causal explainable question answering
  - Causal explainable conversational systems
  - Counterfactual explanations
  - Causal evaluation of explanations
  - Causal multimodal explanations
- **Causal models for Unbiased IR**
  - Causal data debias
  - Causal model debias
  - Causal debias in search
  - Causal debias in recommendation
  - Feedback loops and echo chambers
- **Causal models for Fairness in IR**
  - Counterfactual fairness in search
  - Counterfactual fairness in recommendation
  - Causal evaluation of fairness
  - Fairness-utility trade-off
- **Causal models for Robust IR**
  - Causal anti-spam models
  - Causal shilling attack detection
  - Causal data imputation
  - Counterfactual fake detection
  - Causal sensitivity analysis of IR models
- **Multi-modality Causal Learning**
  - Text-based causal learning
  - Image-based causal learning
  - Knowledge-enhanced causal learning
  - Audio/video-based causal learning
  - Causal learning with heterogeneous information
- **User Behavior and Causal Models**
  - User interaction with causal models
  - Causal attribution of user behaviors
  - Causal click models
  - Causal assumption of user behaviors
- **Evaluation of Causal Models**
  - Treatment effect of causal models
  - Interventional evaluation
  - Counterfactual evaluation
  - User study for causal evaluation
  - New datasets for causal evaluation
- **Causal Modeling for Different Applications**
  - Causal modeling in search engine
  - Causal modeling in recommender systems
  - Causal modeling in e-commerce
  - Causal modeling in social networks
  - Causal modeling in QA systems
  - Causal modeling in conversational systems

3 FORMAT AND ACTIVITIES
The workshop is a full day workshop with two keynote speeches (one research keynote and one industry keynote), a panel discussion about causality in IR (and causality in AI in general), several paper presentations, and a poster session to give researchers opportunity for extensive discussions. The schedule of events include:

- A research keynote speech on causal information retrieval or general causal AI.
- An industry keynote speech to introduce the influence and impact of causal machine learning in practice, society, public decision making, and legal perspectives.
- A panel discussion about causal information retrieval and causal AI in general.
- A mixture of long and short paper presentations.
- A poster session for researchers to communicate and make discussions extensively.

4 ORGANIZERS
Biography of the organizers and their main research experience related to the proposed workshop topic are as follows.

Yongfeng Zhang is an Assistant Professor in the Department of Computer Science at Rutgers University (The State University of New Jersey). His research interest is in Information Retrieval, Economics of Data Science, Explainable AI, Fairness in AI and AI Ethics. In the previous he was a postdoc advised by Prof. W. Bruce Croft in the Center for Intelligent Information Retrieval (CIIR) at UMass Amherst, and did his PhD and BE in Computer Science at Tsinghua University, with a BS in Economics at Peking University. He is a Siebel Scholar of the class 2015. Together with coauthors, he has been consistently working on explainable search
and recommendation models [1–3, 7, 8, 10–12, 14, 22–25, 49–51, 60–62, 64, 67, 69, 75, 83], Fairness-aware machine learning [14, 18, 21, 31, 32, 34, 45, 58, 59], conversational search and recommendation [4, 37–41, 55–57, 65, 84], neural logic/symbolic reasoning [6, 42, 51, 83], economic machine learning [19, 20, 30, 33, 35, 54, 80, 81], efficient/robust machine learning [5, 5, 13, 27, 28, 36, 43, 44, 63, 68, 71–74, 76], knowledge graph embedding [9, 14, 27, 28, 46, 49–51, 84], legal/medical retrieval [29, 47, 48], as well as causal/counterfactual models for information retrieval [52, 53, 58, 82]. His recent research on causality in search and recommendation include causal collaborative filtering, causal explainable recommendation, counterfactual debiasing models, and causal models for mitigating feedback loops in IR systems. He has served as PC members or senior PC members in various WebIR related conferences such as SIGIR, WWW, CIKM, WSDM, ICTIR and CHIIR, and he is serving as the associate editor for ACM Transactions on Information Systems (TOIS). He has experience in organizing tutorials and workshops on Web/IR-related conferences such as SIGIR, WWW, WSDM, ICTIR and IUI [15–17, 26, 66, 70, 77–79].

Xu Chen is an Assistant Professor in the Gaoling School of Artificial Intelligence at the Renmin University of China. Prior to that he was a postdoc researcher in the University of College London, working with Prof. Jun Wang. He obtained his PhD degree from Tsinghua University, Beijing, China. His research aims to build interpretable models to understand user decision making process, especially under dynamic, heterogeneous and interactive environments. In the past few years, he explored to leverage textual, visual and time information to explain the collaborative filtering methods, and applied results into the area of recommender system. His paper “Aesthetic-based clothing recommendation” was awarded as a best paper honorable mention on The Web Conference 2018, and another paper “A Collaborative Neural Model for Rating Prediction by Leveraging User Reviews and Product Images” won the best paper award on AIRS 2017. His recent work on causal modeling for search and recommendation include causal sequential recommendation, causal data augmentation, and causal explainable recommendation.

Yi Zhang is a professor in the School of Engineering, University of California Santa Cruz. Her research interests include large scale information retrieval, recommendation systems, internet advertising, data mining, natural language processing, and applied machine learning. She has published chapters, journal articles, and papers at top conferences in these areas, such as ACM SIGIR, WWW, CIKM, IEEE ICDM, ICMRL, COLINGHS, HLT. She received NSF Faculty Early Career Award in 2010, an Air Force Research Young Investigator Award in 2008, the Best Paper Award at ACM SIGIR in 2002, and several other awards. Her Information Retrieval and Knowledge Management Lab is doing research sponsored by several government agencies and companies (Microsoft, Yahoo, Google, NEC, Bosch, Nokia etc.). She has served as a consultant or technical advisor for companies. She regularly serves on the program committees of the very best conferences in her research areas. She has served as area chair or senior PC member at ACM SIGIR, EMNLP, and ACM Recommender Systems. She has served as conference co-chair in charge of Information Retrieval area at the ACM Conference on Information and Knowledge Management, and tutorial chair for ACM SIGIR. She served as an associate editor for ACM Transaction on Information Systems. Dr. Zhang received her Ph.D. from School of Computer Science at Carnegie Mellon University, specializing in Language and Information Technologies.

Xianjie Chen is a Research Scientist at Facebook AI, where he works on machine learning algorithms and platforms that advance both products and research. His research interest is in Recommendation and Personalization Systems, Computer Vision, and Machine Learning. Prior to joining Facebook, he obtained his PhD degree at the University of California, Los Angeles in 2016, with a focus on machine learning. Prior to PhD, he received his bachelor degree in computer science from Tsinghua University in 2011.

ACKNOWLEDGEMENT

The workshop is supported in part by NSF IIS-1910154 and IIS-2007907. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsors.

REFERENCES

[14] Zushuai Fu, Yikun Xian, Ruoyuan Gao, Jieyu Zhao, Quoqing Huang, Yingqiang Ge, Shuyuan Xu, Shujie Geng, Chiara Shah, Yongfeng Zhang, et al. 2020. Fairness-aware explainable recommendation over knowledge graphs. In SIGIR.

[23] Lei Li, Yongfeng Zhang, and Li Chen. 2020. Generate neural template explanations for recommender systems with Expected Utility in CIKM.


[38] Chen Qun, Liu Yang, W Bruce Croft, Johanne R Trippas, Yongfeng Zhang, and Minghui Qiu. 2018. Analyzing and characterizing user intent in information-seeking conversations. In SIGIR.


[40] C. Qu, L. Yang, W. B. Croft, Y. Zhang, and M. Iyyer. 2019. BERT with history answer embedding for conversational question answering. In SIGIR.


[53] Yongfeng Zhang, Min Zhang, Yi Zhang, Yuqin Liu, and Shaoping Ma. 2014. Understanding the sparsity: Augmented matrix factorization with sampled constraints on unobservables. In CIKM.


