REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
12-11-2020	Final Report	15-Jul-2015 - 14-Sep-2020

4. TITLE AND SUBTITLE	
Final Report: Research Area 5.3.3: Sybil-Resilient Influence Measurement in Microblogging Systems	5a. CONTRACT NUMBER W911NF-15-1-0328
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Arizona State University ORSPA P.O. Box 876011 Tempe, AZ 85287 -6011	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 66275-NS.8

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Yanchao Zhang
UU	UU	UU	UU		19b. TELEPHONE NUMBER 480-727-0039

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 16-Apr-2021

Agency Code: 21XD

Proposal Number: 66275NS
Agreement Number: W911NF-15-1-0328

INVESTIGATOR(S):

Name: Yanchao Zhang
Email: yczhang@asu.edu
Phone Number: 4807270039
Principal: Y

Organization: **Arizona State University**
Address: ORSPA, Tempe, AZ 852876011
Country: USA
DUNS Number: 943360412
EIN: 860196696

Report Date: 14-Dec-2020
Date Received: 12-Nov-2020
Final Report for Period Beginning 15-Jul-2015 and Ending 14-Sep-2020
Title: Research Area 5.3.3: Sybil-Resilient Influence Measurement in Microblogging Systems
Begin Performance Period: 15-Jul-2015
End Performance Period: 14-Sep-2020
Report Term: 0-Other
Submitted By: Yanchao Zhang
Email: yczhang@asu.edu
Phone: (480) 727-0039
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 4
STEM Participants: 9

Major Goals: This report is submitted on behalf of Prof. Huan Liu who is the PI for the second project duration with ARL.

AI-Assisted, Risk-Informed and Reasoning-based Decision Making under Uncertainty

Huan Liu, Ph.D., Professor of Computer Science and Engineering
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, AZ 8587-8809
Tel: (480) 727 7349
Email: huan.liu@asu.edu
POC at ARL: Dr. Adrienne Raglin

In this project, we focus on Army relevance to the pressing need for reasoning under uncertainty for C3I (Command, Control, Communications and Intelligence) and causal learning for real-time decision making. We investigate the application of causality and explainable learning for decision making.

Research with Army Relevance and Objectives

1. Reasoning under Uncertainty
It is critical to enable analysts to quickly provide risk-informed courses-of-action to reduce uncertainty in decision making for continuous and transparent mission command and C3I. The efficacy and advantage provided by machine learning (ML) and artificial intelligence (AI) will be a primary decider in battlefield dominance. However, these ML systems are not without vulnerabilities because uncertainty is inherently embedded in their underlying computational functions. There are challenges for intelligence analysts and commanders, for example, reducing time required by traditional analysis necessitating dependency on advanced technology, filtering large amounts of incomplete data to obtain relevant information, obtaining information from diverse targets of potential threats, and ensuring that a decision maker knows and understands the limitations and risks inherent in advanced technology-produced analysis.

2. Causal Learning for Time Series Decision Making

A mission-critical military network may be composed with heterogenous sensing and networking devices that are equipped for various mission-critical entities (MCEs) such as vehicles, drones, humans (e.g., soldiers), airplanes, satellites, stations and bases, etc. These MCEs generate unprecedented amounts of time series data. Time series are uniquely interesting because they can be essential in addressing questions of causality, trends and the likelihood of future outcomes.

Decision making from time series data provides us with the ability of developing smart outcomes to improve battle-field situation-awareness. However, most decision making algorithms rely on correlations to discover patterns and make suggestions for decision making. One of the key weaknesses of correlation-based machine learning is that there could be many spurious correlations that can be detrimental to decision making. One alternative to correlations learned from data is to learn causations from data as causal relations have distinctive characteristics from correlations and can help eliminate spurious correlations. To this end, we aim to create causality based machine learning frameworks to perform decision making tasks specifically with time series data.

3. Explainable Learning for Decision Making

The recent success of machine learning has led to a surge of Artificial Intelligence (AI) applications. The machine learning based AI systems have excelled in myriad of learning tasks such as image classification, natural language processing, speech recognition, autonomous driving and computer games, to name a few. Despite their empirical success, the effectiveness of these AI systems are often limited due to its black-box nature, and are incapable in explaining why the decisions and the actions are made to human users. To this end, building explainable AI systems are crucial in helping users to comprehend, trust and manage the emerging AI techniques.

Accomplishments: Please see the updated document for details.

Training Opportunities: The team consists of the PI (Huan Liu), PhD students (Raha, Ruocheng Guo, Lu Cheng) and a part-time post-doc with this project (Jiayue Jerry Li who is mainly associated with Professor Dijiang Huang's lab).

RPPR Final Report
as of 16-Apr-2021

Results Dissemination: We list our publications and manuscripts in two categories: (1) Accepted and (2) Under Review.

Accepted Papers:
1) Raha Moraffah, Mansooreh Karami, Ruochen Guo, Adrienne Raglin, and Huan Liu . "Causal Interpretability for Machine Learning-Problems, Methods and Evaluation." ACM SIGKDD Explorations Newsletter 22.1 (2020): 18-33.
2) Lu Cheng, Raha Moraffah, Ruocheng Guo, Kasim Selcuk Candan, Adrienne Raglin, and Huan Liu. A Practical Data Repository for Causal Learning with Big Data. BENCH2019.
3) Raha Moraffah, Lu Cheng, Ruocheng Guo, Kasim Selcuk Candan, Adrienne Raglin, and Huan Liu, CAUSE: A Data Repository for Causal Inference. Demo paper, In SBP-BRiMS 2019.
4) Adrienne Raglin, Anshuman Venkantaswaran, and Huan Liu. Abductive Causal Reasoning for Internet of Things. The 5th IEEE Conference on Internet of People, 2019.
5) Vineeth Rakesh*, Ruocheng Guo*, Raha Moraffah, Nitin Agarwal, and Huan Liu (* Equal Contribution) Linked Causal Variational Autoencoder for Inferring Paired Spillover Effects, in CIKM 2018.
6) Lu Cheng, Ruocheng Guo, and Huan Liu. Robust Cyberbullying Detection with Causal Interpretation. In Proceedings of the WWW'19 CyberSafety Workshop.
7) Lu Cheng, Ruocheng Guo, and Huan Liu. Robust Domain Invariant Representation Learning for Imbalanced Text Classification. SDM 2020
8) Ruocheng Guo, Lu Cheng, Jundong Liu, Richar Hahn, and Huan Liu. "A survey of learning causality with data: Problems and methods." ACM Computing Surveys (CSUR) 53.4 (2020): 1-37.
9) Ruocheng Guo, Jundong Li, Yichuan Liu, Kasim Selcuk Candan,Adrienne Raglin, and Huan Liu. IGNITE: A Minimax Game Toward Learning Individual Treatment Effects from Networked Observational Data. IJCAI 2020.
10) Ruocheng Guo, Jundong Li, and Huan Liu. Counterfactual Evaluation of Treatment Assignment Functions with Networked Observational Data. SDM 2020.
11) Ruocheng Guo, Jundong Li, and Huan Liu. Learning individual causal effects from networked observational data. International Conference on Web Search and Data Mining 2020.
12) Adrienne J Raglin, Dijiang Huang, Huan Liu, James McCabe. ``Smart CCR IoT: Internet of Things Testbed", 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC), pp 232-235, December 12, 2019, Los Angeles, CA.

Manuscripts under Review:

1) Raha Moraffah, Kai Shu, Adrienne Raglin, and Huan Liu. Deep Causal Representation Learning for Unsupervised Domain Adaptation. AAAI2021
2) Raha Moraffah, Bahman Moraffah, Mansooreh Karami, Adrienne Raglin, and Huan Liu. CAN: A Causal Adversarial Network for Learning Observational and Interventional Distributions. AAAI 2021
3) Raha Moraffah, Anchit Bhattacharya, Paras Sheth, Mansooreh Karami, Anique Tahir, Adrienne Raglin, and Huan Liu. Causal Inference for time series: A survey. SIGKDD Exploration

Honors and Awards: Nothing to Report

Protocol Activity Status:

Technology Transfer: Nothing to Report

PARTICIPANTS:

Participant Type: Graduate Student (research assistant)
Participant: Anshuman Venkateswaran
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Postdoctoral (scholar, fellow or other postdoctoral position)
Participant: Jiayue Li

Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Ghazaleh Beigi
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Ruocheng Guo
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Tahora Hossein Nazer
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Jundong Li
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: PD/PI
Participant: Huan Liu
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Deepak Mahudeswaran
Person Months Worked: 4.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Raha Moraffah
Person Months Worked: 14.00 **Funding Support:**
Project Contribution:

RPPR Final Report
as of 16-Apr-2021

National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Kai Shu
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE/ACM Transactions on Networking
Publication Identifier Type: Other Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 9/1/16 12:00AM Date Published: 9/1/16 6:06AM
Publication Location:
Article Title: TrueTrueTop: A sybil-resilient system for user influence measurement on Twitter
Authors: Jinxue Zhang, Rui Zhang, Jinchao Zhang, Yanchao Zhang, Chi Zhang
Keywords: Influence measurement, social networks, Twitter, sybil resilience
Abstract: Influential users have great potential for accelerating information dissemination and acquisition on Twitter. How to measure the influence of Twitter users has attracted significant academic and industrial attention. Existing influence measurement techniques are vulnerable to sybil users that are thriving on Twitter. Although sybil defenses for online social networks have been extensively investigated, they commonly assume unique mappings from human-established trust relationships to online social associations and thus do not apply to Twitter where users can freely follow each other. This paper presents TrueTop, the first sybil-resilient system to measure the influence of Twitter users.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Dependable and Secure Computing
Publication Identifier Type: Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 9/1/16 12:00AM Date Published: 9/1/16 6:11AM
Publication Location:
Article Title: The Rise of Social Botnets: Attacks and Countermeasures
Authors: Jinxue Zhang, Rui Zhang, Yanchao Zhang, Guanhua Yan
Keywords: Social bot, social botnet, spam distribution, digital influence
Abstract: Online social networks (OSNs) are increasingly threatened by social bots which are software-controlled OSN accounts that mimic human users with malicious intentions. A social botnet refers to a group of social bots under the control of a single botmaster, which collaborate to conduct malicious behavior while mimicking the interactions among normal OSN users to reduce their individual risk of being detected. We demonstrate the effectiveness and advantages of exploiting a social botnet for spam distribution and digital-influence manipulation through real experiments on Twitter and also trace-driven simulations. We also propose the corresponding countermeasures and evaluate their effectiveness. Our results can help understand the potentially detrimental effects of social botnets and help OSNs improve their bot(net) detection systems.
Distribution Statement: 1-Approved for public release; distribution is unlimited.
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: International AAAI Conference on Web and Social Media (ICWSM)
Date Received: 01-Sep-2016 Conference Date: 18-May-2016 Date Published: 18-Aug-2016
Conference Location: Cologne, Germany
Paper Title: our Age Is No Secret: Inferring Microbloggers' Ages via Content and Interaction Analysis
Authors: Jinxue Zhang, Xia Hu, Yanchao Zhang, Huan Liu
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE Conference on Communications and Network Security (CNS)
Date Received: 01-Sep-2016 Conference Date: 29-Sep-2015 Date Published: 29-Sep-2015
Conference Location: Florence, Italy
Paper Title: Your Actions Tell Where You Are: Uncovering Twitter Users in a Metropolitan Area
Authors: Jinxue Zhang, Jinchao Sun, Rui Zhang, Yanchao Zhang
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: SIAM International Conference on Data Mining (SDM)
Date Received: 01-Sep-2016 Conference Date: 06-May-2016 Date Published: 06-May-2016
Conference Location: Miami, FL
Paper Title: Exploiting Emotional Information for Trust/Distrust Prediction
Authors: Ghazaleh Beigi, Jiliang Tang, Suhang Wang, Huan Liu
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: nternational AAAI Conference on Web and Social Media (ICWSM)
Date Received: 01-Sep-2016 Conference Date: 18-May-2016 Date Published: 18-May-2016
Conference Location: Cologne, Germany
Paper Title: Signed link analysis in social media networks
Authors: Ghazaleh Beigi, Jiliang Tang, Huan Liu
Acknowledged Federal Support: **Y**

DISSERTATIONS:

Publication Type: Thesis or Dissertation
Institution: Arizona State University
Date Received: 01-Sep-2016 Completion Date: 5/28/16 6:06AM
Title: Secure and Privacy-Preserving Microblogging Services: Attacks and Defenses
Authors: Jinxue Zhang
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 16-Apr-2021

Partners

,

I certify that the information in the report is complete and accurate:
Signature:
Signature Date:

Final Report for the project

AI-Assisted, Risk-Informed and Reasoning-based Decision Making under Uncertainty

Huan Liu, Ph.D., Professor of Computer Science and Engineering
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University, Tempe, AZ 8587-8809
Tel: (480) 727 7349
Email: huan.liu@asu.edu
POC at ARL: Dr. Adrienne Raglin

In this project, we focus on Army relevance to the pressing need for reasoning under uncertainty for C3I (Command, Control, Communications and Intelligence) and causal learning for real-time decision making. We investigate the application of causality and explainable learning for decision making.

1. Research with Army Relevance and Objectives

1.1 Reasoning under Uncertainty

It is critical to enable analysts to quickly provide risk-informed courses-of-action to reduce uncertainty in decision making for continuous and transparent mission command and C3I. The efficacy and advantage provided by machine learning (ML) and artificial intelligence (AI) will be a primary decider in battlefield dominance. However, these ML systems are not without vulnerabilities because uncertainty is inherently embedded in their underlying computational functions. There are challenges for intelligence analysts and commanders, for example, reducing time required by traditional analysis necessitating dependency on advanced technology, filtering large amounts of incomplete data to obtain relevant information, obtaining information from diverse targets of potential threats, and ensuring that a decision maker knows and understands the limitations and risks inherent in advanced technology-produced analysis.

1.2 Causal Learning for Time Series Decision Making

A mission-critical military network may be composed with heterogenous sensing and networking devices that are equipped for various mission-critical entities (MCEs) such as vehicles, drones, humans (e.g., soldiers), airplanes, satellites, stations and bases, etc. These MCEs generate unprecedented amounts of time series data. Time series are uniquely interesting because they can be essential in addressing questions of causality, trends and the likelihood of future outcomes. Decision making from time series data provides us with the ability of developing smart outcomes to improve battle-field situation-awareness. However, most decision making algorithms rely on correlations to discover patterns and make suggestions for decision making. One of the key weaknesses of correlation-based machine learning is that there could be many spurious correlations that can be detrimental to decision making. One alternative to correlations learned from data is to learn causations from data as causal relations have distinctive characteristics from correlations and can help eliminate spurious correlations. To this end, we aim to create causality based machine learning frameworks to perform decision making tasks specifically with time series data.

1.3 Explainable Learning for Decision Making

The recent success of machine learning has led to a surge of Artificial Intelligence (AI) applications. The machine learning based AI systems have excelled in myriad of learning tasks such as image classification, natural language processing, speech recognition, autonomous driving and computer games, to name a few. Despite their empirical success, the effectiveness of these AI systems are often limited due to its black-box nature, and are incapable in explaining why the decisions and the actions are made to human users. To this end, building explainable AI systems are crucial in helping users to comprehend, trust and manage the emerging AI techniques.

2. Research Tasks

We have performed research in five lines to pave way for further research.

Causal Learning for Data Augmentation

Generative Adversarial Networks (GANs) are ubiquitous tools for non-parametric sampling from complicated and high-dimensional distributions. GANs have achieved promising results in generating sharp-looking and realistic images and videos. One well-known extension of GAN is conditional GAN (cGAN) which enables sampling from conditional distributions. Traditional cGANs assume the labels are independent and changing one label does not affect the distribution of other labels. However, this assumption does not hold in many real-world cases and in fact labels often have causal relationships with each other. Ignoring such dependencies could result in generating unrealistic samples. To consider dependencies between labels, we propose a generative Causal Adversarial Network (CAN). Our proposed framework learns the causal relations from the data and generates samples accordingly. In addition to the relationships between labels, our model also learns the label-pixel and pixel-pixel dependencies and incorporate them in sample generation. We quantitatively and qualitatively assess the performance of CAN and empirically show that our model is able to generate both interventional and observational samples without having access to the causal graph for the application of face generation on CelebA data.

Causal Inference for Time Series Decision Making and Interpretablity

Causality is the degree to which one can rule out plausible explanations or competing correlations. Causal learning is to learning causations from observational data. It consists of two major tasks i.e., causal discovery and causal treatment effect estimation. Causal discovery is the task of learning the causal relationships and interactions amongst different variables in the system, whereas causal treatment effect estimation aims to learn the effect of a specific variable (a.k.a. treatment) on another one (a.k.a. outcome). Causal discovery helps us obtain insight into the underlying data generation process, interpret the data and build and improve the theories of modeling the systems. It is also crucial for creating interpretable systems. Treatment effect estimation is also a prevalent task in many fields of science.

Recognizing the abundance and importance of time series information for decision making and the unique features of causal learning differing from correlation-based learning, we aim to characterize and understand dynamic time-series information for decision making via causal learning. In addition, time series data, if used with caution, can also assist causal learning.

In this work, we do a comprehensive survey on state of the art methods and compare them with each other for both causal discovery and causal treatment effect approaches for time series data. We also collect the evaluations metrics and benchmark datasets for both tasks.

Machine learning models have had discernible achievements in a myriad of applications. However, most of these models are black-boxes, and it is obscure how the decisions are made by them. This makes the models unreliable and untrustworthy. To provide insights into the decision making processes of these models, a variety of traditional interpretable models have been proposed. To generate more human-friendly explanations, recent work on interpretability tries to answer questions related to causality such as "Why does this model makes such decisions?" or "Was it a specific feature that caused the decision made by the model?". Models that aim to answer causal questions are referred to as causal interpretable models. The existing surveys have covered concepts and methodologies of traditional interpretability. Therefore, we present a comprehensive survey on causal interpretable models from the aspects of the problems and methods. In addition, this survey provides in-depth insights into the existing evaluation metrics for measuring interpretability, which can help practitioners understand for what scenarios each evaluation metric is suitable. We aim to leverage the methodologies introduced in our work to create causal interpretable machine learning algorithms for decision making.

Abductive Causal Reasoning

The Internet of Things (IoT) is defined as a "global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies." This network of heterogeneous devices that are connected can interact and share data as well as aid in decision making. Abductive causal reasoning (ACR) offers a new perspective to learn from data with the capability to support decisions by explaining the context or environment in which the choices are made. In this work, we propose to use ACR in an IoT environment to infer sensible hypotheses from IoT data and explain how decisions can be made or improved. Examples are given to illustrate how ACR can support decisions from the available data. The datasets suitable for the extension of this work are investigated and will be used for our research, development, and evaluation of abductive causal reasoning.

Causal Features, Deconfounding, and Learning to Rank Potential Outcomes

Studies show that the representations learned by deep neural network models can be transferred to similar prediction tasks in other domains for which we do not have enough labeled data. Recent research on deep domain adaptation proposed to mitigate this problem by forcing the deep model to learn more transferable feature representations across domains. In this work, we propose a novel deep causal representation learning framework for unsupervised domain adaptation, in which we propose to learn domain invariant and at the same time more interpretable features by learning the causal representations of the input. This is achieved by simulating a virtual target domain using reweighted samples from the source domain and estimating the causal effect of features on the outcomes. This allows the learned representations to benefit from the deep network and be more transferable. Our extensive experiments show that the proposed model has both notable classification performance for target domains and is able to learn the causal feature representations which are interpretable.

The massive observational data with auxiliary network information draws increasing attention to the problem of learning individual treatment effects across many research areas such as economics and healthcare. To validate causal effects learned from such data, we need to handle confounding bias. A common practice is to assume that all the confounders have been measured in the set of features, which is often untenable. This work investigates if the auxiliary network information can help mitigate confounding bias. We propose a novel causal inference framework, the graph attention deconfounder, to learn representations of confounders by unraveling patterns of hidden confounders from the auxiliary network structure, and conduct extensive experiments for evaluation.

Applying machine learning to personalized treatment decision making using observational data is a problem of growing importance in various fields including medicine, education, and economics. Despite the recent success in handling confounding bias of observational data to infer individual treatment effects, the existing work focuses on finding out which treatment leads to the optimal outcome while the uncertainty of treatment receivers is underexamined. This inspires us to investigate the novel problem of learning personalized treatment decision under uncertainty by incorporating the uncertain behaviors of receivers in the treatment decision process. We propose a novel counterfactual inference framework - Learning to Rank Potential Outcomes (L2RP) to compute the post-treatment representations of both observed and counterfactual instances such that the gap between their distributions can be minimized. Empirically, we demonstrate the efficacy of the proposed L2RP framework over the state-of-the-art approaches on both semi-synthetic and real-world datasets.

A Testbed with Real and Synthetic Data

We have figured a proof-of-concept solution to generate IoT data from heterogeneous sources. Currently, a docker-based testbed is set up to automatically generate pollution information (.csv) with timestamps based on an AQI metric (see the screenshot below as an example). It is expected to generate large-scale heterogeneous data streams by modifying the configuration of the data and specifying objective metrics.

ozone	particullate_matter	carbon_monoxide	sulfure_dioxide	nitrogen_dioxide	longitude	latitude	timestamp
103	81	33	31	40	10.1165896654129	56.2257947825602	2014-08-01 00:05:00
99	79	35	36	44	10.1165896654129	56.2257947825602	2014-08-01 00:10:00
102	80	33	36	47	10.1165896654129	56.2257947825602	2014-08-01 00:15:00
107	82	33	36	45	10.1165896654129	56.2257947825602	2014-08-01 00:20:00
102	84	28	38	42	10.1165896654129	56.2257947825602	2014-08-01 00:25:00
103	82	29	42	41	10.1165896654129	56.2257947825602	2014-08-01 00:30:00
101	83	29	43	36	10.1165896654129	56.2257947825602	2014-08-01 00:35:00
98	80	28	39	36	10.1165896654129	56.2257947825602	2014-08-01 00:40:00
100	80	32	37	37	10.1165896654129	56.2257947825602	2014-08-01 00:45:00
105	80	37	39	35	10.1165896654129	56.2257947825602	2014-08-01 00:50:00
107	77	41	35	38	10.1165896654129	56.2257947825602	2014-08-01 00:55:00
111	75	43	32	38	10.1165896654129	56.2257947825602	2014-08-01 01:00:00
106	74	38	30	43	10.1165896654129	56.2257947825602	2014-08-01 01:05:00
108	75	39	27	43	10.1165896654129	56.2257947825602	2014-08-01 01:10:00
108	72	38	32	47	10.1165896654129	56.2257947825602	2014-08-01 01:15:00

3. Team, Summary, and Future Plan

The team consists of the PI (Huan Liu), PhD students (Raha, Ruocheng Guo, Lu Cheng) and a part-time post-doc with this project (Jiayue Jerry Li who is mainly associated with Professor Dijiang Huang's lab).

We plan to extend the development of a distributed data collection/processing model based the BlueLight Project at ASU. The data collection/processing model includes a distributed edge cloud data collection and data processing framework that is built on an edge-computing framework established on ASU campus. We plan for an illustrative application scenario, i.e., distributed video surveillance application, based on distributed ML approaches, enabling the use of the distributed data collection and processing to support distributed ML algorithms. The prototype is expected to be ready in next 3 to 4 months. In the meantime, we are working on an emulation approach to incorporate simulation-based solutions to simulate a large sensing system that involves a large number of heterogenous sensors with considerations of their communication, networking, and mobility situations. Moreover, we will explore the distribute machine learning algorithms suitable to run in such a dynamic environment to support military applications.

The Data Mining and Machine Learning lab has proven track record in producing research with impact and values. The research team will continue closely working on causal learning with data with collaborators at ARL on the proposed research tasks ensuring Army relevance. Regular online and face-to-face meetings will continue to exchange ideas and findings. Joint reports and papers will be produced for result dissemination.

4. Publications and Manuscripts in Preparation

We list our publications and manuscripts in two categories: (1) Accepted and (2) Under Review.
Accepted Papers:

1) Raha Moraffah, Mansooreh Karami, Ruochen Guo, Adrienne Raglin, and Huan Liu . "Causal Interpretability for Machine Learning-Problems, Methods and Evaluation." *ACM SIGKDD Explorations Newsletter* 22.1 (2020): 18-33.
2) Lu Cheng, Raha Moraffah, Ruocheng Guo, Kasim Selcuk Candan, Adrienne Raglin, and Huan Liu. A Practical Data Repository for Causal Learning with Big Data. BENCH2019.
3) Raha Moraffah, Lu Cheng, Ruocheng Guo, Kasim Selcuk Candan, Adrienne Raglin, and Huan Liu, CAUSE: A Data Repository for Causal Inference. Demo paper, In SBP-BRiMS 2019.
4) Adrienne Raglin, Anshuman Venkantaswaran, and Huan Liu. Abductive Causal Reasoning for Internet of Things. The 5th IEEE Conference on Internet of People, 2019.
5) Vineeth Rakesh*, Ruocheng Guo*, Raha Moraffah, Nitin Agarwal, and Huan Liu (* Equal Contribution) Linked Causal Variational Autoencoder for Inferring Paired Spillover Effects, in CIKM 2018.
6) Lu Cheng, Ruocheng Guo, and Huan Liu. Robust Cyberbullying Detection with Causal Interpretation. In Proceedings of the WWW'19 CyberSafety Workshop.
7) Lu Cheng, Ruocheng Guo, and Huan Liu. Robust Domain Invariant Representation Learning for Imbalanced Text Classification. SDM 2020
8) Ruocheng Guo, Lu Cheng, Jundong Liu, Richar Hahn, and Huan Liu. "A survey of learning causality with data: Problems and methods." *ACM Computing Surveys (CSUR)* 53.4 (2020): 1-37.
9) Ruocheng Guo, Jundong Li, Yichuan Liu, Kasim Selcuk Candan,Adrienne Raglin, and Huan Liu. IGNITE: A Minimax Game Toward Learning Individual Treatment Effects from Networked Observational Data. IJCAI 2020.
10) Ruocheng Guo, Jundong Li, and Huan Liu. Counterfactual Evaluation of Treatment Assignment Functions with Networked Observational Data. SDM 2020.
11) Ruocheng Guo, Jundong Li, and Huan Liu. Learning individual causal effects from networked observational data. International Conference on Web Search and Data Mining 2020.
12) Adrienne J Raglin, Dijiang Huang, Huan Liu, James McCabe. ``Smart CCR IoT: Internet of Things Testbed", 2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC), pp 232-235, December 12, 2019, Los Angeles, CA.

Manuscripts under Review:

1) Raha Moraffah, Kai Shu, Adrienne Raglin, and Huan Liu. Deep Causal Representation Learning for Unsupervised Domain Adaptation. AAAI2021
2) Raha Moraffah, Bahman Moraffah, Mansooreh Karami, Adrienne Raglin, and Huan Liu. CAN: A Causal Adversarial Network for Learning Observational and Interventional Distributions. AAAI 2021
3) Raha Moraffah, Anchit Bhattacharya, Paras Sheth, Mansooreh Karami, Anique Tahir, Adrienne Raglin, and Huan Liu. Causal Inference for time series: A survey. SIGKDD Exploration