Artificial Intelligence Research For High Energy Physics (DE-FOA-0002705)

Using Generalized Adversarial Networks to ensure trusted science results and maximize science reach within the dark matter community

Amy Roberts (Physics), Farnoush Banaei-Kashani (Computer Science)

Abstract

SuperCDMS is a direct-detection dark matter search that aims to answer a top-priority question of the DOE[1], "What is dark matter?" It is one of a suite of experiments that will begin data taking in the next two years. All these dark matter experiments, including SuperCDMS, are facing unprecedented challenges in interpreting their data. Our work looks for a way to test and improve the trustworthiness of these results and addresses two DOE Primary Research Directives: massive scientific data analysis and intelligent decision-support for complex systems[2].

Massive scientific data analysis (AI for HEP)

The upcoming dark matter experiments will be the cleanest, quietest, rare-event detectors ever built. But the dark matter community doesn't yet know how to use this data to effectively search for dark matter. One challenge faced by the dark matter community is how to perform a "blinded" analysis on incoming data. Scientists use blinded analyses to reduce the effect of human bias on the final science result. Blinded analysis is the gold standard within the physics community, but it remains unclear how to effectively perform blinded analysis on the upcoming generation of dark matter data.

Generalized Adversarial Networks (GANs) could make a promising blinding technique – "salting" – feasible for use with dark matter data. "Salting" adds simulated signal to a data set. This allows analyzers to look at the full data set, making it ideal for dark matter analysis. However, creating realistic simulated signal currently takes years. GANs specialize in generating new data that is indistinguishable from the training set and have the potential to create realistic simulated signal within weeks. Using GANs for data salting could provide a much-needed analysis tool to the dark matter community.

Massive scientific data analysis (HEP for AI)

Initial efforts to create "salt" are promising, but it may be possible to improve on initial results by incorporating physics constraints into the GAN model. This is an under-developed area of AI and could result in improved methods for GANs.

Supporting decisions in complex systems (AI for HEP)

The upcoming dark matter experiments together represent millions of dollars of investment from the scientific community. Previous dark matter experiments consistently lost half of all science data because humans cannot analyze that data in real-time.

The work on GANs has the potential to identify changes in incoming data fast enough for scientists to identify and fix issues with the experiment before weeks or months of costly data has been lost. Like the issue of data blinding, improving the time scale of data-quality monitoring would impact the entire dark matter field.

[1] DOE, "Basic Research Needs for Dark Matter Small Project New Initiatives," 2018.

[2] N. Baker *et al.*, "Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence," Feb. 2019.

NSF BIOGRAPHICAL SKETCH

NAME: Roberts, Amy

ORCID: 0000-0001-8538-9155

POSITION TITLE & INSTITUTION: Assistant Professor, University of Colorado Denver

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
University of Notre Dame	Notre Dame, IN	Physics	Ph.D.	2013
Stony Brook University	Stony Brook, NY	Physics, Mathematics	B.S.	2004

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2018 - present Assistant Professor, University of Colorado Denver, Physics, Denver, CO

2014 - 2018 Postdoctoral researcher, University of South Dakota, Physics, Vermillion, SD

2013 - 2014 Postdoctoral researcher, Los Alamos National Laboratory, Sub-atomic Physics, Los Alamos, NM

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Agnese R, et al. Production rate measurement of Tritium and other cosmogenic isotopes in Germanium with CDMSlite. Astroparticle Physics. 2019 January; 104:1-12. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0927650518301774 DOI: 10.1016/j.astropartphys.2018.08.006
2. Aralis T, et al. Constraints on dark photons and axionlike particles from the SuperCDMS Soudan experiment. Physical Review D. 2020; 101(5):-. Available from: https://link.aps.org/doi/10.1103/PhysRevD.101.052008 DOI: 10.1103/PhysRevD.101.052008
3. Agnese R, et al. Projected sensitivity of the SuperCDMS SNOLAB experiment. Physical Review D. 2017; 95(8):-. Available from: http://link.aps.org/doi/10.1103/PhysRevD.95.082002 DOI: 10.1103/PhysRevD.95.082002
4. Agnese R, et al. Search for low-mass dark matter with CDMSlite using a profile likelihood fit. Physical Review D. 2019; 99(6):-. Available from: https://link.aps.org/doi/10.1103/PhysRevD.99.062001 DOI: 10.1103/PhysRevD.99.062001
5. Amaral D, et al. Constraints on low-mass, relic dark matter candidates from a surface-operated SuperCDMS single-charge sensitive detector. Physical Review D. 2020 November 13; 102(9):-. Available from: https://link.aps.org/doi/10.1103/PhysRevD.102.091101 DOI: 10.1103/PhysRevD.102.091101

Other Significant Products, Whether or Not Related to the Proposed Project

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. Editor, Journal of Open Source Science
2. Co-convener, SNOWMASS Computational Frontier (End-User Analysis)
3. XSEDE Campus Champion
4. SuperCDMS Software Working Group Chair
5. SuperCDMS-SNOLAB Data Quality Technical Coordinator

Farnoush Banaei-Kashani

Education and Training

Sharif University of Technology, Tehran, Iran	Computer Engineering	BS, 1996
University of Southern California, Los Angeles, CA	Electrical Engineering	MS, 2001
University of Southern California, Los Angeles, CA	Computer Science	PhD, 2006
Integrated Media Systems Center (NSF ERC), Los Angeles, CA	Computer Science	Postdoc, 2008

Research and Professional Experience

2021-present Associate Professor, Computer Science and Engineering Department, University of Colorado Denver | Anschutz Medical Campus
2014-2021 Assistant Professor, Computer Science and Engineering Department, University of Colorado Denver | Anschutz Medical Campus
2018-present Director, Data-Driven Cybersecurity, GAANN PhD Fellowship
2017-present Director, Data Science in Biomedicine, MS Program Concentration on Data Science in Biomedicine
2015-present Director, Big Data Science and Engineering, GAANN PhD Fellowship
2014-present Director, Big Data Management and Mining, Computer Science and Engineering Department, University of Colorado Denver | Anschutz Medical Campus
2013-2014 Lecturer, Computer Science Department, University of Southern California
2009 - 2014 Research Scientist, Integrated Media Systems Center (NSF ERC), University of Southern California
2009-2014 Associate Director of Research and Education, Integrated Media Systems Center (NSF ERC), University of Southern California
2007 - 2008 Lecturer, Computer Science Department, University of Southern California

Publications

- Nguyen, A., Alqurashi, R., Raghebi, Z., Banaei-Kashani, F., Halbower, A. C., & Vu, T., "A Lightweight And Inexpensive In-ear Sensing System For Automatic Whole-night Sleep Stage Monitoring", ACM Sensys 2016. **[Best Paper award]**

- S. Helmi, F. Banaei-Kashani, "Spatiotemporal Range Pattern Queries on Large-scale Co-movement Pattern Datasets", IEEE Big Data 2017.

- Farnoush Banaei-Kashani, Cyrus Shahabi, Seon Ho Kim, Luciano Nocera, Giorgos Constantinou, Ying Lu, Yinghao Cai, Gérard G. Medioni, Ramakant Nevatia, "Janus - Multi Source Event Detection and Collection System for Effective Surveillance", Journal of Information Processing Systems (JIPS), Vol 10, No 1, pp. 1-22.

- Farnoush Banaei Kashani, Gérard Guy Medioni, Khanh Nguyen, Luciano Nocera, C. Shahabi, Ruizhe Wang, Cesar E Blanco, Yian Chen, Yuchen Chung, "Monitoring Mobility Disorders at Home using 3D Visual Sensors and Mobile Sensors", Wireless Health 2013.

- F. Banaei-Kashani, M. Asghari, M Rahmani, C. Shahabi, Lisa Brenskelle, "SDPF: A Framework for Online, Real-time Cleansing of Upstream Operating Data", 2013 SPE Western Regional Meeting. **US Patent App**. 13/781,62

- Lucas Marzec, Sridharan Raghavan, Farnoush Banaei-Kashani, Seth Creasy, Edward L. Melanson, Leslie Lange, Debashis Ghosh, Michael A. Rosenberg, “Device-measured physical activity data for classification of patients with ventricular arrhythmia events: A pilot investigation”, PLOSOne, October 2018.

- Farnoush Banaei-Kashani. 2015. Efficient K-nearest neighbor search in time-dependent spatial networks. **U.S. Patent 9,062,985**, issued June 23, 2015.

- H. Shirani-Mehr, F. Banaei-Kashani and C. Shahabi, “Reachability Query in Large Evolving Contact Networks”, VLDB 2012.

- Akdogan, U. Demiryurek, F. Banaei-Kashani and C. Shahabi, “Voronoi-based Geospatial Query Processing with MapReduce”, IEEE CloudCom 2010 **[Best Paper award]**

- Shahabi, M. Jahangiri and F. Banaei-Kashani, “ProDA: An end-to-end wavelet-based OLAP system for massive datasets”, IEEE Computer, Vol. 41, No. 4, April 2008.

- Farnoush Banaei-Kashani, Parisa Ghaemi, Bahman Movaqar, Seyed Jalal Kazemitabar, Efficient maximal reverse skyline query processing, Geoinformatica (2017) 21:549–572.

Synergistic Activities

1. Course Development:

   I have developed and taught a pipeline of seven courses in data management, data mining and data science:
   - Database System Concepts (undergraduate level)
   - Database Systems (graduate level)
   - Data Mining (graduate level),
   - Big Data Mining (advanced graduate level)
   - Data Science (undergraduate and graduate level)
   - Big Data Science (advanced graduate level)
   - Big Data Systems (undergraduate and graduate level)

2. Professional Society Services:
   - Conference Chair: ACM SIGSPATIAL 2018-2019
   - Workshop Chair: IEEE Workshop on Big Spatial Data (BSD), 2016-2018, International Workshop on GeoStreaming (IWGS), 2012-2017
   - Technical Program Committee Member: ACMGIS 2015, GeoRich 2015 (in conjunction with SIGMOD/PODS), ACMGIS 2014, GeoRich 2014 (in conjunction with SIGMOD/PODS), ACMGIS 2013, MDM 2013 (Demo Track), MobiDE 2013 (in conjunction with SIGMOD), ACMGIS 2012; SensorKDD 2012 (in conjunction with KDD), ACS 2012 (in conjunction with SDM), SensorKDD 2011 (in conjunction with KDD), AP2PS 2011, ICDE 2010, AP2PS 2010, AP2PS 2009, SNAS 2009
   - Referee: TODS, TPDS, TKDE, ACM-MM, SIGMOD, VLDB, ICDE, SIGCOMM, SIGKDD, CIKM, DEXA, SSTDM, USENIX