Aryan Arbabi

Department of Computer Science
10 King's College Road, Rm. 3302
Toronto, Ontario M5S 3G4, CANADA

Email: arbabi@cs.toronto.edu

Education

- **University of Toronto**, Toronto, Canada — Sep. 2016 - present
  PhD student in Computer Science
  - Also affiliated with Vector Institute, Sickkids hospital and University Health Network (UHN)
  - Thesis: "Machine Learning for Generating and Understanding Medical Text"
  - Adviser: Prof. Michael Brudno
- **University of Toronto**, Toronto, Canada — Sep. 2013 - Jun. 2015
  MSc in Computer Science
  - Thesis: "Cell Free DNA Fragment-Size Distribution Analysis (FSDA) for Non-Invasive Prenatal CNV Prediction"
  - Adviser: Prof. Michael Brudno
- **Sharif University of Technology**, Tehran, Iran — Sep. 2009 - Aug. 2013
  BSc in Computer Engineering
  - Thesis: "ARYANA: A Sequence Aligner for Next Generation Sequencing Data"

Work Experience

- **Google Research**, Palo Alto, CA, USA
  Research Intern — July. 2019 - Oct. 2019
  Research on generating descriptive text snippets conditioned on structured source data.
- **Google Research**, New York, NY, USA
  Software Engineering Intern — Jun. 2018 - Sep. 2018
  Research on Variational Autoencoders for unsupervised information extraction and representation learning for natural language.
- **The Hospital for Sick Children**, Toronto, ON, Canada
  Research Assistant — Jun. 2015 - Jun. 2020
  Research on Machine Learning and NLP methods to facilitate diagnosis of rare genetic diseases.
- **University of Toronto, Department of Computer Science**, Toronto, ON, Canada
  Research Software Engineer — Jun. 2015 - Aug. 2016
  Designed and developed Machine Learning methods for medical text processing.

Research Interests and Experience

- **Natural Language Processing.** Interested in techniques improving performance in NLU and NLG tasks with limited labeled data. Also interested in tasks involving graphs and natural language. Some of my projects in this direction:
  - Neural Concept Recognizer tool (NCR), improving detection of biomedical concepts mentioned in text, by utilizing the hierarchical relations between the concepts.
  - Unsupervised method based on Variational Autoencoders (VAEs), to detect informative words, such as proper nouns, in task oriented dialogue systems, without any annotated text.
  - BERT-based method that generates textual descriptions, conditioned on structured input such as knowledge graphs.
- **Graph Neural Networks.** Other than tasks involving graphs and natural language, I have recently worked on utilizing Graph Neural Networks on protein-protein-interaction (PPI) networks, to predict functions for novel proteins. The task can be also viewed as a special case of Recommender Systems.
- **Bayesian Deep Learning.** Interested in both theoretical and practical aspects of bayesian deep learning, approximate inference, deep generative models and information theoretic analysis of deep learning methods. Recently have been exploring using VAEs and EBMs for Recommender Systems.

Publications

- **Aryan Arbabi** and Michael Brudno. Learning to propagate labels for protein function prediction. *under preparation*, 2021
- **Aryan Arbabi**, Mingqiu Wang, Nan Du, Laurent El Shafey, and Izhak Shafran. R2d2: Relational text decoding with transformers. *arXiv preprint arXiv:2105.04645*, 2021
- **Aryan Arbabi**, David R Adams, Sanja Fidler, and Michael Brudno. Identifying clinical terms in free-text notes using ontology-guided machine learning. In *International Conference on Research in Computational Molecular Biology*, pages 19–34. Springer, 2019
- Emily Fertig, **Aryan Arbabi**, and Alexander A Alemi. β-vaes can retain label information even at high compression. *Bayesian Deep Learning Workshop at NeurIPS*, 2018
- Yanwei Xi[†], **Aryan Arbabi**[†], Amy JM McNaughton, Alison Hamilton, Danna Hull, Helene Perras, Tillie Chiu, Shawna Morrison, Claire Goldsmith, Emilie Creede, et al. Noninvasive prenatal detection of trisomy 21 by targeted semiconductor sequencing: A technical feasibility study. *Fetal Diagnosis and Therapy*, 2017
  [†] These authors contributed equally to the work
- **Aryan Arbabi**, Ladislav Rampášek, and Michael Brudno. Cell-free DNA fragment-size distribution analysis for non-invasive prenatal CNV prediction. *Bioinformatics*, 32(11):1662–1669, 2016
- Ladislav Rampášek, **Aryan Arbabi**, and Michael Brudno. Probabilistic method for detecting copy number variation in a fetal genome using maternal plasma sequencing. *Bioinformatics*, 30(12):i212–i218, 2014
- Milad Gholami[†], **Aryan Arbabi**[†], Ali Sharifi-Zarchi, Hamidreza Chitsaz, and Mehdi Sadeghi. ARYANA: Aligning Reads by Yet Another Approach. *BMC bioinformatics*, 15(Suppl 9):S12, 2014
  [†] These authors contributed equally to the work
- **Aryan Arbabi**, Milad Gholami, Mojtaba Varmazyar, and Shervin Daneshpajouh. Fast cpu-based DNA exact sequence aligner. In *Formal Methods and Models for Codesign (MEMOCODE), 2012 10th IEEE/ACM International Conference on*, pages 95–98. IEEE, 2012

Honors and Awards

- **14$^{\text{th}}$ team** in the ACM-ICPC East Central North America Regional Contest, Nov. 2013, Windsor, Canada.
- **1$^{\text{st}}$ team** in the Memocode 2012 Design Contest, Performance per Cost Section.
- Recipient of the **grant for undergraduate studies** from the Iranian National Elites Foundation, for outstanding academic success, 2009 - 2013.
- **Gold Medal** in the Iranian National Olympiad in Informatics (INOI), 2008.

Teaching Experience

- **University of Toronto**, Toronto, Canada
  Teaching Assistant Sep. 2013 - present
  Courses: Probabilistic Learning and Reasoning, Introduction to Machine Learning, Natural Language Computing, Data Structures and Analysis, Introduction to Theory of Computation, Software Design, Introduction to Computer Programming
- **Sharif University of Technology**, Tehran, Iran
  Teaching Assistant Sep. 2011 - Apr. 2013
  Courses: Design and Analysis of Algorithms, Theory of Languages and Automata, Discrete Structures, Advanced Programming
- **Young Scholars Club**, Tehran, Iran
  Teacher Summer 2010
  Taught algorithms and programming to high school students accepted to Iranian Olympiad in Informatics (INOI) summer camp

Technical Skills

- Python, TensorFlow, PyTorch, Java, C/C++