Email: arun@maiya.net
Web: http://arun.maiya.net

Arun S. Maiya

Citizenship: **USA**

Summary

Arun Maiya is a computer scientist at the Institute for Defense Analyses (IDA), a federally-funded research and development center. His research interests broadly focus on computational methods to extract meaning from raw data and include the areas of machine learning, natural language processing, computer vision, data mining, and network science. His work has been published in top-tier computer science venues, and he has served as a program committee member and reviewer for many academic conferences and journals such as the ACM SIGKDD conference and Nature Communications. He has contributed to national-level strategic-planning activities and R&D roadmaps. Dr. Maiya is winner of the Andrew J. Goodpaster Award for Excellence in Research, the Larry D. Welch Award for Best External Publication, and the AFEI Award for Excellence in Enterprise Information. A builder of data tools, he has created software packages used to solve a wide-range of problems across different disciplines. He holds a Ph.D. in Computer Science from the University of Illinois at Chicago.

Research Interests

applied machine learning, data science, natural language processing (NLP), network science, computer vision

Education

Ph.D. in Computer Science (University of Illinois at Chicago)
M.S. in Computer Science (DePaul University)
B.S. in Psychology (University of Illinois at Urbana-Champaign)

Employment

2011-Present	**Research Leader** Institute for Defense Analyses – Alexandria, VA USA
2007-2011	**Researcher** University of Illinois at Chicago (Includes research assistantship, fellowship tenure, thesis research, and other research projects.)
2002-2007	**Vice President** MTR Imports, Inc. – Darien, IL USA
2003-2005	**Research and Development Lead** R&D Division, Hostway Corporation – Chicago, IL USA
2002-2003	**Director of Software Development** EGXpress – Aurora, IL USA
2000-2001	**Member of Technical Staff** Tellabs, Inc. – Naperville, IL USA

Software

- **ktrain:** The ktrain library is an open-source software package that makes machine learning and AI more accessible and easier to apply. Featuring out-of-the-box support for different data types including text, images, and graphs, ktrain has been used for a wide range of different applications in industry, government, and academia. Examples include analyses for the U.S. Economic Census, financial crime analytics at Big 4 accounting firms, intelligence analyses, and `CoronaCentral.ai`, a machine-learning-enhanced search engine for coronavirus publications at Stanford University. For more information: `https://arxiv.org/abs/2004.10703`
- **IDATA:** IDATA (IDA Text Analytics) is a suite of software capabilities designed to facilitate search, exploration, and analyses of very large document sets through state-of-the-art information retrieval, natural language processing, and machine learning. It has been used for a variety of different use cases in the federal government including cyber damage assessments, biosurveillance, and policy analyses. For more information: `https://arxiv.org/abs/1308.2359`
- **CausalNLP:** CausalNLP is a practical toolkit for causal inference from observatonal data that includes text in addition to categorical and numerical variables. Supports text as treatment, outcome, or "controlled-for" variable. For more information: `https://arxiv.org/abs/2106.08043`
- **OnPrem.LLM:** OnPrem.LLM is a simple Python package for generative AI that makes it easier to run large language models (LLMs) on your own machine using non-public data. For more information: `https://amaiya.github.io/onprem/`

Publications

- **A.S. Maiya**. ktrain: A Low-Code Library for Augmented Machine Learning. *Journal of Machine Learning Research (JMLR)*. May 2022. [code: `https://github.com/amaiya/ktrain`]
- **A.S. Maiya**. CausalNLP: A Practical Toolkit for Causal Inference with Text. *arXiv preprint arXiv:2106.08043*. June 2021. [code: `https://github.com/amaiya/causalnlp`]
- **A.S. Maiya**. A Framework for Comparing Groups of Documents. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP '15)*. Lisbon, Portugal. September 2015.
- **A.S. Maiya**, D. Visser, and A. Wan. Mining Measured Information from Text. *Proceedings of the 38th International ACM SIGIR conference on research and development in Information Retrieval (SIGIR '15)*. Santiago, Chile. August 2015.
- **A.S. Maiya** and R.M. Rolfe. Topic Similarity Networks: Visual Analytics for Large Document Sets. *Proceedings of the 2014 IEEE International Conference on Big Data (IEEE BigData '14)*. Washington, D.C. October 2014.
- **A.S. Maiya**, J.P. Thompson, F. Loaiza-Lemos, and R.M. Rolfe. Exploratory Analysis of Highly Heterogeneous Document Collections. *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '13)*. Chicago, IL. August 2013.
- **A.S. Maiya** and T.Y. Berger-Wolf. Expansion and Decentralized Search in Complex Networks. *Journal of Knowledge and Information Systems (KAIS)*. First published online January 2013.
- **A.S. Maiya**, F. Loaiza-Lemos, and R.M. Rolfe. Supervised Learning in the Wild: Text Classification for Critical Technologies. *Proceedings of the IEEE Military Communications Conference (MILCOM '12)*. Orlando, FL. October 2012.
- **A.S. Maiya** and T.Y. Berger-Wolf. Benefits of Bias: Towards Better Characterization of Network Sampling. *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '11)*. San Diego, CA. August 2011.
- M.C. Crofoot, D.I. Rubenstein, **A.S. Maiya**, and T.Y. Berger-Wolf. Aggression, Grooming, and Group-level Cooperation in White-faced Capuchins (Cebus capucinus): Insights from Social Networks. *American Journal of Primatology*. First published online May 2011.
- **A.S. Maiya** and T.Y. Berger-Wolf. Expansion and Search in Networks. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*. Toronto, Canada. October 2010.

- **A.S. Maiya** and T.Y. Berger-Wolf. Online Sampling of High Centrality Individuals in Social Networks. *Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD '10)*. Hyderabad, India. June 2010.
- **A.S. Maiya** and T.Y. Berger-Wolf. Sampling Community Structure. *Proceedings of the 19th ACM International Conference on the World Wide Web (WWW '10)*. Raleigh, NC. April 2010.
- **A.S. Maiya** and T.Y. Berger-Wolf. Inferring the Maximum Likelihood Hierarchy in Social Networks. *Proceedings of the 12th IEEE International Conference on Computational Science and Engineering (CSE '09)*. Vancouver, Canada. August 2009.
- M. Lahiri, **A.S. Maiya**, R. Sulo, Habiba, and T.Y. Berger-Wolf. The Impact of Structural Changes on Predictions of Diffusion in Networks. *ICDM '08 Workshop on Analysis of Dynamic Networks*. Pisa, Italy. December 2008.

Service

- Ad hoc Reviewer
  Nature Communications, Nature Scientific Reports, Physica A, ACM Transactions on Knowledge Discovery from Data (TKDD), ACM Transactions on the Web (TWEB), Social Network Analysis and Mining (SNAM)
- Program Committee Member
  KDD '23, KDD '22, DSAA '22, KDD '21, ICDM '21, DSAA '21, SDM '21, KDD '20, ICDM '20, DSAA '20, KDD '19, ICDM '19, DSAA '19, KDD '18, ECML '18, ICDM '17, SDM '17, KDD '16, ICDM '16, SDM '16, ECML PKDD '16, KDD '16 ODD Workshop, KDD '15, ICDM '15, SDM '15, ECML PKDD '15, KDD '15 ODD Workshop, KDD '14, ICDM '14, KDD '14 ODD Workshop, CIKM '13, SIGMOD '13 DyNetMM Workshop, CIKM '12, AAAI '12, KDD '12 SOMA Workshop, ICDM '11
- Panel Participation:
  Invited Expert, "Performance/Reslience of Defense-Informed Giant Models (PARADIGM)", DARPA/ISAT (2023)
  Panel on "Automated Content Understanding", National Archives and Records Administration (2013)
  Invited Expert, "Democratizing Machine Learning", DARPA/ISAT (2013)
  Panel on "Advanced Computing Capabilities in U.S. National Labs", STPI (2012)
- Contributor to Wiley Encyclopedia of Operations Research and Management Science (2009)

Awards and Honors

2021 Winner, General Andrew J. Goodpaster Award for Excellence in Research (IDA)

> **IDA Staff Notice 22-11:** "I am pleased to announce that Arun Maiya of the Information Technology and Systems Division won the 2021 Goodpaster Award for Excellence in Research ... Arun is a pioneer in applied research and is focused on the use of computational methods to extract meaning from raw data. He has led important, high-impact research and development efforts to help sponsors make sense of huge amounts of data. Arun has deep technical expertise in artificial intelligence, natural language processing, machine learning, computer vision and network science. His major software contributions since joining IDA in 2011 include IDA Text Analytics (IDATA) and ktrain ... Arun sets the standard for the attributes that IDA most values in its researchers: rigorous analytic abilities, trusted expertise, innate leadership skills, and excellent presentation and writing skills."
>
> - General Norton A. Schwartz

2016 Winner, General Larry D. Welch Award for best external publication (IDA)
2015 AFEI Excellence in Enterprise Information Award (AFEI)
2013-2015 Finalist (3 years running), General Larry D. Welch Award for best external publication (IDA)
2007-2008 NSF IGERT Fellowship (National Science Foundation)
2008 NSF Travel Grant (National Science Foundation)
2001 Key Contributor Award (Tellabs, Inc.)