Rahul Krishna
http://rkrsn.us ◇ Github: github.com/rahlk

Email : i.m.ralk@gmail.com
Mobile : +1-919-396-4143
LinkedIn: linkedin.com/in/rkrsn

Technical Skills

General Expertise: Software Engineering, Machine Learning, Statistics, Natural Language Processing, Algorithms, Data structures;
Programming: Python, C, C++, Java, R, Javascript, Scala, Matlab;
Frameworks: Tensorflow, Pytorch, LLVM/Clang, Scikit-learn, Hadoop, Spark, Apache Jena, AWS-EMR, Livy, Elasticsearch, GIS;

Work Experience

IBM │ Research Staff Member Feb 2021 - *current*
Languages: Java, Python; Frameworks: Openshift, Kubernetes;

- **Application Modernization**: Part of the Mono2Micro team working on AI based application modernization and refactoring to migrate applications to Kubernetes.

Columbia University │ Postdoctoral Researcher June 2019 - Feb 2021
Languages: Python, C, C++; Frameworks: Tensorflow, Pytorch, LLVM/Clang, NIVDIA Jetson SoC, Hadoop, Spark;

- **Diagnosing and Fixing Performance Faults in Real-time Systems using Causal Inference**: Developed a causal inference based diagnostics and repair tool, called CADET, for diagnosing root-cause of performance faults in real-time embedded systems. CADET's diagnosis outperformed expert recommendations and other state-of-the-art techniques by $28-52\%$ while doing so up to $40\times$ faster.
- **Scalable Performance Optimization for Big Data systems using MCMC**: Developed a Markov Chain Monte Carlo based configuration optimization prototype, called ConEX, for big data systems. Demonstrated that ConEX can *frugally find near optimum configuration settings* for Hadoop and Spark with over $17\times$ better latency gain than competing ML based approaches.
- **Transferable Graybox fuzz testing with Multi-task Neural Networks**: Designed, implemented, and deployed a Multitask Neural Network based fuzz testing framework, called MTFuzz, to expose vulnerabilities in C/C++ programs. MTFuzz creates a *compact embedding* of the program space to find 11 unseen vulnerabilities and increase code coverage by $2-3\times$.
- **Vulnerability prediction using GGNN and Representation Learning**: Developed a vulnerability prediction tool called REVEAL using GGNN architecture to improve prediction performance in real-world software systems such as Chromium, Debian OS, FFMpeg, and QEMU. REVEAL produces a 33.57% boost in precision and 128.38% boost in recall over state-of-the-art methods.

Phase Change LLC │ Software Engineer Intern June 2018 - August 2018
Languages: Python, Java; Frameworks: Tensorflow, NLTK, Scikit-Learn, Gensim, Apache Jena

- **Extracting domain knowledge embedded in COBOL.** Developed an ontology based vector embedding model to automatically summarize financial domain knowledge embedded in monolithic COBOL code. Achieved confidence scores $\geq 80\%$ on determining banking concepts in COBOL code as measured by surveying subject matter experts and active COBOL developers.

LexisNexis │ Data Scientist Intern May 2016 - August 2016 │ May 2017 - August 2017
Languages: Python, Java, Scala; Frameworks: Tensorflow, Spark, Livy, AWS−EMR, NLTK;

- **Streamlining document review using Machine Learning**: Designed a sandbox application for automated document retrieval in Technology Assisted Review (TAR) and E-Discovery. Demonstrated the use of Kernel-SVM and active learning in assisting the recovery of relevant documents from a highly skewed sample set with $\approx 20\%$ improved accuracy.
- **Summarizing legal documents into concise headnotes**: Developed and deployed a text summarization framework for generating "headnotes" in more than 1 Million legal documents. Designed a scalable clustering and a classification scheme with *doc2vec* to categorize documents into specific legal topics reducing document review time by $\geq 50\%$ according to in-house user surveys.

Education & Training

Postdoc in Computer Science June 2019 − *current*
Columbia University

PhD in Computer Science June 2015 − December 2019
North Carolina State University

MS in Electrical Engineering August 2013 − May 2015
North Carolina State University

BE in Electronics & Communication August 2009 − May 2013
Ramaiah Institute Of Technology

Teaching Experience

Theory of Programming Languages & Foundations of Software Engineering Fall 2016, Fall 2018, Spring 2019
Senior Level Graduate and undergraduate Courses

- Assisted teaching a class on programming languages to a class of 40 graduate and undergraduate students.
- Offered guest lectures on some popular abstractions for coding in various commonly used programming languages, data-driven software engineering, artificial intelligence, and applied data science.

Latest Publications

(For a full list of publications, see: https://scholar.google.com/citations?user=WGggocoAAAAJ)

[1] Krishna, R.[†], Iqbal, M S.[†], Javidian, M., Ray, B., Jamshidi, P., *"CADET: A Systematic Method For Debugging Misconfigurations using Counterfactual Reasoning"* In **NeurIPS (ML for Systems), 2020**. Link: https://arxiv.org/pdf/2010.06061.pdf

[2] She, D., Krishna, R., Yan, L., Jana, S. and Ray, B., *"MTFuzz: Fuzzing with a Multi-Task Neural Network."* In **Intl. Conference on Foundations of Software Engineering (ECSE/FSE), 2020**. Link: https://arxiv.org/pdf/2005.12392.pdf

[3] Krishna, R.[†], Tang, C.[†], Sullivan, K. and Ray, B., *"ConEx: Efficient Exploration of Big Data System Configurations for Better Performance."* In **IEEE Transactions on Software Engineering (TSE), 2020**. Link: https://arxiv.org/pdf/1910.09644

[4] Chakraborty, S., Krishna, R., Ding, Y. and Ray, B., *"Deep Learning based Vulnerability Detection: Are We There Yet?"*. In **IEEE Transactions on Software Engineering (TSE), 2020** *(Under review)*. Link: https://arxiv.org/abs/1703.06218

[5] Krishna, R., Nair, V., Jamshidi, P. and Menzies, T., *"Whence to Learn? Transferring Knowledge in Configurable Systems using BEETLE."* In **IEEE Transactions on Software Engineering (TSE), 2020**. Link: https://arxiv.org/abs/1703.06218

[6] Krishna, R., and Menzies, T., *"Learning actionable analytics from multiple software projects."*. In **Empirical Software Engineering (EMSE), 2020**. Link: https://arxiv.org/pdf/1708.05442.pdf

[7] Wang, J., Yang, Y., Krishna, R., Menzies, T. & Wang, Q., *"Effective Automated Decision Support for Managing Crowdtesting"*. In **Intl. Conference on Foundations of Software Engineering (ICSE), 2019, (Best paper award)**, Link: https://arxiv.org/pdf/1805.02744.pdf

[8] Krishna, R. & Menzies, T., *"Bellwethers: A Baseline Method For Transfer Learning"*. In **IEEE Transactions on Software Engineering (TSE), 2018**. Link: https://arxiv.org/abs/1703.06218

[9] Krishna, R., Agrawal, A., Rahman, A., Sobran, A., & Menzies, T. *"What is the Connection Between Issues, Bugs, and Enhancements? (Lessons Learned from 800+ Software Projects)"*. **Intl. Conf. Software Engineering (ICSE), 2018 SEIP**. Link: https://arxiv.org/abs/1710.08736

[10] Agrawal, A., Rahman, A., Krishna, R., Sobran, A. & Menzies, T. *"We Don't Need Another Hero? The Impact of 'Heroes' on Software Development"*. **Intl. Conf. Software Engineering (ICSE), 2018 SEIP**. Link: arXiv:1710.09055

[11] Rahman, A., Agrawal, A., Krishna, R., Sobran, A. & Menzies, T. *"Characterizing The Influence of Continuous Integration. Empirical Results from 250+ Open Source and Proprietary Project"*. **Intl. Conference on Foundations of Software Engineering (ICSE), 2018 SWAN**. Prerint: https://arxiv.org/pdf/1711.03933

[12] Chen, D., Fu, W., Krishna, R., & Menzies, T. *"Applications of psychological science for actionable analytics"*. In **Intl. Conference on Foundations of Software Engineering (ECSE/FSE), 2018**. Link: arXiv:1803.05067

[13] Chen, J., Nair, V., Krishna, R., & Menzies, T. *"Sampling as a Baseline Optimizer for Search-based Software Engineering"*. In **IEEE Transactions on Software Engineering (TSE), 2018**. Link: arXiv:1608.07617

[14] Krishna, R., Menzies, T., & Layman, L. *"Less is more: Minimizing code reorganization using XTREE"*. In **Information and Software Technology (IST), 2017**. Link: https://arxiv.org/abs/1609.03614

[15] Krishna, R., Menzies, T., & Fu, W. *"Too much automation? The Bellwether Effect and its Implications for Transfer Learning."* In **Intl. Conference on Automated Software Engineering (ASE), 2016**. Link: https://doi.org/10.1145/2970276.2970339