Khoa D. Doan

https://khoadoan.github.io khoadoan@vt.edu +1-240-381-9197

ABOUT

I am interested in solving computational problems in various domains, including computational advertising, information retrieval and generative modeling. I also like to understand the gap between existing theory and practice, and scale out/up algorithms in real-world environments.

EDUCATION

Ph.D in Computer Science — **Virginia, USA**
Virginia Tech (VT) — *Expected Graduation July 2021*

MS in Computer Science — **College Park, Maryland, USA**
University of Maryland, College Park (UMCP) — *August 2010-May 2015*

BS in Computer Science with a Minor in Mathematics — **Saint Louis, Missouri, USA**
Webster University — *August 2003-May 2006*
GPA: 4.0/4.0

EMPLOYMENT

Research Scientist – Baidu Research USA, Seattle, WA *2020-*

- Perform fundamental research in Generative Modeling, especially Generative Adversarial Models (GANs) and Deep Energy-based Generative Models, Deep Hashing Models and AI Security.

Researcher – Criteo AI Lab, Palo Alto, CA *2019-2020*

- Perform research in Generative Modeling, especially Generative Adversarial Models (GANs). Besides investigating their theoretical aspects, I studied the applications of generative models into several problem domains, including computational look-alike modeling in advertising, learning to hash for text and images, and regression.

Graduate Research Assistant – Virginia Tech *2016-*

Senior Data Scientist – Verve Mobile *2014-2019*

- Be the principle data scientist behind Verve's Momentum-product [1] which build highly-scalable statistical models of measuring the effectiveness of advertising campaigns. This is a fully-automatic platform that aggregates user-interaction data from multiple, large-scale data sources, build counterfactual models to estimate the *incremental KPIs* of the selected advertising campaigns, and exposes the analytical results to Verve's customers.
- Be a part of the Audience Modeling Team which is responsible for building production-ready predictive models such as Age, Gender, Home Associations, etc... Responsible for modeling the core algorithms behind Verve's Audience System.
- Be a part of the team that builds the core, patented Home Association algorithm at Verve (see below).
- Be a part of the Data Infrastructure Team that design the underlying architecture of the company's data-warehouse/data-analytic platform on Amazon AWS, which is later migrated to Google GCP.

[1] https://www.verve.com/products/momentum/

Faculty Research Associate – University of Maryland, College Park *2013-2015*

- Participate in high performance system research.
- Involve in Automated Event Service (AES) and Data Environment For Rapid Exploration And Characterization Of Hydrometeorological Organized Systems (DERECHOS) research projects at NASA.
- Responsible for development of large scale data (statistical) analysis algorithms in distributed environments such as Hadoop, Spark, Cassandra, and SciDB.

Data Scientist (Consultant) – B3Intelligence *2014-2015*

- Develop a statistical NLP topic model to categorize different types of survey text responses.
- Develop an attribution model to understand the relationship between the topics being discussed by the respondents and the expected sentiment in these responses.

Research Associate – NASA *2012-2014*

- Participate in high performance, distributed system research on NASA large-scale, remote-sensing data.
- Perform research in manifold learning in large-scale environments.
- Involve in Automated Event Service (AES) and Data Environment For Rapid Exploration And Characterization Of Hydrometeorological Organized Systems (DERECHOS) research projects at NASA.
- Responsible for development of large scale data (statistical) analysis algorithms in distributed environments such as Hadoop, Spark, Cassandra, and SciDB.

Teaching Assistant – University of Maryland, College Park *2010-2011*

Senior Java Developer – US Department of Energy *2010-2012*

- Lead application development activities.
- Integration of Open Source API's into existing system architecture.
- Lead maintenance of existing system.

Senior Software Developer – Aquilent *2008-2014*

- Enrollment System Redesign Project is the new implementation of the current standalone Enrollment System at VA. The effort is to establish a fully integrated environment which will replace not only the current Enrollment System, but also all the standalone systems at VA. The development environment at VA is: BEA Weblogic 8.1 and 10.0, Oracle 10g, J2EE, Spring, Hibernate, Struts, BizFlow BPM, ILog Rule Engine, Jasper Report.

HONORS AND SCHOLARSHIPS

Criteo Research Award – Virginia Tech *2018*

NSF Urban Computing Fellowship – Virginia Tech *2016*

Graduation Honor, Summer Cum Laude – Webster University *2006*

Regional ACM Collegiate Computing Contest Honor – Webster University *2004*

University Scholarship – Webster University *2003*

PUBLICATIONS

16. **K. D. Doan**, S. Tan, W. Zhao, & P. Li, "Fast Neural Learning-to-Hash Ranking under Neural Network based Measures". *Under Submission* .

* Equal Contribution

15. **K. D. Doan**, Y. Lao, & P. Li, "Poisoning Attacks with Latent Modification". *Under Submission*.

14. **K. D. Doan**, Y. Lao, W. Zhao, & P. Li, "LIRA: Learnable, Imperceptible and Robust Backdoor Attacks". *2021 IEEE International Conference on Computer Vision*.

13. **K. D. Doan**, J. Xie, W. Zhao, & P. Li, "Generative Cooperative Hashing Network". *Under Submission*.

12. **K. D. Doan**, F. Wang, S. Manchanda, S. Selvaraj, A. Bhowmilk & CK. Reddy, "Image Generation via Minimizing Fréchet Distance in Discriminator Feature Space. *Under Submission*.

11. **K. D. Doan**, S. Badirli, & CK. Reddy, "Generative Hashing Network". *Under Submission*.

10. **K. D. Doan**, S. Manchanda, S. Badirli, & C. K. Reddy. "Image Hashing by Minimizing Discrete Component-wise Wasserstein Distance". *Under Submission*.

9. **K. D. Doan***, S. Manchanda*, S. Mahapatra, & CK. Reddy, "Interpretable Graph Similarity Computation via Differentiable Optimal Alignment of Node Embeddings", *In Proceedings of International ACM SIGIR conference on research and development in Information Retrieval, 2021*.

8. S. Badirli, X. Liu, **K. D. Doan**, Z. Xing, A. Bhowmik & SS. Keerthi. Gradient Boosting Neural Networks: GrowNet. *https://arxiv.org/abs/2002.07971*.

7. S. Manchanda, **K. Doan** & SS Keerthi. Regression via Implicit Models and Optimal Transport Cost Minimization. *https://arxiv.org/abs/2003.01296*.

6. **K. Doan** & C. K. Reddy. Efficient Implicit Unsupervised Text Hashing using Adversarial Autoencoder. *In Proceedings of The Web Conference, 2020*.

5. S. Manchanda, P. Yadav, **K. Doan**, & K. Sathiya. Targeted display advertising: the case of preferential attachment. *In Proceedings of International Conference on Big Data, 2019*.

4. **K. Doan**, P. Yadav & C. K. Reddy. Adversarial Factorization Autoencoder for Look-alike Modeling. *In Proceedings of ACM International Conference on Information and Knowledge Management, 2019*.

3. **K. Doan**, G. Yang & C. K. Reddy. An Attentive Spatio-Temporal Neural Model for Successive Point Of Interest Recommendation. In Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2019.

2. **K. Doan**, A. O. Oloso, K. S. Kuo, T. L. Clune, H. Yu, B. Nelson, & J. Zhang. Evaluating the impact of data placement to Spark and SciDB with an earth science use case. *In Proceedings of IEEE International Conference on Big Data, 2016*.

1. **K. Doan**, A. O. Oloso, K. S. Kuo & T. L. Clune (2014, December). Performance comparison of big-data technologies in locating intersections in satellite ground tracks. *In Proceedings of ASE BigData Conference, 2014*.

ACADEMIC SERVICE

Conference Reviewer

- ACM SIGKDD International Conference on Knowledge discovery and data mining (KDD): 2017, 2018, 2019
- ACM International Conference on Information and Knowledge Management (CIKM): 2017, 2018, 2019

- ACM International Conference on Web Search and Data Mining (WSDM): 2017, 2018, 2019
- The Web Conference (WWW): 2017, 2018, 2019
- International Joint Conference on Artificial Intelligence (IJCAI): 2017, 2018, 2019
- AAAI Conference on Artificial Intelligence (AAAI): 2020, 2021

Journal Reviewer

- ACM Transactions on Internet Technology (TOIT): 2020
- ACMD Transactions on Knowledge Discovery from Data (TKDD): 2018, 2019, 2020, 2021

Program Committee

- Conference on Neural Information Processing Systems (NeurIPS): 2020, 2021
- International Conference on Machine Learning (ICML): 2020, 2021
- Conference on Computer Vision and Pattern Recognition (CVPR): 2020, 2021
- International Conference on Computer Vision (ICCV): 2021
- European Conference on Computer Vision (ECCV): 2020
- IEEE International Conference on Big Data (BigData): 2020
- 1st International Workshop on Industrial Recommendation Systems (IRS): 2020, 2021
- AAAI Conference on Artificial Intelligence (AAAI): 2021

PRESENTATIONS

- **K. Doan**. Generative models meet similarity search (*Seminar*). Baidu Cognitive Computing Lab, USA, 2020.
- **K. Doan**, B. Cave & C. K. Reddy (*Poster*). CrimeLab: A data-driven approach. Virginia Tech Urban Computing Day, 2017.
- H. Avik, G. Takahara & **D. Khoa** (*Talk*). Social Media Analytics using Bayesian Multistate Intensity Models. 43rd Annual Meeting of the Statistical Society of Canada. `https://bit.ly/3extrE`

PATENTS

B. E. Crook, **K. Doan**, G. K. Ng, C. G. Nicotra, M. J. Wrona. Systems, methods, and apparatus for reverse geocoding. US Patent US20160330592A1.

TEACHING

Data Analytics – Virginia Tech	*2019*
Deep Learning – Virginia Tech	*2018*
Operating Systems – University of Maryland, College Park	*2012*
Organization of Programming Language – University of Maryland, College Park	*2011*
Computer Architecture – University of Maryland, College Park	*2011*
Discrete Mathematics – University of Maryland, College Park	*2010*
Compiler Analysis – University of Maryland, College Park	*2010*