CV: Khoi-Nguyen Tran

Contact Information

IBM Research Australia
Level 22, 60 City Road
Southbank, VIC 3006
Australia

Mobile: +61 424 316 544
Email: kndtran@gmail.com
WWW: kndtran.github.io

Objective

To apply my research skills in industry by developing practical AI solutions for business.

Research Interests

Data Science, Machine Learning, Deep Learning, Text Mining, AI for Business, Cyber-security.

Employment

IBM Research Australia

- Research Scientist. Cognitive Analytics. Jun 2018 – Current
  - Leading research into machine understanding of USPTO patents. Ongoing and expected outputs are analytical tools using text mining, NLP, machine learning, and deep learning methods, with UIs and APIs to aid inventors and evaluators in their invention review process.
  - Led a client project on prediction of human effort in asset creation for business processes. Features end-to-end machine learning pipelines with a graph representational model using structured data, and focusing on paralellism and extensibility in hyper-parameter search.
  - Led development of an online chatbot service proof-of-concept for collecting financial documents for analysis. Features background coordination methods with Watson Assistant services for triggering front-end UI functionalities.
  - Publications: 2 pending review.
  - Patents: 3 filed. Reached 1st plateau.
  - Defensive Patent Publications: 5 published.
- Postdoctoral Research Scientist. Cognitive Analytics. Jun 2016 – Jun 2018
  - Developed document chunking solutions and classification models for IBM Global Process Services (for Watson Education). Research work recognized as division led expertise and is one of the core features of an IBM offer (C3 – http://microsite-us.mybluemix.net/).
  - Developed various end-to-end machine/deep learning solutions for unstructured text data for Watson Education. Particularly, document classification to improve unsupervised document alignment models.
  - Transitioned inherited research assets from Watson Education into production ready API services deployed on the IBM Cloud. Replicated and improved original published state-of-the-art research results. Simplified and parallelized assets for scalable generation of document alignment models.
  - Contributed to the development of a novel deep learning architecture for geolocation of Twitter tweets.
  - Publications: [3], [15].
  - Patents: 2 filed.
  - Defensive Patent Publications: 3 published.

Australian Federal Government

- Data Scientist. Research and Development Team. May 2015 – Jun 2016
  - Developed modelling, profiling, and text analytics solutions in R and SQL.
  - Developed code from analysts' observations and business knowledge, producing data visualizations and automated on-demand reports in production environments.

Australian National University

- Research Assistant & Co-founder. Cybercrime Observatory. Mar 2013 – May 2015
  - Projects: Investigating Malicious Spam Emails, History of Cybercrime Activities.
  - Developed a novel detection technique for malicious content (attachments or URLs) using only email text.
  - Publication: [7].
- Research Assistant. Research School of Computer Science. Jan 2011 – May 2015
  - Project: Detecting Abnormal Text Values.
    - Researched techniques to automatically detect abnormal text values from large databases using only the distribution of text data within those databases.
  - Project: New Objective Functions for Social Collaborative Filtering.
    - Developed a Facebook app to collect participants' data and recommend content from our novel collaborative filtering algorithms.
  - Project: Synthetic Data Generation and Corruption.
    - Developed a user interface for a novel synthetic data generation method that mimics real-world errors through a variety of data corruption methods.
  - Publications: [1], [4], [16].

Education

Australian National University

- Ph.D. in Engineering and Computer Science. Feb 2010 – Jul 2015
  - Area of Study: Machine Learning Applications.
  - Thesis Title: *Detecting Vandalism on Wikipedia across Multiple Languages.*
  - Supervisors: Prof. Peter Christen, Dr. Scott Sanner, Dr. Lexing Xie.
  - Publications: [6], [8], [9], [10], [11].
- Bachelor of Computer Science, with First Class Honours. Feb 2006 – Dec 2009
  - GPA: 6.75 / 7. Overall course average of High Distinction.
  - Honours Thesis Topic: *Semantic Sensor Composition.*
    - Supervisor: Dr. Michael Compton.
    - Publications: [2], [5], [12].
  - Individual Research Projects: *Detecting Network Anomalies.*
    - Supervisor: Dr. Huidong (Warren) Jin.
    - Publications: [13], [14].

Skills

Languages:

- Experienced: Java, Python, R, SQL, HTML, JavaScript, Unix Shell Scripting.
- Familiar: Scala, CSS.

Software:

- Experienced: Scikit-learn, Anonconda, pandas, numpy, dplyr/tidyverse, JQuery, Flask, MySQL, Eclipse, PDFBox, GitHub, ZenHub, IBM Cloud, LaTeX, Word, Excel, Powerpoint, Ubuntu Linux, Windows, macOS.
- Familiar: Tensorflow, Docker, Kubernetes, NLTK, Stanford NLP, Watson Assistant, Watson Studio.

Professional Activities

- Invited Researcher, 6th Heidelberg Laureate Forum, 2018.
  - 200 young researchers in mathematics and computer science are selected to participate each year from around the world.
- Invited Talk, RMIT Vietnam, 2017.
- Invited Participant, Future Shapers Forum, 2017.
- Invited Delegate, Australia-Vietnam Young Leadership Dialogue, 2017.
  - 19 delegates were selected from 375 applications.
- Volunteer, Australasian Data Mining Conference (AusDM), 2013.
- Volunteer, Open Source Developers' Conference (OSDC), 2011.
- Microsoft Student Ambassador, 2011–2012.

- Volunteer and President, ANU Computer Science Students' Association, 2007–2012.
  - Winner, Event of the Year 2012 and Runner-Up, Large Club of the Year 2012.

References Available upon request.

Cited Publications Full list available at Google Scholar:
http://scholar.google.com.au/citations?user=ihFcT5QAAAAJ
As of Jan 31: 275 citations, h-index: 8, i10-index: 8.

[1] P. Christen, R. W. Gayler, K.-N. Tran, J. Fisher, and D. Vatsalan. Automatic Discovery of Abnormal Values in Large Textual Databases. *Journal of Data and Information Quality (JDIQ)*, 2016.

[2] M. Compton, H. Neuhaus, K. Taylor, and K. Tran. Reasoning about Sensors and Compositions. In *Proceedings of the 2nd International Workshop on Semantic Sensor Networks (SSN)*, 2009.

[3] J. H. Lau, L. Chi, K.-N. Tran, and T. Cohn. End-to-end Network for Twitter Geolocation Prediction and Hashing. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*, 2017.

[4] J. Noel, S. Sanner, K.-N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad, and N. D. Penna. New Objective Functions for Social Collaborative Filtering. In *Proceedings of the 21st International World Wide Web Conference (WWW)*, 2012.

[5] K.-N. Tran. Semantic Sensor Composition, 2009. Honours Thesis, Australian National University.

[6] K.-N. Tran. *Detecting Vandalism on Wikipedia across Multiple Languages*. PhD thesis, Australian National University, 2015.

[7] K.-N. Tran, M. Alazab, and R. Broadhurst. Towards a Feature Rich Model for Predicting Spam Emails containing Malicious Attachments and URLs. In *Proceedings of the 11th Australasian Data Mining Conference (AusDM)*, 2013.

[8] K.-N. Tran and P. Christen. Cross Language Prediction of Vandalism on Wikipedia Using Article Views and Revisions. In *Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2013.

[9] K.-N. Tran and P. Christen. Identifying Multilingual Wikipedia Articles based on Cross Language Similarity and Activity. In *Proceedings of the 22nd ACM Conference of Information and Knowledge Management (CIKM): Poster*, 2013.

[10] K.-N. Tran and P. Christen. Cross-Language Learning from Bots and Users to detect Vandalism on Wikipedia. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2015.

[11] K.-N. Tran, P. Christen, S. Sanner, and L. Xie. Context-Aware Detection of Sneaky Vandalism on Wikipedia across Multiple Languages. In *Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2015.

[12] K.-N. Tran, M. Compton, J. Wu, and R. Goré. Short Paper: Semantic Sensor Composition. In *Proceedings of the 3rd International Workshop on Semantic Sensor Networks*, 2009.

[13] K.-N. Tran and H. Jin. Fusion of Decision Tree and Gaussian Mixture Models for Heterogeneous Data Sets. In *Proceedings of the International Conference on Multimedia Technology (ICIMT)*, 2009.

[14] K.-N. Tran and H. Jin. Detecting Network Anomalies in Mixed-Attribute Data Sets. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD)*, 2010.

[15] K.-N. Tran, J. H. Lau, D. Contractor, U. Gupta, B. Sengupta, C. J. Butler, and M. Mohania. Document Chunking and Learning Objective Generation for Instruction Design. In *Proceedings of the 11th International Conference on Education Data Mining (EDM)*, 2018.

[16] K.-N. Tran, D. Vatsalan, and P. Christen. GeCo - An Online Personal data Generator and Corruptor. In *Proceedings of the 22nd ACM Conference of Information and Knowledge Management (CIKM): Demo*, 2013.