Xuchao Zhang

Ph.D. Candidate @ Virginia Tech

7054 Haycock Road, Falls Church, VA, 22043, Northern Virginia Center

(+1) 571-253-1231 | xuczhang@vt.edu | https://xuczhang.github.io | xuczhang

Research Interests

In the broad area of data mining, machine learning, and natural language processing, my research interests include:

- Robust model learning in massive data sets under adversarial data corruption.
- Natural language processing tasks such as document revision summarization, document classification, dynamic topic modeling, and event extraction.
- Interdisciplinary applications in areas such as multi-factor personality prediction, spatiotemporal event forecasting in hyper-local price data.

Education

Virginia Tech

Falls Church, VA, USA

Ph.D. in Computer Science (**GPA**: 3.9/4.0) *Jan. 2015 - PRESENT*

- Committee: Chang-Tien Lu (Chair), Naren Ramakrishnan, Ing-ray Chen, Chandan Reddy, Arnold P. Boedihardjo
- Expected conferral of degree: December 2018

Shanghai Jiao Tong University

Shanghai, China

B.E. in Software Engineering *Sept. 2005 - July. 2009*

- Minister of Student Association, the School of Software
- **Awards & Honors**: People Scholarship (Top 10%), Xiukang Scholarship (Top 10%), Award of Excellence for Student Quality Development (Top 15%)

Professional Experience

Microsoft Research AI

Redmond, WA

Research Intern *May. 2018 - Present*

- Take the research internship in the Knowledge Technologies Group in Microsoft Research AI (MSR AI), advised by Sujay Kumar Jauhar and Michael Gamon.
- Anchor the positions of document revisions based on the user annotations; map relevant comments in the process of document editing.
- Work on the task of comment summarization based on the revision history of before-editing and after-editing documents. A new data set based on Wikipedia revision history will be released for this task.

Department of Computer Science, Virginia Tech

Falls Church, VA

Research Assistant *Jan. 2015 - Present*

- Designed and implemented AutoGSR project to extract the protest and military action events from massive text data via an attention-based sequence-to-sequence model. The project is affiliated to EMBERS system, which is sponsored by a contract for over $22M from the Intelligence Advanced Research Projects Activity (IARPA).
- Led and developed a storytelling framework to connect the dots between diverse entities such as people and organizations in social networks, which can help to form hypotheses and uncover latent relationships between entities. The project is sponsored by the U. S. military Research Laboratory.
- Collaborated with ARL, ND, and GMU in multiple research topics such as self-paced network embedding, multi-factor personality prediction, and dynamic topic modeling.

Department of Computer Science, Virginia Tech

Falls Church, VA

Teaching Assistant *Jan. 2015 - Present*

- Course CS5604 Information Storage and Retrieval: Led the project startup and project evaluation sessions.
- Course CS6604 Spatial Data Management: Led the teaching and lab sessions.

Publications

Research Papers In Conference Proceedings

Xuchao Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, Naren Ramakrishnan, "Online and Distributed Robust Regressions under Adversarial Data Corruption", Proceedings of the IEEE International Conference on Data Mining (**ICDM'18**), Singapore, Nov. 17-20, 2018. [acceptance rate: 19.94%].

Zhiqian Chen, Feng Chen, Rongjie Lai, **Xuchao Zhang**, and Chang-Tien Lu, "Rational Neural Networks for Approximating Jump Discontinuities of Graph Convolution Operator", Proceedings of the IEEE International Conference on Data Mining (**ICDM'18**), Singapore, Nov. 17-20, 2018. [acceptance rate: 8.86%].

Xuchao Zhang, Liang Zhao, Zhiqian Chen, Chang-Tien Lu, "Distributed Self-Paced Learning in Alternating Direction Method of Multipliers", Proceeding of the 27th International Joint Conference on Artificial Intelligence (**IJCAI'18**), Stockholm, Sweden, July 13-19, 2018. [acceptance rate: 20.6%].

Xuchao Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, Naren Ramakrishnan, "Online and Distributed Robust Regressions under Adversarial Data Corruption", Proceedings of the IEEE International Conference on Data Mining (**ICDM'17**), pages 625-634, New Orleans, Louisiana, Nov. 18-21, 2017. [acceptance rate: 9.25%].

Xuchao Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, Naren Ramakrishnan, "Spatiotemporal Event Forecasting from Incomplete Hyper-local Price Data", Proceedings of the 26th ACM International Conference on Information and Knowledge Management (**CIKM'17**), pages 507-516, Singapore, Nov. 6-10, 2017. [acceptance rate: 21%].

Xuchao Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, "Robust Regression via Heuristic Hard Thresholding", Proceeding of the 26th International Joint Conference on Artificial Intelligence (**IJCAI'17**), pages 3434-3440, Melbourne, Australia, August 19-25, 2017. [acceptance rate: 26%].

Zhiqian Chen, **Xuchao Zhang**, Arnold P. Boedihardjo, Jing Dai, Chang-Tien Lu, "Multimodal Storytelling via Generative Adversarial Imitation Learning", Proceeding of the 26th International Joint Conference on Artificial Intelligence (**IJCAI'17**), pages 3967-3973, Melbourne, Australia, August 19-25, 2017. [acceptance rate: 26%].

Xuchao Zhang, Liang Zhao, Zhiqian Chen, Arnold Boedihardjo, Dai Jing, Chang-Tien Lu, "Trendi: Tracking Stories in News and Microblogs via Emerging, Evolving and Fading Topics", Proceedings of the IEEE International Conference on Big Data (**BigData'17**), Boston, MA, Dec. 11-14, 2017.

Xuchao Zhang, Zhiqian Chen, Liang Zhao, Arnold Boedihardjo, Chang-Tien Lu, "TRACES: Generating Twitter Stories via Shared Subspace and Temporal Smoothness", Proceedings of the IEEE International Conference on Big Data (**BigData'17**), Boston, MA, Dec. 11-14, 2017.

Ting Hua, **Xuchao Zhang**, Wei Wang, Chang-Tien Lu, Naren Ramakrishnan, "Automatical Storyline Generation with Help from Twitter", Proceedings of the 25th ACM International Conference on Information and Knowledge Management (**CIKM'16**), Indianapolis, IN, Oct. 24-28, 2016. [acceptance rate: 28.8%].

Xuchao Zhang, Zhiqian Chen, Weisheng Zhong, Arnold P. Boedihardjo, Chang-Tien Lu, "Storytelling in Heterogeneous Twitter Entity Network based on Hierarchical Cluster Routing", Proceedings of the IEEE International Conference on Big Data (**BigData'16**), pp. 1522-1531 Washington, DC, Dec. 5-8, 2016.

Research Papers Under Review

Xuchao Zhang, Chang-Tien Lu and Naren Ramakrishnan, "Mitigating Uncertainty in Document Classification", (*Conference Paper Under Review*), 2018.

Lei Zhang, Liang Zhao, **Xuchao Zhang**, David Stillwell, Michal Kosinski and Chang-Tien Lu, "Situation-Based Interaction Learning for Personality Prediction on Facebook", (*Conference Paper Under Review*), 2018.

Xuchao Zhang, Liang Zhao, Zhiqian Chen, Chang-Tien Lu, "Self-Paced Robust Learning for Leveraging Clean Labels in Noisy Data", (*Conference Paper Under Review*), 2018.

Xuchao Zhang, Lei Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, "Robust Multi-Factor Personality Prediction with Correlated Data Corruption in Social Media", (*Conference Paper Under Review*), 2018.

Taoran Ji, Kaiqun Fu, Liang Zhao, **Xuchao Zhang**, Chang-Tien Lu, Naren Ramakrishnan, "Multi-task Feature Learning for Cybersecurity Event Detection in Social Media", (*Conference Paper Under Review*), 2018.

Xuchao Zhang*, Bingsheng Wang*, Chang-Tien Lu, Feng Chen, "Water Disaggregation via Shape Features based Bayesian Discriminative Sparse Coding", * These two authors contributed equally. (*Journal Paper Under Review*), 2018.

Xuchao Zhang, Liang Zhao, Arnold P. Boedihardjo, Chang-Tien Lu, "Robust Regression via Heuristic Corruption Thresholding and Its Adaptive Estimation Variation", (*Journal Paper Under Review*), 2018.

Professional Services

- Program Committee (Research Track) of KDD'18
- Journal Review of Advances of Computer Science for Geographic Information Systems (GeoInformatica)
- External Review: IJCAI'18, AAAI'18, KDD'17/16, ICDM'17/16, SDM'17/16, ICTAI'15, PKDD'15

Honors & Awards

2017	IEEE International Conference on Data Mining Travel Award	*NSF*
2017	GSA Travel Fund Program Award	*Virginia Tech*
2013	Staff Award of Excellence	*Microsoft*
2011	Excellent Employee Award in Business One	*SAP SE*
2008	People Scholarship (10%)	*Shanghai Jiao Tong University*
2008	Xiukang Scholarship (15%)	*Shanghai Jiao Tong University*
2007	Award of Excellence for Student Quality Development	*Shanghai Jiao Tong University*

Work Experience

Microsoft
Shanghai, China

Technical Support Engineer — *Sep. 2012 - Jan. 2015*

- Tackled break-fix issues regarding architecture, performance, deployment, retail solution, BI & Analytics, workflow of Microsoft Dynamics AX.
- Resolved advisory cases from the clients in Asia Pacific and Great China and help them with their customization features and additional feature requests.

SAP SE
Shanghai, China

Software Development Engineer — *Jun. 2010 - Sep. 2012*

- Led and developed a Master Data Cleanup project to archive the data based on the relationship between business objects.
- Maintained and developed the new feature for infrastructure topics, such as Master data cleanup, Data Archive, Numbering, and Business Flow.
- Accomplished a Web based project on SBO UIAPI and realized displaying Windows client-based applications on web browser via kernel code mapping, and Put forward the solutions of grid paging and dynamic image loading.
- Recruiting and training program: interviewed candidates for technical positions and provided training sessions for new recruits on FU/FU+ Module

Telenav Corp. (Start-up company) *Shanghai, China*

Software Engineer *Jul. 2009 - Jun. 2010*

- Programmed the core map engine compilation for a new generation of Map Engines.
- Developed map-related tools and conducted BBC-Proxy performance tuning and maintenance.
- Developed smart comparison tools for regression testing.
- Improved hit rate of cache to over 90% and raised routing performance by 70% via preload predicted map algorithm.

Morgan Stanley *Shanghai, China*

IT Analyst and Developer (Intern) *Nov. 2008 - Jun. 2009*

- Developed and tuned Asia Limit Report programs based on the usage of stored procedure and dynamic data loading.
- Accomplished data acquisition module, web server deployment, extjs-based report, and added BI Analysis function.

Google *Shanghai, China*

Software Developer (Summer Intern Program) *Jun. 2008 - Sep. 2008*

- Programmed the Sitemaps generator plugin for Google Apps Engine.

Skills

Programming	Python, Java, MATLAB, C/C++, Perl, Shell
Machine Learning	Robust Statistical Models, Deep Learning, Numerical Optimization, Self-paced Learning, Topic Modeling
Tools/Softwore/System	Tensorflow, Keras, MATLAB, Linux/Unix, Database, LAMP stack, Performance Tuning, Design Pattern