Chase Geigle

Updated on March 7, 2018

Permanent Address: 1503 Village Park Way / Savoy, IL / 61874 • Citizenship: United States Citizen
Phone: 309-912-6213 • E-mail: geigle1@illinois.edu • Website: https://chara.cs.illinois.edu/sites/cgeigle/

Research Interests

I am broadly interested in techniques for mining both user behavior data and text data. Recently, I have been focused on using text mining and generative modeling techniques for improving both the scalability and quality of online educational environments.

Education

PhD in Computer Science (expected) Fall 2018

- University of Illinois at Urbana-Champaign, College of Engineering
- Advisor: Dr. ChengXiang Zhai

BS in Computer Science Fall 2013

- University of Illinois at Urbana-Champaign, College of Engineering
- GPA: 3.95

Awards and Honors

National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) 2015

- Five-year fellowship, three years of complete funding

Outstanding Teaching Assistant Spring 2017

Illinois Technology Association Fall Challenge First Place Winner Fall 2013

- Press release: `http://www.marketwired.com/press-release/a-1845938.htm`

Illinois Technology Association Fall Challenge Finalist Fall 2012

- Press release: `http://www.marketwired.com/press-release/a-1728087.htm`

Data Sciences Summer Institute (DSSI) Best Project Summer 2012

James Honors Scholar Fall 2009–Fall 2013

Teaching Experience

CS 510 Advanced Topics in Information Retrieval (Teaching Assistant) Fall 2017

- Gave six lectures: (1) Class-based n-gram language models, (2) word embeddings, (3) Bayesian inference for mixture language models (three lectures), and (4) a tutorial on the MeTA toolkit. Deployed a cloud-based virtual lab that was used to run all four programming assignments.

CS 598CXZ Advanced Topics in Information Retrieval (Teaching Assistant) Fall 2016

- Gave two lectures on word embedding methods and designed accompanying assignments and examination questions for this new topic

CS 591txt Text Mining Seminar (Co-organizer) Spring 2016–Fall 2016

CS 225 Data Structures and Programming Principles (Instructor) Summer 2014 & 2015

- Updated curriculum to include C++14 instruction, developed a suite of pre-lecture activities to employ a "flipped-classroom" style for lectures.
- **2014** `https://chara.cs.illinois.edu/sites/su14-cs225`
  2015 `https://chara.cs.illinois.edu/sites/cs225`

CS 528 Object Oriented Programming and Design (Instructor) Summer 2013, 2014, & 2015

- Covers design patterns and object oriented principles through the use of Smalltalk
- Responsible for all aspects of the course (assignments, quizzes, projects) except for lectures
- Website: `https://chara.cs.illinois.edu/sites/cs528/index.html`

CS 296 Advanced C++ (Instructor) Spring 2013

- Brand new course designed from scratch to cover a variety of advanced C++ concepts (template metaprogramming, design patterns, C++11)
- Website: `https://chara.cs.illinois.edu/cs296/index.html`

CS 225 Data Structures and Programming (Undergraduate CA) Spring 2011–August 2013

- Large undergraduate introductory course ($\geq$ 400 students)
- Developed several course machine programming problems, helped implement an improved autograding system, developed a course competition framework, co-designed and implemented a parallelism intro curriculum with a focus on visualization, and co-led the honors section

CS 232 Computer Architecture II (Unergraduate CA) Spring 2011

- Designed and implemented an honors curriculum about caches where students developed methods for cache visualization used as teaching tools in the regular course

CS 125 Intro to Computer Science (Undergraduate CA) Spring 2010–Fall 2010

- Led a lab section and held office hours

Industry Experience

Piazza Technologies, Inc. (Software Engineer) Summer 2011–Fall 2011

- Educational start-up company focusing on educational communication
- Was company's fifth full time employee
- One of two feature development engineers (full stack development)
- Project management, training/onboarding, hiring
- Presented at conferences and trade shows to gather adoption (hands-on demos)

UIUC Technology Services Group (Windows System Administrator) Fall 2010–Spring 2011

- Administered Windows based machines in the Computer Science Department at UIUC
- Used Windows Deployment Server for image creation and deployment

Publications

Books

1. ChengXiang Zhai, **Chase Geigle,** *Statistical Language Models for Text Data Retrieval and Analysis*. The Information Retrieval Series. Springer International Publishing, forthcoming.

Refereed Journal Articles

2. **Chase Geigle**, ChengXiang Zhai. "Modeling MOOC Student Behavior with Two-Layer Hidden Markov Models". In: *Journal of Educational Data Mining* 9.1 (Sept. 2017), pp. 1–24.

Book Chapters

3. **Chase Geigle**, Qiaozhu Mei, ChengXiang Zhai. "Feature Engineering for Text Data". In: *Feature Engineering for Machine Learning and Data Analytics*. Ed. by Gouzhu Dong and Huan Liu. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, Mar. 2018, forthcoming. ISBN: 9781138744387.

Refereed Conference Proceedings

4. **Chase Geigle**, Ismini Lourentzou, Hari Sundaram, ChengXiang Zhai. "CLaDS: A Cloud-Based Virtual Lab for the Delivery of Scalable Hands-on Assignments for Practical Data Science Education". In: *Proceedings of the 23rd Annual Conference on Innovation and Technology in Computer Science Education*. ITiCSE'18. Larnaca, Cyprus: ACM, July 2018, to appear.
5. Himel Dev, **Chase Geigle**, Qingtao Hu, Jiahui Zheng, Hari Sundaram. "The Size Conundrum: Why Online Knowledge Markets Can Fail at Scale". In: *Proceedings of WWW 2018: The Web Conference*. Lyon, France: ACM, Apr. 2018, to appear.
6. Assma Boughoula, **Chase Geigle**, ChengXiang Zhai. "A Probabilistic Approach for Discovering Difficult Course Topics Using Clickstream Data". In: *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. L@S '17. Cambridge, Massachusetts: ACM, Apr. 2017, pp. 303–306.
7. **Chase Geigle**, ChengXiang Zhai. "Modeling MOOC Student Behavior with Two-Layer Hidden Markov Models". In: *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. L@S '17. Cambridge, Massachusetts: ACM, Apr. 2017, pp. 205–208.
10. Kenneth Heafield, **Chase Geigle**, Sean Massung, Lane Schwartz. "Normalized Log-Linear Language Model Interpolation is Efficient". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Lingustics*. Berlin, Germany: ACL, Aug. 2016, pp. 876–886.
11. Sean Massung, **Chase Geigle**, ChengXiang Zhai. "MeTA: A Unified Toolkit for Text Retrieval and Analysis". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Lingustics: System Demonstrations*. Berlin, Germany: ACL, Aug. 2016, pp. 91–96.
8. **Chase Geigle**, ChengXiang Zhai. "Scaling up Online Question Answering via Similar Question Retrieval". In: *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*. L@S '16. Edinburgh, Scotland, UK: ACM, Apr. 2016, pp. 257–260.
9. **Chase Geigle**, ChengXiang Zhai, Duncan Ferguson. "An Exploration of Automated Grading of Complex Assignments". In: *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*. L@S '16. Edinburgh, Scotland, UK: ACM, Apr. 2016, pp. 351–360.
12. Swarup Kumar Sahoo, John Criswell, **Chase Geigle**, Vikram Adve. "Using Likely Invariants for Automated Software Fault Localization". In: *Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems*. ASPLOS '13. Houston, Texas, USA: ACM, Mar. 2013, pp. 139–152.

Workshop Proceedings

13. Lane Schwartz, Bill Bryce, **Chase Geigle**, Sean Massung, Yisi Liu, Haoruo Peng, Vignesh Raja, Subhro Roy, Shyam Upadhyay. "The University of Illinois submission to the WMT 2015 Shared Translation Task". In: *Proceedings of the Tenth Workshop on Statistical Machine Translation*. Lisbon, Portugal: ACL, Sept. 2015, pp. 192–198.

Open-Source Software

MeTA (ModErn Text Analysis) Toolkit Spring 2013–Present

- Website: `https://meta-toolkit.org`
- Commits: `https://github.com/meta-toolkit/meta/commits?author=skystrife`
- Co-author of an open-source C++11/14 toolkit for data science with a focus on textual information
- Supports information retrieval, topic modeling, machine learning (classification and regression), NLP (language modeling, tagging, parsing, word embeddings), and graph algorithms
- Used each year in UIUC's *Text Information Systems* class
- Used by Coursera course *Text Retrieval and Search Engines* in programming assignments available to thousands of students
- Used by Coursera course *Text Mining and Analytics* in its programming assignments
- Showcased in the textbook *Text Data Analysis and Management: A Practical Introduction to Text Mining and Information Retrieval* by ChengXiang Zhai and Sean Massung

kenlm Spring 2015–Fall 2015

- Commits: `https://github.com/kpu/kenlm/commits?author=skystrife`
- Added support for efficient log-linear language model interpolation to one of the most popular language modeling toolkits used in machine translation

cpptoml Spring 2013–Present

- Commits: `https://github.com/skystrife/cpptoml/commits?author=skystrife`
- Developed an open-source C++11 library for parsing TOML configuration files
- Used by MeTA for its configuration files
- Used to create the R language bindings for TOML files: `http://dirk.eddelbuettel.com/code/rcpp.toml.html`

procxx Spring 2014–Present

- Commits: `https://github.com/skystrife/procxx/commits?author=skystrife`
- Developed an open-source C++11 library for process execution and management on POSIX systems
- Supports creating process pipelines, reading from process streams, and resource limits via `rlimit`

ATALANTA Toolkit Fall 2012

- Commits: `https://bitbucket.org/skystrife/atalanta-toolkit/commits/all`
- Developed a C++11 toolkit of pluggable distributed systems components

Educational Support Software

CLaDS: A Cloud-based Virtual Lab for Data Science Spring 2017–Present

- Deployed a platform for running data-science assignments at-scale in the cloud
- Used for four separate course offerings to provide more practical hands-on assignments
- Developed a competition leaderboard for information retrieval and classification experiments

Chara Sites Spring 2014–Present

- Developed a platform for hosting and automatically updating course websites using Jekyll
- Used by nine separate course offerings (five distinct courses) since created

Chara Gradebook Fall 2012–Present

- Website: `https://chara.cs.illinois.edu/gb`
- Developed an online grade visualization platform
- Allows students to perform grade prediction
- Allows instructors to mine their gradebook data, determine struggling and outstanding students, etc.

Chara Lab Queueing Application Spring 2012–Present

- Website: `https://chara.cs.illinois.edu`
- Developed real-time queueing system for students at office hours, used in CS 225
- Gathering data of interaction quality for mining (correlation analysis, etc.)
- Used each semester by 4–5 courses in 2 departments, and for CS department undergraduate advising

Chara Contest Framework Spring 2012–Present

- Example: `https://chara.cs.illinois.edu/contest/sp13-psort/`
- Developed a web-based competition leaderboard to facilitate teaching parallel programming
- Interacts with the Monad autograding/unit testing framework

Professional Activities and Service

Publication co-chair, the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016)

Program committee member, the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2018, demo track)

Program committee member, the 11th ACM International Conference on Web Search and Data Mining (WSDM 2018, demo track)

Program committee member, the ACL 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP 2017, demo track)

Ad-hoc reviewer, Information Processing & Management