Sarah Wiegreffe

Atlanta, GA
swiegreffe6@gatech.edu
sarahwie.github.io

EDUCATION

Georgia Institute of Technology (Georgia Tech)   **August 2017 - present**
PhD in Computer Science
Advisor: Professor Jacob Eisenstein

Research interest in natural language processing and machine learning with clinical applications. Specific interests include machine reading and information extraction, hierarchical representations of text, and text generation/summarization. Interested in temporal/hierarchical modeling of hospital clinical notes for the prediction of patient diagnosis and treatment.

Honors College at the College of Charleston   **August 2013 - May 2017**
Bachelor of Science in Data Science, Summa Cum Laude
Minors in Mathematics and International Studies

University of Tartu, Estonia   **January - June 2015**
Visiting Student, Faculty of Mathematics and Computer Science

PUBLICATIONS

Mullenbach, J., **Wiegreffe, S.**, Duke, J., Sun, J. and Eisenstein, J. *Explainable Prediction of Medical Codes from Clinical Text*. NAACL Human Language Technologies 2018. Full paper, accepted for oral presentation. New Orleans, LA, June 2018.

Wiegreffe, S., Anderson, P. and Obeid, J. *Can Classifications of Publications by Translational Categories be Automated?*. Proceedings of the American Medical Informatics Association (AMIA) Joint Summits on Translational Science. Extended abstract, accepted for poster presentation. San Francisco, CA, March 2017.

RESEARCH

Computational Linguistics Lab at Georgia Tech   **August 2017 - present**

Projects include tying convex optimization to word embeddings to learn under hierarchical constraints as well as domain-knowledge grounding for end-to-end learning of effective representations of clinical text via deep learning. Worked with Professor

Research and Development Intern, Sutter Heath   **May 2018 - December 2018**

Worked with the Research, Development, and Dissemination group and Professor Jimeng Sun (Georgia Tech) to develop deep learning methodology for disease prediction from clinical text.

Anderson Lab at the College of Charleston　　　　　　　**January 2016 - May 2017**

Researched extensions to Google's Word2Vec algorithm used to generate word embeddings for variable-length documents. Investigated performance of the algorithm when used directly as a classifier, and whether this technique, along with similarly created ensemble methods, could outperform benchmark preprocessing and machine learning pipelines on topic recognition tasks.

Medical University of South Carolina　　　　　　　**May - September 2016**
Biomedical Informatics Center

Researched automation methodology for identifying translational research patterns in medicine. Built datasets of unlabelled medical abstracts and trained machine learning algorithms to classify by translational type. Presented at AMIA Informatics Summit.

Medical University of South Carolina　　　　　　　**June - December 2015**
Office of the Chief Informatics Officer

Designed a novel model forecasting the number of operating room cases in the university hospital on a given day, in order to improve hospital efficiency. Implemented the model as an automated visualization tool using Tableau which generated daily 30-day forecasts for departments across the hospital to assess.

PROFICIENCIES

Daily Use	**Languages:** Python (Pytorch/sklearn/pandas/nltk/numpy/multiproc), Bash. **Tools:** Git, TeX.
Past Use	**Languages:** R, Java, SQL, SAS, Octave. Fluency in French. **Tools:** Dynet, Oracle RDBMS, MongoDB, Tableau.

REVIEWING

NeurIPS Machine Learning for Healthcare Workshop (2017, 2018), **AMIA Informatics Summit** (2018, 2019), **ACL Subreviewer** (2018).

AWARDS

Computing Research Association (CRA-W) Grad Cohort for Women Attendee (2018).

Phi Kappa Phi Graduate Fellowship Recipient (2017).

Data Science Major of the Year, Departmental Honors (2017). College of Charleston.

Grace Hopper Scholar, the Anita Borg Institute (2016). Grace Hopper Celebration of Women in Computing Attendee (2015, 2016).

Crosby Computer Science Award (2014). Awarded by professorial nomination to the most promising student in an introductory computer science course at the College of Charleston.

William Aiken Fellow (2013-2017). A fellowship representing the top 1% of students at the College of Charleston.

TEACHING AND EMPLOYMENT

Graduate Research Assistant, Georgia Institute of Technology   **Fall 2017 - present**

Teaching Assistant, Charleston Digital Corridor   **Fall 2014**
MongoDB database course for adult professionals.

Math Lab Tutor, College of Charleston Center for Student Learning   **Fall 2014**
Subjects including calculus, statistics, business math, precalculus and algebra.

Medical University of South Carolina Biomedical Informatics Center   **Summer, Fall 2014**
Used relational database and SQL querying skills to test and improve the data warehouse for South Carolina's hospitals. Worked in a team environment to develop a testing process and write scripts to test web applications against the database.

Women in Computing Club at the College of Charleston   **2014 - 2017**
President (2017), Vice-president (2016) and Treasurer (2015-16). Worked to promote diversity in computer science, host professional development workshops, and conduct community outreach.

OTHER PROJECTS

Bachelor's Thesis   **College of Charleston, May 2017**
Titled *A Survey of Word2Vec Inversion Methods in Topic Recognition Tasks*. Advised by Dr. Paul Anderson.

Senior Capstone Project   **College of Charleston, Spring 2017**
A medical provider ranking algorithm and webapp, using Coley's algorithm, for healthcare API company PokitDok.