BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Fries, Jason

eRA COMMONS USER NAME: JAFRIES

POSITION TITLE: Research Engineer

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of Iowa, Iowa City, IA	B.A.	12/2009	Computer Science, English
University of Iowa, Iowa City, IA	Ph.D.	08/2015	Computer Science
Stanford University, Stanford, CA	Postdoctoral Fellowship	12/2017	Computer Science/ Biomedical Data Science

A. Personal Statement

My research explores methods for building large-scale, biomedical machine learning models without relying on human-labeled training data. In the era of big biomedical data, predictive models need to unify highly heterogeneous, unstructured data, from medical images and clinical time series to the text found in electronic health records (EHR) and scientific publications. Under current approaches, it is intractable to manually construct training sets for every task that requires machine learning. My research focuses on developing new ways to encode domain knowledge into software and programmatically generate labeled data for *supervising* machine-learning models. This approach allows us to train machine-learning models by applying dramatically weaker, noisier forms of supervision to large collections of unlabeled data. For example, in recent work I demonstrate how a single domain expert can use standard biomedical ontologies to train a near-state-of-the-art biomedical concept extraction model over the course of a single day *without using human-labeled training data*. These findings create a foundation for ongoing research into building tools that "democratize" machine learning for non-computer scientists, enabling faster, more intuitive ways of injecting domain knowledge into model development.

Analyzing large-scale data from the EHRs is a central application area in my research. EHR data includes unstructured data (e.g., text, images) and medical codes, however coded data only captures a small fraction of a patient's medical history and is highly biased by billing practices. Moreover, medical data alone only captures part of the human health experience; researchers must now incorporate information from wearable devices, social media, and other sources to quantify daily human activities. Reasoning over these heterogeneous, often-unstructured data requires rapid methods for training machine learning models. My graduate work focused on extracting health data from unstructured web data; my current research focuses on ways to do this quickly across multiple text modalities. During my postdoctoral training, I am co-mentored by Professor Chris Ré in computer science, Professor Scott Delp in bioengineering, and I collaborate closely with Professor Nigam Shah in biomedical informatics. As a researcher, I am well positioned to bridge the gap between developing cutting-edge computer science methods for information extraction and applying these innovations to large-scale medical informatics challenges, such as monitoring post-surgical outcomes of joint replacements, or conducting nation-wide medical device surveillance.

B. Positions and Honors

Positions and Employment

2007-2008 Undergraduate Research Assistant, Department of Computer Science, University of Iowa
2009-2013 Research Assistant, Department of Computer Science, University of Iowa
2013-2015 Instructor, University of Iowa
2015-2017 Complimentary Faculty Appointment, University of Iowa
2015-pres. Postdoctoral Fellow, Department of Computer Science and Mobilize Center, Stanford University

Honors

2008 Iowa Research Experience for Undergraduates Grant Recipient
2008 Iowa Research Experience for Undergraduates: Poster Presentation Award Winner
2009 Tippie College of Business Elevator Pitch Seed Grant Winner at the University of Iowa
2013 Strategic Initiative Fund Fellowship at the University of Iowa
2015 Distinguished Postdoctoral Fellowship at Stanford University

Service and Professional Memberships

2009-2013 Member of the International Society for Disease Surveillance (ISDS)
2013-2014 Reviewer for the PLOS One
2016 Co-organizer for "Machine Learning for Health" at Neural Information Processing Systems
2017 Co-organizer for "Learning to Run" competition at Neural Information Processing Systems

Invited Lectures

2015 Positive Psychology Center, University of Pennsylvania
2015 Department of Bioengineering and the Mobilize Center, Stanford University

Conference Podium Presentations

2010 9th Annual Conference of the International Society for Disease Surveillance, Park City, UT
2011 21st Annual Scientific Meeting of the Society for Healthcare Epidemiology of America, Dallas, TX
2011 10th Annual Conference of the International Society for Disease Surveillance, Atlanta, GA
2012 ASE/IEEE 4th International Conference on Social Computing. Amsterdam, Netherlands
2014 IEEE International Conference on Healthcare Informatics, Verona, Italy

C. Contributions to Science

1. Training Biomedical Machine Learning Models with Weak Supervision

My current research focuses on developing new methods for training machine learning models without human-labeled data, instead relying on sources of noisy or "weak" supervision. I co-developed a lightweight information extraction framework for training models using noisy heuristics (b). In a critical subtask of natural language processing (NLP), named entity recognition, I showed that by leveraging noisy signals from standard biomedical ontologies, it is possible to build state-of-the-art extraction models with dramatically less human effort (a). This enables building sophisticated text mining applications over the course of days instead of weeks. In related application work with electronic health records (EHR), I showed how new representation learning ("deep learning") methods removes the need for any manual feature engineering in several complex temporal relationship extraction tasks (c).

a. **Fries, J**, Wu, S, Ratner, A, Ré, C. "SwellShark: A Generative Model for Biomedical Named Entity Recognition without Labeled Data." Jan 2018. arXiv:1704.06360

b. Ratner, A, Bach, S, Ehrenberg, H, **Fries, J**, Wu, S, & Ré, C. "Snorkel: Rapid Training Data Creation with Weak Supervision" Proceedings of the VLDB Endowment, 11(3), 269-282, 2017

c. **Fries, J.** "Brundlefly at SemEval-2016 Task 12: Recurrent Neural Networks vs. Joint Inference for Clinical Temporal Information Extraction." *Proceedings of SemEval (2016)*: 1274-1279. June 2016.

2. Syndromic Surveillance of Sexual Health Behaviors on the Internet

My doctoral work explored text mining and machine learning methods for web-based syndromic surveillance of high-risk sexual behaviors. Healthcare encounters only capture a small portion of the story of human health. Social media is a promising source of additional health information, especially for measuring aggregate behaviors that are difficult to collect in standard healthcare settings. In this work, extracted public health information on high-risk sexual behaviors, analyzing the text of anonymous requests for sexual encounters found on websites like Craigslist. Public health instruments like surveys are traditionally used to capture this data, but suffer from reporting delays and are limited in geographic scope. Using a 4-year collection of 200 million anonymous ads, I showed that many survey questions used in public health clinics mapped directly to text classification problems, allowing us to build Internet surveillance tools that operate in near-real time and cover the entire United States. I examined methods for automatically learning terminology characterizing high-risk sexual behaviors, including activities like unprotected sex and illegal drug use (c-d). I developed semi-supervised approaches for extracting demographic variables like race and ethnicity from text (a) and identifying movement patterns of anonymous individuals (b).

a. **Fries, J**, Segre, A, Polgreen, P. "Mining the Demographics of Craigslist Casual Sex Ads to Inform Public Health Policy." *IEEE International Conference on Healthcare Informatics*. September 2014.

b. **Fries, J**, Segre, A, Polgreen, P. “Using Online Classified Ads to Identify the Geographic Footprints of Anonymous, Casual Sex-seeking Individuals.” *ASE/IEEE 4th International Conference on Social Computing*. September 2012.

c. **Fries, J**, Segre, A, Polgreen, L, Polgreen, P. "Using Craigslist Messages for Syphilis Surveillance." *International Meeting on Emerging Diseases and Surveillance*. February 2011.

d. **Fries, J**, Segre, A, Polgreen, L., Polgreen, P. "The Use of Craigslist Posts for Risk Behavior and STI Surveillance." *9th Annual Conference of the International Society for Disease Surveillance*. December 2010.

3. Improving Hand Hygiene Compliance with Real-time Surveillance

My undergraduate research focused on developing and deploying wearable sensor technology for tracking health care worker (HCW) movement in health care facilitates. I helped develop algorithms for predicting HCW spatial locations, constructing social networks of patient/HCW interactions, and automatically measuring rates of hand hygiene compliance through instrumented alcohol rub dispensers (b-c). Using these devices, I helped conduct a 2-week observational study where we monitored all HCW movement within a medical intensive care unit at the University of Iowa Hospitals and Clinics. Using this movement data, I developed agent-based simulations for empirically assessing and optimizing hand hygiene policy decisions (a). I co-developed a hand-hygiene monitoring mobile app, iScrub, which provided real-time feedback on compliance rates within hospital units via a publically visible analytics dashboard. Conducting a 6-month longitudinal intervention study, we showed a statistically significant 23% increase in mean compliance rate across the unit (d).

a. **Fries J**, Segre AM, Thomas G, Herman T, Ellingson K, Polgreen PM. “Monitoring hand hygiene via human observers: how should we be sampling?” *Infection Control and Hospital Epidemiology*, v. 33(07), pp. 689-695. July 2012. **PMID: 22669230**

b. Herman, T, Pemmaraju, S, Segre, A, Polgreen, P, Curtis, D.E., **Fries, J**, Hlady, C, Severson, M. "Wireless Applications for Hospital Epidemiology." *ACM International Workshop on Medical-grade Wireless Networks*. May 2009.

c. **Fries, J**, Hlady, C, Herman, T, Polgreen, P, Segre, A. "A Low-Cost Non-RFID Based Method for Automated Monitoring of Hand-Hygiene Compliance." *19th Annual Scientific Meeting of the Society for Healthcare Epidemiology of America*. March 2009.

d. Hlady, C, Curtis, D, **Fries, J**, Yang, M, Segre, A, Polgreen. P. "iScrub: A Pilot Intervention with Feedback from a Companion Website." 21st Annual Scientific Meeting of the Society for Healthcare Epidemiology of America. April 2011.

Complete List of All Publications and Peer-Reviewed Conference Abstracts:
Personal Website: http://stanford.edu/~jfries/

D. Research Support

ACTIVE

NIH U54 EB020405 *Delp (PI)* *09/29/2014 – 09/30/2018*
Mobility Data Integration to Insight
Vast amounts of data characterizing human movement are available from research labs, clinics, and millions of smartphones and wearable sensors, but integration and analysis of this large quantity of mobility data are extremely challenging. We have established the Mobilize Center (http://mobilize.stanford.edu) to harness these data to improve human mobility and help lay the foundation for using data science methods in biomedicine. The Center will develop new data science approaches, share data and validated software tools, and train thousands of researchers.
Role: Postdoctoral Fellow