

Unlocking Big Data via End-User Interaction with Machine Learning

James Fogarty

Computer Science & Engineering
University of Washington

jfogarty@cs.washington.edu

<http://homes.cs.washington.edu/~jfogarty/>

Big data increasingly defines modern computing, but advances in our ability to collect, store, and process big data have far outpaced our ability to actually understand or interact with that data. The power and potential of big data is therefore limited to organizations that can afford teams of highly-trained experts. Just as the past twenty years have seen the emergence of keyword search as a fundamental and empowering tool amidst information explosion on the Web, I believe the next twenty years must see new methods for empowering everyday people to understand and interact with big data across the many domains where computing is and will be fundamental.

Machine learning offers our best opportunity for scaling human attention to big data, but remains cryptic and challenging to apply. My research focuses on developing new methods for advancing interaction with machine learning. This includes both: (1) supporting software developers as they *create* applications that incorporate machine learning, and (2) supporting everyday people as they *use* and *control* applications that incorporate machine learning.

Software developers currently attempt to manage big data by implementing machine learning techniques to provide automated processing and decision-making in applications. Despite breakthroughs achieved in many domains, the adoption and impact of machine learning remains dramatically limited by a lack of appropriate tools. We have studied why machine learning systems are hard to implement, how current tools fail to support machine learning, and how to design new tools that can demystify machine learning and help bring its power to typical software developers. Our goal is for integrated development environments to provide rich support for machine learning analogous to that currently provided for other common tasks (e.g., web development). Machine learning is fundamental to the future of computing, and improved developer tools will be critical to getting it out of elite labs and into broad practice.

Fully supporting understanding and interacting with big data also requires giving everyday people control over how an application's machine learning components interpret underlying data. Without this control, end-users are limited to distinctions in the data that were pre-provided by the developer. We have examined how end-users can interactively train classifiers, then use those classifiers in their applications. Although traditional active learning approaches are often optimal from the perspective of a learning algorithm, our work reveals other important challenges around which examples to present during training, how to embed training in interaction, and differences in the commitment a person is willing to make to the training process.

We have pursued this research in domains including gesture recognition, document sentiment analysis, web information extraction, web image search, social network group creation, and sensor-based activity recognition, and we recently begun collaborations in computational biology. However, these problems are fundamental and advances will impact many more domains.