CAREER : Machine Learning and Event Detection for the Public Good

With the critical importance of addressing global policy problems such as disease, crime, and terrorism, and the continuously increasing mass of available data, machine learning has become increasingly essential for the development of new, practical technologies that can be applied for the public good. We propose an integrated plan for research and education which will facilitate broad, successful innovations at the intersection of machine learning and public policy.

The *research component* of this proposal will create and explore novel methods for detection of emerging events in massive, complex real-world datasets. By creating general, flexible, scalable, and powerful detection methods which directly address the fundamental challenges of policy problems, we will advance the state of the art both in the theory of event detection and in the practical application of these methods for the public good (e.g. disease surveillance and crime prevention).

The *educational component* of this proposal will incorporate machine learning into the public policy curriculum through development of courses and seminars, Workshops in Machine Learning and Policy Research and Education, and establishment of a new joint Ph.D. program in Machine Learning and Policy. This multi-pronged Machine Learning and Policy (MLP) initiative has the enthusiastic support of CMU's Heinz College and Machine Learning Department.

Integration of research and education will be achieved in numerous ways, including advising student course projects and independent research. The majority of the requested funding is to support Ph.D. students who will become leaders in the MLP field; the project will draw on the unique expertise of students in the MLP Ph.D. program and will contribute to multiple new courses for the joint curriculum. Finally, the proposed workshops will foster discussion aimed at addressing fundamental challenges and general principles for MLP research and education.

Intellectual Merit. The proposed work will address challenging and important research questions:
- How can we incorporate model learning from user feedback into event detection, creating an integrated learning and detection framework which can improve performance over time, discover previously unknown events, and focus attention on the most relevant patterns?
- How can the "linear-time subset scanning" property enable event detection methods to scale up to large and high-dimensional data without sacrificing accuracy or flexibility?
- How can we extend fast subset scanning methods to general multivariate datasets, and incorporate search constraints such as proximity, connectivity, and self-similarity?
- How can we augment event detection methods with novel and useful methodological tools for event characterization, explanation, visualization, investigation, and response?
- How can we use data from emerging, transformative technologies such as mobile phones and online search to enhance event detection, addressing the fundamental scalability challenges?

Broader Impacts. The proposed work will have broad impacts in advancing the core science of event detection, and in the practical application of methods to deployed systems which enhance the public good. By enabling rapid and effective response to emerging disease outbreaks, prediction and prevention of crime, and improved quality of patient health care, our novel detection methods will have transformative and wide-ranging impacts on public health, safety, and security. Research results and educational innovations will be widely disseminated through conference presentations, publications, tutorials, system deployments, technology transfer, and widespread distribution of freely available software. The PI's MLP initiative will have vast impacts by training future policy-makers to understand the benefits of machine learning in the public sector, building bridges between machine learning and policy researchers, and educating future researchers with the skills needed to tackle challenging research questions at the intersection of machine learning and policy. We hope to eventually build a "critical mass" of MLP researchers focused on transforming society through the development and application of machine learning for the public good.