"Crowdsourcing: Quality Management and Scalability"

Abstract
I will discuss the use of crowdsourcing for building machine learning models quickly and under budget constraints, with a focus on the case where humans are noisy and the of "labels" provided by humans for data items are imperfect. I will present strategies of managing quality in a crowdsourcing environment, showing in parallel how to integrate data acquisition with the process of learning machine learning models. I illustrate the results using real-life applications drawn from the field of online advertising. Time permitting, I will also discuss our latest results showing that mice and Mechanical Turk workers are not that different after all.

1 Panos Ipeirotis is an Associate Professor at the Department of Information, Operations, and Management Sciences at the Stern School of Business of New York University. His recent research interests focus on crowdsourcing and on mining user-generated content on the Internet. He received his Ph.D. in Computer Science from Columbia University in 2004. He has received three "Best Paper" awards (IEEE ICDE 2005, ACM SIGMOD 2006, WWW 2011), two "Best Paper Runner Up" awards (JCDL 2002, ACM KDD 2008), and is also a recipient of a CAREER award from the National Science Foundation. He also maintains the blog "A Computer Scientist in a Business School" where he blogs about crowdsourcing, user-generated content, and other random facts, and his blogging activity seems to generate more interest and recognition than any of the other activities mentioned in this bio.