

# Social Media Predictive Analytics

*Svitlana Volkova, Benjamin Van Durme, David Yarowsky, and Yoram Bachrach*

The recent explosion of social media services like Twitter, Google+ and Facebook has led to an interest in social media predictive analytics – automatically inferring hidden information from the large amounts of freely available content. It has a number of applications, including: online targeted advertising, personalized marketing, large-scale passive polling and real-time live polling, personalized recommendation systems and search, and real-time healthcare analytics etc.

In this tutorial, we will describe how to build a variety of social media predictive analytics for inferring latent user properties from a Twitter network including demographic traits, personality, interests, emotions and opinions etc. Our methods will address several important aspects of social media such as: dynamic, streaming nature of the data, multi-relationality in social networks, data collection and annotation biases, data and model sharing, generalization of the existing models, data drift, and scalability to other languages.

We will start with an overview of the existing approaches for social media predictive analytics. We will describe the state-of-the-art static (batch) models and features. We will then present models for streaming (online) inference from single and multiple data streams; and formulate a latent attribute prediction task as a sequence-labeling problem. Finally, we present several techniques for dynamic (iterative) learning and prediction using active learning setup with rationale annotation and filtering.

The tutorial will conclude with a practice session focusing on walk-through examples for predicting latent user properties e.g., political preferences, income, education level, life satisfaction and emotions emanating from user communications on Twitter.

**Svitlana Volkova** is a Ph.D. Candidate in Computer Science at the Center for Language and Speech Processing, Johns Hopkins University. She works on machine learning and natural language processing techniques for social media predictive analytics. She develops batch and streaming (dynamic) models for automatically inferring psycho-demographic profiles from social media data streams, fine-grained emotion detection and sentiment analysis for under-explored languages and dialects in microblogs, effective interactive and iterative rationale annotation via crowdsourcing.

**Benjamin Van Durme** is the Chief Lead of Text Research at the Human Language Technology Center of Excellence, and an Assistant Research Professor at the Center for Language and Speech Processing. He works on natural language processing (specifically computational semantics), predictive analytics in social media and streaming/randomized algorithms.

**David Yarowsky** is a Professor at the Center for Language and Speech Processing, Johns Hopkins University. His research interests include natural language processing and spoken language systems, machine translation, information retrieval, very large text databases and machine learning. His research focuses on word sense disambiguation, minimally supervised induction algorithms in NLP, and multilingual natural language processing.

**Yoram Bachrach** is a researcher in the Online Services and Advertising group at Microsoft Research Cambridge UK. His research area is artificial intelligence (AI), focusing on multi-agent systems and computational game theory. Computational game theory combines the theoretical foundations of economics and game theory with creative solutions from AI and computer science.