

# Analyzing Text and Social Network Data with Probabilistic Models

Padhraic Smyth

University of California, Irvine, USA

smyth@ics.uci.edu

<http://www.ics.uci.edu/~smyth/>

**Abstract.** Exploring and understanding large text and social network data sets is of increasing interest across multiple fields, in computer science, social science, history, medicine, and more. This talk will present an overview of recent work using probabilistic latent variable models to analyze such data. Latent variable models have a long tradition in data analysis and typically hypothesize the existence of simple unobserved phenomena to explain relatively complex observed data. In the past decade there has been substantial work on extending the scope of these approaches from relatively small simple data sets to much more complex text and network data. We will discuss the basic concepts behind these developments, reviewing key ideas, recent advances, and open issues. In addition we will highlight common ideas that lie beneath the surface of different approaches including links (for example) to work in matrix factorization. The concluding part of the talk will focus more specifically on recent work with temporal social networks, specifically data in the form of time-stamped events between nodes (such as emails exchanged among individuals over time).

## Bio

Padhraic Smyth is a Professor at the University of California, Irvine, in the Department of Computer Science with a joint appointment in Statistics, and is also Director of the Center for Machine Learning and Intelligent Systems at UC Irvine. His research interests include machine learning, data mining, pattern recognition, and applied statistics and he has published over 150 papers on these topics. He was a recipient of best paper awards at the 2002 and 1997 ACM SIGKDD Conferences, received the ACM SIGKDD Innovation Award in 2009, and was named a AAAI Fellow in 2010. He is co-author of *Modeling the Internet and the Web: Probabilistic Methods and Algorithms* (with Pierre Baldi and Paolo Frasconi in 2003), and co-author of *Principles of Data Mining*, MIT Press (with David Hand and Heikki Mannila in 2001). Padhraic has served in editorial and advisory positions for journals such as the *Journal of Machine Learning Research*, the *Journal of the American Statistical Association*, and the *IEEE Transactions on Knowledge and Data Engineering*. While at UC Irvine he has received research funding from agencies such as NSF, NIH, IARPA, NASA, and DOE, and from companies such as Google, IBM, Yahoo!, Experian, and Microsoft. In addition to his academic research he is also active in industry consulting, working with companies such as eBay, Yahoo!, Microsoft, Oracle, Nokia, and AT&T, as well as serving as scientific advisor to

local startups in Orange County. He also served as an academic advisor to Netflix for the Netflix prize competition from 2006 to 2009. Padhraic received a first class honors degree in Electronic Engineering from National University of Ireland (Galway) in 1984, and the MSEE and PhD degrees (in 1985 and 1988 respectively) in Electrical Engineering from the California Institute of Technology. From 1988 to 1996 he was a Technical Group Leader at the Jet Propulsion Laboratory, Pasadena, and has been on the faculty at UC Irvine since 1996.