Scalable, General-purpose Semantic Learning

In this talk, I will present recent work on two semantic tasks: (1) all-of-vocabulary sense distribution learning, where we show that topic models can be used to learn sense distribution data at a level comparable to or better than SemCor, but over a much larger vocabulary; and (2) open-class lexical relation classification, where we explore the general utility of vector differences over word embeddings to capture the relation between ordered word tuples.

Biosketch:
Tim Baldwin is a Professor in the Department of Computing and Information Systems, The University of Melbourne, and an Australian Research Council Future Fellow. He has previously held visiting positions at Cambridge University, University of Washington, University of Tokyo, Saarland University, NTT Communication Science Laboratories, and National Institute of Informatics. His research interests include text mining of social media, computational lexical semantics, information extraction and web mining, with a particular interest in the interface between computational and theoretical linguistics. Current projects include web user forum mining, monitoring and text mining of Twitter, and text analytics for the creative industries.

Tim completed a BSc(CS/Maths) and BA(Linguistics/Japanese) at The University of Melbourne in 1995, and an MEng(CS) and PhD(CS) at the Tokyo Institute of Technology in 1998 and 2001, respectively. Prior to joining The University of Melbourne in 2004, he was a Senior Research Engineer at the Center for the Study of Language and Information, Stanford University (2001-2004).