A Spatial Factor Model for Discrete Data

James McInerney  
Columbia University  
12/8/2015 at 10:30 am  
Core A (Room 301)

Abstract

Grouped collections of discrete data are the focus of interest for many machine learning applications (e.g., topic modelling for text, collaborative filtering, genetic studies). Geographical data are often available for such data sets. The typical ways to incorporate spatial information are partial pooling (i.e., sharing parameters between nearby groups of data) and post-processing (i.e., training a model before considering location). We show that both perform badly when the data are very sparse and describe a new model that performs soft pooling, based on the idea that spatially-dependent parameters (e.g., topics in text corpora, venue attributes in collaborative filtering) can be characterized by globally shared functions of location. We compare our method against existing approaches on two sparse data sets: geographically tagged messages from Twitter and venue check-ins from a location-based social network.

Bio

James McInerney is a postdoctoral research scientist with David Blei at Columbia University in the City of New York, working on probabilistic modelling and inference algorithms. He obtained his PhD in Mar 2014, advised by Nick Jennings and Alex Rogers on the ORCHID project at the University of Southampton, UK. He has an MSc in Computing (Artificial Intelligence) from Imperial College London and a BA in Computer Science from Oxford University.

Faculty Host: Tina Eliassi-Rad