THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

SIHAN CHEN

Department of Statistics
The University of Chicago

Topic Modeling and Its Relation to Word Embedding

THURSDAY, July 28, 2016, at 11:00 AM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Topic Modeling has been a popular area of research for more than a decade. Each vocabulary in a document is assigned to a topic, which is a distribution of words, and thus, each article becomes a mixture of different kinds of topics. As a result, we can group similar articles based on the topic assignment of their words. Even though a large number of models has been applied to topic modeling, I mainly focus on two methods of topic modeling: the Latent Dirichlet Allocation (LDA), and Spherical Hierarchical Dirichlet Process (sHDP). The LDA model is based on the method Variational Bayes; the sHDP, on the other hand, needs some extra information gained from Word Embedding. In this paper, I will compare these two methods by computing their Pointwise Mutual Information(PMI).