



The University of Chicago  
Department of Statistics

MASTER'S THESIS PRESENTATION

---

**MARK DAVENPORT**

Department of Statistics  
The University of Chicago

**Identifying Congressional Political Affiliation Using  
Finite Mixture Modeling**

**FRIDAY, February 10, 2012, at 9:00AM**

110 Eckhart Hall, 5734 S. University Avenue

**ABSTRACT**

I apply machine learning techniques to unstructured data sets in order to cluster members of the 109th Congress based on the words used by individual members in floor speeches. In doing so, I study the ability of statistical models to identify political party based on word choice. I develop Multinomial and Dirichlet finite mixture models for clustering from discrete data. The discrete data include raw word counts and topic counts from speeches. Topics are defined according to a probabilistic model called latent Dirichlet allocation, which I develop and apply to my data set. I also explore sparse simulation from a Dirichlet distribution in order to mimic the Congressional data set with which I am working. My results show that clustering Congressional members based on topic counts outperforms clustering on raw word counts. Furthermore, in the best case, clustering is able to correctly identify political affiliation 84% of the time.

---

For information about building access for persons with disabilities, please contact Matt Johnston at 773.702-0541 or send an email to [mhj@galton.uchicago.edu](mailto:mhj@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.