Bayesian Modeling of the Bureaucratic Role in Policy-Making

Keywords: policy-making, computational social science, Bayesian modeling

The United States government is losing popularity. With protests from liberals in the Occupy Movement, action from conservatives in the Tea Party, and approval ratings of congress reaching all-time lows,[1] it appears that something is not working. One recent embarrassing occasion was last summer's debt-ceiling crisis. Congress risked a government shut down by spending weeks debating whether to raise the cap on national debt. This crisis was just one example of government inefficiency. **How could our political system be more efficient?**

One aspect of political efficiency is the process by which legislative agendas are set. This issue has received a great deal of attention, [1, 6] but is typically investigated using data from macro-level behavior such as surveys or financial reports. **I intend to use statistical methods for text and network analysis to investigate how micro-level behavior affects what agendas are chosen and whether those agendas are productive**.

A major theory of policy-making from the political science literature, which I will refer to as the triad theory, states that in order for an issue to receive attention from the legislature, it needs to be supported by a legislator, a bureaucratic agency, and an interested third party. [4] I currently have access to the emails of the county department managers from several counties in North Carolina and the legislative agendas for at least one of those counties. If this triad theory of policy-making is true and email is a good representative of overall communication, there should be some indication of the coordination between these three parties in these administrators' emails. **My plan is to develop novel statistical methods to identify what bureaucratic communication strategies are most effective at getting issues onto the legislative agendas and into the law.** There are five steps to this project:

1. Data Validation: A simple way to check that the triad theory is evident in the email data would be to take particular keywords from certain agenda items, find the subsets of the emails containing those keywords, and identify whether those subsets contain communication between at least one supportive bureaucrat, legislator, and third party.

2. Model Development: My modeling goal is to track the paths of specific issues that arise in the email communications of these bureaucrats and identify which of those issues are addressed in the county agendas and in ordinances that are passed. The first model I will develop will be based on latent Dirichlet allocation (LDA) [2]. LDA is a probabilistic model of text that automatically identifies meaningful "topics," sets of related words, within a collection of documents by treating those topics as model parameters whose values can be estimated without human guidance. My model will associate each "issue" with a set of LDA topics. Every email, agenda, and ordinance will then be associated with a particular set of issues that determines the topics of those documents. In addition, at every time point, each actor in the email network will have a set of "interests" in particular issues. An actor's interests determine what the actor talks about and who the actor talks with. There is precedent in the literature for this type of model: The text aspect is similar in spirit to the role-author-recipient topic model of McCallum et al. [5], and the recipient-prediction aspect is related to many recent network models [3].

---

3. Model Validation: I will validate both the network component and the text component of every model I develop. To validate the network components, I will predict unseen email recipients, and I will generate synthetic networks based on the models and compare the properties of these simulated networks to the observed email networks. To validate the text components, I will enlist human participants (such as political science graduate students) to judge whether the learned topics are intuitively meaningful.

4. Exploratory Analysis: The point of developing descriptive models of our email data is to be able to generate hypotheses about what communication structures are effective. Some important questions these models could help answer are: Do the issues that reach the county agendas appear in emails involving one general manager or involving specific department managers? Is it beneficial for an issue to get attention from the press? Or does the best communication strategy depend on the issue at hand?

5. Experimental Intervention: In the final stage of this project, I will work with political scientists to identify individual bureaucrats or entire bureaucratic agencies that are using sub-optimal communication strategies according to our models. Ideally, we will then advise any willing participants on how to improve their strategies, and observe whether this intervention has an effect on the output of each of their organizations.

My proposed research is a natural extension of my current research. My current research focuses on understanding what makes particular individuals effective, but what is good for an individual may not necessarily be good for the organization as a whole. In my proposed research, I want to understand the performance of an entire organization. This different angle requires fundamentally different models. In addition, my proposed model is more technically sophisticated than my previous work since it incorporates change over time.

If successful, my proposed research would clearly have high impact. It could benefit politicians, bureaucrats, and the general public by suggesting what communication strategies could help make our political system more efficient. In addition, the research will offer novel methodological contributions to the interdisciplinary field of computational social science. There is little previous research on jointly modeling text and network attributes in communication networks. Longitudinal models of how the content and structure of communication networks change over time could be beneficial in problems as diverse as designing the communication channels for an organization, understanding how information spreads through social networks, and structuring classroom activities. I would be grateful for the NSF to afford me the opportunity to contribute my part to these exciting areas.

[1] K. Bawn. Political control versus expertise: congressional choice about administration procedures. *American Political Science Review*, 1995.

[2] D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003.

[3] A. Goldenberg, A. Zheng, S. Fienberg, and E. Airoldi. A survey of statistical network models. *Cambridge Handbook of Computational Cognitive Modeling*, 2009.

[4] J. Kingdon. Agendas, alternatives, and public policies. *HarperCollins College Publishers*, 1995.

[5] A. McCallum, X. Wang, and A. Corrada-Emmanuel. Topic and role discovery in social networks with experiments on Enron and academic email. *Journal of Artificial Intelligence Research*, 2007.

[6] M. McCubbins, R. Noll, and B. Weingast. Structure and process, politics and policy: administrative arrangements and the political control of agencies. *Virginia Law Review*, 1989.