

Lisa Gandy

Research Statement

We are currently in an era where data concerning a wide array of subjects is abundant and easy to obtain. During the course of my graduate studies I have worked on two major projects, Congressional Close Up and News at Seven, which both seek to utilize this abundant data in new and interesting ways. Congressional Close Up makes the vast amount of voting and campaign contribution data more accessible to the everyday citizen. News at Seven automatically creates news shows by analyzing and skillfully combining online news and reader comments to create a unique user experience.

Congressional Close Up

There are currently sixteen nations, which are committed to making government data open and accessible to the public [1]. However this data is usually in a format that is difficult for the average person to sift through. Therefore it seems the potential of this data is wasted. Congressional Close Up is a computer system that seeks to narrow the “data divide” between open government data and the average citizen.

To begin Congressional Close Up makes the assumption that all United States Senators will vote with their party. This is a valid assumption, as today’s congress is increasingly partisan [2]. Therefore when a Senator votes against party the system seeks to explain this anomalous behavior. The system automatically finds relevant campaign contributions, local industries, committee memberships and other influences, which might have influenced the unusual vote.

Most senate votes are on whether to pass or table a bill or amendment. Therefore in order to make a connection between a Senate vote and possible influences on that vote, it is important to know the topic of the bill or amendment that the vote dealt with. Luckily there is a classification system for bills that exists: the Jones and Baumgartner codebook [3][4]. This codebook lists 19 topics and 226 subtopics that can classify any bill. Thousands of bills from prior sessions of congress have been hand coded by topic. Congressional Close Up uses Naïve Bayes to classify new bills and amendments as they arrive. Then the system identifies senators who vote against party, and then possible influences that are matched to the current topic of the relevant bill or amendment. A report is generated and is then publicly available. To view Congressional Close Up at work please visit <http://www.ccu-dev.com/>.

News at Seven

As a new graduate student, I worked as part of a team that concentrated on the News at Seven system [5]. News at Seven is a machine-generated content system, which creates an automatically generated audio/visual news show complete with animated anchors, and text-to-speech generated dialogue. News at Seven consists of several "dynamics" that create a certain type or style of presentation. I specifically worked on the *Shout Out* [6][7] dynamic. In Shout Out, one anchor presents a news story, while another anchor

regularly interrupts with his own short thoughts and opinions about the story.

The Shout Out dynamic began by monitoring a RSS feed for news articles. For each news article, the system extracted reader comments and the text of the article. The system then clustered the reader comments to prevent duplicate comments and then used the cosine similarity between the comment and each paragraph of the news article to create a dialog between the two news anchors. After the dialog was created, a flash presentation with speech and background images was created. The Shout Out dynamic can be used for any type of news story and has been used for entertainment and sports news specifically. For a demonstration of the Shout Out dynamic, and News at Seven in general, please see <http://infolab.northwestern.edu/projects/news-at-seven/>.

Future Research

In the immediate future I would like to focus on using Congressional Close Up to examine state wide voting behavior. I would also like to focus on open government data in general and the concept of making this data transparent and easy to use.

I would also like to make a general-purpose tool centered on Congressional Close up. Currently Congressional Close Up looks at a finite set of influences for each senator. One can imagine that if a system was created that let any user upload relevant data for a particular vote, then this system would become even more useful and relevant. We could imagine that if a vote was centered on regulating mortgage practices, then a user could upload data based on each state's foreclosure rate. Or if a vote dealt with regulating water/air pollution a user could upload air quality data.

I would also enjoy assisting political scientists and other scientists who hand code data with automatic classification of data. With today's machine learning tools, it is completely possible to step away from hand coding data. In a practical sense this would allow political scientists and other social scientists to focus on more important research.

I have just listed very specific research interests, but in general I am interested in information retrieval and utilizing the vast amount of open data in new and interesting ways. I have focused on politics in the past, but am open to any area where open data is available.

References

- [1] Lichtenstein, J. Why Open Data is Not Enough. Wired Magazine, July 2011.
- [2] Ramirez, M. The Dynamics of Partisan Politics on Congressional Approval. American Journal of Political Science. Vol 53, Issue 3. 681-694. July 2009.
- [3] John, P. The Policy Agendas Project: A Review. Journal of European Public Policy. Vol 13, Issue 7, 975-986. 2006.
- [4] Topic Codebook. Policy Agendas Project. <http://www.policyagendas.org/page/topic-codebook>.

- [5] Nichols, N. and Hammond, K. Machine-Generated Multimedia Content. Proceedings of the Second International Conference on Advances in Computer-Human Interactions, 2009.
- [6] Gandy, L. and Hammond, K.. Creating Conversations: an Automated Dialog System. International Conference of Weblogs and Social Media, 2011.
- [7] Gandy, L., Nichols, N. and Hammond, K. Shout Out: Integrating News and Reader Comments. International World Wide Web Conference, 2010.