

## Abstract

Polarizing opinions about political and social controversies take place commonly in mass and more recently user-generated media. A functional democratic society builds on civic discussions among people holding different beliefs on an issue. However, so far, few computer technologies have been devoted to facilitate mutual understanding, and arguably could have worsened the situation.

We envision a computer system that can automatically understand different ideological viewpoints on an issue and identify biased news stories, blog posts, and television news. Such a computer system will raise news readers' awareness of individual sources' biases and encourage them to seek news stories from different viewpoints.

- Computer understanding of ideological perspectives, however, has been long considered almost impossible. In this thesis, we show that ideology, although very abstract, exhibits a concrete pattern when it is communicated among a group of people who share similar beliefs in written text, spoken text, television news production, and web video folksonomies. This emphatic pattern in ideological discourse opens up a new field of automatic ideological analysis, and enables a large amount of ideological text and video to be automatically analyzed.
- We develop a new statistical model, called Joint Topic and Perspective Models, based on the emphatic pattern in ideological discourse. The model combines two essential aspects of ideological discourse: topic matters and ideological biases. The simultaneous inference on topics and ideological emphasis, however, poses a computational challenge. We thus develop an approximate inference algorithm for the model based on variational methods.
- The emphatic pattern in ideological discourse and the Joint Topic and Perspective Model enable many interesting applications in text analysis and multimedia content understanding. At the corpus level, we show that ideological discourse can be reliably distinguished from non-ideological discourse. At the document level, we show that the perspective from which a document is written or a video is produced can be identified with high accuracy. At the sentence level, we extend the model to summarize an ideological document by selecting sentences that strongly express a particular perspective.