

Kenneth R. Fleischmann, Thomas Clay Templeton, and **Jordan Boyd-Graber**. **Modeling Diverse Standpoints in Text Classification: Learning to Be Human by Modeling Human Values**. *iConference*, 2011.

```
@inproceedings{Fleischmann:Templeton:Boyd-Graber-2011,  
Author = {Kenneth R. Fleischmann and Thomas Clay Templeton and Jordan Boyd-Graber},  
Booktitle = {iConference},  
Year = {2011},  
Location = {Seattle, Washington},  
Title = {Modeling Diverse Standpoints in Text Classification: Learning to Be Human by Modeling Human Values},  
}
```

Modeling Diverse Standpoints in Text Classification: Learning to Be Human by Modeling Human Values

Kenneth R. Fleischmann
University of Maryland
4105 Hornbake Building, South Wing
College Park, MD 20742-4345
kfleisch@umd.edu

Thomas Clay Templeton
University of Maryland
4110 Hornbake Building, South Wing
College Park, MD 20742-4345
clayt@umd.edu

Jordan Boyd-Graber
University of Maryland
4105 Hornbake Building, South Wing
College Park, MD 20742-4345
jbg@umiacs.umd.edu

ABSTRACT

An annotator's classification of a text not only tells us something about the intent of the text's author, it also tells us something about the annotator's standpoint. To understand authorial intent, we can consider all of these diverse standpoints, as well as the extent to which the annotators' standpoints affect their perceptions of authorial intent. To model human behavior, it is important to model humans' unique standpoints. Human values play an especially important role in determining human behavior and how people perceive the world around them, so any effort to model human behavior and perception can benefit from an effort to understand and model human values. Instead of training humans to obscure their standpoints and act like computers, we should teach computers to have standpoints of their own.

Categories and Subject Descriptors

H.1.1 [Models and Principles] Systems and Information Theory – *information theory*; User/Machine Systems – *human factors*; J.4 [Social and Behavioral Sciences] Sociology.

General Terms

Theory.

Keywords

Value sensitive computing, machine learning, framing theory, standpoint epistemology, diversity.

1. MODELING DIVERSE STANDPOINTS

Each person views the world from a different standpoint. Reading a document entails viewing it from that standpoint. Consequently, a reader's reaction to a text tells us something not only about the author's intent but also about the reader's standpoint. Here, we argue that the insights of framing theory and standpoint epistemology can be usefully applied to machine learning-based text classification within natural language processing.

The concept of framing helps to demonstrate the influence of a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iConference 2011, February 8-11, 2011, Seattle, WA, USA
Copyright © 2011 ACM 978-1-4503-0121-3/11/02...\$10.00

diversity of standpoints on interpretations of texts. "The major premise of framing theory is that an issue can be viewed from a variety of standpoints and be construed as having implications for multiple values or considerations" [1, p. 104]. As such, there is no one right way to interpret a text, but instead multiple potentially equally valid or likely interpretations depending on one's standpoint.

Figure 1 presents a traditional model of textually mediated communication. Supervised text classification from natural language processing traditionally adopts one view toward this situation: a statistical model of the writing process is developed in order to infer underlying properties of the text, such as author intent. In contrast, research guided by framing theory builds models of the reading process that link framing devices in the text with effects on reader response. This paper seeks to combine the theoretical insights of framing theory with the practical applications of supervised text classification to propose an approach that can have implications for understanding how diverse audiences use information and to apply that understanding to produce classifiers that model diverse standpoints.

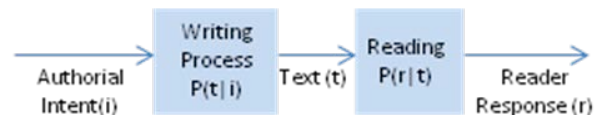


Figure 1. Model of Textually Mediated Communication

Research within the fields of library science and American studies on the history of reading demonstrate that reading is an important activity that affects different individuals in different ways. Indeed, one of the key driving research questions within this research area is: "Are gender, race, class, age, creed, sexual orientation (etc.) differences evident in how readers select and read library materials?" [2, p. 381]. Researchers are now focusing increasingly on the impact of reading on the lives of individuals, as documented by "detailed records of what [individuals] read and what it meant to them" [3, p. 47]. There is a need to "uncover the specific reading practices of actual readers" [4, p. 143] in order to better understand how texts shape individuals and communities. In this spirit, it is indeed quite relevant to consider how people read texts not only to understand the intentions of the author(s) who wrote the text or the features of the text itself but also to study the attitudes and beliefs of the reader(s) of the text.

The approach of supervised text classification is philosophically compatible with the concept of situated knowledges from the field

of science and technology studies [5]. Situated knowledges is a variation on standpoint epistemology [6], which holds that each person has a unique and valuable standpoint from which they view the world that shapes how they perceive the world, and that a strong objectivity can be built by summing those standpoints. Situated knowledges are the product of those standpoints, the knowledges generated by particular positions from which to view the world. Thus, instead of attempting to achieve machine learning of a generic human standpoint, it is instead valuable to think about embedding these situated knowledges within software, and teaching machines to also view the world from situated standpoints.

2. MODELING HUMAN VALUES

Human behavior is significantly influenced by values, which can be expressed through communication and can spread across social networks. Values can be defined as “what a person or group of people consider important in life” [7, p. 349]. “Values are determinants of virtually all kinds of behavior that could be called social behavior or social action, attitudes and ideology, evaluations, moral justifications and justifications of self to others, and attempts to influence others” [8, p. 5]. As such, socially intelligent computing must include awareness of and sensitivity to values as a core component. Values can be expressed through communication, such as writing, and as such, techniques from natural language processing can be used to detect both the values of the author of a text and the values of the readers of a text, leading to a computational understanding of the relationship between author and audience and how values are expressed and perceived. Thus, much can be gained from efforts to annotate human values in texts [9], automate the classification of human

Amy Weinberg. This paper is based in part on work supported by National Science Foundation grants IIS-0729459 and IIS-0734894.

4. REFERENCES

- [1] Chong, D. and Druckman, J.N. 2007. Framing theory. *Annu. Rev. Polit. Sci.* 10, 1 (Jun. 2007), 103-126.
- [2] Wiegand, W.A. 2003. To reposition a research agenda: What American studies can teach the LIS community about the library in the life of the user. *Libr. Quart.* 73, 4 (Oct. 2003), 369-382.
- [3] Kaestle, C.F. 1991. The history of readers. In *Literacy in the United States: Readers and Reading Since 1880*, C.F. Kaestle, H. Damon-Moore, L.C. Stedman, K. Tinsley, and W.V. Trolinger, Jr., Eds. Yale University Press, New Haven, CT, 33-72.
- [4] Pawley, C. 2002. Seeking ‘significance’: Actual readers, specific reading communities. *Book Hist.* 5 (2002), 143-160.
- [5] Haraway, D. 2003. Situated knowledges: The science question in feminism and the privilege of partial perspective. In *The Feminist Theory Reader: Local and Global Perspectives*, C.R. McCann and S.-K. Kim, Eds. Routledge, New York, NY, 391-403.
- [6] Harding, S. 1991. *Is Science Multicultural: Postcolonialisms, Feminisms, and Epistemologies*. Indiana University Press, Bloomington, IN.

values in texts [10], and to measure the relationship between annotators’ perceptions of texts and their values [11].

One particular application of the ability to predict how diverse individuals with different values will interpret texts could be the development of a “focus group in a box” that allows for the simulation of different individuals’ or types of individuals’ reactions to texts (or, perhaps eventually, other media). This would of course be of tremendous interest and relevance to social scientists, who could have fast, inexpensive, and non-invasive access to a pool of “individuals” without needing to go through the human subject process or worry about the potential harm to individuals. Diplomats could benefit from the ability to test messages on different audiences without risking a diplomatic crisis. Marketing analysts could cheaply and easily try out new sales pitches on diverse audiences. Finally, political strategists could test new political campaign themes without worrying about the wrong themes “going viral” unexpectedly and being spread without permission or control, since simulated audiences post no blogs (at least for now).

To ensure that computers are socially intelligent in ways that are compatible with human intelligence, it is important that the machine learning community considers the importance of standpoint, and develops the capability to develop artificial intelligences that are aware of and sensitive to human values [12].

3. ACKNOWLEDGMENTS

This paper has benefited significantly from conversations with colleagues including An-Shou Cheng, Mark Goldberg, Emi Ishita, Malik Magdon-Ismael, Doug Oard, Philip Resnik, Asad Sayeed, Al Wallace, Jan Wiebe, and

- [7] Friedman, B., Kahn, P.H., Jr., and Borning, A. 2006. Value sensitive design and information systems. In *Human-Computer Interaction and Information Systems*, P. Zhang and D. Galletta, Eds. M.E. Sharp, Armonk, NY, 348-372.
- [8] Rokeach, M. 1973. *The Nature of Human Values*. Free Press, New York, NY.
- [9] Cheng, A.-S., Fleischmann, K.R., Wang, P., Ishita, E., and Oard, D.W. 2010. Values of stakeholders in the Net neutrality debate: Applying content analysis to telecommunications policy. In *Proceedings of the 43rd Hawai’i International Conference on Systems Sciences* (Koloa, HI, Jan. 5-8, 2010).
- [10] Ishita, E., Oard, D.W., Fleischmann, K.R., Cheng, A.-S., and Templeton, T.C. 2010. Investigating multi-label classification for human values. In *Proceedings of the 73rd Annual Meeting of the American Society for Information Science and Technology* (Pittsburgh, PA, Oct. 22-27, 2010).
- [11] Templeton, T.C., Fleischmann, K.R., and Boyd-Graber, J. 2011. Comparing values and sentiment using Mechanical Turk. In *Proceedings of the 6th iConference* (Seattle, WA, Feb. 8-11, 2011).
- [12] Fleischmann, K.R., Oard, D.W., Cheng, A.-S., Wang, P., and Ishita, E. 2009. Automatic classification of human values: Applying computational thinking to information ethics. In *Proceedings of the 72nd Annual Meeting of the American Society for Information Science and Technology* (Vancouver, BC, Canada, Nov. 6-11, 2009).