Computing Political Preference among Twitter Followers

Jen Golbeck
College of Information Studies
Human Computer Interaction Lab (HCIL)
University of Maryland
golbeck@cs.umd.edu

Derek L. Hansen
College of Information Studies
CASCI, HCIL
University of Maryland
dlhansen@umd.edu

ABSTRACT

Through anecdotal evidence and a variety of methods, claims are constantly being made about the bias of media outlets. As many of those outlets create online personas, we seek to measure the political preferences of their audience, rather than of the outlet itself. In this paper, we present a method for computing political preferences of an organization's Twitter followers using congressional liberal/conservative ADA scores as a seed. We apply this technique to characterize the political preferences of major news media Twitter audiences. We discuss how this technique can be extended or used to create personalized recommendations and insights for news seekers and social media analysts and marketers.

Author Keywords

Twitter, politics, liberal, conservative, news, journalism

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

As major media outlets establish online presences in social media, understanding the characteristics of those audiences is an important task. It has implications for how information is presented in this environment where personalization is expected. Furthermore, it can provide valuable information to marketers and social media analysts.

As a first step toward understanding audiences, we present a technique for establishing audience preferences in a given domain on Twitter. We use politics as our motivating example and show our method applied to determining the political preference of media outlets' audiences.

We begin by reviewing other computational strategies that analyze political preferences and biases of news outlets. We next describe our novel method for inferring political preferences of Twitter audiences, and illustrate it's effectiveness by showing political preferences of major news outlets that correspond with expectations. Finally, we describe the interface design and information presentation implications of this work as well as its application to other domains.

BACKGROUND

While we are not studying media bias, but rather the political preferences of audiences, it is worth briefly discussing the extensive research on analyzing media bias. A subset of this work uses automated methods to infer liberal/conservative bias of news stories and outlets. These automated methods do not depend on subjective measurements of bias, although the specific techniques used to infer bias can be problematic and are highly contested. One approach is to compute a media bias score based on citations in the news story – news outlets that cite "think tanks" that are also cited by Congressperson's with a known liberal bias are assumed to be more liberal [7]. Another approach is to compare keywords and phrases used by Congresspeople of known political persuasions with those used in news articles – news outlets that use terms like "death tax" and "illegal immigration" are more likely to be conservative [5]. A final approach assigns a liberal/conservative score to web documents based on the number of times they are co-cited with other web documents that have a known political bias [3].

In contrast to these approaches, we characterize the political preferences of news outlet audiences, not the news outlet content itself.

Our strategy is similar to [7] in that we use Congresspeople's American for Democratic Action's (ADA) scores as a starting point for our scoring; however, we use Twitter Follow relationships rather than article citations. Using Follow relationships avoids the concern with [7] (see [8]) related to citation practices of journalists and Congresspeople. Our approach does not require coding of data (as in [7]) or access to large corpuses of news stories and congressional speeches; it relies instead on freely available and open access data from Twitter.

METHOD AND SAMPLING

Unlike [7], we are not interested in predicting media bias. Instead, we are interested in understanding the political preference of the *audience* of different media outlets and organizations by using social networking sites like Twitter. Our examples and applications are in the political domain,

but we believe the technique is generalizable when the right background information is available.

Our approach includes the following steps:

Step 1: Apply known scores to a seed group, in this case congresspeople using Twitter. The base data of liberal/conservative scores is obtained from Americans for Democratic Action (ADA), who puts out an annual report that considers the voting record of members of Congress [1]. ADA defines a key set of votes that indicate liberal and conservative positions, and use the Congressperson's voting record to assign each a score. The most liberal score is a 1.0, and the most conservative is 0.0. This is a widely accepted measure of political position. We apply the 2009 ADA liberal/conservative ratings of current congresspeople.

Step 2: Map the scores of the seed group onto their followers. We collect the list of followers for each member of Congress on Twitter – we'll hereafter call them Congress Followers. An inferred political preference score (P Score) for each Congress Follower is computed as the simple average of the ADA scores for all Congresspeople he or she follows. We tested this assumption by surveying a convenience sample of 40 subjects who follow politicians on Twitter. Of those, 62.5% follow politicians who mostly share their political views, while only 5% follow politicians who mostly hold political views that oppose their own. The rest reported following a mix of both, with an average of 65% of the politicians they follow sharing their own views. Thus, overall, users tend to follow politicians with similar views; even when there is a mix of political views, those that match the user's tend to dominate. These preliminary results suggest that our assignment of a P Score to congressional followers is reasonable.

Step 3: Map the inferred scores of the seed group followers (i.e., Congress Followers) onto the target of the investigation, in this case the Twitter accounts of media outlets. A simple approach is to assign the average of the liberal/conservative scores of all Congress Followers who also follow the target media outlet. However, this approach raises a problem: Twitter users may not represent the population well.

Indeed, Republicans significantly outnumber Democrats on Twitter (127 to 103), though Democrats outnumber Republicans in Congress. Furthermore, Republican Twitter users tend to have disproportionally more followers than Democrats. Even excluding John McCain, who has over 1.7 million followers as a result of the 2008 Presidential election (more than 30 times the next most followed member of Congress and nearly double the total number of followers of all other Congresspeople), the sum of Republican followers is 581,997, compared to 291,050 for Democrats. As a result, the pool of Congress Followers is significantly biased toward conservatives.

Without any adjustment, the audiences of news outlets will incorrectly appear more conservative. Our solution is to selectively sample Congress Followers so they more closely match the general population, minimizing the effect of the initial selection bias. Below is an explanation of how we implemented this approach.

To account for the bias in our full data set, we created a sample that balanced Democrats and Republicans. John McCain was excluded because of the significant impact he could have on scores, and because his Presidential run makes him a particularly abnormal data point.

Congresspeople were broken into groups by the number of followers they had: over 10,000, 5,000, 1,000, 500, 100, 50, and 10. Within each group, we randomly selected equal numbers of Republican and Democratic representatives until we had the maximum number for the least represented group. We chose equal numbers from each party since in recent elections there are roughly equal numbers in Congress. For each selected Congressperson, we randomly selected the number of followers equal to the group where they were a member (e.g. for a congressman from the "over 1,000" group, we selected 1,000 random followers).

To compute the final media audience P Scores, we used a 10-fold validation; we drew 10 samples using this technique, computed the P Score for each media audience using each sample, and averaged the scores over all samples.

CONGRESS FOLLOWER SCORES

Figure 1 shows a distribution of scores for all Congress Followers, as well as the distribution of ADA scores for all Congresspeople (not just those on Twitter).
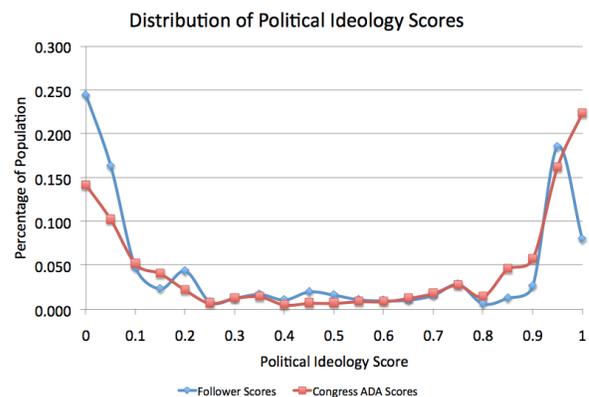
Figure 1. The distribution of P Scores among all Congress Followers and ADA scores for Congresspeople. A score of 0.5 is moderate, 0.0 is very conservative and 1.0 is very liberal.

The distribution of scores among the Congress Followers and the sampled follower population is strongly bimodal, since it is based upon the congressional ADA scores which are bimodal. This is in contrast to other evidence that suggests the distribution of political ideologies among the general public is a more normal distribution, with most of the population as moderate [5]. One practical effect of this bimodal distribution is that the spread of the audience scores is more pronounced than it would be if preference

Media Outlet	Twitter User ID	Total Followers	Avg. % of Total Followers sampled	Avg. P Score
Fox News	foxnews	266,121	7.33%	0.26
The Drudge Report	Drudge_Report	102,981	20.52%	0.27
Washington Times	washtimes	13,545	33.75%	0.29
Wall Street Journal	WSJ	392,332	5.75%	0.41
US News & World Report	usnews	7,836	28.60%	0.49
The L.A. Times	latimes	72,296	9.91%	0.51
USA Today	USATODAY	63,714	9.90%	0.51
Good Morning America	gma	1,698,875	0.72%	0.51
The News Hour	NewsHour	54,787	12.79%	0.52
CBS News	CBSNews	1,578,599	0.94%	0.53
Newsweek	Newsweek	1,256,536	1.25%	0.54
Washington Post	washingtonpost	154,400	9.51%	0.54
The Today Show	todayshow	629,088	1.89%	0.55
The Early Show	theearlyshow	12,628	11.16%	0.55
Time Magazine	TIME	2,134,411	0.99%	0.57
ABC World News	abcworldnews	12,973	14.93%	0.57
CNN Breaking News	Cnnbrk	3,314,716	1.40%	0.58
NBC Nightly News	nbcnightlynews	27,137	16.85%	0.58
The New York Times	nytimes	2,553,291	1.40%	0.60
Morning Edition (NPR)	MorningEdition	6,146	17.22%	0.66

Table 1. The average audience P Scores for the twenty popular media outlets studied in [6]. Averages are computed over 10 randomly drawn samples using the sampling method described above. Results are sorted from most conservative to most liberal.

scores is more pronounced than it would be if based on the general populations' self-reported political preferences.

We now turn to the results of our analysis of media outlet audience preferences.

MEDIA AUDIENCES' POLITICAL PREFERENCE

Our first application of this method was to compute the audiences' political preferences, through their P Scores, of the same popular media outlets evaluated in [7]. Table 1 shows the audience bias scores along with the percentage of each outlet's followers who were considered when computing the value. Both values are averages over 10 samples.

For traditionally conservative outlets [9], such as Fox News [2], we found audiences with correspondingly conservative P Scores: Fox News (0.256), the Drudge Report (0.265), and the Washington Times (0.290). There are no outlets with audiences that have P Scores that are liberal to the same extent that these are conservative, but some liberal preference is visible in the audiences of outlets like the New York Times (0.604) (considered to be liberal leaning [9, 11]) and NPR's Morning Edition (0.659).

The vast majority of these media outlets' audiences – 15 out of 20 – fall between the moderate scores of 0.4 and 0.6. Half are even closer to the midpoint, falling within 0.05 of the perfect moderate 0.5 value. These moderate audiences

are found despite the fact the P Scores over the population of followers tend to be either very liberal or very conservative.

Our findings of audience preference are similar to the estimates of political orientation of media outlets found in [3], which used co-citation of hyperlinks to infer political orientation of web documents and their associated news outlets. Although, the scales differ, like [3] we found that the Wall Street Journal and The New York Times equally deviated from the middle in opposite directions (conservative and liberal respectively). This similarity may come from the fact that [3] relies partially on news outlet audiences by using their website linking behavior. Our method provides a more direct measure of audience political preferences by focusing on follower relationships.

Note that these scores do not imply that the outlets themselves present news in a way that reflects their audience's political preferences. However, pursuing this line of investigation is an interesting space for future work.

DISCUSSION
There are a number of implications and areas for future work that follow from these results. The expectation with social media is that the information presented to users will be personalized. Some of that comes from the structure of the services themselves (e.g. who users choose to follow on Twitter affects which tweets they see). However, there are

many opportunities to further personalize and enhance the way information is presented to users.

Understanding the political preference of an audience can be important for presenting tailored information and personalizing the user's experience. For example, an audience's political preference can be used as input into recommender systems. In collaborative filtering systems, items are recommended by finding people with tastes similar to the user and recommending things those people like. In this context, if we know a user's political preferences, we can find media outlets that have audiences with a similar preference, mimicking the basic idea behind collaborative filtering. Tweets (or information provided on other social media sites) can be highlighted, filtered out, or sorted based on the similarity of their audiences' political preferences to those of the user. Alternatively, our method could be used as input to systems that recommend content read by audiences with diverse political opinions [10]. Finally, marketers and social media analysts can use our method to get a sense of their Twitter reach within different political markets, helping them know if they are reaching their intended audience.

Outside of personalization, this technique may have applications for studying media bias in social media. While we have set out to measure audience political preference, not media bias, previous work has shown that news consumers have a significant preference for like-minded media outlets [5] and use new social media tools to actively seek out those with similar views [4]. This implies that people may choose a media outlet because its presentation of the news reflects their own political beliefs, and thus the preference of an audience may generally reflect the bias of the outlet. While we do not have evidence to support this connection, looking at the potential applicability of audience preference to understanding media bias is an area for future research.

Finally, while we have used political preference and media outlets as our example case, we believe this technique is applicable in other domains. We have begun to use this technique to understand the political preferences in the Twitter audiences of government agencies, think tanks, political organizations, and individuals. It could also be used for non-political analysis. For example, a similar analysis could be done using the Green Scores that rate how environmentally responsible companies are. These could be used to create an environmental score for followers which could then be applied to understand the audience of organizations, politicians, and other entities on Twitter.

CONCLUSION
In this paper, we have presented a technique for estimating the preferences of Twitter user followers. Using the political domain as our motivating example, we present a method that uses the follower connections in the Twitter network to propagate liberal/conservative scores from members of Congress to Congress Followers and then to audiences of media outlets.

Our results show that the political preferences of media outlets' audiences reflect the liberal/conservative leanings of the media organizations as presented in prior literature.

The results have potential applications for motivating new interface personalization techniques, understanding media bias, and for detecting other types of audience preferences in different domains. There is much future work to be done in this space and we hope this initial work serves as motivation to pursue those issues.

REFERENCES
1. Americans for Democratic Action. Annual Voting Records, 2009. Available at http://www.adaction.org/

2. DellaVigna, S. and Kaplan, E. The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics, 122*, 3 (2007), 1187-1234.

3. Efron, M. 2004. The liberal media and right-wing conspiracies: using cocitation information to estimate political orientation in web documents. In *Proceedings of CIKM'04.* (Washington, D.C., November 2004), ACM Press, 390-398.

4. Garret, K. Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate. *Journal of Communications, 59*, 4 (December 2009), 676-699.

5. Gelman, A. *Red state, blue state, rich state, poor state: why Americans vote the way they do*. Princeton Univ Press, Princeton NJ, 2008.

6. Gentzkow, M. and Shapiro, J. What drives media slant? Evidence from US daily newspapers. *Econometrica, 78*, 1 (2010), 35–71.

7. Groseclose, T. and Milyo, J. A Measure of Media Bias. *The Quarterly Journal of Economics, 120*, 4 (2005), 1191–1237.

8. Media Matters for America. Former fellows at conservative think tanks issued flawed UCLA-led study on media's "liberal bias". (2005). Available at http://mediamatters.org/research/200512220003

9. Mondo Times: The worldwide news media directory. Available at http://www.mondotimes.com

10. Munson, S. A. and Resnick, P. 2010. Presenting diverse political opinions: how and how much. In *Proceedings of CHI'10*. (Atlanta, GA, April 2010). ACM Press, 1457-1466.

11. Puglisi, R. Being the New York Times: The Political Behaviour of a Newspaper. Available at SSRN: http://ssrn.com/abstract=573801