Social Proof: The Impact of Author Traits on Influence Detection

Sara Rosenthal
IBM Research*
Yorktown Heights, NY, USA
sjrosenthal@us.ibm.com

Kathleen McKeown
Columbia University
Computer Science Department
NY, NY, USA
kathy@cs.columbia.edu

Abstract

It has been claimed that people are more likely to be influenced by those who are similar to them than those who are not. In this paper, we test this hypothesis by measuring the impact of author traits on the detection of influence. The traits we explore are age, gender, religion, and political party. We create a single classifier to detect the author traits of each individual. We then use the personal traits predicted by this classifier to predict the influence of contributors in a Wikipedia Talk Page corpus. Our research shows that the influencer tends to have the same traits as the majority of people in the conversation. Furthermore, we show that this is more pronounced when considering the personal traits most relevant to the conversation. Our research thus provides evidence for the theory of social proof.

1 Introduction

The psychological phenomenon of *social proof* suggests that people will be influenced by others in their surroundings. Furthermore, social proof is most evident when a person perceives the people in their surroundings to be similar to them (Cialdini, 2007). This tendency is known as *homophily*. One manner in which people can be similar is through shared author traits such as the demographics age (year of birth), gender (male/female), and religion (Christian/ Jewish/ Muslim/ Atheist), as well as political party (Republican/Democrat).

In this paper, we explore the impact of social proof via author traits in detecting the most influential people in Wikipedia Talk Page discussions. We present an author trait detector that can detect a suite of author traits based on prior state-of-the art methods developed for individual author traits alone, and use it to classify individuals along four author traits: age, gender, religion, and political party. We train the classifier using automatically labeled or prior existing datasets in each trait. Our classifier achieves accuracy comparable to or better than prior work in each demographic and political affiliation. The author trait classifiers are used to automatically label the author traits of each person in the Wikipedia Talk Page discussions.

An influencer is someone within a discussion who has credibility in the group, persists in attempting to convince others, and introduces topics/ideas that others pick up on or support (Biran et al., 2012; Nguyen et al., 2013b). We use supervised learning to predict which people in the discussion are the influencers. In this paper we use the demographics and political affiliation of the authors in the Wikipedia Talk Page as features in the classifier to detect the influencers within each discussion. This is known as situational influence. In contrast, global influence refers to people who are influential over many discussions. It is important to explore situational influence because a person can be quite influential in some Wikipedia Talk Page discussions but not at all in others. We show that social proof and homophily exists among participants and that the topic of the discussion plays a role in determining which author traits are useful. For example, religion is more indicative of influence in discussions that are religious in nature such as a discussion about the Catholic Church.

In the rest of this paper we first discuss related work in influence detection. We then describe our author trait classifier, related work, and the datasets used to train the models. All of our datasets are publicly available at http://www.cs.columbia.edu/~sara/data.php. Next, we discuss the Wikipedia Talk Page (WTP) dataset and how they were labeled for influence. Afterwards we discuss our method for detecting influence, the experiments and results. Finally, we conclude with a discussion of the impact of author traits on influence detection.

* Work completed as graduated student at Columbia University

Proceedings of 2016 EMNLP Workshop on Natural Language Processing and Computational Social Science, pages 27–36,
Austin, TX, November 5, 2016.

2 Related Work

Influence detection has been explored in conversations and social networks. We discuss both types of influence in more detail in this section.

2.1 Influence in Conversations

Several authors have detected influencers in a single conversation using the actual discussion (Quercia et al., 2011; Nguyen et al., 2013b; Biran et al., 2012). This work has explored detecting influencers using features such as dialog structure, agreement, persuasion, sentiment, and topic control in several online corpora such as WTP, Twitter, Presidential Debates, and the ICSI meeting corpus (Janin et al., 2003). This work did not, however, explore the impact of author traits in detecting influence.

There has also been work exploring influence on the utterance level (Young et al., 2011) within hostage negotiation transcripts. The utterances were labeled for influence using Robert Cialdini's weapons of influence (Cialdini, 2007), including social proof. However, they define social proof differently as: 1) an utterance that is a reference to a social norm (e.g. referring to a way a person could be influential) and 2) an appeal to the group regarding how they should proceed. Our use of social proof is based on shared author traits. Furthermore, they do not distinguish between the weapons of influence in their results making it impossible to determine their performance on social proof alone. Other related work has looked at analyzing the interactions of persuasive arguments using dialog structure, style, and textual features in the Reddit's ChangeMyView discussions (Tan et al., 2016).

A closely related area of research has been predicting power relations in dialog (Prabhakaran and Rambow, 2014; Danescu-Niculescu-Mizil et al., 2012; Strzalkowski et al., 2013). This includes several types of power relationships, such as hierarchical and administrative power, as well as influence. Most relevant among this work, Prabhakaran et al (2012) have explored the role of gender in hierarchical power within the Enron e-mail corpus. They find that female superiors use less displays of power than male superiors, and subordinates in female environments use more conventional language than any other group. Finally, they use the actual gender of the participants to improve the accuracy of predicting who is the subordinate and who is the superior given a pair of people.

2.2 Influence in Social Networks

There has been a lot of work that has explored influence in social networks (e.g. (Watts and Dodds., 2007; Bakshy et al., 2011; Barbieri et al., 2013; Huang et al., 2012; Goyal et al., 2011; Myers et al., 2012)) by analyzing how it spreads through the network.

Bamman et al (2012) explore the effect of gender identity and homophily on how information spreads. Aral and Walker (2002) analyze the impact of demographics on influence by identifying influential people on Facebook. They find influential people by examining how a viral message spreads through the network. They found interesting patterns among demographics: Men are more influential than women, older people tend to be more influential than younger people and that people are the most influential to their peers. We have similar findings in age and gender in our analysis. In contrast to our work, they did not use the demographics to predict influence nor do they predict influence within a discussion. Similarly, Dow et al (2013) investigate how photos on Facebook are shared and which demographics are more likely to share a particular photo.

3 Author Trait Detection

We implemented an author trait detection system that uses lexical, and lexical-style features to automatically detect author traits such as demographics and political affiliations. We also include features related to online behavior. In particular, we include the time and day of posting, but avoid features that are not available on all online discussion forums such as number of friends, interests, comments, likes/favorites, and hashtags. Several of these features are available on the datasets used in author trait detection: LiveJournal (interests, comments, friends), Blogger (comments, friends), and Twitter (friends, favorites, hashtags). However, none of them are available in WTP discussions, the dataset we use to detect influence.

3.1 Related Work

Prior work in demographic detection has used classic features such as n-grams (1-3 words), Part-of-Speech (POS) tags (e.g. is the word a noun or verb), and stylistic features (e.g. (Schler et al., 2006; Rao et al., 2010; Mukherjee and Liu, 2010)), as well as domain specific features such as hashtags and the social network in Twitter (Nguyen and Lim, 2014; Burger et al., 2011; Conover et al., 2011; Zamal et al., 2012) and friends and interests in LiveJournal (Rosenthal and McKeown, 2011). In this work we aim to make our author trait detector as general as possible and therefore only use features available in all online discussion forums by excluding genre specific features. Thus, we compare our system's results to the results in prior work that exclude genre specific features.

Prior work in age detection has explored classification based on age groups in blogs and tweets (Schler et al., 2006; Goswami et al., 2009; Rao et al., 2010; Rosenthal and McKeown, 2011) and exact age using regression (Nguyen et al., 2011; Nguyen et al., 2013a) in blog and tweets. Gender detection too has been classified in

author trait	source	label	size
age	blogger.com	year of birth	19098
	livejournal.com	year of birth	21467
gender	blogger.com	Male	9552
		Female	9546
	livejournal.com	Male	4249
		Female	3287
political party	Twitter.com	Republican	1247
		Democrat	1200
religion	Twitter.com	Christian	5207
		Islam	1901
		Atheist	1815
		Judaism	1486

Table 1: The size (in users) of each trait corpus

blogs (Schler et al., 2006; Mukherjee and Liu, 2010; Goswami et al., 2009; Nowson and Oberlander, 2006) and Twitter (Rao et al., 2010; Burger et al., 2011; Bamman et al., 2012). Predicting political orientation or ideologies has focused on predicting political views as left-wing vs right-wing in Twitter (Conover et al., 2011; Cohen and Ruths, 2013) or debates (Iyyer et al., 2014; Gottipati et al., 2013). There is little work on predicting religion with the only known prior work found to be on the prediction of Christian vs Muslim Twitter users (Nguyen and Lim, 2014) and work on classifying documents by Islamic ideology (e.g Muslim Brotherhood) and organization (e.g. Hamas) (Koppel et al., 2009).

3.2 Data

Our author trait data comes from two different types of online sources; weblogs for age and gender and microblogs for politics and religion. All of our datasets are publicly available at `http://www.cs.columbia.edu/~sara/data.php`.

3.2.1 Age and Gender

We use the publicly available blogger.com authorship corpus (Schler et al., 2006) and the LiveJournal age corpus (Rosenthal and McKeown, 2011) to detect age and gender. The Blogger corpus is annotated for age and gender while the LiveJournal corpus provides the date of birth for each poster. We use these annotations as gold labels for predicting age and gender. For uniformity, we converted the blogger age in the authorship corpus to the date of birth based on the time of download (2004). For example, a 22 year old in 2004 was born in 1982. We then automatically generated gender labels for the LiveJournal corpus internally. We generate gender labels by looking at the first name of the blogger if it was provided. We used the Social Security Administration lists[1] to determine the appropriate gender based on the popularity of the name. If the name is predominantly male or female at a 2:1 ratio we assign it that gender. Otherwise, we exclude the blogger from the gender corpus. The size of the age and gender corpora are shown in Table 1.

3.2.2 Politics and Religion

There are several websites that either automatically generate (tweepz.com), or allow users to self-label (twellow.com and wefollow.com) their Twitter account into categories. Previous work (Zamal et al., 2012) has used the labels from wefollow.com to automatically download Twitter users related to desired categories. We follow this approach to download Twitter users based on political party (Republican/Democrat), and religion (christian, jewish, muslim, atheist). After downloading the list of users we performed some post-processing to exclude non-English speakers based on the language in their bio. We excluded any users whose bios contained many (40%) foreign characters and non-english words. Additionally, we discarded users that appeared in more than one category within a single author trait (e.g. a person cannot be labeled as Republican *and* Democrat).

We then used the Twitter API to download the last 100 tweets of each user on 11/4/2014. Downloading on this date was desirable because it ensured that the data was rich in political information because it was election day in the US. Our political party tweets consists of Republican and Democrat. We downloaded tweets pertaining to the four most popular religions in the United States[2]: Christianity, Judaism, Islam, and Atheism. The full data statistics are provided in Table 1.

3.3 Method

We present a supervised method that draws on prior work in the area as discussed in the prior section. We experimented with several classifiers in Weka (Hall et al., 2009) and found that SVM always performs the same or better than the other methods. We use this single classifier to build several models which detect each author trait by training and testing on the relevant data (e.g. the classifier is trained using the age data to build a model to predict age). The only exception is that we use Linear Regression to predict the exact age of each user using year of birth. We apply χ^2 feature selection to all groups of features in the training data to reduce the feature set to the most useful features. The features are generated by looking at the past 100 tweets or 25 blogs per user. We also limit the text to 1000 words per user to improve processing time. We include three type of features: lexical, lexical-stylistic, and online behavior.

3.3.1 Lexical Features

We include three kinds of lexical features: n-grams, part-of-speech (POS) (using Stanford Core NLP (Man-

[1] http://www.ssa.gov/oact/babynames/limits.html

[2] www.census.gov/compendia/statab/cats/population/religi-on.html

Author Trait	Majority	Accuracy
Age	57.1	79.6
Gender	51.9	76.4
Political Party	51.3	75.2
Religion	50.0	78.3

Table 2: The author trait results of SVM classification using accuracy

ning et al., 2014)), and collocations which have all been found to be useful in prior work (Schler et al., 2006; Rao et al., 2010; Mukherjee and Liu, 2010; Rosenthal and McKeown, 2011). We keep the top 1000 features of each type. n-grams refers to a count of 1-2 word phrases. POS features refer to the counts of POS tags. Collocations are bigrams that take the subject/object (S/O) relationship of terms into account. We implement this using Xtract (Smadja, 1993). We ran our own implementation of Xtract on the most recent 100 blog posts or tweets per user. In the Twitter datasets we run Xtract on all the text. Due to the large size of the blog corpora, we limit it to the 2,000 most recent words per user. We include the S/O bigrams (e.g. voting Democrat), POS bigrams (e.g. we VB) and S/O POS bigrams (e.g. vote NN) generated from Xtract as features.

3.3.2 Lexical-Stylistic Features

We include two types of lexical-style features: general and social media. General features can be found in any genre, such as the number of capital words, exclamation points, and question marks. Social Media features are those common in online discussions such as word lengthening (e.g. looooooong), emoticons, and acronyms. Younger people may be more likely to use such features. We also include the Linguistic Inquiry Word Count (LIWC) categories (Tausczik and Pennebaker, 2010) as features as in prior work (Schler et al., 2006). The LIWC classifies words as belonging to one or more broad categories (e.g., work, family, religion, negative emotion). These different categories can be very indicative of author traits. For example, men may talk more about work and Atheists will be less likely to talk about religion.

3.3.3 Online Behavior

While we do exclude all features that don't occur in all datasets (e.g. comments, friends, and hashtags), there is one online behavior feature that is found in all discussions. That is a time-stamp indicating when the person posted. We use this to generate two features, the most common hour (0-24 GMT) and most common day of the week (Sunday-Saturday) that the person posts. For example this could be useful in predicting age as younger people may post later in the evening than older people.

3.4 Results

We trained our classifier on each author trait. The classifier was tuned using cross-validation and all results are shown on a held-out test set of 10% of the data. All datasets were kept unbalanced. The results are shown in Table 2. The gender, religion, and political party demographics were classified using SVM.

We classified age using two models. First, we tried to predict the exact year of birth using Linear Regression; we achieved a mean absolute error (MAE) of 5.1 from the year of birth and a .55 correlation (r) which is slightly better than the results in prior work (Nguyen et al., 2011) when avoiding blog-specific features. The next approach we took was performing binary classification using 1982 as the splitting point. This year of birth was found to be significant in prior work (Rosenthal and McKeown, 2011).

Our results on gender detection are slightly worse than leading methods (Schler et al., 2006; Mukherjee and Liu, 2010). However, we think this is due to prior work using cross-validation as opposed to a held-out test set. In fact, our cross-validation results were 82.5%, slightly better than Schler et al (2006) . It is more difficult to compare to the work of Mukherjee and Liu (Mukherjee and Liu, 2010) as the datasets are different and much smaller in size. Mukherjee and Liu have a collection of blogs from several websites (e.g. technorati.com and blogger.com) and only 3100 posts. In contrast we generate our model with blogs from livejournal.com and blogger.com (Schler et al., 2006) and over 25,000 blogs labeled with gender.

Prior work in detecting politics on tweets tends to combine Republican and conservative to "right-wing" and Democrat and liberal to"left-wing" and use Twitter-specific features such as political orientation of friends to achieve high accuracy making it difficult to compare against them. Although not directly comparable due to different datasets, our results are similar or better than the results in prior work where Twitter-specific features are excluded.

Finally, the prior work in religion is two-way classification of Muslim vs Christian, making it difficult to compare against their results.

In some cases our results are better than prior work or on a new area of classification. Our system is competitive or better than prior state-of-the-art classifiers with good accuracy in detecting each trait. In addition, we are the only one to use the same system to generate four models to predict the author traits (In the past only age and gender have been detected in this manner (Schler et al., 2006)).

4 Influence Detection

In this section we describe the data, method, and experiments in detecting influence in WTP discussions using

	Train	Dev	Test	Total
# discussions	410	47	52	509
# posts	7127	730	892	8749
# participants	2536	277	317	3130
# influencers	368	41	47	456
# files w/o influencers	62	8	6	76

Table 3: Data statistics for the Wikipedia Talk Page Influence corpus

Topic	G	A	R	P
Abortion	35	8	9	3
Catholic Church	27	9	553	7
George W. Bush	4	0	8	68
Israel	4	18	623	12
Michael Jackson	2	42	4	0

Table 4: A list of topics and the occurrence of issues associated with them in **A**ge, **G**ender, **R**eligion, and **P**olitics. An occurrence > 5 indicates it is an issue relevant to that topic.

the author traits described in the previous section.

4.1 Data

We use the author trait detector to explore the impact of social proof in detecting influencers in WTP. Each page on Wikipedia is generated by user contribution, and thus discussion is needed to avoid conflict from different contributors. This discussion occurs in the Wikipedia Talk Pages[3]. They are rich in content and argumentative in nature making it an ideal dataset for detecting influence.

Our dataset is an extension of the Wikipedia dataset described in prior work (Biran et al., 2012) and contains 509 discussions ranging over 99 different topics. It is important to note that although there may be some overlap among authors across the dataset, we find the influencer within each discussion individually. This is known as situational influence. Detecting global influence would be an interesting extension in future work. The WTP discussions were annotated for influence by four different people with an average inter annotator agreement using Cohen's κ of .61. The annotators were given guidelines similar to those described in prior work (Biran et al., 2012; Nguyen et al., 2013b): An influencer is someone who has credibility in the group, persists in attempting to convince others, and introduces topics/ideas that others pick up on or support. Typically there is one influencer in each discussion, and on rare occasion two (20/509 or 3.9%). Since our goal is detecting influence, we excluded the 76 discussions without influencers from the experiments resulting in 433 discussions. Of the 3130 participants, 456 of them were found to be influential. The data was broken down into a training (80%), development (10%), and test set (10%). The statistics for each set is shown in Table 3.

4.2 Method

Our method involves four groups of features. The first is single features related to each author trait; the second is features indicating if the author trait is the majority in the discussion, and the third is a combination of author traits. We also include features related to the issue being discussed in the Wikipedia Page. We will describe the features in greater detail in the rest of this section.

In addition, as a baseline feature we include the number of words the participant has written. This feature is important because in addition to indicating the likelihood of someone being influential (if someone barely participates in the discussion it reduces their chances of being influential), the odds of the predicted author trait being correct decreases if the provided text is minimal.

4.2.1 Single Features

We explore the occurrence of influence in each author trait as an indication of what type of people are more influential. Each author trait is represented as a binary feature during classification. The breakdown of each feature by influence in the training set is shown in Figure 1. There tend to be more old people in Wikipedia, but there is also a clear indication that older people are more influential. We have similar findings with the male gender, the Republican political party, and the Jews and Christians in religion. We suspect that the tendency towards an author trait may be dependent on the topic of the Wikipedia article as discussed in the following section. For example, political party may play a more important role in a discussion regarding abortion and religion may play a more important role in a discussion regarding Israel. Finally, we also have a feature indicating the exact year of birth that was predicted for each author (e.g. 1983).

4.2.2 Topic Features

The topic in a discussion can indicate what kind of issues will be addressed. This in turn can indicate a stronger presence of different author traits. We use the title of each discussion to infer its topic. For example, a Wikipedia article with the title "The Catholic Church" will be more likely to be edited by religious people than an article about the pop star Michael Jackson. This in turn can indicate the author trait tendencies of the people in the WTP. In order to analyze the impact of topic on influence and author traits we automatically inferred the author traits that were likely to be related to the Wikipedia article.

We implemented this by counting the occurrence of the labels and related synonyms of each author trait within the Wikipedia article. For example, male and female are gender labels. This alone was sufficient for our task since we want high precision and care less about recall. It is important to stress, that we did not do this in the WTP

[3] http://en.wikipedia.org/wiki/Wikipedia:Tutorial/Talk_pages

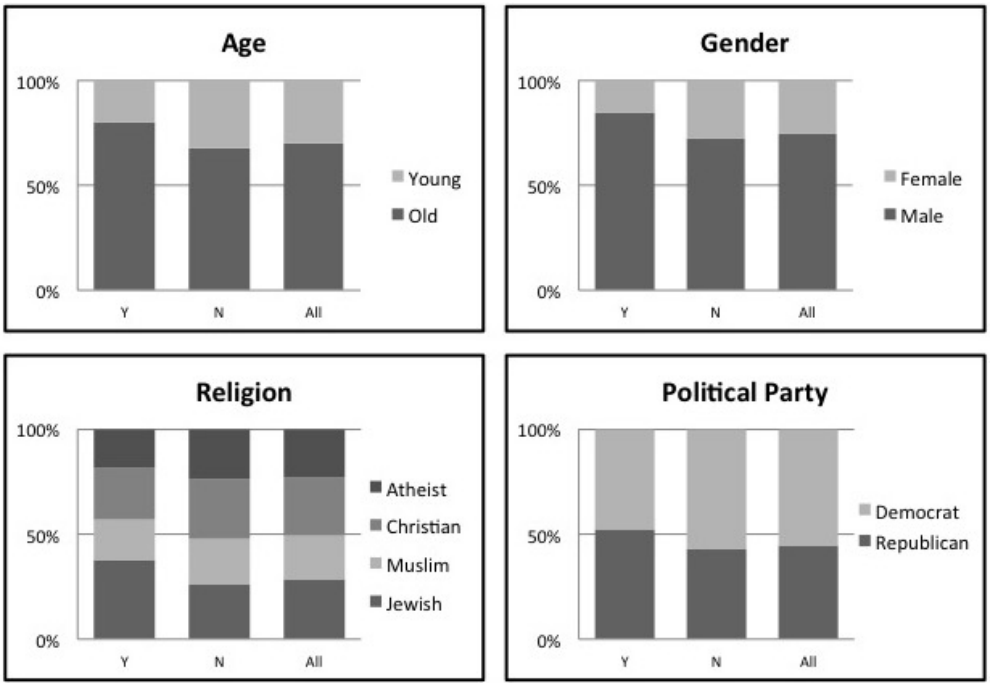

Figure 1: Breakdown of the binary features by influence (Y/N) and overall (All) in the training set.

discussions but rather in the *actual Wikipedia article*. If an author trait term occurred more than five times[4] it was considered to be an issue related to that author trait to ensure the occurrence was more than happenstance. Table 4 lists an example of topics and the occurrence of issues within the Wikipedia article. Using this method, there were 38 age, 42 gender, 66 religious, and 58 political articles. Most articles overlap on more than one author trait issue. There are a total of 99 topics with one or multiple discussions from the WTP associated to the topic.

We use each issue as a feature which is true if that topic is associated with the article and false if it is not. For example, the gender, age, and religion issues would be true for Abortion Talk Pages.

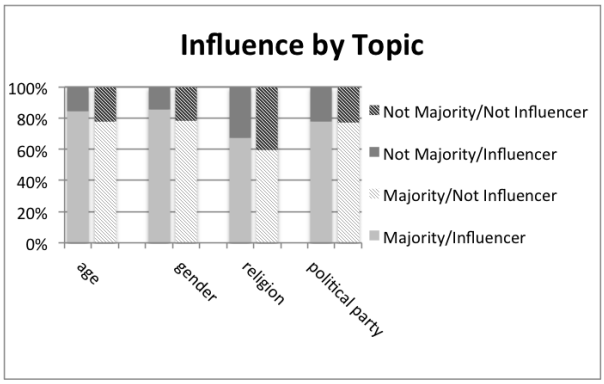

Figure 2: The breakdown of the users being in the majority within their document for each author trait with topic being taken into account.

4.2.3 Majority Features

Social proof indicates that people will be influenced by those that are like them. We measure this per author trait by determining if a person is predicted to be in the majority within the discussion and have a majority feature corresponding to each author trait. For example, if the majority of the people in a discussion are predicted to be Republican, we expect that the influencer is likely to be predicted to be Republican as well. Furthermore, we expect this to be most evident when the discussion is relevant to the particular author trait. For example, a discussion on abortion would be relevant to religion, politics, and gender. Figure 2 illustrates that influencers are in the majority more than non-influencers when the issue is relevant in the Wikipedia article. In general all people tend to be in the majority author trait in a discussion, but there is a stronger tendency towards being in the majority when a person is an influencer. The results displayed take the topic of the document into account in that only documents applicable to each author trait are shown in the chart. For example, discussions on abortion are only included in the bars on religion, politics, and gender. We also include features to indicate whether the participant is in the majority in *all* author traits or in *no* author traits.

[4]The split of terms among documents is such that documents have no terms whatsoever most often and fewer than 6 terms related to an issue 48.5% times whereas 51.5% of the issues have 6 or more terms.

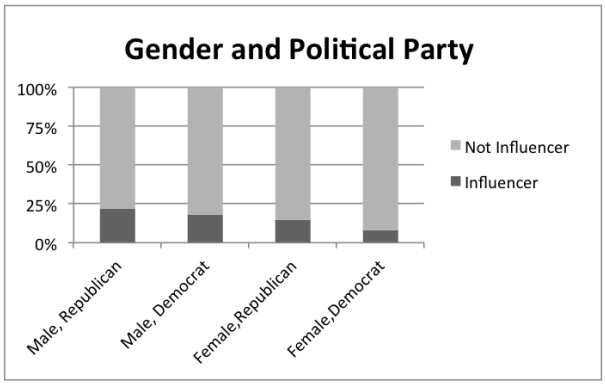

Figure 3: The breakdown of influencers and non-influencers in the training data based on the binary combination feature of gender and political party.

In order to determine whether the majority features should be useful, in addition to using the single features, we needed to verify whether there were enough cases where the overall minority author trait was still the majority author trait within a reasonable amount of discussions. We find that in the training data, in 84.1% of the discussions the majority is older people and in 88.5% of the discussions the majority is male. These percentages are in line with the trends found in the single features as shown in Figure 1. However, there still are many discussions where the majority is female (11.5%) or younger people (15.9%). In contrast to our findings in the single features shown in Figure 1, where overall there were slightly more Republicans than Democrats, we found that in 55.8% of the discussions in the training data the majority is Democrat whereas slightly more editors are Republican. In terms of religion we found that in 41.5%, 16.6%, 18.8%, and 23.2% of the discussions the majority is Jewish, Muslim, Christian, and Atheist respectively. Although Christianity is the most commonly predicted religion overall (see Figure 1), we expect that in the discussions the majority is Judaism due to the many articles that are controversial to the state of Israel (e.g. regarding Gaza and the Israeli Defense Force). This indicates that, in particular, using the majority religion feature should have a positive impact on predicting influencer in addition to the single religion features.

4.2.4 Combination Features

In addition to looking at a single author trait of a person at a time, we also explore whether combining author traits is beneficial. Many studies have shown that certain tendencies towards an issue are based on several author traits. In particular, this applies to combining demographics and politics. For example, women tend to vote for Democrats[5] and Christians tend to vote for Republicans[6].

As one example, we find that indeed in our dataset women are 53% more likely to be Democrat. However, we find that women that are Republican are more likely to be influential than women who are Democrat as shown in the breakdown of the <gender,political party> feature in the training data in Figure 3.

4.3 Experiments and Results

All results were predicted using the SVM classifier in Weka (Hall et al., 2009) built with a polynomial kernel, complexity tuned towards the development set ($C = 10$), and logistic models to provide confidence values. We experimented with other classifiers (e.g. Naive Bayes, Logistic Regression) but SVM consistently performed better or the same as other classifiers. Rather than balancing the training set using downsampling, we balance the class weights of the influencer examples based on their occurrence in the training data. This ensures that the classifier knows we are more interested in finding influencers without incurring a considerable loss in data.

Influencers are rare in discussions. Therefore, the standard measure of accuracy does not appropriately describe the success of the system. This is because predicting that no one is an influencer will have a high accuracy, but will not address our goal of finding influencers. Instead, we present results for predicting influence using F-score on the influencer class. We compare our experiments to two baselines, picking everyone as an influencer (all-yes baseline), and the number of words a person wrote in the discussion (num-words baseline).

In addition to using the results provided by the classifier, we also use the confidence of the classifier as a second prediction which we consider to be the experiment with *ranking*. Since we know that there is at least one influencer in each discussion, we choose the person given the highest confidence by the classifier as the influencer. It is important to note that it is still possible for more than one person to predicted to be the influencer. This approach only applies for discussions where no influencer was chosen. Using ranking to predict the influencer can outperform the equivalent system without ranking. In the future we would like to adjust the annotation method to rank all of the people in the discussion based on influence instead of just choosing the influencer(s).

All results are shown in Table 5. All results following the baselines include the number of words and topic features unless otherwise mentioned. The system using just the best majority features gives 2.4 points improvement in F-score compared to using just the number of

[5]http://www.pewresearch.org/fact-tank/2014/11/05/as-gop-celebrates-win-no-sign-of-narrowing-gender-age-gaps/

[6]http://www.pewforum.org/2014/11/05/how-the-faithful-voted-2014-preliminary-analysis/

Experiment	Conf. Matrix	P%	R%	F%
all-influencer	$\begin{bmatrix} \mathbf{47} & 0 \\ 235 & 0 \end{bmatrix}$	16.7	**100.0**	28.7
num words	$\begin{bmatrix} 24 & 46 \\ 23 & 189 \end{bmatrix}$	34.3	51.0	41.0
majority best	$\begin{bmatrix} 28 & 54 \\ 19 & 181 \end{bmatrix}$	34.1	59.6	43.4^R
single best	$\begin{bmatrix} 26 & 45 \\ 21 & 190 \end{bmatrix}$	36.6	55.3	44.1
majority+single best	$\begin{bmatrix} 20 & 51 \\ 18 & 184 \end{bmatrix}$	36.3	**61.7**	45.7^R
best w/o topic	$\begin{bmatrix} 27 & 51 \\ 20 & 184 \end{bmatrix}$	34.6	57.5	43.2^R
best	$\begin{bmatrix} \mathbf{29} & 50 \\ 18 & 185 \end{bmatrix}$	**36.7**	**61.7**	$\mathbf{46.0}^R$

Table 5: The results of all groups of features on influence detection using author traits. The confusion matrix is filled, by row, as [TP FN] and [FP TN]. ^{R} indicates that ranking was used in the results. The best results are highlighted in bold.

words in a sentence (row 3) using all of the majority features. Ranking was also useful in this system. In row 4, we show that the best system using just single features achieves a 3.1 points improvement in F-score compared to using just the number of words in the sentence. This system uses gender, religion, and political party. The best system using single and majority features combined (row 5) gave an improvement of 4.7 points in F-score overall. These features are the exact age and distance from mean age, and religion single features, and the majority, gender, religion, political party, always-the-majority, and never-the-majority features as well as using ranking. Finally, in the last row, the best set of combination and majority features had a 5.0 points improvement in F-score using the same features as in the single and majority system in addition to combination features: majority <political party, gender>, and single <religion, gender> and uses ranking. This provides evidence that homophily and social proof are both important in predicting influencers. Finally, as a comparison, we show the best system without using the topic features. In row 6, we show that excluding topic features causes a reduction in performance.

5 Discussion

Our goal in this paper is not to produce the best system for influence detection, but rather to analyze the impact of social proof in influence detection. Our results show that social proof is important in being influential. This is indicated by the usefulness of the majority features and a 5.0 boost in F-score using the best group of features.

It is interesting to note that even when the author trait of a person may be predicted incorrectly, certain tendencies are found in discussions on different issues. This indicates that topic is important. For example, the majority religion in most articles regarding the Catholic Church is predicted to be Christian.

We believe that the biggest drawback to our author trait predictions in the WTP discussions is due to the limited amount of text available for some people. Roughly half of the participants write less than 100 words within the discussion indicating a higher likelihood of incorrectly predicting their author traits. We included the number of words as a feature to help address this issue. The classifier should use this feature to learn that the author trait features are less reliable when the author has written less. We would like to explore combining the text written by each person throughout the entire corpus (most authors appear in more than one article) to improve the author trait predictions.

The author trait models are trained on different corpora than Wikipedia and as a result we do not know how accurate the author trait predictions on Wikipedia are. We do find that there are similar trends in our predictions in the Wikipedia training data in comparison to reported statistics of Wikipedia Editor demographics [7]. For example, in a 2013 study it was found that 83% of the Wikipedia editors were male. In Figure 1, we find that approximately 75% of the users are predicted to be male. The reported demographics on age indicate that there are more old people than young people and that the 50% split occurs somewhere between 1980-1989. Similarly, we find that the majority of users are born before 1982 (See Figure 1), indicating they are older and that 1982 is likely a good split for Wikipedia. Finally, the most popular religions of contributors on Wikipedia in 2012 are Christianity (35%), no religion (36%), Judaism (9%), and Islam (6%). In our predictions, we find that Christianity is the most common with Judaism following next. We expect the discrepancy with atheism is because it is a subset of no religion. Statistics on the political party of Wikipedia editors could not be found. The relationships between the trends in our training data and the most recent reported statistics are encouraging and indicative of positive labeling of author traits in our dataset. In the future, we would also like to have the discussions annotated for author traits to analyze the upper bound impact of author traits on influence prediction.

Finally, does being in the minority indicate that it will be harder to be influential? For example, as shown, men are more influential than women in this dataset (see Figure 1). Does this mean that women have no hope of being influential, particularly in a male dominant setting? On the surface, yes. Women may have to work harder to be influential in a male dominant setting. We, however, do not have to lose hope if we are in the minority!

[7] en.wikipedia.org/wiki/Wikipedia:Wikipedians#cite_note-UNU-M-6, meta.wikimedia.org/wiki/List_of_Wikimedians_ by_religion

There are many traits and their importance varies across discussions. Gender may not play an important role in some discussions. For example, political party may be more important. In other words, if the majority of people in a political discussion are democrats it would be better to be a female democrat than a male republican. Social proof does, however, indicate that if a person has nothing in common with the other participants in the discussion being influential will be nearly impossible. The key then is to find something, no matter how small, that can help one relate to others in a discussion. This connection can then be exploited to become influential.

6 Conclusion

In this paper, we present an author trait detection system which predicts four different author traits: age, gender, religion, and political party. We show that influencers tend to have certain of author traits within the WTP dataset. These are particularly dependent on the issue being discussed. We also show that influencers tend to be aligned with the majority of the other participants in the conversation. This indicates that social proof is a useful measure for detecting influence. Including such features gives a 5.0 improvement compared to using the number of words of each participant in the discussion for an F-score of 46.0%.

In the future, we would like to use the different author traits to help improve each of the individual author trait results. For example, using the predicted age and gender to improve the model for predicting political party. To improve our result in influence detection, we would like to use the content per author across the corpus for author trait prediction at once. When available, the increase in content would allow us to more accurately predict the correct author traits of a person. We would also like to annotate the influencer corpus for gold author trait labels to gain a stronger grasp of the importance of author traits in influence prediction. In addition, we would like to explore the impact of detecting influence with author traits and other features used in prior work such as agreement, dialog structure, and persuasion. Finally, we would also like to explore using word embeddings and deep learning.

7 Acknowledgements

This paper is based upon work supported by the DARPA DEFT Program. The views expressed are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

References

[Aral and Walker2012] Sinan Aral and Dylan Walker. 2012. Identifying Influential and Susceptible Members of Social Networks. *Science*, 337(6092):337–341, July.

[Bakshy et al.2011] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on WSDM*, WSDM '11, NY, NY, USA. ACM.

[Bamman et al.2012] David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2012. Gender in twitter: Styles, stances, and social networks. *CoRR*.

[Barbieri et al.2013] Nicola Barbieri, Francesco Bonchi, and Giuseppe Manco. 2013. Topic-aware social influence propagation models. *Knowledge and Information Systems*, 37(3):555–584.

[Biran et al.2012] Or Biran, Sara Rosenthal, Jacob Andreas, Kathleen McKeown, and Owen Rambow. 2012. Detecting influencers in written online conversations. In *Proceedings of the LSM 2012 Workshop*, Montreal, June.

[Burger et al.2011] John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on twitter. In *Proceedings of the Conference on EMLNP*, EMNLP '11.

[Cialdini2007] Robert B. Cialdini. 2007. *Influence: The Psychology of Persuasion (Collins Business Essentials)*. Harper Paperbacks, January.

[Cohen and Ruths2013] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on twitter: Its not easy! In *International AAAI Conference on Weblogs and Social Media*.

[Conover et al.2011] M Conover, B Gonçalves, J Ratkiewicz, A Flammini, and F Menczer. 2011. Predicting the political alignment of twitter users. In *Proceedings of 3rd IEEE Conference on Social Computing (SocialCom)*.

[Danescu-Niculescu-Mizil et al.2012] Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st International Conference on WWW*, WWW '12, pages 699–708, NYC, USA. ACM.

[Dow et al.2013] P. Alex Dow, Lada A. Adamic, and Adrien Friggeri. 2013. The anatomy of large facebook cascades. In *ICWSM*.

[Goswami et al.2009] Sumit Goswami, Sudeshna Sarkar, and Mayur Rustagi. 2009. Stylometric analysis of bloggers' age and gender. In *International AAAI Conference on Weblogs and Social Media*.

[Gottipati et al.2013] Swapna Gottipati, Minghui Qiu, Liu Yang, Feida Zhu, and Jing Jiang. 2013. Predicting user's political party using ideological stances. In *SocInfo*, volume 8238 of *Lecture Notes in Computer Science*, pages 177–191. Springer.

[Goyal et al.2011] Amit Goyal, Francesco Bonchi, and Laks V. S. Lakshmanan. 2011. A data-based approach to social influence maximization. *Proc. VLDB Endow.*, 5(1):73–84, September.

[Hall et al.2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The weka data mining software: An update.

[Huang et al.2012] Junming Huang, Xue-Qi Cheng, Hua-Wei Shen, Tao Zhou, and Xiaolong Jin. 2012. Exploring so-

cial influence via posterior effect of word-of-mouth recommendations. In *Proceedings of the Fifth ACM International Conference on WSDM*, WSDM '12, pages 573–582, NY, NY, USA. ACM.

[Iyyer et al.2014] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *ACL*.

[Janin et al.2003] Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, and Chuck Wooters. 2003. The icsi meeting corpus. pages 364–367.

[Koppel et al.2009] Moshe Koppel, Navot Akiva, Eli Alshech, and Kfir Bar. 2009. Automatically classifying documents by ideological and organizational affiliation. In *ISI*, pages 176–178. IEEE.

[Manning et al.2014] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

[Mukherjee and Liu2010] Arjun Mukherjee and Bing Liu. 2010. Improving gender classification of blog authors. In *Proceedings of the 2010 Conference on EMLNP*, EMNLP '10, Stroudsburg, PA, USA. ACL.

[Myers et al.2012] Seth A. Myers, Chenguang Zhu, and Jure Leskovec. 2012. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 33–41, NY, NY, USA. ACM.

[Nguyen and Lim2014] Minh-Thap Nguyen and Ee-Peng Lim. 2014. On predicting religion labels in microblogging networks. In *Proceedings of the 37th International ACM SIGIR Conference*, SIGIR '14, pages 1211–1214, NY, NY, USA. ACM.

[Nguyen et al.2011] Dong Nguyen, Noah A. Smith, and Carolyn P. Rosé. 2011. Author age prediction from text using linear regression. In *Proceedings of the 5th Workshop on LaTeCH*, LaTeCH '11. ACL.

[Nguyen et al.2013a] Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. 2013a. " how old do you think i am?" a study of language and age in twitter. In *ICWSM*.

[Nguyen et al.2013b] Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, Deborah Cai, Jennifer Midberry, and Yuanxin Wang. 2013b. Modeling topic control to detect influence in conversations using nonparametric topic models. In *Machine Learning*, pages 1–41. Springer.

[Nowson and Oberlander2006] Scott Nowson and Jon Oberlander. 2006. The identity of bloggers: Openness and gender in personal weblogs. In *AAAI*. American Association for Artificial Intelligence.

[Prabhakaran and Rambow2014] Vinodkumar Prabhakaran and Owen Rambow. 2014. Predicting power relations between participants in written dialog from a single thread. In *Proceedings of the 52nd Annual Meeting of the ACL (Volume 2: Short Papers)*, pages 339–344. ACL.

[Quercia et al.2011] Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft. 2011. In the mood for being influential on twitter. In *SocialCom/PASSAT*, pages 307–314. IEEE.

[Rao et al.2010] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying latent user attributes in twitter. In *Proceedings of the 2Nd International Workshop on SMUC*, SMUC '10, pages 37–44, NY, NY, USA. ACM.

[Rosenthal and McKeown2011] Sara Rosenthal and Kathleen McKeown. 2011. Age prediction in blogs: A study of style, content, and online behavior in pre- and post-social media generations. In *proceedings of ACL-HLT*.

[Schler et al.2006] J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. 2006. Effects of age and gender on blogging. In *AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*.

[Smadja1993] Frank Smadja. 1993. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19:143–177.

[Strzalkowski et al.2013] Tomek Strzalkowski, Samira Shaikh, Ting Liu, George Aaron Broadwell, Jennifer Stromer-Galley, Sarah M. Taylor, Veena Ravishankar, Umit Boz, and Xiaoai Ren. 2013. Influence and power in group interactions. In *SBP*, pages 19–27.

[Tan et al.2016] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of WWW*.

[Tausczik and Pennebaker2010] Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Language and Social Pyschology*.

[Watts and Dodds.2007] D. J. Watts and P. S. Dodds. 2007. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34:441–458.

[Young et al.2011] Joel Young, Craig Martell, Pranav Anand, Pedro Ortiz, and IV Henry Gilbert. 2011. A microtext corpus for persuasion detection in dialog. In *AAAI*.

[Zamal et al.2012] Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *ICWSM*.