Utilizing Facebook pages of the political parties to automatically predict the political orientation of Facebook users
Esther David Maayan Zhitomirsky-Geffet Moshe Koppel Hodaya Uzan

Article information:
To cite this document:
Esther David Maayan Zhitomirsky-Geffet Moshe Koppel Hodaya Uzan , (2016),"Utilizing Facebook pages of the political parties to automatically predict the political orientation of Facebook users", Online Information Review, Vol. 40 Iss 5 pp. 610 - 623
Permanent link to this document:
http://dx.doi.org/10.1108/OIR-09-2015-0308

Downloaded on: 14 September 2016, At: 08:23 (PT)
References: this document contains references to 58 other documents.
To copy this document: permissions@emeraldsight.com
The fulltext of this document has been downloaded 38 times since 2016*

Users who downloaded this article also downloaded:

Access to this document was granted through an Emerald subscription provided by emeraldsrm:140257 []

For Authors
If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldsight.com/authors for more information.

About Emerald www.emeraldsight.com
Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.
Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.
Utilizing Facebook pages of the political parties to automatically predict the political orientation of Facebook users

Esther David
Department of Computer Science, Ashkelon Academic College, Ashkelon, Israel

Maayan Zhitomirsky-Geffet
Department of Information Science, Bar-Ilan University, Ramat Gan, Israel, and

Moshe Koppel and Hodaya Uzan
Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel

Abstract

Purpose – Social network sites have been widely adopted by politicians in the last election campaigns. To increase the effectiveness of these campaigns the potential electorate is to be identified, as targeted ads are much more effective than non-targeted ads. Therefore, the purpose of this paper is to propose and implement a new methodology for automatic prediction of political orientation of users on social network sites by comparison to texts from the overtly political parties’ pages.

Design/methodology/approach – To this end, textual information on personal users’ pages is used as a source of statistical features. The authors apply automatic text categorization algorithms to distinguish between texts of users from different political wings. However, these algorithms require a set of manually labeled texts for training, which is typically unavailable in real life situations. To overcome this limitation the authors propose to use texts available on various political parties’ pages on a social network site to train the classifier. The political leaning of these texts is determined by the political affiliation of the corresponding parties. The classifier learned on such overtly political texts is then applied on the personal user pages to predict their political orientation.

To assess the validity and effectiveness of the proposed methodology two corpora were constructed: personal Facebook pages of 450 Israeli citizens, and political parties Facebook pages of the nine prominent Israeli parties.

Findings – The authors found that when a political tendency classifier is trained and tested on data in the same corpus, accuracy is very high. More significantly, training on manifestly political texts (political party Facebook pages) yields classifiers which can be used to classify non-political personal Facebook pages with fair accuracy.

Social implications – Previous studies have shown that targeted ads are more effective than non-targeted ads leading to substantial saving in the advertising budget. Therefore, the approach for automatic determining the political orientation of users on social network sites might be adopted for targeting political messages, especially during election campaigns.

Originality/value – This paper proposes and implements a new approach for automatic cross-corpora identification of political bias of user profiles on social network. This suggests that individuals’ political tendencies can be identified without recourse to any tagged personal data. In addition, the authors use learned classifiers to determine which self-identified centrists lean left or right and which voters are likely to switch allegiance in subsequent elections.

Keywords Social networks, Machine learning, Automatic political profiling, Cross-corpora classification, Text categorization

Paper type Research paper
1. Introduction
In the past decade social network sites have been exploited as a source of social capital (Ellison et al., 2007; Steinfeld et al., 2008; Burke et al., 2011), especially, for career promotion and marketing campaigns by private users and enterprises (Pesonen, 2011; Weinberg, 2011). In addition, users employ social networks for political discussions and communication (Stieglitz and Dang-Xuan, 2013). Politicians have also discovered the great potential of social network sites and use them for their political campaigns (Williams and Gulati, 2013; Kim, 2011; Baek, 2015). Nowadays, every political party and leader maintains an account on Facebook, Twitter and/or other social network sites, where they publish their agenda. Recently, many politicians and their strategists use social networks as an effective platform to influence the agendas of professional journalists and to appeal to strong supporters (Kreiss, 2014). They also attempt to employ social networks to directly communicate with their electorate and to build community support (Gunn and Skogerbo, 2013; Hong and Nadler, 2011; Kavanaugh et al., 2011; Paris and Wan, 2011).

To this end, politicians need to identify the potential electorate among different users on the network site. Moreover, politicians might be interested in early identifying “swing voters” who are likely to change allegiance in subsequent elections. Therefore, political institutions (parties, strategists and leaders) need to collect, monitor and analyze a large amount of data on social networks which is mostly unstructured and is not blatantly political to extract relevant for them political information. To this end, automatic and semi-automatic techniques of text analysis are to be applied.

The challenge of automatically identifying distinct political topics, sentiments, influential users and groups gathering and aggregating them is being tackled by social media analytics (Zeng et al., 2010; Agrawal et al., 2011; Leskovec, 2011; Nagarajan et al., 2011). Stieglitz and Dang-Xuan (2013) propose a framework for systematic social network analysis for in political context. There are two goals of monitoring and analysis in their framework: self-reputation management and general monitoring. For each of these goals they propose the following approaches: identification of emergent political topics/issues/trends, sentiment/opinion recognition in discussions of topics and candidates, and structural to identify relevant leaders and communities with political influence on social networks. In this context, the current study contributes a new approach to the above framework as part of the general monitoring goal: automatic identification of users with a certain political orientation. Previous studies (Chan, 2011; Brumbaugh et al., 2002) have shown that targeted ads are more effective than non-targeted ads leading to substantial saving in the advertising budget. Therefore, determining the political orientation of users on social network sites is crucial for targeting political messages, especially during election campaigns. Hence, in this study, we explore the use of automated text categorization methods to determine the political orientation of Facebook users by their texts.

Automatic text categorization have been widely employed for a variety of author profiling tasks (Argamon et al., 2009), typically for the purpose of identifying authors’ demographic characteristics such as age, gender or native language. However, the application of these methods for the determination of political orientation is especially challenging. First of all, unlike demographic characteristics, an individual’s political orientation may vary over time and is often complex and thus not easily captured by a single simplistic label such as left or right. Furthermore, conventions of public expression often dictate that political views are stated in a subtle manner, if at all. A number of papers (Laver et al., 2003; Efron, 2004; Mullen and Malouf, 2006;
Hassanali and Hatzivassiloglou, 2010) have considered the automatic identification of political tendency for overtly political documents, such as political blogs. However, the problem is even more difficult in contexts of personal pages on social networks, where the discussion is mostly not intended to be political at all. Recent studies on political identification of social network users (Kosinski et al., 2013; Rao et al., 2010; Conover et al., 2011) applied automatic text categorization methods based on training with labeled data where political orientation for each text was pre-supplied. As a result of the training process the optimal classifier is learned, which then can be applied to classify new texts. However, in practice, such labeled data can be easily induced for overtly political texts, but typically there is no available politically labeled data for apolitical or semi-political texts (such as Facebook profiles of individual users). In addition, human-labeled data might be biased and rather subjective.

To overcome the above challenges, we propose and implement a new methodology for automatic prediction of political preferences of text with unknown political orientation with no need in labeled training examples. The underlying idea is training an automatic classifier on an “easy case” texts, for which labeled data are easily accessible: political parties’ Facebook pages, and applying it to classify personal Facebook pages, where such labeled data are hard to get. This is in contrast to the previous work where the classifier is trained and applied on the same corpus. Our approach is based on the assumption that similar features (terms) are discriminative of political tendencies in diverse types of texts. From the theoretical perspective, the validity of this assumption might indicate how well political parties’ self-presentation on Facebook is suited to their potential voters.

Thus, the main research questions addressed in this research are:

RQ1. Whether it is possible to effectively employ a classifier trained on features from an overtly political corpus (“easy classification and labeling case”) to automatically determine the political orientation of texts from the corpus of Facebook user pages (“hard case”) which most likely do not contain terms with explicit political content?

RQ2. Whether it is possible to recognize indefinite political position of the users, such as centrist voters and predict voters likely to switch allegiances (“swing voters”) by their writing style in texts such as Facebook personal pages?

The paper’s outline is as follows. In the next section, we describe related work. Then we present the corpora used in the paper and follow that with an outline of our methodology and experiments. The two sections after that include detailed presentation of our results and some conclusions.

2. Related work

Numerous studies have been performed in the area of automatic recognition of an author’s demographic profile. Text categorization methods have been used to identify an anonymous author’s gender (Argamon et al., 2003; Burger et al., 2011; Filippova, 2012), age (Koppel et al., 2006), native language (Koppel et al., 2005) and personality (Pennebaker et al., 2003). It has been shown that such demographic profiling can also be done on personal Facebook pages (Otterbacher, 2010; Popescu and Grefenstette, 2010; Gosling et al., 2011).

A survey of automated demographic profiling is presented in (Argamon et al., 2009).

Several studies have considered ways in which additional available information can be used to enhance purely text-based features to improve demographic profiling.
Thus, for example, it has been found that text-based gender classification of authors can be improved using additional information such as names (Burger et al., 2011) and social network topology (Filippova, 2012). Similar such methods have been used to improve automated classification according to location and educational level (Rao et al., 2010; Gillick, 2010) and age (Rosenthal and McKeown, 2011). Others have considered patterns of social network activity to determine personality type (Bachrach et al., 2012; Gosling et al., 2011; Ross et al., 2009).

A number of studies have considered the problem of automatically determining an author’s political preference (left, right). For example, Laver et al. (2003), Efron (2004), Mullen and Malouf (2006), Hassanali and Hatzivassiloglou (2010) use text categorization methods for determining the political orientation of political blogs. Grefenstette et al. (2004) explore the same problem for websites by considering the aggregate of documents found on a site. Yu et al. (2008) classify US congressional speeches according to party affiliation. In general, these studies deal with overtly political texts, in which labeled texts are relatively easy to find. In this study, we wish to classify texts in genres for which political opinion of the authors is unknown and examples labeled according to political tendency are hard to come by, such as private social network profiles.

Numerous recent studies have explored the role of social networks in shaping political communication around the world (e.g. Aday et al., 2010; Benkler, 2006; Bennett, 2003; Farrell and Dreznier, 2008; Sunstein, 2002; Tumasjan et al., 2010). The last US presidential campaigns have shown that social networks have become increasingly important for political communication and persuasion (Wattal et al., 2010). Another example is the “Twitter revolutions” in totalitarian countries. For example, Gaffney (2010) found that Twitter helped protesters during the 2009 Iran elections by tracking the use of the #IranElection hashtag. Larsson and Moe (2011) show that Twitter was used during the 2010 Swedish general election for disseminating political contents and not for political dialog. Recently, Rao et al. (2010) and Conover et al. (2011) extended the work in (Grefenstette et al., 2004) to identify political orientation of Twitter accounts that were not necessarily blatantly political. Kosinski et al. (2013) have shown that political views, among other personal characteristics, can be predicted from a user’s “likes” on Facebook.

Jungherr (2015) discussed the data on Twitter as a potential source for analysis of political information and election result prediction. The author examined the relationship between the political parties’ mentions in Twitter messages and their actual vote shares in the elections. He analyzed messages which included political hashtags. As opposed to the previous studies (Tumasjan et al., 2010; Conover et al., 2011) he found that the Twitter metrics seem to be more suitable to be used as a mirror of political interests and attention and to analyze political controversies than as a forecast medium for political behavior in elections. Jungherr (2015) concludes that “it is hard if it all possible to use digital trace data to draw valid inferences on the political opinions and voting intentions of Twitter users or the public at large” (p. 6).

However, all the above studies were based on human-labeled training data and explore an easier case where the automatic classifier is trained (learned) and then tested (applied) on the same corpus. Such manually labeled training data are typically unavailable in real life situations. As opposed, in this study, we consider the possibility of learning the classifier on the manifestly political corpus and then applying it to the apolitical or semi-political corpus, thus sparing the need for human-labeled training data for the latter corpus.
3. Methods and materials

3.1 Corpora

The texts we consider here are Hebrew texts written by Israelis. This presents a number of challenges and opportunities specific to this linguistic and political context. Since we use only lexical features, the morphological quirks of Hebrew will not present any special challenges. However, Israel’s purely proportional single-region parliamentary election system presents one interesting opportunity. Unlike winner-take-all regional elections, which typically result in only two major parties, there are many medium-sized parties in Israel. While each of these parties can rather easily be identified as left, right or center, the parties differ widely in terms of the demographic group to which they appeal. In particular, because there are a number of self-declared centrist parties, we will consider two-class (right/left) experiments, as well as three-class (right/center/left) experiments. We will also explore which self-identified centrist voters are closer to the right and which are closer to the left. In this study, we consider two Hebrew corpora, one of which is explicitly political and the other is not. As in most previous work, both corpora were collected in the pre-election period—a few weeks before the Parliament elections in the end of January 2013:

(1) Posts on the Facebook pages of nine major Israeli political parties. Each party is labeled as left/center/right, with three parties assigned to each category. In particular, for the right wing we considered the pages of “Habayit Hayehudi,” “Likud-Beiteinu” and “Otzma le-Israel.” For the center wing we considered the pages of Yesh Atid, Kadima, and Hatnua. Finally, for the left wing we considered pages of “Haavoda,” “Meretz” and “Hadash.” While the assignments to categories are uncontroversial, the parties in each category are diverse in terms of their demographic appeal. The corpus consists of 646 posts, including over 550,000 words (229 posts for the right wing, 208 posts for the left wing and 209 posts for the center wing). For our purposes, chronologically consecutive posts are concatenated until they exceed 1,000 words in aggregate.

(2) Personal Facebook pages of 450 random Israeli individuals, divided evenly among those whom self-identified as right wing, left wing or centrist. Each page contained approximately 1,200-1,700 words. These pages were collected by distribution of the viral application especially implemented for our research which started from the authors’ personal friends. To obtain users’ self-identification we asked them to fill in a short online questionnaire on their political orientation. Only the first 150 profiles for a given political wing were stored to create a balanced corpus with even distribution of profiles among the political wings. Users were also asked for permission to download their personal Facebook page data (as anonymous texts for research use only). Each individual’s text included all status updates, as well as the titles of “liked” pages. These pages were mostly not political, only few of them included posts that refer to politics.

3.2 Experimental setup

We begin by introducing the basic concepts from text categorization that we use here. First, each text in a set of labeled example texts is represented as a numerical vector reflecting the frequencies in the text of each feature in a specified feature set. Some machine learning algorithm is then used to learn a classifier that best distinguishes among training examples in different classes. These classifiers can then be used to classify new texts. The effectiveness of this method can be measured by applying a
learned classifier to labeled test texts for which the correct answer is given. A related method is that of $k$-fold cross-validation. We divide the training set into $k$ roughly equal parts, train on $k-1$ parts and test on the holdout set, repeating this $k$ times with a different part held out each time.

In this context, we perform all the following experiments:

- For each of our corpora, we perform tenfold cross-validation experiments to determine the accuracy with which we can train a political preference (right, left) classifier for a given corpus.

- We train a classifier on training data in political corpora (party Facebook pages) and check its effectiveness for classifying personal Facebook pages.

- We use learned two-class (left/right) political preference classifiers to determine whether self-identified centrist voters are closer to the left or to the right. We also learned three-class (left/center/right) classifier to determine the accuracy with which we can distinguish between left-, center- and right-leaning texts in each corpus in separate.

- Finally, we perform tenfold cross-validation experiments to determine the accuracy with which we can identify “swing” voters who intend to switch allegiances (right to left or vice versa) in upcoming elections.

In all such experiments each text is represented as a numerical vector (histogram) of features encoding the frequency in the text. For each experiment we only use features that appear in the relevant corpus at least three times. This selection criterion resulted in feature vectors of approximately 10,000 individual words and 2,000 word bigrams for each corpus. Except in the case of $k$-fold cross-validation experiments, we applied Student’s $t$-test to select only those features for which frequency differences between classes are significant on the training examples with significance at $p < 0.05$. We use sequential minimal optimization (SMO) (Platt, 1998), an efficient implementation of support vector machine (Joachims, 2002), a state-of-the-art machine learning algorithm. Other machine classification algorithms implemented in the Weka system (Hall et al., 2009), such as multi-layered perceptron, Bayesian multinomial regression and Winnow were used as well in our preliminary experiments but they yielded close but slightly worse results. At the training phase the algorithm uses the above feature vectors of the training set to learn an optimal classifier which finds the accurate boundary between texts with different political orientation. Then, the learnt classifier is applied to classify new texts.

4. Results

4.1 Individual corpora

In our first experiment, we consider for each of our corpora individually the accuracy with which we can classify out-of-sample examples as having left or right political orientation. Thus, in these experiments for each corpus in separate we learned the classifier on a subset of texts and then applied it to classify the rest of the texts from the same corpus. This was repeated ten times for different subsets of texts (as part of tenfold cross-validation). Ground truth in each case is as described in Section 3 above. As noted, our feature set consists of all word unigrams and bigrams that appear in the corpus at least three times and we use SMO as our learning method. Results of tenfold cross-validation experiments on each of the corpora are shown in Figure 1. As can be seen, result accuracy in each case exceeds 90 percent.
4.2 Learning across corpora

Next, we apply the classifier trained on easily-identified explicitly political texts to predict the political slant of the non-necessarily political corpus. As above, our feature set consists of all word unigrams and bigrams that appear in the training corpus at least three times. In this case, we filter the feature set by considering a feature only if its difference in frequency across classes (in the training set) is significant at $p < 0.05$.

Using the Facebook party pages as the training set yields the accuracy of 82.0 percent for the personal Facebook pages. This result can be explained by the relatively high resemblance between the most characteristic features in both the personal users’ and parties’ Facebook pages. In both corpora, the right is characterized by references to religion, patriotism, positive attitudes as well as first-person pronouns, while the left is characterized by references to social protest, rights, minorities and third-person pronouns.

The significance of this result is that it suggests that the use of easily-assembled inherently-tagged data like party Facebook pages is sufficient for classifying user profile pages. This also spares us the need to gather and manually label personal pages as training examples. Note that in the cross-corpora learning experiment we used the political self-identification of Facebook users only at the test phase (and not for training the classifier) in order to evaluate the accuracy of our approach.

4.3 Distinguishing features

Consideration of the main distinguishing features for each experiment (as measured by Student’s $t$-test) yields insight into why successful classification is possible for each corpus. We now consider a more detailed comparison of the key features per corpus. All mentions of “significant” differences are at $p > 0.05$.

Party Facebook pages: right-wing party posts make significantly more frequent mention of religious concepts (Rabbi, Torah, God, Sabbath, Amen) and positive attitudes (love, beloved, good luck, be strong), while left-wing party posts make significantly more frequent mention of particular politically-loaded terms (rights, social protest, Palestinians, two states, refugees) and third person (e.g. he, they) and female pronouns (e.g. she, her).

Personal Facebook pages: all the differences found in the first corpus are found even more strongly in the personal pages. Self-identified right-wingers use significantly more terms reflecting positive attitudes (love, good luck, happy, good week, good news, smile, be blessed) and religious terms (God, Sabbath, Holy), while self-identified left-wingers use all the politically-loaded terms associated with the left in the other corpus. Left-wingers also make many more references to university life (education,
university), possibly reflecting demographic differences. In addition, the right-wingers use more first-person pronouns (e.g. I, we, us), while the left-wingers use more third-person pronouns (e.g. he, she, her, they, them).

In summary, our feature analysis reveals that many predominant distinguishing features were not solely from the political domain, such as pronouns, positive attitudes, religion, student life and social aspects. Interestingly, these similar types of non-political features were appeared in both corpora, which allowed for quite accurate classification of personal Facebook pages.

4.4 Centrist voters
So far we have classified texts to one of two classes: left and right. Now we wish to characterize centrist voters for every corpus in isolation. First, we wish to determine the extent to which we can identify an author as being left, right or center. Second, we wish to determine, for each of the respective corpora in separate, to which of the two political wings self-identified centrists are more similar.

For the three-class classification experiment, our feature set consists of all word unigrams and bigrams that appear in the corpus at least three times. The classifiers in these experiments were learned and tested on the same corpus. Results of tenfold cross-validation experiments on each of the corpora are shown in Figure 2.

As can be seen, personal Facebook pages prove to be challenging to classify. Examining the confusion matrix in Table I indicates that right-wing and left-wing users are rarely confused, as we found in the two-class problem considered above, but center is more frequently confused with right.

We now revert to the two-class (left/right) classifiers for the respective corpora and use them to determine what percentage of centrists is assigned to each class. Results are shown in Figure 3. It can be observed that for personal Facebook pages the centrist texts are distributed quite equally between right and left, while just over 60 percent of posts of centrist parties on Facebook were classified as right. The centrist political parties tend to post right-leaning messages probably as an attempt to attract more right-wing electorate, or to increase their popularity since the government was led by the right wing at that period of time.

4.5 Swing voters
Finally, we consider using personal Facebook pages to identify potential “swing votes.” To this end, as part of their political self-identification, participating Facebook users
were asked for which party they had voted in the previous elections and for which party they intended to vote in (at the time) upcoming elections. Subjects who switched allegiance among left/right/center parties were marked as “swing voters.” Identifying such voters is critical for political campaigns.

Our corpus includes 63 individuals who report that they intended to switch from a non-right-wing party to the right-wing party, and 52 users who report that they intended to switch from a non-left-wing party to a left-wing party. In each case, we randomly choose an equal number of non-switchers (who are plentiful). As usual, we use as our feature set all unigrams and bigrams that appeared at least three times in the corpus. In tenfold cross-validation experiments, we obtain accuracy of 72 percent for classifying Facebook profiles of non-right-wing voters as switching to the right or not. For non-left-wing voters, we obtain accuracy of 77 percent for distinguishing switchers from non-switchers. Interestingly, in both cases we find that switchers use significantly more terms related to education (education, university, research, school), and significantly fewer terms reflecting positive attitudes than the non-switchers. These results suggest that swing voters are best identified not by political terms characteristic of right and left, but rather by references to education and the relative absence of references to positive attitudes.

5. Discussion and conclusions
Profiling according to political orientation is an important element of targeted political campaigns. Previous studies have focussed each on a specific corpus and shown that useful classifiers can be learned for it. The main contribution of this research is that we proposed and empirically evaluated a new cross-corpora approach for automatic prediction of political tendency of Facebook users. Our findings show that the same classifier can be effectively used to classify texts from different corpora, and particularly for the “hard case” corpus with no overtly political orientation, such as Facebook users’ personal pages. This is due to the fact that in both corpora (the party

<table>
<thead>
<tr>
<th>Table I. Confusion matrix for three-class Facebook profiles classification experiment</th>
<th>Left</th>
<th>Center</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>132</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Center</td>
<td>22</td>
<td>91</td>
<td>37</td>
</tr>
<tr>
<td>Right</td>
<td>6</td>
<td>24</td>
<td>120</td>
</tr>
</tbody>
</table>

![The number of centrist users’ pages classified as right for the two individual corpora](image)

![The number of centrist users’ pages classified as right for the two individual corpora](image)
Facebook corpus and the personal Facebook corpus), the right is characterized by references to religion and positive attitudes, as well as first-person pronouns, while the left is characterized by references to education, social rights and third-person pronouns. Such a use of pronouns might be partially explained by the fact that at the time of collecting the corpora the government was led by the right-wing parties, while the left-wing parties (and their electorate) were protesting and criticizing its decisions from the opposition.

These findings have important theoretical implication. That is, in general, communication of Israeli political parties is aligned with the individual user’s communication on Facebook, as they use a language which is similar to the language of their potential voters. Thus, we conclude that from the linguistic perspective Israeli parties do a good job in self-representation on Facebook.

In particular, we have found the following:

1. Political views of the authors of personal Facebook pages may be automatically recognized by their statistical properties with very high accuracy (over 90 percent), when the classifier is trained on other personal Facebook pages from the same corpus.

2. Classifiers trained on political genre can be used to effectively classify non-political (or at least not necessarily political) personal Facebook pages. This reduces the need for manually annotating personal Facebook pages for training data.

3. Similar political and non-political terms are discriminative of the author’s political bias in different corpora.

4. Centrist users’ texts split roughly equally between right and left, but centrist parties texts are leaned to the right.

5. Similar non-political terms prove to be useful for identifying potential “swing voters” (independently from the direction of change in their views) with moderate accuracy. This can be helpful for efficient use of campaign resources.

Previous studies have shown that targeted ads are more effective than non-targeted ads leading to substantial saving in the advertising budget. Therefore, our approach for automatic determining the political orientation of users on social network sites might be beneficial for targeting political messages, especially during election campaigns.

We note that this research has some limitations. It was based solely on Israeli Facebook users and party pages, and the data were collected in the particular period of time (the post-election period) when the government was led by the right wing. Therefore, in future work we intend to employ the proposed methodology to explore whether similar phenomena occur in different political scenario and also for different political systems in other countries.

References


Further reading


About the authors

Dr Esther David received her PhD Degree in Computer Science from the Bar-Ilan University, Israel in 2003. Dr David has worked for three years (2003-2006) as a Senior Researcher at the Southampton University under the supervision of Professor Nicholas Jennings in the UK. Since 2006 she has been a Senior Lecturer of the Computer Science Department at Ashkelon Academic College. Her research is primarily rooted in electronic commerce, mechanism design, game theory and auction theory. Her recent research includes also machine learning applications as building intelligent tutoring system for enhancing abilities in the domain of reading comprehension; and author profiling for political tendency. In the last six years she has been one of the organizers of the Agents Mediated Electronic Commerce Conference (AMEC) which is jointly held with the AAMAS Conference (one of the top AI and agent conferences).

Dr Maayan Zhitomirsky-Geffet received her PhD in Computer Science from the Hebrew University in Jerusalem, Israel. In her PhD thesis she explored automatic methods for ontological relationship recognition from large corpora and from the web. Currently, Dr Zhitomirsky-Geffet is an Assistant Professor in the Department of Information Science in the Bar-Ilan University and her main research fields include the semantic web, internet research, social networks and web-based information retrieval. Dr Maayan Zhitomirsky-Geffet is the corresponding author and can be contacted at: maayan.geffet@gmail.com

Professor Moshe Koppel conducts research on a variety of machine learning applications including text categorization, image processing, speaker recognition and automated game playing. He is best known for his contributions to the branch of text categorization concerned with authorship attribution. More recently, he has begun researching fundamental problems in social choice theory.

Hodaya Uzan is an MA Student at the Department of Computer Science at the Bar-Ilan University in Israel. Her main areas of interest include: automatic text categorization, internet research and social networks.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
This article has been cited by: