Using demographics in predicting election results with Twitter

Eric Sanders, Michelle de Gier, and Antal van den Bosch

CLS/CLST, Radboud University, Nijmegen, the Netherlands,
e.sanders@let.ru.nl

Abstract. The results of two Dutch elections are predicted by counting political party mentions from tweets. In an attempt to improve the predictions, gender and age information from the Twitter users is automatically derived and used to adapt the party counts to the demographics in the election turnout. The prediction improves only slightly in one of the elections where the correlation between election outcome and Twitter-based prediction was relatively lower to begin with (0.86 versus 0.97). The relatively inaccurate estimation of Twitter user age may hinder a larger improvement.

Keywords: Twitter, political election prediction, demographics

1 Introduction

1.1 Twitter as predictor of political election outcome

The social media web platform Twitter[1] has been used as a basis for predicting many different types of events and outcomes of processes, among which the outcome of political elections. Traditionally the result of elections are predicted by questioning a representative part of society, which is costly and time consuming. If these polls could be replaced by predictions based on tweets, that are available anyway, costs could be largely reduced.

Tumasjan et al were one of the first to report on election predictions based on tweets [17]. Although some flaws in their study were exposed [8], other researchers have used their procedure of predicting by simply counting tweets that mention political parties, with varying success. There are reports on this procedure for elections in a number of countries [7]. For the Netherlands, this method has been conducted by Tjong Kim Sang and Bos [14] and by Sanders and Van den Bosch [13]. In this latter study we compared the mentions of political parties in tweets of ten days before the Dutch parliamentary elections of 2012 with the polls and the election results. We found a very high correlation with both the election results (0.95) and the polls (0.96), although the polls still outperform the prediction based on tweet counts: the correlation between election results and polls was 0.98.

[1] http://twitter.com

Gayo-Avello discusses a number of aspects that are usually not taken into account in the election predictions based on tweets [3], among which demographics. Barber and Rivero show the inbalance in distribution of gender in tweets about the 2011 Spanish legislative elections and the 2012 US presidential elections [1]. Tjong Kim Sang and Bos attempted to incorporate demographics in their study [14], but they did not have demographic data of Twitter users. Wang et al forecasted the 2012 USA presidential elections based on highly non-representative polls conducted on the Xbox using multilevel regression and post-stratification [19]. In this paper we report on a case study on two recent political elections in the Netherlands in which we take demographic distributions of Twitter users into account when predicting election results from the Twitter stream. The demographics are estimated using machine learning methods [11, 12] and are therefore error-prone.

For using demographics[2] in the prediction of the election results based on tweets to be meaningful, there should be a difference in voting behavior of different demographic groups. Furthermore, there should be a difference in demographics between voters and Twitter users for the demographic data to have a correcting effect on the basic count-based estimates. These two conditions are discussed in Section 2. In Section 3 we introduce our data, and in Section 4 we show how we obtain demographics in the voter and tweeter populations. In Section 5 we explain how we adapt our predictions to the demographic distributions and show the results. We finish with conclusions in Section 6 and a discussion in Section 7.

1.2 Elections used in this study

Two Dutch elections have been investigated in this study: 1) The national parliamentary elections of 12 September 2012, which are the most important elections in the Netherlands; 2) The provincial elections of 18 March 2015, where the parliaments of the twelve Dutch provinces are elected. Because the Dutch national senate is elected in these elections indirectly (as the provincial electees elect the senate members), they were also important at the national level.

2 Demographic bias

2.1 Voting behaviour for different gender and age groups

Different voting behaviour of different age and gender groups is assumed as a boundary condition for this study. There is ample evidence in the literature though that different voting behaviour across demographic groups is real. For instance, Inglehart and Norris show in their study in 19 countries that men and women have different political orientations [6]. Also, men show more interest

[2] In this study we only studied gender and age because these are the two most basic demographic data and because these are the only two that are automatically retrievable to a certain extent.

in politics than women [18]. Webster et al. show that people tend to vote for people close to their age [20] while Goerres argues that people tend to vote more when they are older [5]. Furthermore people in Great Britain tend to vote more conservative when they age [15].

2.2 Demographic distribution on Twitter

Users on social media in general and Twitter specifically do not mirror the people in society in all demographic aspects [9]. More specifically, people are younger and there are slightly more men on Twitter than women [10]. Duggan and Brenner did surveys for different social media and found out that Twitter is especially appealing for adults from 18 to 29 [2].

3 Political tweets, polls and election results

In this study we compare counts of political parties mentioned in tweets with election results and polls. In this section we explain how we acquired the data.

The tweets that were retrieved for this study were taken from TwiNL, a large collection of about 40% of all Dutch tweets [16]. TwiNL filters Dutch tweets on typical Dutch words that rarely appear in other languages, and follows a dynamically updated group of Dutch Twitter users that post often. Political tweets were selected from TwiNL by looking for party names in the tweets using regular expressions of the 11 political parties in parliament (see [13] for details). A tweet that matches one of the political party regular expressions is called a "political tweet" throughout this paper. The person who posted the tweet is called a "political tweeter". For both elections political tweets were gathered from ten days preceding (and including) election day. In total we collected 159,826 political tweets for 2012 and 183,602 for 2015.

The polling data was retrieved from allepeilingen.com, which is a website that keeps track of all polls from all well known polling institutes in the Netherlands since 2000. The election results were retrieved from kiesraad.nl[3], the official website that presents the election results in detail.

4 Demographics in elections and in tweets

4.1 Demographics

If the demographics of the political tweeters would be the same as that of the electorate that turns up for voting, there would be no need for adaptation. Hence, first we need to compare whether (as the literature suggests) and how these demographics differ. In the next subsections we explain how we retrieved and validated the demographic data for the elections of 2012 and 2015 and for the political tweeters.

[3] http://www.verkiezingsuitslagen.nl/Na1918/Verkiezingsuitslagen.aspx

4.2 Voter demographics

Information about the age and gender distribution of the voters in the 2012 and 2015 elections were obtained from TNS-Nipo[4]. TNS-Nipo is a market research company well-known for election polling. They used a large user panel that was asked for their age, gender and what they voted (among many other things). TNS-Nipo divided their data in three age groups: 18-35, 35-55 and 55+. These age groups are used throughout this paper. Table 1 lists the distributions over age groups and gender for voters in 2012 and 2015.

Table 1. Demographics of elections of 2012 and 2015 according to TNS-Nipo in percentages

	2012			2015		
Age group	Men	Women	Total	Men	Women	Total
18-34	6.1	5.6	11.7	6.2	6.5	12.7
35-54	15.7	17.5	33.2	13.6	15.3	28.9
55+	24.8	30.3	55.1	26.8	31.6	58.4
Total	46.6	53.4	100	46.6	53.4	100

The table shows that the gender balance is about even –with slightly more women than men voting– and over half of the voters is older than 55 years. The demographic distribution of the voters is fairly constant among the two elections. For most of the gender–age groups the differences between the two elections is less than two percent.

4.3 Twitter user demographics

To automatically estimate the gender and age of the political tweeters, an offline version of TweetGenie [11, 12] was used. TweetGenie is a machine-learning system that uses the language in the aggregated set of Dutch tweets posted by a Twitter user to identify the age and gender of the user, based on a training set of users with known ages and gender. TweetGenie was used "as is"; it was not adapted or retrained for the purpose of these experiments.

Table 2 lists the age and gender distributions of users who posted political tweets in the two elections as produced by TweetGenie. Because the age is derived from the most recent tweets of the user in 2016, while the elections were in 2012 and 2015, respectively four and one year are deducted from the age estimated by TweetGenie to arrive at the estimated ages during the elections.

The table shows that three in four of the political tweeters is male, which is in sharp contrast with the election turnout in which male and female voters are almost at a par. Furthermore the political tweeters are on average much younger than the real voters. Notable is the low number of 55+ political tweeters

[4] www.tns-nipo.com

Table 2. Demographics of political tweeters in 2012 and 2015 according to TweetGenie's estimates based on 2016 user accounts, in percentages

	2012			2015		
Agegroup	Men	Women	Total	Men	Women	Total
18-34	51.3	17.6	68.9	28.7	8.1	36.8
35-54	21.3	8.5	29.8	45.5	15.8	61.3
55+	1.0	0.4	1.4	1.4	0.5	1.9
Total	73.6	26.5	100	74.2	25.8	100

according to TweetGenie. It is a known systematic weakness of TweetGenie that it generates too low age estimates for older people [12], so the real number of tweeters in this age category is probably higher. More generally, the statistics in the table highlight that there is a difference between voters and political tweeters and that correcting Twitter-based counts on the basis of differences in the two demographics is in principle a good idea.

4.4 Quality control by human annotators

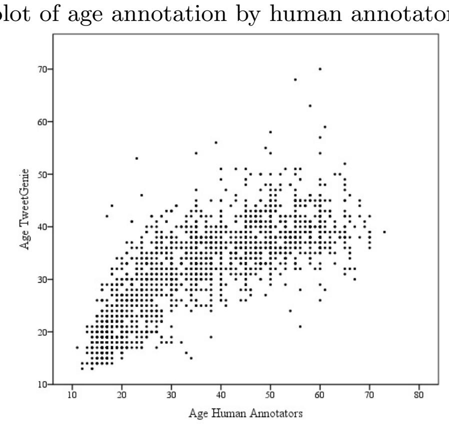
We use the data provided by TweetGenie as the demographics for the political tweeters, but we do not know how accurate the estimation of gender and age by TweetGenie is. In the absence of a golden standard (a set of Twitter users of which we have certainty on their gender and age) we compared the TweetGenie estimates to demographics determined by human annotators. Two annotators guessed the gender and age of 3,000 political tweeters. This was done in a webtool, created for this purpose, that showed the profile from 2012 or 2015 (depending from what year the political tweets was) and a link to the current profile. Sometimes there is age information in the profile description, but most demographics clues were derived from the profile picture. Note that the profile picture in Twitter is often not up to date so that the guessed age based on it might be a several years too young. The demographics of a number of tweeters could not be established by either the annotator or TweetGenie, and about 10% of the tweeters was annotated as non-human (but as an institution, newspaper, etc.). Leaving these out, we ended up with a set of 1,815 Twitter accounts for which the demographics were compared. Table 3 shows the age and gender distributions for the 2012 and 2015 as estimated by the human annotators and TweetGenie. Figure 1 shows a scatter plot with the age as indicated by humans against that of TweetGenie.

For gender, the distributions as annotated by the human annotators and TweetGenie do not differ very much for both elections, which indicates that the gender distribution as produced by TweetGenie is reliable. For age, on the other hand, the differences are larger. TweetGenie assigns some of the youngest group to an older group, and assigns some of the oldest group to a younger group. This same pattern is visible in the scatter plot in Figure 1.

Table 3. Gender and age distributions as annotated by humans and TweetGenie for 2012 and 2015 in percentages

	Human	TweetGenie
Gender		
Men	70.6	74.4
Women	29.4	25.6
Age		
18-	18.7	10.3
18-34	33.4	36.7
35-54	35.8	51.6
55+	14.1	1.3

Fig. 1. Scatterplot of age annotation by human annotators and TweetGenie



To verify that human annotators can guess the age and gender of a tweeter better than TweetGenie, we computed intra-annotator agreement between the two human annotators and TweetGenie. A set of 100 accounts were annotated by both annotators and TweetGenie. After removal of unannotated accounts we ended up with a comparison set of 72 accounts for which gender was annotated and 61 for which age was annotated by the two annotators and TweetGenie. Table 4 lists the Cohen's kappa values for the agreement between the annotators and TweetGenie for age and gender.

Cohen's kappa between the two human annotators is near-perfect for gender and substantial for age, which makes the annotation of the humans reliable for comparison. The kappa between TweetGenie and the humans is moderate for gender, and only slight for age. From this we conclude that the age annotations of the human annotators is more reliable than those of TweetGenie.

Table 4. Cohen's Kappa between two human annotators and TweetGenie for gender and age

	gender		age	
	annotator2	TweetGenie	annotator2	TweetGenie
annotator1	0.97	0.52	0.64	0.09
annotator2	-	0.56	-	0.14

5 Prediction of election results

5.1 Without adaptation

Following Sanders and Van den Bosch [13], the outcome of the elections are predicted by counting how often a political party is mentioned in a tweet in the ten days before and including election day. These counts are compared with the polls of one day before the elections and the outcome of the elections.

5.2 Adaptation of tweet counts based on demographics

Using the distributions in Table 1 and Table 2 we can use post-stratification [4] to compute weights with which we adapt our tweet counts for the prediction of election results, so they reflect the correct demographic distribution. Table 5 and Table 6 present the results with and without adaptation for the 2012 and 2015 elections, respectively. The quality of the prediction is represented in Pearson's correlation over all political parties and the Mean Absolute Error (MAE); the mean absolute difference in percentages.

Table 5 shows that the prediction of the election outcome is very good for 2012. The correlation of 0.97 is only slightly worse than that between the polls and the election outcome, 0.98. The MAE shows a clear difference, though. Importantly, correcting the demographics does not improve the prediction; the MAE ends up slightly worse. For 2015, the prediction based on tweets is clearly worse than the polls. Yet, here the adaptation for gender and age does improve the predictive power of the tweets. The correlation is slightly improved from 0.84 to 0.86, and the MAE is decreased by 21% from 2.9 to 2.3.

6 Conclusions

In this study we attempted to improve the prediction of election results on the basis of counts of mentions of political parties in tweets by using demographic (age and gender) information. We did this for two Dutch national elections. For the parliamentary elections of 2012 the prediction without demographic adaptation was precise to begin with. Adding demographic information did not result in an improvement. For the provincial/senate elections of 2015 the prediction was not as good as for the 2012 election. Here, adding demographic information did improve the prediction. The Pearson's correlation with the election results improved, and the MAE was reduced.

Table 5. Prediction with Twitter, polls and outcome of elections of 2012 with and without demographic adaptation in percentages. At the bottom the correlation (with 95% confidence interval) and mean absolute error with the election results.

Political party	Election result	Polls	Twitter	Twitter adapted
VVD	26.8	23.4	22.7	23.1
PVDA	25.1	22.7	20.1	21.9
SP	9.8	14.3	11.2	11.8
PVV	10.2	11.5	10.2	7.7
CDA	8.6	8.1	8.1	8.7
D66	8.1	7.9	9.0	7.5
GL	2.4	2.7	8.1	8.1
CU	3.2	4.1	3.6	4.6
SGP	2.1	1.8	2.6	1.9
PVDD	2.0	1.6	3.0	3.2
50PLUS	1.9	2.1	1.4	1.6
Correlation elections		0.98	0.97	0.97
		(0.94 - 1.0)	(0.79 - 1.0)	(0.70 - 1.0)
MAE elections		1.3	1.8	1.9

7 Discussion

In a case study in which we compare two national Dutch elections we find a small improvement of adding demographic information in the predictive power of Twitter for election results when the prediction without the demographic information is not very high to begin with. Obviously, correcting for demographics is only one of the possible and perhaps necessary adaptations that could be done to improve the predictive power of tweets, as suggested by Gayo-Avello. In this respect, the results we found are encouraging. Further research is needed to find out whether adapting for demographics really helps in predicting election results based on tweets.

The first avenue for improvement appears to lie in achieving a better estimate of the age distribution of the tweeters. TweetGenie appeared to guess the gender of political tweeters well, but for age the agreement with human annotators was somewhat low, especially for the oldest age category. TweetGenie may be retrained on more data, or its predictions could be calibrated itself to correct its biases.

In this study we only took age and gender of Twitter users into account, while neglecting education, social status, ethnicity etc. Including these demographic factors might increase the predictions, but they will likely be harder to automatically retrieve with sufficient accuracy.

For correcting the Twitter count we used the actual voter turnout demographic distribution of the election we tried to predict the results from. These numbers are not known in advance. However, the distributions for the 2012 and 2015 were fairly similar, so we presume that using the distribution of a recent election will be a good estimate for a new election.

Table 6. Prediction with Twitter, polls and outcome of elections of 2015 with and without demographic adaptation in percentages. At the bottom the correlation (with 95% confidence interval) and mean absolute error with the election results.

Political party	Election result	Polls	Twitter	Twitter adapted
VVD	16.7	16.7	26.1	25.3
PVDA	10.6	10.1	14.1	13.9
SP	12.2	13.4	7.0	8.8
PVV	12.3	15.5	11.2	8.8
CDA	15.4	12.4	9.9	11.6
D66	13.0	15.5	15.6	13.3
GL	5.6	4.1	5.7	6.1
CU	4.2	4.0	3.4	4.3
SGP	2.9	2.6	2.2	2.1
PVDD	3.6	2.9	3.3	3.6
50PUS	3.5	2.9	1.2	2.3
Correlation elections		0.96	0.84	0.86
		(0.89 - 0.99)	(0.68 - 0.99)	(0.75 - 0.98)
MAE elections		1.3	2.9	2.3

8 Acknowledgments

The authors would like to thank Dong Nguyen who provided the TweetGenie data, TNS-Nipo who provided the demographic data of the election turnout, Ruut Brandsma of allepeilingen.com for the polling information and Eline Pilaet, who did part of the demographic annotations of the political tweeters.

References

1. Barberá, P., Rivero, G.: Understanding the political representativeness of twitter users. Social Science Computer Review p. 0894439314558836 (2014)
2. Duggan, M., Brenner, J.: The demographics of social media users, 2012, vol. 14. Pew Research Center's Internet & American Life Project Washington, DC (2013)
3. Gayo-Avello, D.: A meta-analysis of state-of-the-art electoral prediction from Twitter data. Social Science Computer Review (2013)
4. Gelman, A.: Struggles with survey weighting and regression modeling. Statistical Science pp. 153–164 (2007)
5. Goerres, A.: Why are older people more likely to vote? the impact of ageing on electoral turnout in europe. The British Journal of Politics and International Relations 9(1), 90–121 (2007)
6. Inglehart, R., Norris, P.: The developmental theory of the gender gap: Womens and mens voting behavior in global perspective. International Political Science Review 21(4), 441–463 (2000)
7. Jungherr, A.: Predictor of electoral success and public opinion at large. In: Analyzing Political Communication with Digital Trace Data, pp. 189–210. Springer (2015)

8. Jungherr, A., Jürgens, P., Schoen, H.: Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe, im predicting elections with twitter: What 140 characters reveal about political sentiment. Social science computer review 30(2), 229–234 (2012)
9. Mellon, J., Prosser, C.: Twitter and facebook are not representative of the general population: Political attitudes and demographics of social media users. Available at SSRN (2016)
10. Mislove, A., Lehmann, S., Ahn, Y.Y., Onnela, J.P., Rosenquist, J.N.: Understanding the demographics of Twitter users. ICWSM 11, 5th (2011)
11. Nguyen, D., Trieschnigg, D., Meder, T.: Tweetgenie: development, evaluation, and lessons learned. In: Proceedings of the 25th International Conference on Computational Linguistics, COLING 2014. pp. 62–66. Association for Computational Linguistics (August 2014), `http://doc.utwente.nl/94056/`
12. Nguyen, D.P., Gravel, R., Trieschnigg, R., Meder, T.: "how old do you think i am?" a study of language and age in Twitter (2013)
13. Sanders, E., Van den Bosch, A.: Relating political party mentions on Twitter with polls and election results. In: Proceedings of DIR-2013. pp. 68–71 (2013), `http://ceur-ws.org/Vol-986/paper\_9.pdf`
14. Sang, E.T.K., Bos, J.: Predicting the 2011 dutch senate election results with twitter. In: Proceedings of the Workshop on Semantic Analysis in Social Media. pp. 53–60. Association for Computational Linguistics (2012)
15. Tilley, J., Evans, G.: Ageing and generational effects on vote choice: Combining cross-sectional and panel data to estimate apc effects. Electoral Studies 33, 19–27 (2014)
16. Tjong Kim Sang, E., Van den Bosch, A.: Dealing with big data: The case of Twitter. Computational Linguistics in the Netherlands Journal 3, 121–134 (12/2013 2013)
17. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with Twitter: What 140 characters reveal about political sentiment. ICWSM 10, 178–185 (2010)
18. Verge Mestre, T., Tormos Marín, R.: The persistence of gender differences in political interest. Revista Española de Investigaciones Sociológicas (138) (2012)
19. Wang, W., Rothschild, D., Goel, S., Gelman, A.: Forecasting elections with non-representative polls. International Journal of Forecasting 31(3), 980–991 (2015)
20. Webster, S.W., Pierce, A.W.: Older, younger, or more similar? the use of age as a voting heuristic. Tech. rep., working paper (2015)