Towards Passive Political Opinion Polling using Twitter

Nicholas A. Thapen[1], Moustafa M. Ghanem[2]

[1]Department of Computing, Imperial College London
180 Queens Gate London, SW7 2AZ, United Kingdom {nicholas.thapen12@imperial.ac.uk}

[2]School of Science and Technology, Middlesex University, London
The Burroughs, London NW4 4BT, United Kingdom {m.ghanem@mdx.ac.uk}

Abstract. Social media platforms, such as Twitter, provide a forum for political communication where politicians broadcast messages and where the general public engages in the discussion of pertinent political issues. The open nature of Twitter, together with its large volume of traffic, makes it a useful resource for new forms of 'passive' opinion polling, i.e. automatically monitoring and detecting which key issues the general public is concerned about and inferring their voting intentions. In this paper, we present a number of case studies for the automatic analysis of UK political tweets. We investigate the automated sentiment analysis of tweets from UK Members of Parliament (MPs) towards the main political parties. We then investigate using the volume and sentiment of the tweets from other users as a proxy for their voting intention and compare the results against existing poll data. Finally we conduct automatic identification of the key topics discussed by both the MPs and users on Twitter and compare them with the main political issues identified in traditional opinion polls. We describe our data collection methods, analysis tools and evaluation framework and discuss our results and the factors affecting their accuracy.

1 Introduction

Twitter is a social networking service set up in 2006 allowing users to publish a stream of short messages called 'tweets', consisting of 140 characters or less. The author of a tweet may add the "#" symbol as prefix to arbitrary words in its content which become known as 'hashtags'. These hashtags can be regarded as keywords for identifying related messages. Optionally, users may also auto-tag their geo-location to a tweet. The social network is structured so that users can 'follow' each other, thus adding the followed user's tweets to the follower's newsfeed. Unlike other social networks, following a user in Twitter is usually automatic and does not require authorization. It is more like subscribing to an RSS feed or news service than establishing a friendship. Other built-in features are the ability to rebroadcast, or 'retweet', another tweet and the ability to reply to specific users as well as mention them in tweets. These features give Twitter aspects of both a social network and a news medium. In 2012, the service had over 140 million active users, with over 340 million tweets sent daily [19].

© 2013 N. A. Thapen & M. M. Ghanem.
Paper submitted to BCS SGAI Workshop on Social Media Analysis 2013, Cambridge, UK.

1.1 Towards Passive Political Opinion Polling from Twitter

Many politicians have embraced Twitter as a means to reach out directly to the public, bypassing traditional media sources. For example, Barack Obama launched his 2012 re-election campaign on Twitter and his victory tweet became the most retweeted of all time. Moreover, with a large number of users, the service has become an important arena for the dissimulation, discussion and creation of political news and opinion.

The fact that Twitter is seemingly open and accessible[1] makes it much more appealing to use for the purposes of research than other social networks such as Facebook, which have more emphasis on privacy. A tweet is generally intended to be read by a wide audience as part of the public record. While individual tweets contain very little information their brevity means that they are typically focused on a single issue. Moreover the aggregate of thousands or millions of tweets can potentially be fruitfully analyzed to discover different views around the same issue. The fact that political representatives use Twitter along with their constituents allows their language and interaction to be studied in order to discern what they say on Twitter about key issues and to discern whether they are out of step with the population at large. The analysis of what the public is discussing on Twitter could also be used for identifying their key concerns, and potentially also inferring their voting intentions.

Although a number of papers have been published on using Twitter data for predicting election results (See for example [1-3,5,6,8,10-12,14,15,17]), there is little work linking Twitter data with tracking opinion polls analyzing which key issues may be influencing the opinions of the people or the polls themselves. One paper [12] investigated Twitter economic sentiment and US presidential tracking polls. It found significant correlation between their data set and economic sentiment but little correlation on the political polls. Their study used simple keyword matching based on the 'Obama' and 'McCain' keywords. Whether more sophisticated approaches can find a meaningful correlation between the debate on Twitter and the opinion polls is clearly an open research question.

1.2 Motivation and Paper Overview

We start with two key questions:

1. Can we use Twitter to infer the proportion of the general public intending to vote for specific political candidates or parties?
2. Can we use Twitter data to infer the distribution of issues that the general public is concerned about?

We do not attempt to answer either question in this paper. However, a first reasonable step towards answering them is to compare the results of automated analysis of Twitter data with available poll data. This would allow us to gain a better understanding of what is needed to develop appropriate methodologies for conducting 'passive' opinion polling. Starting from two data sets that we collected on UK tweets in

[1] We discuss briefly in Section 3 some of the practical challenges for collecting historical Twitter data

2012/13, we experimented with automatic sentiment analysis tools and used both tweet volume and sentiment-weighted tweet volume as proxies for voting intention. We also developed and investigated the use of automatic topic identification techniques from the tweets and compared the outputs to key issues identified as important in opinion polls. Although our experiments are based on simple approaches, they provide many illuminating results that help in appreciating the questions better.

In Section 2, we review related work on election result prediction from Twitter data and discuss some of its key challenges. In Section 3, we describe the data sets used in our experiments. In Sections 4 and 5 we describe both the sentiment analysis tools and topic detection algorithms used and present and discuss the results for each case. Finally, in Section 6, we present our conclusions and discussion.

2 Related Work

Various researchers have investigated the use of Twitter for election result prediction. However, the successes of the approaches used have shown great variation. In an analysis of the 2009 German federal election [17] the authors were able to predict the vote shares in the election with a Mean Average Error of 1.65%, compared to an average error of 1.14% for six standard opinion polls. A study of the UK 2010 General Elections [18] reported a final average error of 1.75%. However, a study of the 2011 Singapore Elections in 2011 [15] found a greater error rate of 5.23%., whereas a study of the U.S Senate elections in 2010 [10] found far larger errors of around 17%.

Most work used the volumes of tweets mentioning particular candidates or parties as the measure of their popularity. However, some studies also investigated different methods for incorporating automated sentiment analysis of tweets' contents towards the contenders. The German study reported evidence that tweets about parties lying in similar places on the political spectrum contained similar emotional content. The US study reported that the final prediction error was reduced from 17% to 7.6% when sentiment analysis was applied. The German study simply used the Linguistic Inquiry and Word Count (LIWC) software tool to compute word and phrase statistics whereas an investigation of the 2011 Irish General Election [2] trained a classifier on a corpus of manually annotated positive and negative political tweets, then used tweet volume weighted by sentiment to report a final error of 5.85%. Given the prevalence of sarcasm and sophisticated humor in political discussions the reported results are encouraging.

One criticism [5] is that most studies are retrospective, performing backward-looking analysis rather than true prediction, and that their data selection methods arbitrarily influence their conclusions. One paper [8] showed that if the German study had included the German Pirate Party, much favored by Internet activists, they would have been predicted a landslide victory. We note that all studies also vary drastically in terms of data collection methods, sample sizes and how the analysis is conducted. There is usually no attempt at elucidating how the underlying assumptions of the studies may relate to standard opinion polling techniques, such as demographic weighting. It is rare that attempts are made at analyzing the context of the tweets or what is being discussed. In many cases, there is also little attempt to remove the influence of spammers or ‘Twitter bombs’ [10] - deliberate campaigns by political activists sending out thousands of similar tweets in a form of campaign advertising.

Moreover, most studies in this sphere are typically single shot experiments focused on the technological aspects. There is little or no methodological framework describing how they should be repeated and no standard benchmark against which they could be measured or through which their effectiveness could be analyzed time after time.

3 UK Political Tweets and Poll Data

UK Polling Report and Ipsos MORI Issues Index

We retrieved the list of voting intention polls kept by the UK Polling Report website [22]. This list provides all voting intention polls in the UK since June 2012. The polls are from all polling companies, and are thus based on various methodologies, such as phone-polling, internet panels and face to face interviews.

To retrieve a list of the issues that the public is concerned about we used Ipsos MORI [7], which has published a monthly Issues Index for the UK since 1974. This is based on a face-face survey asking around 1,000 British people the following question: “What do you see as the main/other important issues facing Britain today?” Respondents normally give around three categories as being important issues and Ipsos MORI then condense the answers into categories such as ‘Health’ and ‘Economy’.

Taking a list of topics from this source enables us to compare if political discussions on Twitter centre around the same topics or not. For our experiments we retrieved the Ipsos MORI Issues Index for the months of July 2012 - July 2013. To keep our analysis tractable, we consolidated the most frequent issues appearing in the poll data in the past year into 14 main categories, as well as an ‘Other’ category intended to catch all other issues. The categories chosen are:

Crime, Economy, Education, Environment, EU, Foreign Affairs, Government Services, Health, Housing, Immigration, Pensions, Politics, Poverty, Unemployment

In our classification, ‘Economy’ includes Inflation, Tax, Value of pound as well as the Ipsos-MORI Economy category, ‘Foreign Affairs’ includes all defense related matters, ‘Environment’ includes Rural Affairs, ‘Pensions’ includes Adult Social Care, and ‘Politics’ refers to Devolution and Constitutional Reform.

UK MP Twitter Data and Political Discussion Data

In order to identify UK political issues discussed on Twitter automatically we needed to collect a training data set that could be used in learning a lexicon of UK political terms. We focused on UK Members of Parliament (MPs) with the assumption that their tweets would mainly be focused on topical political issues. Moreover, the political orientation of these delegates is known and their tweets can be used to provide sanity checks on automated sentiment analysis methods as described later.

A list of the Twitter accounts of 423 UK MPs, classified by party affiliation, was retrieved from news website Tweetminster [18]. We retrieved 689,637 tweets from the publically available timelines of the MPs on 10th June 2013 using Twitter's REST API [20]. We note that timeline data returned by the API is capped at a maximum of 3,200 tweets for a single user's timeline. Although Twitter holds an archive of all Tweets posted since the service began, these Tweets are not held on the user's timeline and are indexed only by their unique id. Query access to such data is only possible through a number of commercial data providers [20].

In order to collect sample tweets relevant to UK political discussions, we considered collecting data using geo-location queries for the UK and then filtering by political keywords. This would have allowed us to look at geographic topic distributions and voting intentions. However, very few people enable geo-tagging due to privacy concerns. We thus decided to consider Twitter users who had mentioned recently the leaders of the three main political parties in their tweets. Our assumption is that most such users would be UK-based and more interested in UK political discussion than others. We thus retrieved the list of Twitter users who had recently mentioned the leaders of the three main political parties. We removed from those users known news sources to avoid news oriented tweets. We also ensured that none of them were in the existing UK MPs list. We then took a random sample of 600 of the remaining users. Similar to the MP data set, we retrieved the tweets from each user's timeline. This resulted in 1,431,348 tweets; retrieved in August 2013.

4 Sentiment Analysis and Voting Intentions

4.1 Sentiments of the MPs towards different parties

We experimented with different types of automated sentiment analysis techniques. In this paper we report on the results achieved using SentiStrength [16], a freely available sentiment analysis software tool which assigns sentiment scores based on look-ups to keywords in a sentiment polarity lexicon. We applied the tool to both the MP dataset and the political discussion datasets.

First, using the MP dataset, we extracted the tweets where political parties are mentioned. MPs discussing other parties can generally be assumed to be attempting to disparage them in some way, while when discussing their own parties they will usually use a positive spin. We used a keyword set containing the names, nicknames and shortenings of the names of the three main parties and then excluded from the dataset any tweets that mentioned more than one party. This resulted in a data set with 48,140 tweets (Labour: 23,070; Conservative: 18,034; Liberal Democrats: 7,036). The smaller number of Liberal Democrat tweets reflects the small size of the parliamentary party and activist base compared to the two main parties. The tweets were then split into groups depending on the party of the MP who tweeted them.

To identify how accurate the sentiment detection was, 30 tweets were selected at random from each of the nine groups and manually annotated as 'Positive 'or 'Negative' based on the language used and the sense meant. The results are summarized in Table 1.

Table 1. Sentiment Accuracy on Test Data Set

Class	Precision	Recall	F1 Measure
Negative	0.583	0.483	0.528
Positive	0.651	0.737	0.691
Overall	0.617	0.610	0.614

Clearly, the low precision and recall values raise alarms about the accuracy of the results for individual tweets, but overall indicate that the sentiment score could still be usable for overall sentiment detection. To verify, we then applied SentiStrength to the nine data sets (Figure 1). Here the figure shows SentiStrength's average positive and negative classification over each group, on a scale ranging from 1 (least positive/negative) to 5 (most positive/negative). The results back the general hypothesis. The simple methods work over aggregate data and show that MPs from each party tweet more negatively about other parties. Yet, the high level of negative sentiment of MPs concerning their own parties would be a surprise to most of the MPs themselves and their parties, as is the fact that any Labour tweets about Tories and vice-versa were positive.

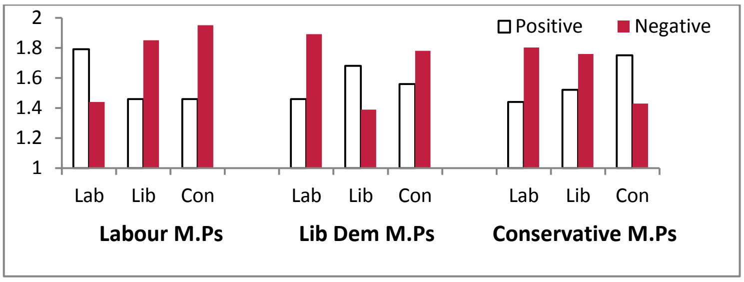
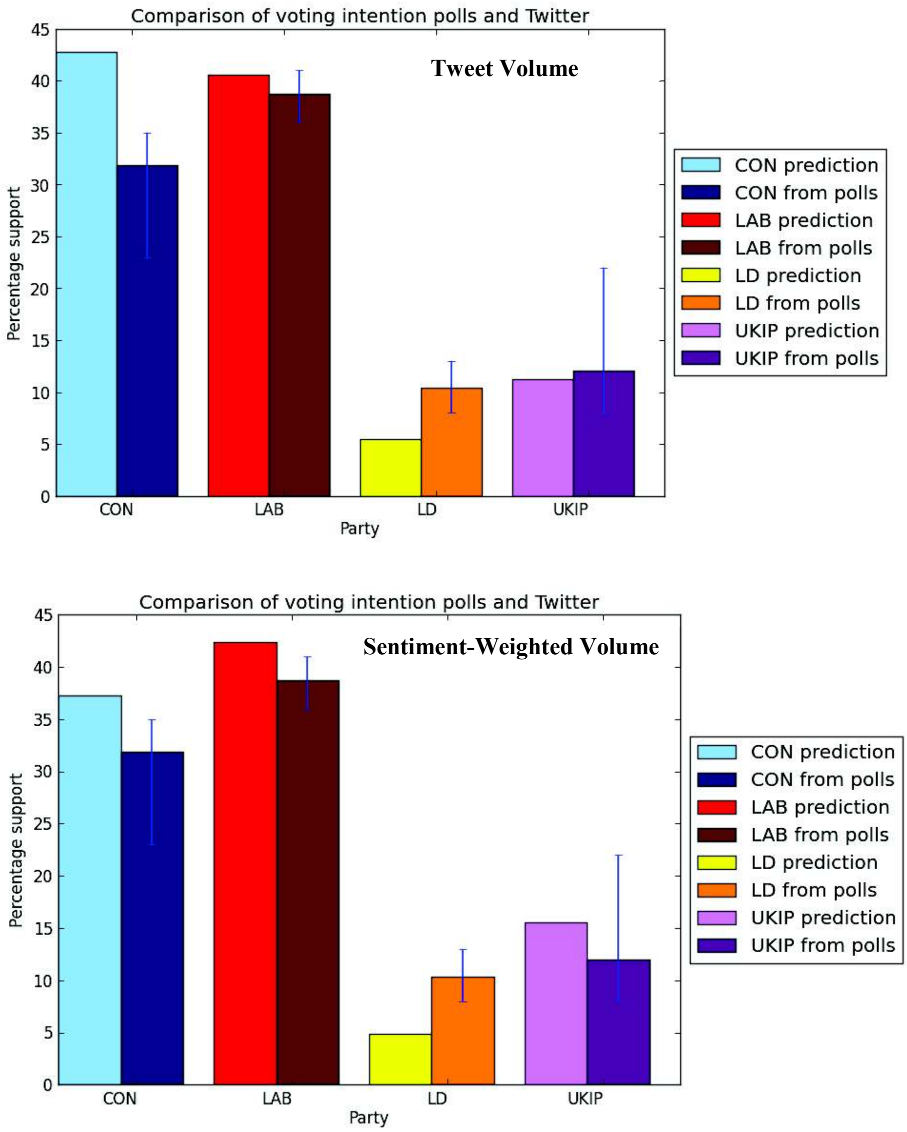


Fig. 1. Sentiment of MPs tweets towards each party based on SentiStrength

Despite both the simplicity of the method and its classification errors, the results show interesting insights. For example, they reveal that the parties tweet more negatively about their main opponents than the third party. Also, despite the coalition Lib Dems and Conservatives are not positive about each other, just less negative.

4.2 Investigating Voting Intentions

We proceeded to investigate whether the tweets in the political discussion dataset can be used as a proxy for inferring voting intentions. Our first experiment was simply to examine the proportion of tweets mentioning each party in the month of July 2013 (for which we had the most data in our data set) and to compare this to the average of the published opinion polls for the party. Firstly we obtained the numbers of tweets mentioning each party in that month. We excluded tweets which mentioned more than one party, as these would not be useful for the later sentiment analysis steps. We then took the proportion of tweets mentioning each party and compared it to the average predicted share of the vote from the opinion polls.



Figs. 2a and 2b Voting Intentions vs. Tweet Volume and Sentiment-Weighted Volume

The results are shown in Figure 2.a and the error bars give the range of values from the different opinion polls. The comparison between the Twitter prediction and the polls has a Mean Absolute Error of 4.6%, which as a first attempt was a surprisingly high correspondence. As shown in the figure, there is a close match for Labour and UKIP, but the Conservatives are given too much prominence and the Lib Dems too little. The ordering of Labour and the Conservatives is also incorrect.

Since many of the tweets mentioning the Conservatives are presumably negative, as they are the main party of government, we now moved on to weighting the results by sentiment to see if this could improve the fit of the data. In order to do so we adopted the sentiment-weighting methodology described in [12]. Adding in the sentiment weighting improved the error slightly, to 4.52%. More importantly all four parties are now in the correct rank order (Figure 2.b). The weighting was achieved by first running the sentiment analysis against all tweets to split them into positive and negative classes, then calculating sentiment weighted volume as follows:

$$weightedcount = count(party\ mentions) \times \frac{count(positive\ party\ mentions)}{count(negative\ party\ mentions)}$$

The fraction to compare against the polls is then: $\frac{weightedcount\ for\ party}{sum\ of\ weighted\ counts\ for\ all\ parties}$

Investigating Temporal Effects

We then looked at the same figures over the period July 2012 to July 2013. This revealed that the sentiment-weighted tweet data were extremely volatile, especially when looking at the earlier months in the chart. Before April 2013 they fail to match well with the voting intention figures at all. This analysis would seem to suggest that the close match between our data and the opinion polls for a single month could be a coincidence. However, the discrepancy could be accounted for by noting that we had much more data for recent months than for older ones due to the timeline retrieval limitations on Twitter. As mentioned earlier, the Twitter API allows retrieving only the most recent 3,200 tweets for each user. For example in our data set we have 9,979 tweets which mention party names in July 2013, but only 2,830 for April 2013.

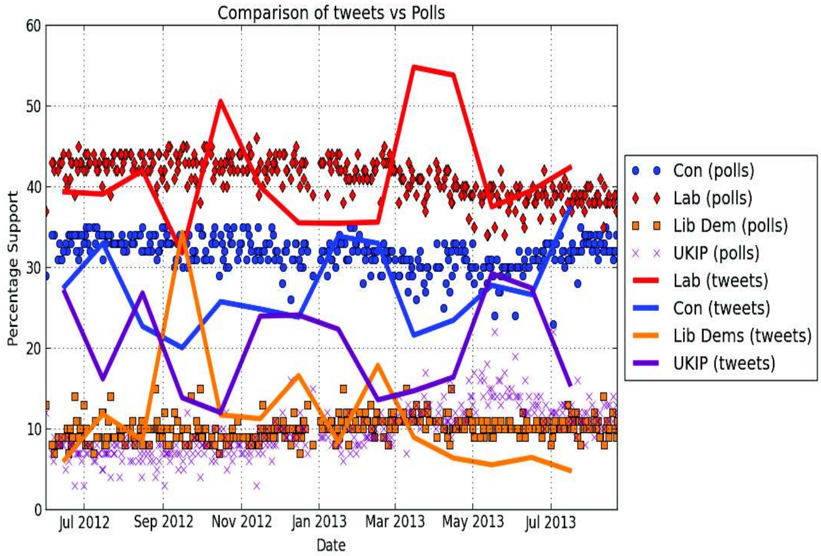


Fig. 3. Comparing Sentiment-Weighted Tweet Volume and Voting Intensions for 12 months

5 Topic Detection and Poll Issues

5.1 Topic Detection Algorithms

Iterative Keyword-based Topic Detection

We used an iterative snowballing method similar to [9], allowing the vocabulary to be built up gradually from the data, to develop the keyword-based classifier. We started with a manually constructed initial keyword list for each topic by consulting Wikipedia and our pre-existing political knowledge. These were used to classify the tweets in the MP dataset into the 14 categories. The keywords for each category were then augmented by additional keywords. This was done by first counting the frequency of 1, 2 and 3-grams in the tweets in each category and then applying a Fisher's Exact Test [21] to find n-grams which occurred most frequently within each category compared to outside the category. Human judgment was then applied to the candidate list to decide on which new keywords should be added. The process was iterated a number of times.

When using the initial keyword list on the MP dataset, the method was able to assign 109,644 tweets into the 14 categories but left 579,993 tweets, or 84% of the dataset, uncategorized. After the 5th iteration it could be seen that diminishing returns were setting in, since the final iteration categorized very few extra tweets. Snowballing allowed categorizing 94,647 extra tweets, leaving 70.4% of the dataset uncategorized. The large number of uncategorized tweets is expected since not all tweets will be discussing the 14 categories. We also experimented with using a Porter stemmer.

This reduced the number of uncategorized tweets to 63.5% of the total. However, it also slightly reduced classification accuracy, and was not taken forward.

To evaluate the keyword classifier, we used a data set of 300 tweets (20 tweets at random from those matched for each topic at the 5th iteration from the political discussion set) and manually annotated them. This resulted in Precision of 87.2%, Recall of 86.4% and F1-measure of 86,8%. These results ignore the large number of uncategorized tweets but indicate that the method is quite precise for our training purposes.

Bayesian Topic Classification

We then developed a Naïve Bayesian multi-label topic classifier that treats each tweet as a bag of words (similar to [4]). However, annotating a sufficient number of tweets for each topic to train the classifier would have been extremely time-consuming. We thus used output labels from the keyword-based classifier as training labels, giving thousands of tweets for each category. Moreover, since the training labels are noisy, the prior probabilities used by the classifier for the class labels were calculated from a probability distribution obtained by sampling 300 tweets and manually annotating them.

We trained a classifier for each topic separately, in order to allow for multi-label classification. If a topic classifier decides that the probability of the topic given the words in the tweet is greater than the probability of not belonging to the topic given the words, then the tweet is included in the topic label. If none of the classifiers assign a label to the tweet then the class with the greatest probability is selected as the single label.

An important caveat is that the distribution from the sample was fed into the Bayesian classifiers as prior knowledge. This means that classifiers are somewhat over-fitted. We thus prepared another randomly selected test data set of 300 tweets that was manually annotated. We then evaluated both classifiers on a randomly sampled manually annotated sample data set of 300 tweets. The results are summarized in Table 2. The results indicate that the Bayesian classifier is more accurate than the keyword-based one. Moreover, its accuracy is reasonable. Also, as can been seen, training the Bayesian classifier on stemmed data slightly improved both precision and recall. Nonetheless, the difference can be assumed not to be statistically significant.

Table 2. Classifier Evaluation on random data set

Classifier	Precision	Recall	F1 Measure
Keyword matching (5^{th} iteration)	0.287	0.279	0.283
Bayesian on non-stemmed data	0.753	0.793	0.773
Bayesian on stemmed data	0.764	0.798	0.781

To gain further confidence in our Bayesian classifier, we used it to label all tweets in the MP dataset and compared the distribution of topics detected to that of yet a new annotated sample set. These results gave close agreement, with a Mean Absolute Error of 0.0138 for the standard and 0.0129 for the stemmed classifier, with most topics falling within the margin of error of the sample. To perform sanity checks, we also compared the distribution of topics based on MPs from different political parties. The results (not shown because of space limitations) were consistent with political expectation. Labour MPs were concerned more with unemployment, poverty and housing than the Conservatives or Liberal Democrats. They also tweet more about the economy, reflecting their strong opposition to the governing coalition's austerity measures. The Conservatives' focus on the E.U compared to other parties is also evident, along with a slight extra interest in foreign affairs. Education receives a lot of emphasis from the Conservatives, perhaps due to their heavy focus on free schools and education reform in this parliament. It is somewhat surprising that Labour focus more on crime than the Conservatives and less on the environment.

Comparing Topic Distribution

We then compared the topic distribution between the MP data set and political one as shown in Figure 4. A greater proportion of the tweets here were identified as 'chatter', 70% rather than the 52% found amongst the MPs. Given that MPs are public figures, it was to be expected that a greater proportion of their tweeting would concern political topics. The higher proportion of tweets in the 'Other' category accounts for part of this, as does the fact that the keywords are explicitly political. The language used by MPs habitually uses many more of the political terms than the users. But, more, importantly, it was the MP data set that was used in training our methods.

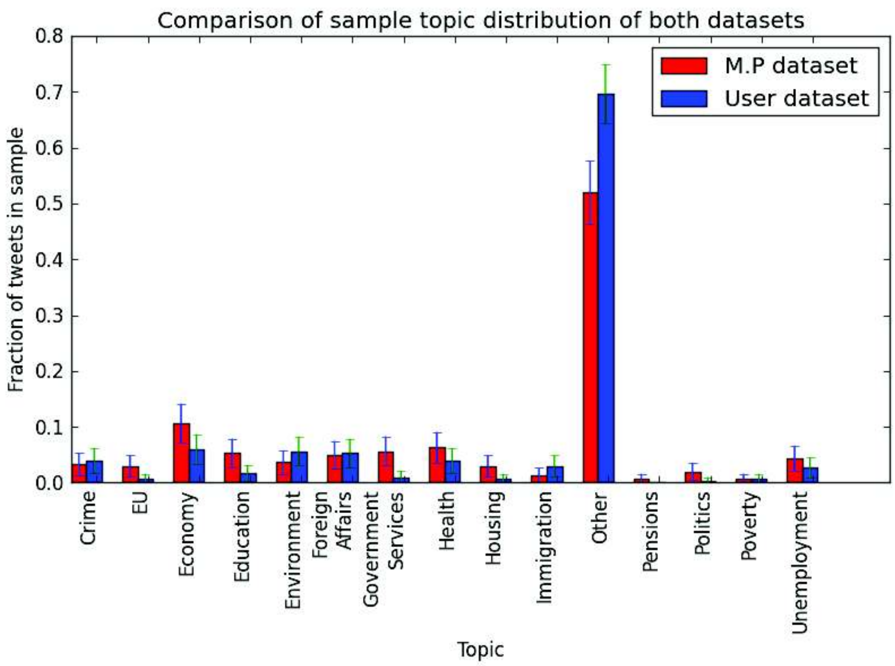


Fig. 4. Comparison between MP and Political Discussion Sets

5.2 Comparison with Polling Data

Finally, we proceeded to comparing the Ipsos MORI Issues Index to the topics extracted by the Bayesian classifier on both datasets. Again, we initially focused on the month of June, for which we had most data. The results are summarized in Figure 5. The Mean Absolute Error was 0.098 for the MPs and 0.11 for the users. One could interpret this slight difference in favour of the MPs being slightly more in touch with the concerns of the ordinary population than general Twitter users, since their topic distribution is found to be slightly closer to the polls. However, one must also notice that it was the MP data set used in the training of the classifier.

We do note the discrepancies between the polls and both the MPs and normal users in several categories; specifically, 'Economy', 'Immigration', 'Pensions' and 'Unemployment'. They all seem to be much less discussed on Twitter than in the poll. Analyzing the reasons of the mismatches is beyond the scope of this paper, but we cannot avoid making some comments. For example, one could also argue that normal users may not discuss the immigration issue too much over Twitter if they would be seen as racist by others. They could, however, be more likely to admit worries about it in private to a pollster than to broadcast them over Twitter.

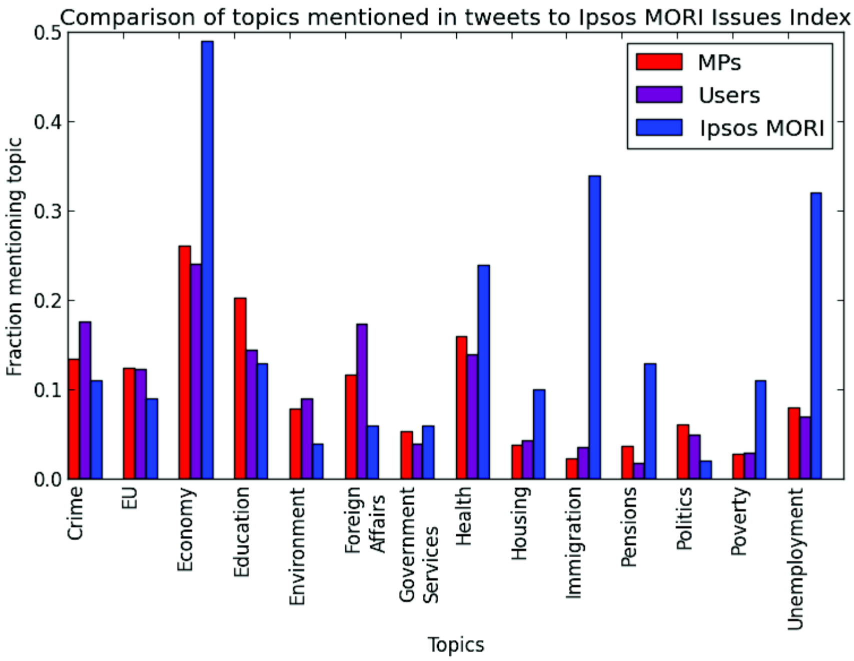


Fig. 5. Comparison of Twitter topics and Ipsos-MORI Issues for the month of June 2013

The demographics of Twitter users could potentially have had a big impact on the results. One could argue that Twitter users could be a much younger age group, and possibly one that is more affluent, than the broader spectrum taking part in the Ipsos MORI poll. However, there are no demographics in our Twitter data so we therefore examined the breakdown for the poll data itself for the 18-34 year old ABC1 group. This a social grade used by polling companies and market research organizations in the UK representing the segment of the population in high managerial down to administrative and professional positions, and is approximately 57% of the UK population [7]. We do not present the results here, but summarize that our experiments could not find any closer match in issues of this segment to the topics discussed on Twitter.

Investigating Temporal Effects

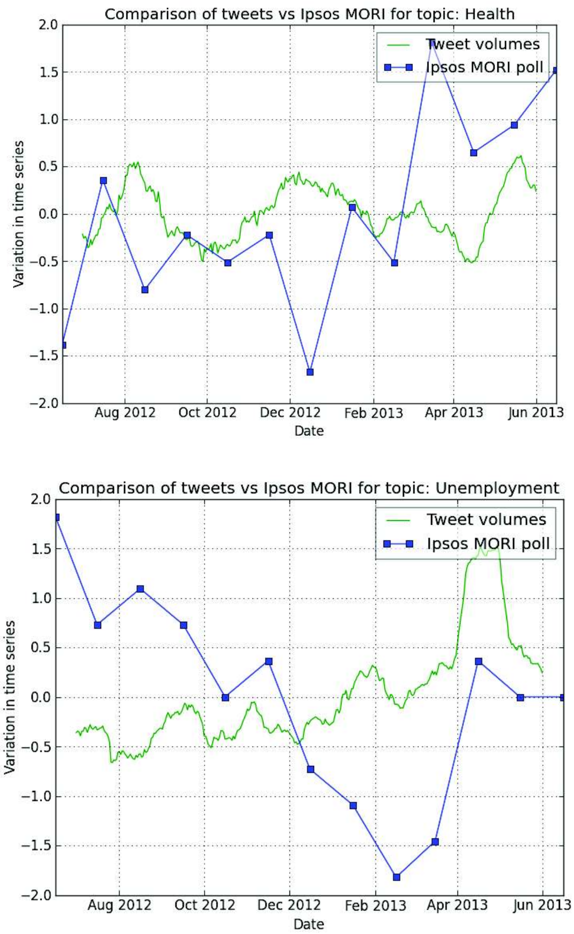


Fig. 6.a and 6.b 12-months comparison 'Health' and 'Unemployment' tweets and Ipsos-MORI category

Finally, we examined how the results varied for individual topics for the period June 2012 – June 2013. We provide only two examples here due to space limitations. Figure 6 shows time analysis for the topics ‘Health’ and ‘Immigration’ in the political discussion set vs. the poll data. The visual analysis of the graph does show some correlation between the trends of the respective time series, even if they do not match point-wise. The results are encouraging, but clearly indicate that more work is needed in developing the appropriate comparison methodology.

6 Summary and Discussion

In this paper, we presented our case studies conducted towards automated ‘passive’ political opinion polling from UK tweets. Namely, we looked at comparing volume and sentiment-weighted volume of tweets mentioning political parties with voting intentions reported from traditional polls. We also looked at detecting key political topics discussed over Twitter and comparing them to issues identified in polls. We described the techniques used and presented our results. Overall, the techniques yielded a close match and indeed showed that sentiment-weighted volume showed better matches for recent data. However, they showed volatility for the complete year. When comparing topics discussed in Twitter vs. those identified in polls, the task proved to more difficult, even if still promising.

Throughout the paper we identified all of our assumptions and described how our data collection methods could have influenced the results. The sample of tweets used in our work is not necessarily representative and our results are clearly not statistically rigorous. Our aim was not to conduct political analysis over the Twitter data but to investigate some of the key challenges that need to be addressed when doing so. Further development of the methodology for collecting the data and of the appropriate analysis methods is needed. Also more work is needed to understand how socio-political and demographics issues affect the results.

In the paper, we also showed how we used the known affiliation the MP to provide various sanity checks and also for training our lexicons. Clearly, the known affiliation could also be used in more interesting ways. For example we are currently investigating its use in developing Bayesian analysis techniques that take the context of a tweet into consideration when assigning a sentiment score. Moreover, we are investigating with various other Twitter data selection and sampling methods to avoid issues relating to political campaigning and also to increase the users under consideration.

If real-time, ‘passive’ opinion polling could be perfected it would be possible to cheaply canvas public opinion on a much broader range of issues than traditional polling. It could also potentially avoid ‘framing’ effects where respondents are influenced by the question asked. If such methods could be augmented by a theoretical underpinning, more sophisticated sentiment analysis and techniques such as demographic weighting then they could become a valuable tool in the political forecaster’s arsenal and also for marketing analysts. However, more investigation is still required into developing and evaluating new appropriate methodologies for collecting the required data, developing more sophisticated software tools and algorithms for its analysis and developing standardized methods and benchmarks to evaluate the accuracy of results.

References

1. Ampofo, L. et al. Trust, Confidence and Credibility: Citizen responses on Twitter to opinion polls during the 2010 UK General Election. Information, Communication & Society 14.6 (2011): 850-871.
2. Bermingham, A. & Smeaton, A., 2011. On using twitter to monitor political sentiment and predict election results. Workshop on Sentiment Analysis where AI meets Psychology.
3. Boutet, A. et al., 2012. What's in your tweets? I know who you supported in the UK 2010 general election. The International AAAI Conference on Weblogs and Social Media.
4. Frank, E. & Bouckaert, R.R., 2006. Naive Bayes for Text Classification with Unbalanced Classes. In PKDD'06 Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases.
5. Gayo-Avello, D., 2012. I wanted to predict elections with twitter and all I got was this lousy paper. A balanced survey on election prediction using twitter data. arXiv preprint arXiv:1204.6441 [Accessed June 3, 2013]. Available at: http://arxiv.org/abs/1204.6441
6. Hong, S and Nadler, D. Does the early bird move the polls?: the use of the social media tool Twitter by US politicians and its impact on public opinion." Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times. ACM, 2011.
7. Ipsos MORI [Accessed 4 September 2013] http://www.ipsos-mori.com/.
8. Jungherr, A., et al., 2012. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment." Social Science Computer Review, 30, pp.229–234.
9. McCallum, A and Nigam, K. Text classification by bootstrapping with keywords, EM and shrinkage. ACL 1999 Workshop for the Unsupervised Learning in Natural Language Processing, pp. 52-58, 1999.
10. Metaxas, P.T. & Mustafaraj, E., 2012. Social Media and the Elections. *Science.* [Accessed May 31, 2013]. Available at: http://www.sciencemag.org/content/338/6106/472.full.pdf.
11. Mungiu-Pippidi, A and Munteanu, I Moldova's Twitter Revolution. Journal of Democracy 20.3 (2009): 136-142.
12. O'Connor, B. et al., 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. International AAAI Conference on Weblogs and Social Media.
13. Pew Research, 2012. Assessing the Representativeness of Public Opinion Surveys. [Accessed June 7, 2013]. Available at: http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.
14. Sang, E et al. Predicting the 2011 Dutch senate election results with twitter. Proceedings of the Workshop on Semantic Analysis in Social Media. Association for Computational Linguistics, 2012.
15. Skoric, M. et al., 2012. Tweets and Votes: A Study of the 2011 Singapore General Election. 2012 45th Hawaii International Conference on System Sciences, pp.2583–2591.
16. Thelwall, M. et al., 2010. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), pp.2544–2558.

17. Tumasjan, A. et al., 2010. Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment. Word Journal of The International Linguistic Association, 280(39), pp.178–185.
18. Tweetminster, 2010. Tweetminster predicts. [Accessed September 4, 2013]. Available at: http://www.scribd.com/doc/31208748/Tweetminster-Predicts-Findings.
19. Twitter, 2012. Twitter turns six. @Twitter. [Accessed September 4, 2013]. Available at: https://blog.twitter.com/2012/twitter-turns-six.
20. Twitter, 2013. Twitter API [Accessed December 1, 2013], Available at https://dev.twitter.com/
21. Weisstein, E.W., 2013. Fisher's Exact Test. *MathWorld--A Wolfram Web Resource*. Available at: http://mathworld.wolfram.com/FishersExactTest.html [Accessed August 28, 2013].
22. Wells, A., 2013. Voting Intention since 2010. [Accessed September 4, 2013]. Available at: http://ukpollingreport.co.uk/voting-intention-2.