ElectionWatch: Detecting Patterns in News Coverage of US Elections

Saatviga Sudhahar, Thomas Lansdall-Welfare, Ilias Flaounas, Nello Cristianini
Intelligent Systems Laboratory
University of Bristol
(saatviga.sudhahar, Thomas.Lansdall-Welfare,
ilias.flaounas, nello.cristianini)@bristol.ac.uk

Abstract

We present a web tool that allows users to explore news stories concerning the 2012 US Presidential Elections via an interactive interface. The tool is based on concepts of "narrative analysis", where the key actors of a narration are identified, along with their relations, in what are sometimes called "semantic triplets" (one example of a triplet of this kind is "Romney Criticised Obama"). The network of actors and their relations can be mined for insights about the structure of the narration, including the identification of the key players, of the network of political support of each of them, a representation of the similarity of their political positions, and other information concerning their role in the media narration of events. The interactive interface allows the users to retrieve news report supporting the relations of interest.

1 Introduction

U.S presidential elections are major media events, following a fixed calendar, where two or more public relation "machines" compete to send out their message. From the point of view of the media, this event is often framed as a race, with contenders, front runners, and complex alliances. By the end of the campaign, which lasts for about one year, two line-ups are created in the media, one for each major party. This event provides researchers an opportunity to analyse the narrative structures found in the news coverage, the amounts of media attention that is devoted to the main contenders and their allies, and other patterns of interest.

We propose to study the U.S Presidential Elections with the tools of (quantitative) narrative analysis, identifying the key actors and their political relations, and using this information to infer the overall structure of the political coalitions. We are also interested in how the media covers such event that is which role is attributed to each actor within this narration.

Quantitative Narrative Analysis (QNA) is an approach to the analysis of news content that requires the identification of the key actors, and of the kind of interactions they have with each other (Franzosi, 2010). It usually requires a significant amount of manual labour, for "coding" the news articles, and this limits the analysis to small samples. We claim that the most interesting relations come from analysing large networks resulting from tens of thousands of articles, and therefore that QNA needs to be automated.

Our approach is to use a parser to extract simple SVO triplets, forming a semantic graph to identify the noun phrases with actors, and to classify the verbal links between actors in three simple categories: those expressing political support, those expressing political opposition, and the rest. By identifying the most important actors and triplets, we form a large weighted and directed network which we analyse for various types of patterns.

In this paper we demonstrate an automated system that can identify articles relative to the 2012 US Presidential Election, from 719 online news outlets, and can extract information about the key players, their relations, and the role they play in the electoral narrative. The system refreshes its information every 24 hours, and has already analysed tens of thousands of news articles. The tool allows the user to browse the growing set of news articles by the relations between actors, for example retrieving all articles where Mitt Romney

praises Obama[1].

A set of interactive plots allows users to explore the news data by following specific candidates and also specific types of relations, to see a spectrum of all key actors sorted by their political affinity, a network representing relations of political support between actors, and a two-dimensional space where proximity again represents political affinity, but also they can access information about the role mostly played by a given actor in the media narrative: that of a subject or that of an object.

The ElectionWatch system is built on top of our infrastructure for news content analysis, which has been described elsewhere. It has also access to named entities information, with which it can generate timelines and activity-maps. These are also available through the web interface.

2 Data Collection

Our system collects news articles from 719 English language news outlets. We monitor both U.S and International media. A detailed description of the underlying infrastructure has been presented in our previous work (Flaounas, 2011).

In this demo we use only articles related to US Elections. We detect those articles using a topic detector based on Support Vector Machines (Chang, 2011). We trained and validated our classifier using the specialised Election news feed from Yahoo!. The performance of the classifier reached 83.46% precision, 73.29% recall, validated on unseen articles.

While the main focus of the paper is to present Narrative patterns in elections stories, the system presents also timelines and activity maps generated by detected Named Entities associated with the election process.

3 Methodology

We perform a series of methodologies for narrative analysis. Figure 1 illustrates the main components that are used to analyse news and create the website.

Preprocessing. First, we perform co-reference and anaphora resolution on each U.S Election article. This is based on the ANNIE plugin in GATE (Cunningham, 2002). Next, we ex-

tract Subject-Verb-Object (SVO) triplets using the Minipar parser output (Lin, 1998). An extracted triplet is denoted for example like "Obama(S)–Accuse(V)–Republicans(O)". We found that news media contains less than 5% of passive sentences and therefore it is ignored. We store each triplet in a database annotated with a reference to the article from which it was extracted. This allows us to track the background information of each triplet in the database.

Key Actors. From triplets extracted, we make a list of actors which are defined as subjects and objects of triplets. We rank actors according to their frequencies and consider the top 50 subjects and objects as the key actors.

Polarity of Actions. The verb element in triplets are defined as actions. We map actions to two specific action types which are endorsement and opposing. We obtained the endorsement/opposing polarity of verbs using the Verbnet data (Kipper et al, 2006)).

Extraction of Relations. We retain all triplets that have a) the key actors as subjects or objects; and b) an endorse/oppose verb. To extract relations we introduced a weighting scheme. Each endorsement-relation between actors a, b is weighted by $w_{a,b}$:

$$w_{a,b} = \frac{f_{a,b}\left(+\right) - f_{a,b}\left(-\right)}{f_{a,b}\left(+\right) + f_{a,b}\left(-\right)} \quad (1)$$

where $f_{a,b}(+)$ denotes the number of triplets between a, b with positive relation and $f_{a,b}(-)$ with negative relation. This way, actors who had equal number of positive and negative relations are eliminated.

Endorsement Network. We generate a triplet network with the weighted relations where actors are the nodes and weights calculated by Eq. 1 are the links. This network reveals endorse/oppose relations between key actors. The network in the main page of ElectionWatch website, illustrated in Fig. 2, is a typical example of such a network.

Network Partitioning. By using graph partitioning methods we can analyse the allegiance of actors to a party, and therefore their role in the political discourse. The Endorsement Network is a directed graph. To perform its partitioning we first omit directionality by calculating graph $B = A + A^T$, where A is the adjacency matrix of the Endorsement Network. We computed eigenvectors of the B and selected the eigenvector that
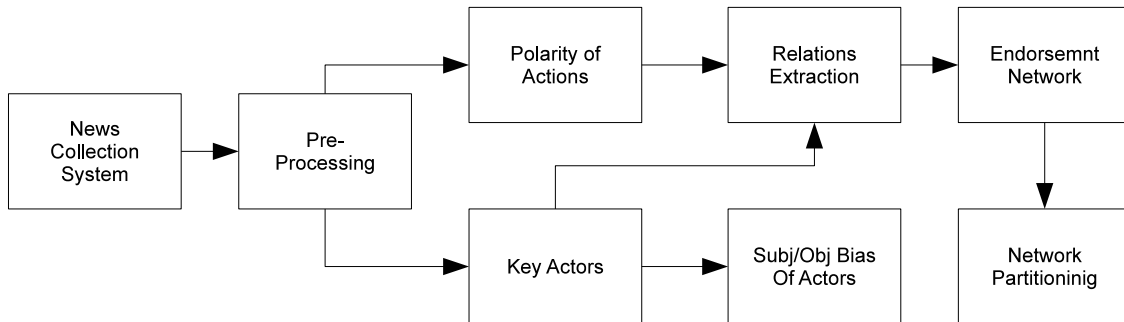

Figure 1: The Pipeline

correspond to the highest eigenvalue. The elements of the eigenvector represent actors. We sort them by their magnitude and we obtain a sorted list of actors. In the website we display only actors that are very polarised politically in the sides of the list. These two sets of actors correlate well with the left-right political ordering in our experiments on past US Elections. Since in the first phase of the campaign there are more than two sides, we added a scatter plot using the first two eigenvectors.

Subject/Object Bias of Actors. The Subject/Object bias S_a of actor a reveals the role it plays in the news narrative. It is computed as:

$$S_a = \frac{f_{Subj}(a) - f_{Obj}(a)}{f_{Subj}(a) + f_{Obj}(a)} \tag{2}$$

A positive value of S for actor a indicates that the actor is used more often as a subject and a negative value indicates that the actor is used more often as an object.

4 The Website

We analyse news related to U.S Elections 2012 every day, automatically, and the results of our analysis are presented integrated under a publicly available website[2]. Figure 2 illustrates the homepage of ElectionWatch. Here, we list the key features of the site:

Triplet Graph – The main network in Fig. 2 is created using the weighted relations. A positive sign for the edge indicates an endorsement relation and a negative sign indicates an opposition relation in the network. By clicking on each edge in the network, we display triplets and articles that support the relation.

Actor Spectrum – The left side of Fig. 2 shows the Actor Spectrum, coloured from blue for Democrats to red for Republicans. Actor spectrum was obtained by applying spectral graph partitioning methods to the triplet network.Note, that currently there are more than two campaigns that run in parallel between key actors that dominate the elections news coverage. Nevertheless, we still find that the two main opposing candidates in each party were in either sides of the list.

Relations – On the right hand side of the website we show the endorsement/opposition relations between key actors. For example, "Republicans Oppose Democrats". When clicking on a relation the webpage displays the news articles that support the relation.

Actor Space – The tab labelled 'Actor Space' plots the first and second eigenvector values for all actors in the actor spectrum.

Actor Bias The tab labelled 'Actor Bias' plots the subject/object bias of actors against the first eigenvector in a two dimensional space.

Pie Chart – Pie Chart on the left bottom in the webpage shows the share of each actor with regard to the total number of articles mentioning an endorse/oppose relation.

Map – The map geo-locates articles related to US Elections and refer to US locations.

Bar Chart – The bar chart tab, illustrated in Fig. 3, plots the number of articles in which actors were involved in a endorse/oppose relation. The height of each column reveals the frequency of it. The default plot focuses on only the first five actors in the actor spectrum.

Timelines & Activity Map – We track the activity of each named entity in the actor spectrum within the United States and present it in a timeline. The activity map monitors the media atten-

---

[2]ElectionWatch: http://electionwatch.enm.bris.ac.uk

84

Figure 2: Screenshot of the home page of ElectionWatch

Figure 3: Barchart showing endorse/oppose article frequencies for actor "Obama" with other top actors.

tion for Presidential candidates in each state in the Unites States. At present we monitor this activity for Mitt Romney, Rick Perry, Michele Bachmann, Herman Cain and Barack Obama.

5 Discussion

We have demonstrated the system ElectionWatch that presents key actors in U.S election news articles and their role in political discourse. This builds on various recent contributions from the field of Pattern Analysis, such as (Trampus, 2011), augmenting them with multiple analysis tools that respond to the needs of social sciences

investigations.

We agree on the fact that the triplets extracted by the system are not very clean. This noise can be ignored since we perform analysis on only filtered triplets containing key actors and specific type of actions, and also it's extracted from huge amount of data.

We have tested this system on data from all previous six elections, using the New York Times corpus as well as our own database. We use only support/criticism relations revealing a strong polarisation among actors and this seems to correspond to the left/right political dimension. Evaluation is an issue due to lack of data but results on the past six election cycles on New York Times always seperated the two competing candidates along the eigenvector spectrum. This is not so easy in the primary part of the elections, when multiple candidates compete with each other for the role of contender. To cover this case, we generate also a two-dimensional plot using the first two eigenvalues of the adjacency matrix, which seems to capture the main groupings in the political narrative.

Future work will include making better use of the information coming from the parser, which

goes well beyond the simple SVO structure of sentences, and developing more sophisticated methods for the analysis of large and complex networks that can be inferred with the methodology we have developed.

Acknowledgments

References

Chang C.C., and Lin C.J. 2011. *LIBSVM: a library for support vector machines*. ACM Transactions on Intelligent Systems and Technology 2(3):1–27

Cunningham H., Maynard D., Bontcheva K. and Tablan V. 2002. *GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications*. Proc. of the 40th Anniversary Meeting of the Association for Computational Linguistics 168–175.

Earl J., Martin A., McCarthy J.D., Soule S.A. 2004. *The Use of Newspaper Data in the Study of Collective Action*. Annual Review of Sociology, 30:65–80.

Flaounas I., Ali O., Turchi M., Snowsill T., Nicart F., De Bie T., Cristianini N. 2011. *NOAM:News Outlets Analysis and Monitoring system*. Proc. of the 2011 ACM SIGMOD international conference on Management of data, 1275–1278.

Franzosi R. 2010. *Quantitative Narrative Analysis*. Sage Publications Inc, Quantitative Applications in the Social Sciences, 162–200.

Kipper K., Korhonen A., Ryant N., Palmer M. 2006. *Extensive Classifications of English verbs*. 12th EURALEX International Congress, Turin, Italy.

Lin D. 1998. *Dependency-Based Evaluation of Minipar*. Text, Speech and Language Technology 20:317–329.

Sandhaus, E. 2008. *The New York Times Annotated Corpus*. Linguistic Data Consortium

Trampus M., Mladenic D. 2011. *Learning Event Patterns from Text*. Informatica 35