Paper SAS349-2014

Summarizing and Highlighting Differences in Senate Race Data Using SAS® Sentiment Analysis

Hilke Reckman, Michael Wallis, Richard Crowell, Linnea Micciulla, Cheyanne Baird

SAS Institute Inc.

ABSTRACT

Contrasting two sets of textual data points out important differences. For example, consider social media data that have been collected on the race between incumbent Kay Hagan and challenger Thom Tillis in the 2014 election for the seat of US Senator from North Carolina. People talk about the candidates in different terms for different topics, and you can extract the words and phrases that are used more in messages about one candidate than about the other. By using SAS® Sentiment Analysis on the extracted information, you can discern not only the most important topics and sentiments for each candidate, but also the most prominent and distinguishing terms that are used in the discussion. Find out if Republicans and Democrats speak different languages!

INTRODUCTION

This paper shows how to use SAS® Text Analytics to see an overview of the differences between two collections of text, in terms of topics and sentiment. Example data comes from the North Carolina senate race between incumbent Democrat Kay Hagan, US Senator for North Carolina since 2009, and Republican challenger Thom Tillis, Speaker of the North Carolina House of Representatives since 2011. The ever-increasing use of social media and other unstructured texts to discuss public figures and political issues provides a wealth of data that can be used for textual analysis. In the essentially two-party system that prevails in the United States, a contrastive analysis between candidates from each party can provide useful information about the social media presence of each candidate in the context of an election race. Performing a contrastive analysis between the candidates can surface additional insights beyond what you might learn from performing sentiment analysis on each candidate individually.[1]

DATA

Kay Hagan, a Democrat, is the current senator for North Carolina, and she is running for reelection. Her main competitor is Thom Tillis, a Republican. The data include discussion about both candidates in social media and other web sources, including blogs, forums, news stories, and Twitter. All data collected are from the time period beginning July 1, 2013, and ending August 6, 2013. The volume of discussion about Hagan is higher than it is about Tillis (presumably because Hagan is the incumbent), and at the time of the data collection the elections are still more than a year in the future.

This experiment uses the following query to collect data from Twitter and a data aggregation service:

(Hagan OR Tillis) AND (senate OR senator OR candidate OR election)

A number of issues are considered in processing the data. One issue is how to handle documents that exist in both the Hagan and Tillis document sets—that is, documents in which both candidates are mentioned. Because SAS Sentiment Analysis can correctly assign sentiment to a target based on distance and syntax, it is possible to retain documents in which both candidates are mentioned for analyses of entire documents. However, for the purpose of associating phrases with candidates at the document level, a subset of the original data set is created. This subset retains only documents that mentioned either Kay Hagan or Thom Tillis, but not both.

As a final step, duplicate and near-duplicate documents are excluded from the data set. In the Twitter data set, a number of tweets are identical except for their timestamp. In the news data set, a number of stories were republished as similar or identical texts. These duplicate texts contain fragments that occur repeatedly, thus showing up in the results as very long phrases that span multiple sentences. Although this is representative of the actual Twitter and web traffic, the large number of duplicate texts in a small sample can skew the results so that they represent a very small number of the original texts, especially in such a small sample. Table 1 shows the final document counts.

---

[1] All results are considered experimental. Results are not authoritative and do not represent the opinions of SAS or the authors.

Source	Hagan	Tillis	Total
Blogs	189	162	351
Forums	74	25	99
News	339	386	725
Twitter	331	98	429
Total	933	671	1,605

Table 1. Numbers of Documents per Candidate by Source, after Filtering

METHOD

EXTRACTING PHRASES

A collection of phrases can provide a more accurate impression of the data than can a collection of only single words. Sentiment analysis also works better when it takes phrases rather than individual words into consideration. Expressions of sentiment often consist of more than one word and can contain negations or other modifiers (for example, *not so bad*). Therefore, the first step is to extract phrases that are sequences of words that tend to occur together. If two words show a strong tendency to occur together, the word boundary between them is ignored, and they are treated as one unit for future processing. Longer phrases have strong connections between each two-word sequence within them. For example, if the combinations *speaker Thom* and *Thom Tillis* each have a strong connection, then *speaker Thom Tillis* (if it occurs) is also treated as a phrase.

Take the phrase *North Carolina* as an example. In the news, blog, and forum data, if you count the words *North* and *Carolina* in the original data set (2,497 instances and 2,246 instances, respectively) in addition to the total number of words (827,706), you can calculate how often you expect them to occur together to form *North Carolina*, assuming words are randomly distributed. The expected value is calculated by multiplying the frequency of the word *North* by the frequency of the word *Carolina* and then dividing by the total number of works in the corpus. Using the previous numbers, the expected value has an approximate value of 7 (2,497 * 2,246 / 827,706). In reality, the phrase occurs 2,117 times in the data, which is much higher than expected. By calculating the chi-square statistic, you can show that this difference between expected and observed co-occurrences is significant.

Chi-square (χ^2) is a metric of how much the four observed values (O) in Table 2 differ from the expected values (E) in Table 3.

Observed	+carolina	–carolina	Total
+north	2,117	380	2,497
–north	129	825,080	825,209
Total	2,246	825,460	825,706

Table 2. Observed Frequencies

Expected	+carolina	–carolina	Total
+north	7	2,490	2,497
–north	2,239	822,970	825,209
Total	2,246	825,460	825,706

Table 3. Expected Frequencies

For the *North Carolina* example, χ^2 = 639341.81, which is well above any sensible significance threshold and is one of the strongest collocations in this data set. Therefore, the expression is treated as a phrase rather than as two single words. Note that the reverse order *Carolina North* occurs only once, which is much less than expected. Therefore, *Carolina North* is not treated as a phrase.

You can experiment with different thresholds to determine which combinations should be treated as phrases. The lower the threshold, the longer the phrases you will see in your results. At the very least, you want the associations to be highly significant. Table 4 shows which thresholds correspond to various levels of significance. However, you might find that setting a substantially higher threshold is preferable.

p-value	Chi-Square
0.01	6.63
0.005	7.88
0.001	10.83
0.0005	12.12

Table 4. Minimal Chi-Square Values for Different Significance Levels

Typical stop word lists are not useful here, because function words such as *the* and *of* are part of larger phrases (for example, *the United States of America*).

For low-frequency words, it is important to adjust the counts. For example, if a word occurs only once in the data, there is not enough information available to determine whether that word has a special relationship with its neighbors. Linguistic data always has many low-frequency words. The χ^2 itself does not take this lack of information into account. If a word occurs only once, then the probability of it occurring together with any other word is extremely low; so even though that combination occurs only once, it will appear to be much more frequent than expected. A simple way to solve this problem is to adjust the counts. This experiment uses a smoothing technique called Simple Good-Turing (Gale and Sampson 1995) for frequency estimation. This technique enables you to assign probability mass both to words that have been seen and to words that have not been seen.

CONTRASTING WORDS AND PHRASES IN TWO COLLECTIONS OF TEXT

To find out which phrases are associated with which candidate, you can use the same association metric. You count how often the phrase occurs in documents about a particular candidate and compare that frequency to the frequency of that phrase in the rest of the data.

Because you have two candidates in this experiment, you compile two lists: one with expressions that occur much more frequently in the documents about Tillis, and one with expressions that occur much more frequently in the documents about Hagan. You order each list by association strength—that is, the most frequently occurring phrases on top.

You can use the same threshold as you used for extracting the phrases, or you can use a different one. The lower your threshold, the longer your list of results. Be sure to keep your threshold above a reasonable significance level.

The contrastive aspect makes stop word lists largely superfluous. Stop words—terms that are used to filter out documents from further analysis—normally have similar frequencies in all portions of the data and therefore do not surface as associated with either candidate. If the stop words do not surface as associated with either candidate, something interesting, such as a difference in discourse style, might be the cause. For example, if you contrast news data with Twitter data instead of contrasting data for two political candidates, the stop words that show up in your results will give you a sense of the different styles that are used in one genre versus the other.

TEXTUAL EXTRACTION AND TOKENIZATION

In order to obtain the strongest phrases that are associated with a particular candidate, a combination of the TGFILTER and TGPARSE procedures in SAS® Text Miner are used for textual extraction and parsing of candidate data, respectively. A per-document processing approach is taken, and a temporary directory is created to hold a single candidate data file. This directory serves as the source directory through which PROC TGFILTER searches. Although a single invocation of PROC TGFILTER can examine multiple documents, the approach taken in this example is designed to ease the separation of per-document results and the subsequent *n*-gram analysis. Successful execution of the TGFILTER procedure generates a data set that contains the results of the textual extraction process, which are used later in text parsing.

Tokenization of candidate data is performed using the TGPARSE procedure with various options that impact what is included in the final tokenization result data set. For this example, stemming is included and punctuation is excluded. Stemming is the reduction of an inflected term down to its root form (for example, *elections* reduces to *election*). When a term is stemmed, it is mapped to a parent term. Parent terms take on the root form of the original term and serve as a single representation for all term variations.

The following statements invoke four SAS procedures: TGFILTER, TGPARSE, SORT, and EXPORT.

```
proc tgfilter
 out=documents
 srcdir="C:\senate\data\source"
 numchars=32767;
run;

proc tgparse
 data=documents
 key=parsekey
 out=parseout
 outoffset=parseoutoffset
 stemming=yes;
 var text;
 select "punct" / drop;
run;

proc sort
 data=parsekey;
 by Key;
run;

proc sort
 data=parseoutoffset;
 by _OFFSET_;
run;

proc export
 data=parsekey
 outfile="C:\senate\data\parse_key.csv"
 dbms=csv;
run;

proc export
 data=parseoutoffset
 outfile="C:\senate\data\parse_out_offset.csv"
 dbms=csv;
run;
```

PROC TGFILTER is configured with the name of the output data set, a source directory that corresponds to the location where a single temporary file is placed for text extraction, and a value for the number of characters to use in the text variable that is contained in the output data set. The procedure is invoked once for each file that resides at the location specified in the source directory. This file, which contains textual data that was provided by either Twitter or the data aggregator service.

PROC TGPARSE is configured with the name of the data set that is used as the output data set from the TGFILTER procedure, the name (KEY) of the data set that contains the summary information about the terms in the document, the name (OUT) of the data set that contains the compressed term-by-document matrix, and the name (OUTOFFSET) of the data set that contains the byte offset for each term. The stemming option is configured with either YES or NO, which indicates whether stemming should be applied. Punctuation is dropped by using the SELECT statement with the PUNCT part-of-speech label, followed by the DROP keyword directive.

After PROC TGPARSE has tokenized the data, the SORT procedure sorts the data sets that were created by PROC TGPARSE. The PARSEKEY data set is sorted in ascending order by the variable name Key and the PARSEOUTOFFSET data set is sorted in ascending order by the variable name _OFFSET_. The Key variable uniquely identifies a term from the PARSEKEY data set, and the _OFFSET_ variable corresponds to the byte offset of the term in the originating document. Both data sets are then exported to comma separated file (CSV) file formats by using PROC EXPORT.

The original ordering of terms is reconstructed using both data sets together. For each observation in the PARSEOUTOFFSET data set, the _TERMNUM_ variable is matched against the Key variable in the PARSEKEY data set. When a match occurs, a test is performed to determine whether the term has been stemmed. If the term has been stemmed, the associated Parent_id variable is used as a Key value to look up in the same PARSEKEY data

set. The term identified by the Parent_id is then used as the term in place of the stemmed term. If the term has not been stemmed, the term is used as is.

SENTIMENT ANALYSIS

SAS Sentiment Analysis recognizes expressions of sentiment in text. It can also recognize words that belong to a predefined topic. This example builds a sentiment model for the domain of political campaigns. The sentiment model tracks mentions of each candidate and mentions of the most important topics in this race: civil rights, defense and veterans, economy and fiscal policy, education, energy and environment, healthcare, immigration, and transportation. Each of these topics is considered to be a product so that the prominence information provides insight into which topics are most prominent for which candidate.

The sentiment model for the two US Senate candidates, Kay Hagan and Thom Tillis, is customized by compiling a list of possible names, nicknames, noun phrases, and pronouns that could be used to refer to each of the candidates and by creating rules to identify these words in the input texts. Some examples of rules written to identify Kay Hagan include the following:

```
CLASSIFIER:Hagan
CLASSIFIER:@SenatorHagan
PREDICATE_RULE:(ORDDIST_100, "Hagan", "_a{Kay}")
PREDICATE_RULE:(UNLESS,
 "_def{DEFCANDIDATE1Competitors}",
 (ORDDIST_15,"Hagan","_a{she}"))
PREDICATE_RULE:(UNLESS,
 "_def{DEFCANDIDATE1Competitors}",
 (ORDDIST_30, "Hagan","_a{the Senator's}"))
```

The classifier rules identify the strings “Hagan” and “@SenatorHagan” as referring to Kay Hagan whenever they occur in the text. The first predicate rule identifies the string “Kay” whenever it occurs within 100 words after the string “Hagan.” The second and third predicate rules identify the strings “she” and “the Senator’s” whenever they occur within a certain distance after the string “Hagan” provided that no other potential referents occur between them. More than 100 rules were designed to identify text strings that refer to Kay Hagan. A similar process was followed to create rules to identify mentions of Thom Tillis.

In a similar vein, in order to detect mentions of the eight topics to be tracked, a set of linguistic rules was built to identify these eight topics, along with a ninth topic: reputation, which identifies comments about a candidate’s reputation in general. For example, “*He is a good candidate*” would belong to the topic of a candidate’s reputation without necessarily being related to one of the other eight topics. Rules for several of the topics used in this model of political campaigns already exist in other domain-specific sentiment models; these rules are incorporated into the political campaign model, making adjustments as needed.

Linguistic rules that use both syntactic and distance considerations are used to associate topics with the candidates. Rules are used to associate sentiment with a topic. The phrase extraction approach described previously can be an excellent source of phrases you can incorporate into rules. In addition, the generic, domain-independent rules that are designed to recognize positive and negative sentiment are supplemented by additional rules that are based on language patterns that are frequently found in political discourse.

The sentiment model was run on the original corpus and on the output files. Table 5 shows the total number of mentions of each topic for each candidate, and Figures 1 and 2 show their distributions.

RESULTS

SENTIMENT ANALYSIS WITHOUT PHRASE EXTRACTION

Figures 1 and 2 show that discussion of two candidates is similarly distributed when raw documents are processed without phrase extraction. The highlighting shows topic-related terms in blue, positive expressions in green, and negative expressions in red. This color-coding in Figures 1 and 2 enables you to see at a glance which topic-related terms occur in the documents and phrases that are associated with each candidate and whether those terms are expressed positively or negatively.

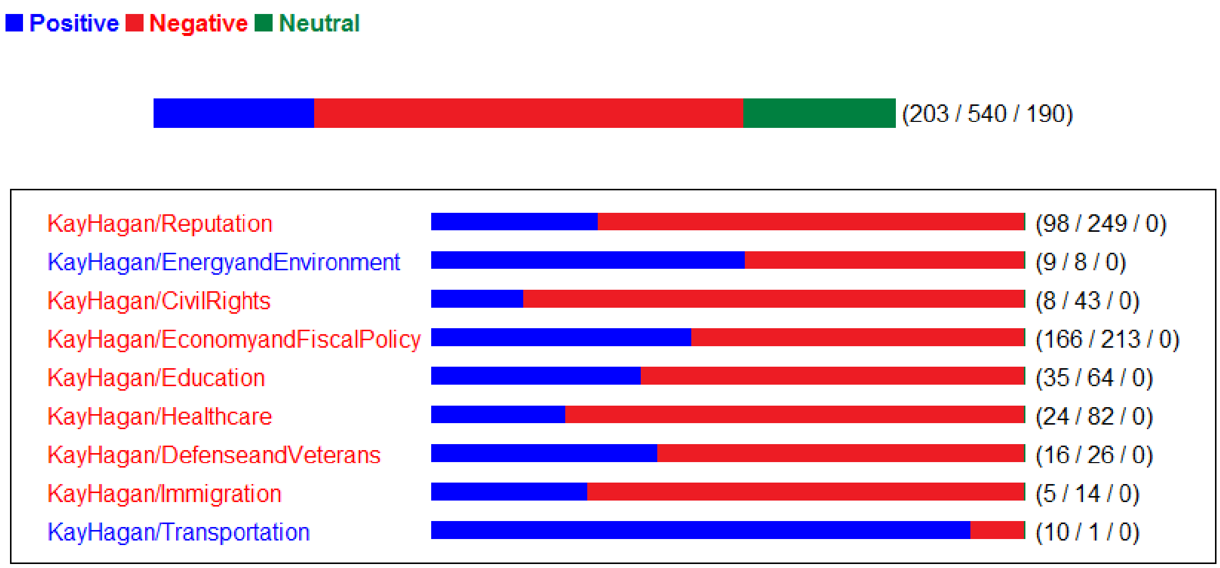

Figure 1. Sentiment Analysis for Hagan

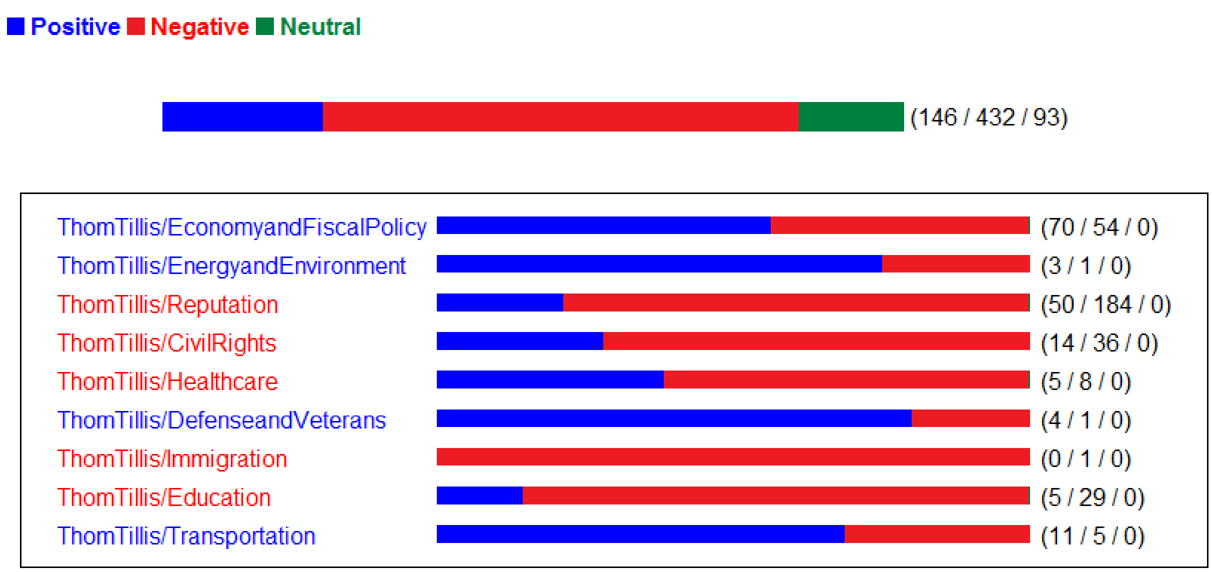

Figure 2. Sentiment Analysis for Thom Tillis

For both Hagan and Tillis, Reputation and EconomyandFiscalPolicy are the topics most often addressed. Education is far more prominent in the Hagan documents than in the Tillis documents, whereas CivilRights figures more prominently for Tillis. Immigration is rarely mentioned in conjunction with either candidate. The DefenseandVeterans, EnergyandEnvironment, Transportation, and Immigration topics are mentioned in less than 5% of documents for either candidate, suggesting that these topics might be of little interest to constituents who participate in social media and news discussion. In the Tillis corpus, Healthcare and Education are also mentioned in less than 5% of documents.

Hagan			Tillis			Topic
POS	NEG	TOTAL	POS	NEG	TOTAL	
166	213	**379**	70	54	**124**	EconomyandFiscalPolicy
98	249	**347**	50	184	**234**	Reputation
35	64	**99**	5	29	34	Education
24	82	**106**	5	8	13	Healthcare
8	43	51	14	36	**50**	CivilRights
16	26	42	4	1	5	DefenseandVeterans
9	8	17	3	1	4	EnergyandEnvironment
10	1	11	11	5	16	Transportation
5	14	19	0	1	1	Immigration

Table 5. Topic Mentions in Each Corpus

For the Kay Hagan corpus, negative sentiment outweighs positive sentiment in all high-frequency topics. Although positive sentiment outweighs negative sentiment for the EnergyandEnvironment and Transportation topics, there are very few mentions of these topics in the corpus.

The Tillis corpus shows stronger positive than negative sentiment on the EconomyandFiscalPolicy topic, a prominent topic in that corpus. Three other topics, DefenseandVeterans, EnergyandEnvironment, and Transportation also have positive sentiment, but these topics are underrepresented in this particular corpus.

PHRASE EXTRACTION ANALYSIS

When you augment the sentiment analysis by phrase extractions, you can analyze the relative frequency of phrase association, and you gain additional insights by the contrastive analysis of the candidates.

Contrastive Phrase Association

Using the steps outlined in the section "Textual Extraction and Tokenization," 31,351 phrases were extracted for Hagan, and 30,381 phrases were extracted for Tillis. The top 25 associated phrases for each candidate are shown in Table 6.

Ranking	Hagan	Tillis
1	s	jul.
2	senatorhagan	(919)
3	hagan	a.m.
4	obama	p.m.
5	loan	the state
6	senate	tillis
7	ď	thom tillis
8	a yes vote be	1974
9	31,	mccrory
10	senator	house speaker thom tillis
11	kay hagan	raleigh
12	aug	1929
13	democrat	today update
14	student	the house
15	washington	26,
16	2014	thomtillis
17	furlough	and they need to be compensate for

18	i	2013
19	Ll	teacher
20	gop	tax reform
21	president obama	copyright 2013 by capitol broadcasting company all right reserve this material may not be publish broadcast rewrite or redistribute read more posts from this blog
22	for	the house health care committee say friday the legislation will be consider tuesday senate republicans be criticize by abortion right advocate and democrat for add abortion regulation to an unrelated bill
23	student loan	this week that will implement new state rule on abortion
24	sen kay hagan	north carolina house panel be schedule to review a bill approve by the senate
25	face	north carolina

Table 6. Top-Ranked Phrases for Hagan and Tillis, by Association Strength

In this small, experimental sample, a picture emerges of each candidate. Several of the phrases in the Hagan corpus refer to national concepts (such as *American,* rank=26; and *the white house,* rank=78*),* national figures (such as *obama,* rank=4; *mcconnell,* rank=52; and *reid*, rank=136*)*, and various states (including *georgia*, rank=29; *kentucky*, rank=32; and *montana*, rank=114*)*, all of which indicate a stronger association between Hagan and national politics relative to Tillis. On the other hand, several of the Tillis phrases (such as *raleigh*, rank=11; *north carolina*, rank=25; *raleigh n c,* rank = 34; *mccrory*, rank=9; *city*, rank=42; and *state government,* rank=*59)* point to an association with local politics. This is to be expected, because Hagan is currently a US Senator and Tillis is currently the state House Speaker. Although this small study is experimental, it might be advisable for Hagan to increase her association with state politics and local concerns if a similar pattern emerges in a larger data set. Similarly, Tillis could interpret this output as an indication that a stronger national presence could help his campaign.

Both candidates have dates associated with them. The dates that are more highly associated with Hagan in this data set (*2014*, rank=16; *2012*, rank=40; and *2001*, rank=54) are recent or future dates relative to the dates of publication of the data. On the other hand, the dates most strongly associated with Tillis (*1974*, rank=8; *1929*, rank=12; and *2013*, rank=18) include some that are significantly in the past. The dates that are associated with Hagan are related to electoral politics and recent financial turmoil. They are accompanied by terms such as *new* (rank=36) and *last year* (rank=43). The 1929 and 1974 dates in the Tillis refer to a historical civil rights issue, and occur with phrases such as *eugenics* (rank=136), *to receive 10 million in compensation* (rank=182), and *sterilization* (rank=209). If this pattern emerges in a larger data sample, it could indicate that Hagan has a tendency to be associated with more modern concerns, or that Tillis is typically associated with an awareness of historical events.

There is a strong association displayed between Hagan and her political party, with *democrat* (rank=13) being highly associated with Hagan. It is interesting that the words *republicans* (rank=82) and *gop* are both associated more with Hagan than when they are with Tillis. Tillis does not have a strong association with *republican* in the data set, unless it is in the context of a longer phrase: *by the full house and senate before it can go to mccrory 's desk for his signature republican have comfortable margin in each chamber so passage be likel*y (rank=51). These results might be a reflection of Hagan's current role in national politics as opposed to Tillis's current position at the state level.

Both Hagan and Tillis are associated with financial concerns; the extracted associated phrases show that they are associated with different financial issues. In this data set, Hagan is strongly associated with banking (*loan*, rank=5; *student loan*, rank=23; *rate*, rank=73; *mortgage*, rank=178; and *subsidize loan,* rate=182), whereas Tillis is associated with other financial areas, especially tax concerns (*tax reform,* rank=20*; economic development,* rank= 52; and *will energize the state economy and give tax relief to,* rank=53).

Sentiment

Sentiment analysis for each candidate was performed with each phrase in a separate document. The model used generic features, because candidate names are not usually mentioned in individual phrases.

The phrases that are associated with Hagan contain 1,686 phrases that were identified as positive, 3,681 phrases that were identified as negative and 25,984 neutral phrases. EconomyandFiscalPolicy and Education are both assigned negative sentiment by the model, although very few phrases expressed sentiment towards Education in the Hagan corpus. Reputation, EnergyandEnvironment, HealthCare, CivilRights, and DefenseandVeterans were all assigned negative sentiment. Transportation and Immigration were not assigned sentiment by the model in any of the phrases associated with Hagan. The overall sentiment of the phrases associated with Hagan in the contrastive analysis is negative.

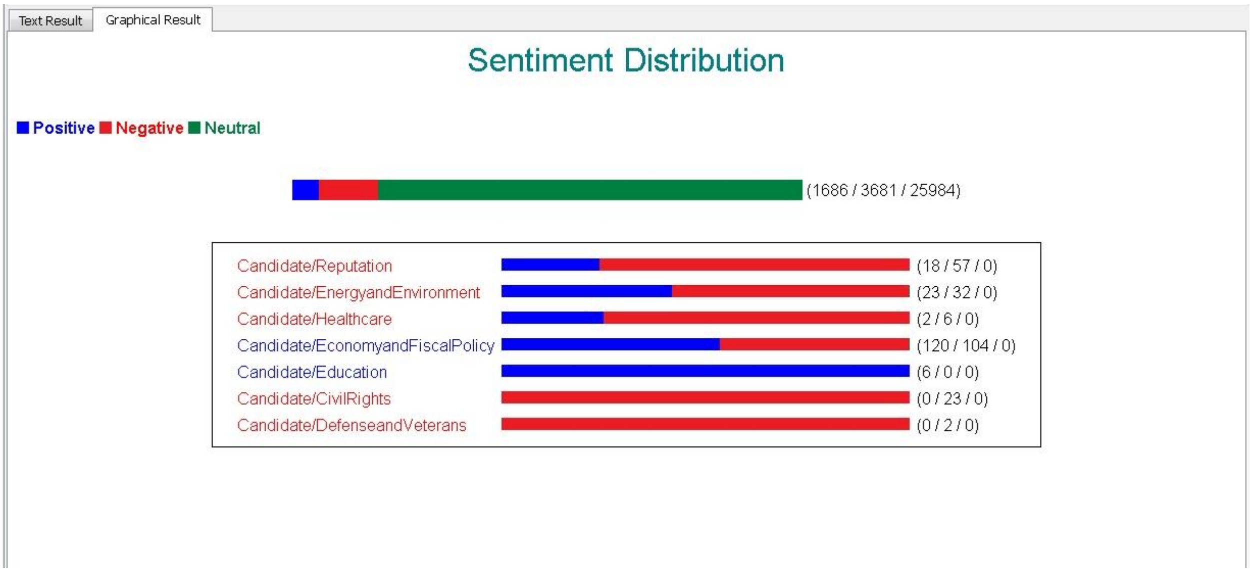

Figure 3. Sentiment Analysis of Hagan Phrases

The contrastive phrase analysis adds to the profile of Hagan that was performed using sentiment analysis on the raw document set. For example, document-level analysis of the Hagan corpus showed overall negative sentiment towards the EconomyandFiscalPolicy topic, but the sentiment is positive overall in that topic's phrases that are associated strongly with Hagan in contrast to Tillis. Although the document-level sentiment analysis did not indicate much relevance for the EnergyandEnvironment topic, the contrastive phrase analysis shows terms such as *climate change* (found in 14 phrases) and *epa* (found in 43 phrases) are associated with Hagan rather than Tillis. Candidates might want to use this type of information to determine what efforts they want to make to more accurately reflect their priorities.

The phrases that are associated with Tillis include 1,481 positive phrases, 3,406 negative phrases, and 25,494 phrases classified as neutral. The DefenseandVeterans and Transportation topics have no matches in the phrase list. Among the other topics, EconomyandFiscalPolicy is assigned positive sentiment, and Education, CivilRights, Healthcare, Immigration, and EnergyandEnvironment are unclassified for sentiment. The overall sentiment is negative.

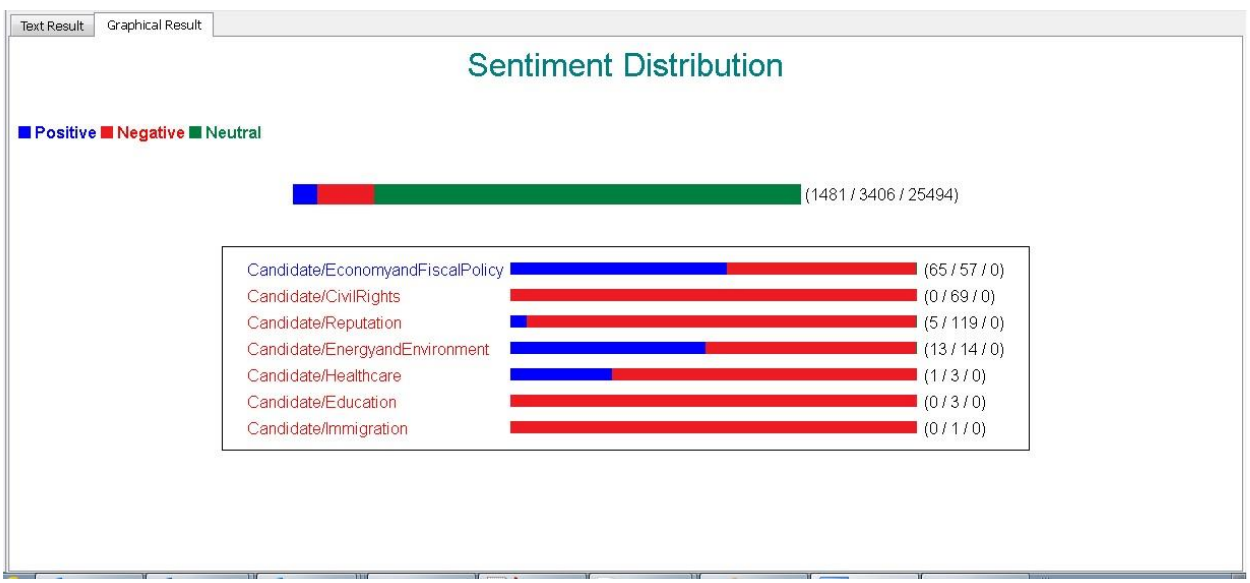

Figure 4. Sentiment Analysis of Tillis Phrases

The phrase level analysis for Tillis followed the patterns found in the document-level analysis, with EconomyandFiscalPolicy standing out as positive. As was the case for Hagan, the document-level analysis did not indicate much relevance for the EnergyandEnvironment topic, but the contrastive phrase analysis shows some

phrases, such as *fracking* (found in 44 phrases) *and landfill* (found in 25 phrases) that are associated with Tillis rather than Hagan.

CONCLUSION

This study finds that using phrase extraction and contrastive association together with sentiment analysis provides a more complete picture of associations of topics with public figures and of sentiment toward those figures. Even with a small, experimental data sample of about 2,000 documents, the topics that are discovered provide useful information about the contrasting candidate profiles to supplement the sentiment analysis of whole documents.

The data that are revealed in this approach to topic detection and sentiment analysis can be used to identify areas of constituent concern. Because the results reflect not only a candidate's own social media profile but also the candidate's social media profile as it contrasts with a rival's profile, they can be useful for suggesting campaign strategies for candidates for public office. Is it desirable to be strongly associated with one's own political party? Or does the candidate prefer to have a more independent profile? Is it beneficial for a candidate to be associated with historical civil rights events? If the candidate is indeed associated with the desired concepts, are these couched in positive or negative terms? How well these phrase associations fit the candidate's platform is one measure of how well they are communicating their message.

This type of analysis shows candidates a snapshot of how they are framed during a particular time period and can be a useful tool for reviewing and shaping a candidate's image to match the ideas that are important to the campaign.

REFERENCES

Bane, M. 2011. Python Implementation of "Gale and Sampson's 'Simple Good-Turing' Algorithm." Available at https://raw.github.com/maxbane/simplegoodturing/master/sgt.py. GitHub. 21 June 2011. Accessed on February 20, 2014.

Gale, W. and Sampson, G. 1995. "Good-Turing Frequency Estimation Without Tears." *Journal of Quantitative Linguistics*, 2(3), pp. 217–237.

Reckman, H., Baird, C., Crawford, J., Crowell, R., Micciulla, L., Sethi, S., and Veress, F. 2013. "Teragram: Rule-Based Detection of Sentiment Phrases Using SAS Sentiment Analysis." *International Workshop on Semantic Evaluation at *SEM, NAACL 2013*. Atlanta, Georgia.

Sykes, J. 2012. "Predicting Electoral Outcomes with SAS Sentiment Analysis and SAS Forecast Studio." *Proceedings of the SAS Global Forum 2012 Conference*, Cary, NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings12/131-2012.pdf.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. You may contact the authors at:

Michael D. Wallis
SAS Institute, Inc.
600 SAS Campus Drive
Cary, NC 27513
1-919-531-5886
michael.wallis@sas.com
http://www.sas.com