# Agreement and Disagreement:
# Comparison of Points of View in the Political Domain

**Stefano Menini**
University of Trento
Fondazione Bruno Kessler
Trento, Italy
menini@fbk.eu

**Sara Tonelli**
Fondazione Bruno Kessler
Trento, Italy
satonelli@fbk.eu

## Abstract

The automated comparison of points of view between two politicians is a very challenging task, due not only to the lack of annotated resources, but also to the different dimensions participating to the definition of agreement and disagreement. In order to shed light on this complex task, we first carry out a pilot study to manually annotate the components involved in detecting agreement and disagreement. Then, based on these findings, we implement different features to capture them automatically via supervised classification. We do not focus on debates in dialogical form, but we rather consider sets of documents, in which politicians may express their position with respect to different topics in an implicit or explicit way, like during an electoral campaign. We create and make available three different datasets.

## 1 Introduction

When it comes to evaluate whether the statements of two persons are in agreement or not about a topic, several past works approach the problem by classifying the single statements as supporting or opposing the topic, considering the task as a variant of sentiment analysis (Somasundaran and Wiebe, 2010).

These approaches proved to be reliable in specific settings, where the goal of the statements was to express support or opposition w.r.t. the topic. However, when applied to the political domain, they often result into an oversimplified representation of the dynamics involved in the comparison of two positions. In our view, several aspects contribute to the assessment of agreement and disagreement in the political domain, requiring to be properly addressed. As an example, let us consider two excerpts uttered by Kennedy and Nixon in 1960 about the situation in *Cuba* under the *Castro regime*:

> **Kennedy:** *"There is not any doubt we had great influence in Cuba, and I think it is unfortunate that we did not use that influence more vigorously to persuade Castro to hold free, open elections, so that the people of Cuba could have made the choice."*

> **Nixon:** *"What we must remember too is that the United States has the military power - and Mr. Castro knows this - to throw him out of office tomorrow or the next day or any day that we choose."*

The two examples show that neither Nixon nor Kennedy support the Castro regime and that they both share the same negative sentiment about it in their speeches. But beside being on the same side in contrasting the regime, their points of view on it are, from a pragmatic perspective, very different: while Kennedy supports free elections, Nixon does not hesitate to remark United States military power and the possibility to remove Castro in any moment. Overall, the two positions are in disagreement. This example shows that the sentiment and the proposed solution are two relevant aspects to define a politician's attitude w.r.t. a topic, and that the contribution of these two aspects need to be better studied when comparing different points of view: while one could argue that disagreement is expressed by

different sentiment and different semantic content, our pilot analysis shows that the boundaries are not so clear-cut.

The main contribution of this work lies in the presentation of a first feasibility study on manually annotated data from the political domain (Section 4), where we decompose the notion of agreement / disagreement, and in the presentation of a novel system (Section 5), which takes into account the insight gained during the feasibility study. We evaluate our approach on three datasets created for this task, which we make freely available[1], and show that our approach is effective on each of them: the one created for the feasibility study presented in Section 4.1, an extended version of the same dataset, and a larger one extracted from Debatepedia presented in Section 5.2.

## 2   Related Work

Given the highly polarized nature of political debates, we can find in the literature many works focused on classifying political statements as supporting or opposing a debated topic. This classification can be approached in different ways, for instance by emphasizing the role of sentiment polarity in a statement, or using topic modeling to define each position.

In (Somasundaran and Wiebe, 2009; Somasundaran and Wiebe, 2010) the authors propose a way to classify stances. They gather posts on different topics from online forums, and classify the statements from these debates as in favor or against the debated issue. They use the MPQA corpus to automatically generate a lexicon of entries indicative of a positive and negative argument and add information about the use of modal verbs and sentiment-based features. A similar task is proposed in Abbott et al. (2011) to recognize disagreement in online political forums between quoted text and a given response. The goal was achieved by using word-based features (for example discourse markers) as well as meta-information. Another work investigating stance classification of online posts was presented in Anand et al. (2011): there, the authors attempt to improve the unigram baseline by adding more cognitive-motivated features such as contextual information and opinion dependencies to define the target of opinion words. The results show that the use of these features improves classification results for many topics.

Another work from Gottipati et al. (2013) proposes to learn topics and support/opposition from discussions in Debatepedia by using a model based on the probabilistic distribution of the terms over the topics and the sides of the debate. A similar work on Debatepedia has been proposed by Awadallah et al. (2012), to classify quotes as belonging to a topic and supporting or opposing it.

Other approaches to classification rely on corpus-specific features, as in Thomas et al. (2006), who detect support and opposition to legislation in congressional debates by using speech transcriptions as well as records on voting, information about the speakers and the relations among them. Other works focus on the identification of agreement and disagreement in dialogues, such as (Galley et al., 2004; Hillard et al., 2003). They classify consecutive speech transcription segments produced by different speakers as positive or negative with respect to the discussed topic by using lexical, structural, and prosodic features.

Compared to previous works, our task is different in that we perform pairwise agreement/disagreement detection between two points of view: our focus is on the relation between the two rather than the single stance. Another difference lies in the types of textual units we want to classify: we do not work on single statements but rather on longer snippets including several sentences, based on the assumption that in the political domain a person's position with respect to a topic may not be overtly expressed. Our goal is to generalise over single statements and detect agreement and disagreement based on a broader, but not necessarily explicit, textual context.

## 3   Task Description

In the political domain, public debates in which two opponents discuss their point of view on specific topics, usually suggested by a moderator, are just one of several occasions in which political agendas are described. If we consider an electoral campaign, for instance, candidates issue declarations, mostly in the form of speeches, in which several topics are more or less explicitly discussed, and only towards

---
[1]Available at `https://dh.fbk.eu/resources/agreement-disagreement`

the end of the campaign a direct confrontation between opponents takes place. While past approaches to detect direct support or opposition in dialogues (Galley et al., 2004; Hillard et al., 2003) could be appropriate to analyse such direct confrontations, they may fail to capture the complexity of other types of statements. For example, the limited length of turns in dialogues, the presence of specific emphatic expressions and the need to focus on one topic at a time all contribute to the automated detection of agreement or disagreement in direct confrontations, but this information is not necessarily present when we consider larger documents collections, containing the public declarations of a politician. This second scenario is the focus of the present work. Specifically, given two document collections containing the public declarations and speech transcriptions of two politicians, our goal is to assess whether the two agree or not on a topic. In this scenario, we cannot assume that, each time a topic is mentioned, a position is explicitly expressed, but rather that it can be understood given a set of statements related to the topic. Besides, we cannot assume that two datasets representing two politicians contain direct references or replies between the opponents on a given topic. In this complex scenario, we first define a methodology to process the document collections and extract the text passages to be compared and classified. Then, we propose an approach to classify such pairs based on supervised learning, that takes into account features capturing the relevant dimensions analysed in the pilot study.

## 4 Pilot Study

To investigate which dimensions are involved in the perception of agreement and disagreement in the political domain, we first perform an exploratory study by manually annotating a dataset created for the task.

### 4.1 1960 Presidential Campaign Dataset

We collect the transcription of discourses and official declarations issued by Nixon and Kennedy during 1960 presidential campaign from The American Presidency Project[2]. The corpus includes 881 documents[3] and more than 1,6 million tokens (around 830,000 tokens for Nixon and 815,000 tokens for Kennedy).

We define 38 topics relevant to the electoral campaign with the help of a history scholar and we represent them via a set of manually defined keywords (e.g. a topic about the Agricultural program defined as *[agricultural program, agricultural policy, farmer, farm]*, about Education *[education, school]* or about Atomic energy *[atomic energy, nuclear energy, atomic power, nuclear power]*).

For each topic, for example *Education*, we extract from the two sets of documents all sentences containing at least one of the keywords defining the topic, plus the previous and the following sentence. The decision of extending the selection to three sentences is taken in order to have a more complete portion of text about the topic. These three sentences correspond to a *text excerpt*.

For each topic, we finally pair five random excerpts from Kennedy with five random excerpts from Nixon to create our *snippets*. We used this approach because the annotation task focused on general questions that needed enough context to be answered, and single excerpts may not provide enough information for the task.

### 4.2 Corpus Annotation

We build 350 snippets across all the topics and ask two independent annotators to annotate them using the CrowdFlower web interface[4]. We did not open the task to the public, but we relied on trusted annotators. For each pair, the following questions are asked:

1. *Are Nixon's and Kennedy's statements about the topic in agreement, disagreement or neutral?*

2. *What is Kennedy's sentiment with respect to the topic?*

---

3. *What is Nixon's sentiment with respect to the topic?*

4. *Are the solutions or initiatives proposed by Nixon and Kennedy similar or different?*

For each pair, we collect *i)* a judgment about the *agreement/disagreement* relation between Nixon and Kennedy on the topic (*Question 1*), *ii)* a judgment on Kennedy's and Nixon's *sentiment* w.r.t. the topic (*Question 2 and 3*) being either *Positive*, *Neutral* or *Negative* and *iii)* a judgment on the *solutions* proposed by the two candidates as *similar*, *different* or *neutral*, i.e. no solution proposed (*Question 4*). The inter-annotator agreement as provided by Crowdflower is 73% for Question 1, 69% for Question 2, 70% for Question 3, and 78% for Question 4. Crowdflower re-assigns then each snippet for which there is no agreement to one of the three categories based on a 'confidence score' (we could not obtain the exact formula used to compute this value but we assume it takes into account annotators' reliability).

### 4.3 Data Analysis

Based on the answers to Question 1, the dataset is composed by 203 pairs of snippets where Nixon and Kennedy agree on the topic, 97 pairs in which they disagree and 50 pairs where they neither agree nor disagree. We focus on the pairs in agreement or disagreement and further analyse the contribution of *sentiment* and *proposed solution* to the perception of agreement. Results are reported in Table 1. The analysis shows that being in agreement does not necessarily imply having the same sentiment or suggesting the same solution (no solution is suggested in most of the cases). In cases of disagreement, instead, a different sentiment and a different solution prevail.

Overall, if we cast the agreement / disagreement prediction on the basis of the sentiment, we can correctly guess 221 pairs over 350 (63.1%). Similarly, if we cast the agreement / disagreement prediction based on the fact that same or different solutions are proposed, we can correctly guess 173 pairs over 350 (49.4%). These values show that sharing the same *sentiment*, and proposing the same *solution*, *per se* are not reliable indicators of agreement.

| | Same Sentiment | Same Solution |
|---|---|---|
| Agreement Nixon-Kennedy (203 pairs) | 134/203 (66.0%) | 89/203 (43.8%) |
| | Different Sentiment | Different Solution |
| Disagreement Nixon-Kennedy (97 pairs) | 87/97 (89.7%) | 84/97 (86.5%) |

Table 1: Correspondence between agreement/disagreement and similar/different sentiment or solution.

We further analyse the direct relation between the sentiment and the semantic content (i.e. solution proposed) of the snippet pairs in the dataset with the help of Figure 1. The graph highlights that the pairs in which the two politicians share the same sentiment (on the left side) are split in half between similar and different solutions (on the right side). Instead, most of the pairs with a different sentiment (left side of the graph) end to different solutions, but only half of the pairs with different solutions (right side of the graph) derive from statements with a different sentiment.
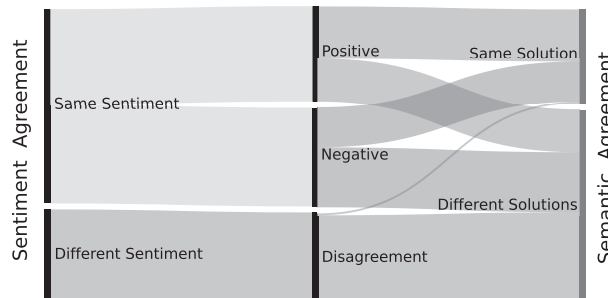


Figure 1: Relation between sentiment (annotations from Questions 2 and 3) and semantic content (annotations from Question 4) in Nixon's and Kennedy's statements.

The main findings of this pilot study can be summarised as follows: *i)* there is not a direct correspondence between the fact that Nixon and Kennedy agree on a topic and the fact that they share the same sentiment, *ii)* there is not a direct correspondence between the fact that Nixon and Kennedy agree on a topic and the fact that they propose the same initiative or solution, and *iii)* the sentiment and semantic content of the statements contribute both to defining agreement/disagreement between the two politicians. This implies that automatically classifying agreement and disagreement in this scenario needs to be addressed with appropriate features that can capture these different dimensions.

## 5 Experimental Setup

Based on the insight gained in the pilot study, we develop a system to automatically classify agreement and disagreement between two politicians (see Question 1 in the annotation task) trying to integrate information related to sentiment and semantic content of the statements (Questions 2-4). We classify pairs of *snippets*, one for each politician. For the task we adopt a supervised machine learning approach using LIBSVM (Chang and Lin, 2011) to train a Support Vector Machine (SVM). We evaluate our approach on the three datasets described in Section 5.2.

### 5.1 Features

We rely on three main categories of features. The contribution of speakers' sentiment to their agreement/disagreement is represented by a set of *sentiment-based features* (e.g. the sentiment of the statements and sentiment of the topic). To capture the differences and similarities in the proposed solutions, we use a set of *semantic features* (e.g word embeddings, cosine similarity and entailment), based on the intuition that two texts proposing the same solution are more likely to be semantically similar than two texts proposing different solutions. Finally, we add a set of *surface features* (e.g. the lexical overlap and the use of negations) that we expect to bear some information on the relation between the two snippets.
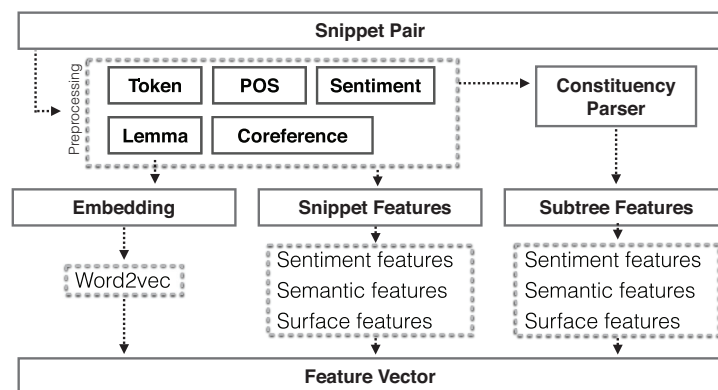


Figure 2: Features Extraction Pipeline

Figure 2 shows the pipeline we implemented for feature extraction. After preprocessing, all the features, except for word embeddings, are extracted at two different levels of granularity. In the first one, we extract the features at snippet level, which provide more information about the context in which the topic is used. In the second one, we focus only on the portion of text directly related to the topic. We use the Stanford Parser (De Marneffe et al., 2006) to extract from the sentences in the snippets all the subtrees containing the keywords representing the topic, and then we extract the features from them. With this pruning, we focus less on the context and more on the information which is directly related to the topic.

For each pair of snippets, we extract the following features:

**Sentiment information:** These features are inspired by previous works using sentiment information to predict a speaker's opinion on a topic (Pang and Lee, 2008; Abbasi et al., 2008). Each pair is represented by four sentiment scores, two scores for each snippet in the pair. We rely on the sentiment analysis module in the Stanford CoreNLP (Socher et al., 2013). We use a global sentiment score for each snippet (obtained by the average of the sentiment score of each sentence in it), and a score for the sentiment in

the subtrees related to the topic (obtained by the average of the sentiment score of each content word in the subtrees).

**Word embeddings:** Past works showed that word embeddings are an effective tool to define ideological positions in political documents (Iyyer et al., 2014). In our case, we do not focus on the sentence level, but rather on the keywords defining the topics debated in our pairs. We treat each snippet separately and we obtain two vectors for each pair: one vector representing the keywords of the topic in the first snippet and one vector for the topic in the second snippet (e.g. a vector for *Castro Regime* in Kennedy and a vector for *Castro Regime* in Nixon). The vectors are extracted using Word2vec (Mikolov et al., 2013) on each snippet (425 words on average with the topic occurring multiple times), with continuous bag-of-word algorithm, a windows size of 8 and a vector dimensionality of 50. We use this feature assuming that, when two people agree (i.e. have a similar point of view) on a topic, their respective vectors are more similar than when they are in disagreement.

**Cosine similarity:** In addition to the representation of the topics based on word embeddings, we use a set of features to quantify the relatedness between the way the two speakers talk about the topic. From each snippet in the pair, we extract two types of semantic vectors based on the co-occurrences (Turney et al., 2010) of topic keywords: one is computed over the entire snippet, while the other considering only the topic subtree. Co-occurrences are extracted from a window of 8 tokens and weighed using local pointwise mutual information. We then compute the cosine similarity between the vectors of the two sides of the snippet.

**Entailment:** The presence of entailment between the two snippets can be relevant to define if the position expressed by a speaker is accepted by the other (Cabrio and Villata, 2012). For this feature, we use the Excitement Open Platform (Magnini et al., 2014). For each pair, we use information about the entailment between the two snippets (in both the directions) and between the text in the subtrees related to the topic (in both the directions).

**Lemma overlap:** Past works showed that lexical overlap contributes to determining topical alignment between two texts (Somasundaran et al., 2009). Therefore, we compute lemma overlap of nouns, verbs and adjectives between two snippets. Although lexical overlap is already integrated in the textual entailment features, we believe that this information can provide useful information also in isolation.

**Negation:** For each snippet, we extract two features related to explicit negation cues (e.g. *not*, *don't*, *never*), adopting the list used in Councill et al. (2010). Using the parse tree of the snippets, we identify the words under the scope of a negation, and then consider as features *i)* the number of negated words in each snippet (normalized to its length) and *ii)* the percentage of the overlapping lemmas that in one snippet are under a negation. We expect that, if the same words are negated in a snippet and not in the other, this information can shed light on the relation between them.

## 5.2 Datasets

We evaluate our approach on three different datasets. The first is the 1960 Presidential Campaign dataset presented in the pilot study (300 snippet pairs). Given the limited number of pairs in this dataset, we create an extended version of it as follows: given a snippet pair related to a topic, we randomly replaced in each snippet two of the five excerpts belonging to it with others from the same politician, on the same topic, and with the same agreement/disagreement. By swapping random pairs of excerpts between snippets, we generated new ones and we were also able to better balance the agreement/disagreement proportion. Overall, the final dataset contains 1,400 pairs, balanced between the agreement/disagreement classes.

We further wanted to measure the impact of our approach on a larger corpus, more suitable for machine learning tasks. Therefore, we extract from Debatepedia[5] pairs of snippets compliant with the structure and the content of the two other datasets. We choose Debatepedia, an online encyclopedia of debates, because it provides statements from two opposing sides debating on well-defined, controversial topics. In particular, for each topic, Debatepedia gathers a set of relevant evidences and statements, mainly from news, that are framed as being in favour or against a specific debate question, e.g. "Is the $700 billion

---

[5]http://www.debatepedia.org/

bailout for the 2008 US financial crisis a good idea?" (Figure 3). In this way we have a large amount of snippets clustered into topics (e.g. individual rights, public safety, clean energy) and already structured as in agreement or not.
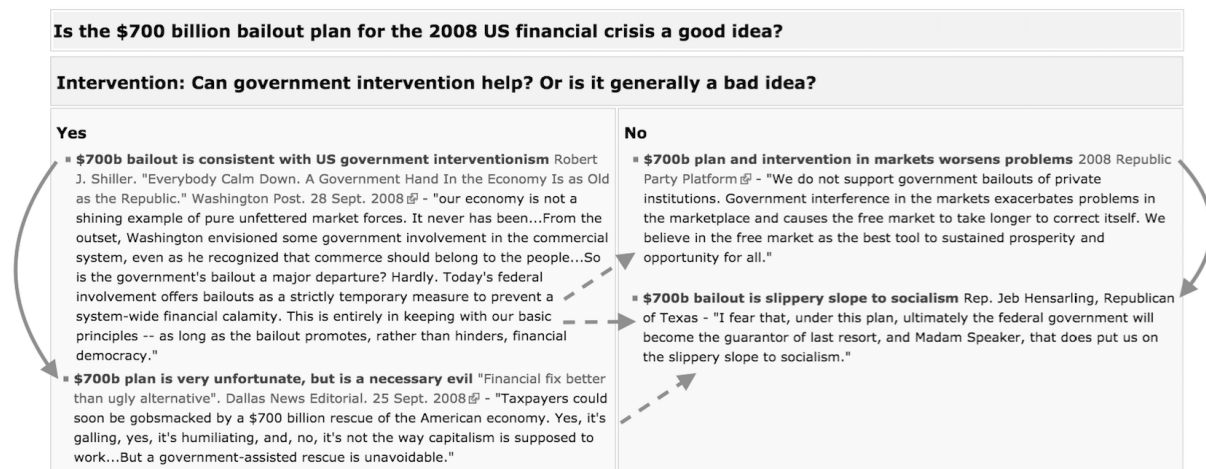


Figure 3: The Debatepedia Structure: the list of statements supporting the topic are on the left side and the list of statements opposing the topic on the right side. The solid arrows connect a pair in agreement, while the dashed lines link pairs in disagreement.

We collect from Debatepedia 646 debates for a total of 17,055 unique statements. To simulate our task, we need to organize the statements into pairs and associate each pair to a topic expressed by a keyword appearing in both statements. The pairs were created as indicated in Figure 3, by coupling two snippets being on the same side (for the agreement cases) or being on two opposite sides of the debate (for the disagreement cases).

In order to identify a keyword shared between two paired snippets, we first select only the pairs sharing at least one noun (74%). Then, we manually revised the 500 nouns most frequently overlapping, keeping only those that are not too generic and are related to the given topic of the debate. This led to a final list of 164 nouns, used as topics, for a total of 29,354 pairs. This final set of pairs, used in the classification task, is still balanced, with 14,042 pairs marked as in agreement and 15,312 in disagreement.

## 6 Evaluation

We report in Table 2 the results obtained on the three datasets using SVM with radial kernel (10-fold cross validation). A random baseline corresponding to the majority class is also reported. Beside evaluating the classifier performance with all features, we also analysed the contribution of single features (i.e. *negation/overlap*, *entailment*, *sentiment*, *cosine*, *word embeddings*) and their combinations to the task. Finally, we perform another evaluation, adding to the features mentioned before the outcome of the coreference resolution system in StanfordCoreNLP (Lee et al., 2011). Our intuition was that, since our snippets usually include more than one sentence, resolving pronouns and coreferential expressions may make the content more explicit, thus enabling a better agreement detection. However, results show that this information causes a slight performance drop. This can be due to coreference resolution errors rather than the feature itself, and we plan to further investigate this issue in the future.

Evaluation results confirm the findings suggested by the feasibility study: considering only sentiment-based features, or those related to semantic content is not effective as combining the two information layers (*Sent+Entailment+Embeddings+Cosine*). Surface features still have an impact on the outcome of every run, although limited.

If we compare the results obtained on the three datasets, we observe that the limited size of the 1960 Elections dataset affects classification, because some configurations yield the same performance as the random baseline. On the two other datasets, instead, the selected features are effective for classification, with a better performance achieved on the Extended 1960 dataset. This is due to the fact that all snippets

| Used Features | 1960 Elections<br>Accuracy | Extended 1960 Elections<br>Accuracy | Debatepedia<br>Accuracy |
|---|---|---|---|
| Random Baseline | 67.6% | 50.0% | 52.2% |
| Negation+Overlap | 67.6% | 52.7% | 54.5% |
| Entailment | 68.7% | 55.7% | 55.6% |
| Sentiment | 67.6% | 56.2% | 54.7% |
| Cosine | 67.6% | 67.5% | 53.0% |
| Word Embeddings | 76.3% | 77.5% | 67.3% |
| Sentiment+Entailment | 68.6% | 60.0% | 57.6% |
| Sent+Entailment+Embeddings | 81.4% | 79.7% | 72.9% |
| Sent+Entailment+Embeddings+Cosine | 81.7% | 79.8% | 73.1% |
| All features | **83.0%** | **80.1%** | **74.0%** |
| All features + coreference | 81.6% | 79.0% | 73.7% |

Table 2: Classification results on the 1960 Elections dataset (300 pairs), Extended 1960 Elections dataset (1,400 pairs) and Debatepedia Dataset (29,354 pairs)

are extracted from documents by Nixon and Kennedy, so that language and style are consistent across all pairs. Debatepedia, instead, relies on a wide range of sources, thus language variation is much higher.

We finally run a last experiment to test our approach on a different task, i.e. the classification of single statements as supporting or opposing a topic. We argue that, even if we chose our features with a focus on *pairwise* comparison, some of them may be effective also for single snippets. We rely again on Debatepedia, with the goal of classifying single arguments as supporting or opposing a topic. To this purpose, we remove the features strictly related to the pairwise comparison (e.g. the lemma overlap or the cosine similarity), and then classify each snippet belonging to one of the 29,354 pairs of our Debatepedia dataset via 10-fold cross-validation. Using SVM, we yield an accuracy of 87.2%. This result shows that the core set of features used to capture the point of view at snippet level are effective both to perform comparisons and to detect support or opposition given a single statement and a topic. Gottipati et al. (2013) perform the same task on Debatepedia data using a topic model-based approach, and achieve 86.0% accuracy. Although the two results are not directly comparable, since their dataset comprising 3,000 snippets is not available, we can conclude that our approach is a reliable solution also for single argument classification and can generalise well over different tasks.

## 7 Conclusion

In this paper, we introduced a study on the different dimensions which contribute to define agreement and disagreement between points of view in the political domain. We presented a pilot study that highlights how agreement w.r.t. a topic is derived both from the sentiment about it and the solution proposed. We used then these two dimensions, complemented by other lexical features, to train a classifier that was tested on data from the 1960 U.S Presidential Election and from Debatepedia. With this approach, we were able to correctly classify agreement and disagreement with good accuracy.

In addition to SVM, we experimented also with Convolutional Neural Networks using the TensorFlow implementation (Abadi et al., 2015), configured with 10 layers, 100 nodes and 100 iterations. So far, the performance achieved with CNN is around 20% lower than with SVM on all datasets, therefore we did not report the details in the experimental description. We plan to further investigate the motivations behind this gap, and to continue experimenting with other TensorFlow configurations.

Another research direction that we plan to pursue is the role of neutral judgments when analysing agreement and disagreement: we collected 50 neutral judgments during the pilot annotation, but we discarded them because we wanted to focus on pairwise agreement or opposition. However, in a real application scenario, it would be very important to add also this class. This extension is currently ongoing. In the future, we also plan to include the module for agreement and disagreement detection in the platform for the analysis of political speeches presented in Moretti et al. (2016).

# References

Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2015. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015. *Software available from tensorflow. org*, 1.

Ahmed Abbasi, Hsinchun Chen, and Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3):12.

Rob Abbott, Marilyn Walker, Pranav Anand, Jean E Fox Tree, Robeson Bowmani, and Joseph King. 2011. How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Languages in Social Media*, pages 2–11. Association for Computational Linguistics.

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 1–9. Association for Computational Linguistics.

Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. 2012. Polaricq: Polarity classification of political quotations. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1945–1949. ACM.

Elena Cabrio and Serena Villata. 2012. Combining textual entailment and argumentation theory for supporting online debates interactions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, pages 208–212. Association for Computational Linguistics.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Isaac G Councill, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing*, pages 51–59. Association for Computational Linguistics.

Marie-Catherine De Marneffe, Bill MacCartney, Christopher D Manning, et al. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, volume 6, pages 449–454.

Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 669. Association for Computational Linguistics.

Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, and Noah A. Smith. 2013. Learning topics and positions from Debatepedia. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1858–1868, Seattle, Washington, USA, October. Association for Computational Linguistics.

Dustin Hillard, Mari Ostendorf, and Elizabeth Shriberg. 2003. Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003–short papers-Volume 2*, pages 34–36. Association for Computational Linguistics.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the Association for Computational Linguistics*.

Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 28–34. Association for Computational Linguistics.

Bernardo Magnini, Roberto Zanoli, Ido Dagan, Kathrin Eichler, Günter Neumann, Tae-Gil Noh, Sebastian Pado, Asher Stern, and Omer Levy. 2014. The excitement open platform for textual inferences. In *ACL (System Demonstrations)*, pages 43–48.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Giovanni Moretti, Rachele Sprugnoli, Stefano Menini, and Sara Tonelli. 2016. ALCIDE: Extracting and visualising content from large document collections to support Humanities studies. *Knowledge-Based Systems*, 111:100–112.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 226–234. Association for Computational Linguistics.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124. Association for Computational Linguistics.

Swapna Somasundaran, Galileo Namata, Lise Getoor, and Janyce Wiebe. 2009. Opinion graphs for polarity and discourse classification. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, pages 66–74. Association for Computational Linguistics.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics.

Peter D Turney, Patrick Pantel, et al. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37(1):141–188.