Political News Sentiment Analysis for Under-resourced Languages

Patrik F. Bakken, Terje A. Bratlie, Cristina Marco, Jon Atle Gulla
Norwegian University of Science and Technology
Trondheim, Norway
`bakken.patrik@gmail.com`, `terje.a.bratlie@gmail.com`,
`{cristina.marco,jag}@ntnu.no`

Abstract

This paper presents classification results for the analysis of sentiment in political news articles. The domain of political news is particularly challenging, as journalists are presumably objective, whilst at the same time opinions can be subtly expressed. To deal with this challenge, in this work we conduct a two-step classification model, distinguishing first subjective and second positive and negative sentiment texts. More specifically, we propose a shallow machine learning approach where only minimal features are needed to train the classifier, including sentiment-bearing Co-Occurring Terms (COTs) and negation words. This approach yields close to state-of-the-art results. Contrary to results in other domains, the use of negations as features does not have a positive impact in the evaluation results. This method is particularly suited for languages that suffer from a lack of resources, such as sentiment lexicons or parsers, and for those systems that need to function in real-time.

1 Introduction

In the rapidly changing World Wide Web, getting informed opinions about facts is becoming increasingly challenging for users. In online news, the same event can be presented from very different perspectives depending on the source. Journalism is supposed to be objective, yet opinions are also expressed in newswire text (Belyaeva and van Der Goot, 2008; Blaz et al., 2009). In this type of texts, opinions are generally expressed in a subtle way, by using language resources other than sentiment vocabulary, such as irony, sarcasm, metaphors, etc., as it is illustrated in the following example, where the use of an ironic comparison *som rundingsbøyer* 'as if they were human buoys' suggests a negative opinion:

> - Den nye forskriften betyr at i vannene der friluftsfolket fortsatt kunne få være i fred, i en liten bortgjemt vik, der er det nå åpent for skuterfolket å bruke turfolket som rundingsbøyer, sier Bjørn Hansen, på vegne av Naturvernforbundet i Finnmark. *(- Given by the new regulation, places near the waters where people still could have piece and quiet, such as small secluded coves, are now open for people on jet skis to make use of the former as if they were human buoys, says Bjrn Hansen, on behalf of the Nature Conservatory of Finnmark.)*[1]

In this context it is difficult for readers to have informed opinions about the events happening in the world. Thus there is a growing need for resources that can help readers filter out news information, so that they can have an informed opinion about the current events. These resources would be particularly useful in the domain of political news, where readers want to be informed about political parties, politicians, and policies. If successfully employing sentiment analysis in the political news domain, possible biases and opinions will be revealed, and readers will get a more complete and transparent news scenario to gain knowledge from, leading to more informed political opinions.

However, the unstructured nature of news articles together with the subtle ways of expressing opinions in these texts make this task particularly challenging for machines. To deal with these challenges, in this paper we propose an unsupervised machine learning approach that takes sentiment-bearing Co-Occurring

---

[1] `https://www.nrk.no/finnmark/_-de-blir-a-bruke-turfolk-som-rundingsboyer-1.12519921`

Terms (COTs) and negation words as features. We argue that this approach is suited for languages other than English where computational linguistics resources for sentiment analysis are scarce, and also for those systems that perform in real-time.

The contents of this paper are as follows. After introducing the state of the art in 2, in 3 we present the method used to deal with sentiment analysis in political news. Lastly, the results are evaluated in 4 and the paper closes with a discussion and conclusion in 5 and 6.

2 Related work

In sentiment analysis, classifiers have been trained to automatically detect the polarity and subjectivity of texts. The former takes into account that not all incoming text is opinionated, and that a system might have to distinguish between subjective and objective texts.

For example, (Yu and Hatzivassiloglou, 2003) focus on separating subjective texts from those that portray factual information. The latter assumes text to be opinionated, and thus classifies the text as falling in one out of two sentiment categories − in general, positive or negative (Pang and Lee, 2008). (Turney, 2002; Pang et al., 2002; Hu and Liu, 2004; Kim and Hovy, 2004, among others) focus on distinguishing an author's positive or negative opinion towards a certain topic or object. To perform these tasks, both supervised and unsupervised approaches have been used (Feldman, 2013).

A combination of these two classification tasks can be seen in (Wilson et al., 2005). Here, a two-step binary classification takes place, which firstly filters out neutral expressions, and secondly, classifies the polarity of the selected set of expressions. As (Mihalcea et al., 2007) suggest, improvements in the more challenging task of subjectivity detection might have a positive impact on polarity classification. By firstly increasing precision and recall in subjectivity detection, the performance of the second task in a live system will also be improved. In this paper we present work in this line, in that we perform a two-step classification task, where first subjective and second positive and negative texts are detected within those classified as subjective.

In both tasks, negation is often considered one of the most important training features. (Wilson et al., 2005) argue that a phrase's sentiment will be better understood if the contextual and prior polarity of the phrase is taken into account. In contrast, other research, such as (Kim and Hovy, 2004; Hu and Liu, 2004; Grefenstette et al., 2004), focuses on local negation - negation terms occurring within sentiment words. Negation is a very complex linguistic issue, with different semantic effects in the sentence depending on the scope of the negative term, see (Lasnik, 1972; Partee, 1992; Ladusaw, 1992, among many others). In order to determine the scope of the negative phrase, most computational approaches make use of a syntactic parser. However, for languages suffering from a lack of such resources, such as Norwegian, this strategy would be too expensive. Due to the lack of resources for the Norwegian languages, in the present paper we present a very simple method where only the presence of negative terms, such as English *not*, are considered.

Initial efforts for sentiment analysis in the newswire domain have been conducted by (Belyaeva and van Der Goot, 2008) and (Blaz et al., 2009). Similar work to that of ours has been completed for Turkish in the political news domain by (Kaya et al., 2012). In this work four supervised machine learning algorithms are evaluated, namely Naïve Bayes, Maximum Entropy, SVM and N-Gram character based Language Model, in which features are unigrams or single words pertaining to relevant morphosyntactic classes. Their approach yields a maximum accuracy of 76.78%. Our approach is different from the latter in that we focus on different machine learning algorithms, such as J48, known to be computationally faster. Besides, we employ a two-step binary classifier and analyze all paragraphs in a collection of newswire text, whereas they look at the overall sentiment of selected newspapers columns.

3 Method

Figure 1 shows a high-level overview of the two-step sentiment analysis approach presented in this paper. Firstly, the input dataset consists of paragraphs annotated with three different classes: positive, negative,
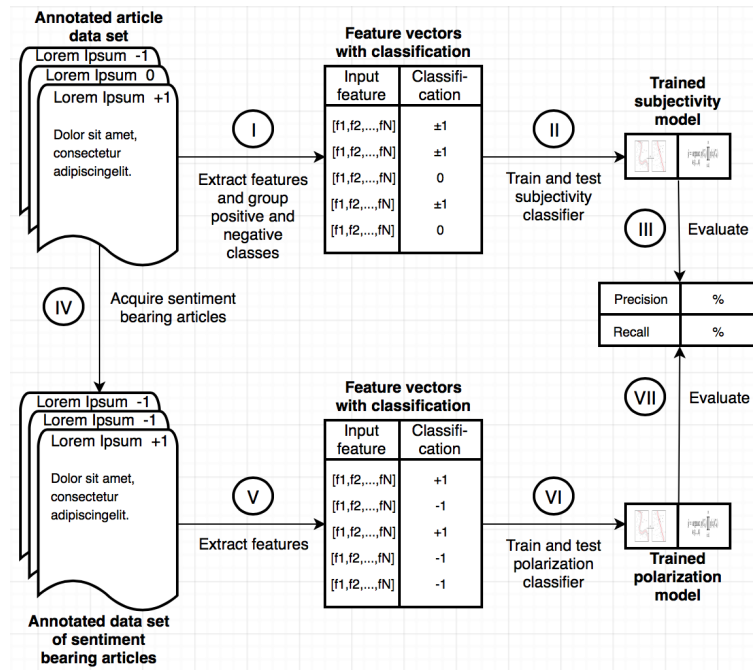
Figure 1: High-level system overview with two-step binary classification during testing and training.

or neutral. Features are then extracted from this dataset and used for training the subjectivity model. The precision of the resulting model is then evaluated using 10-fold crossvalidation. In the second phase, only the sentiment-bearing paragraphs from the dataset are used, and then the same procedure is followed for training and evaluation. We will now explain in further detail the different parts of our system.

3.1 Data and annotation

A total number of 3961 paragraphs within the political category was selected from the Norwegian online news sources *NRK*[2] and *VG*[3]. This dataset includes news articles over a span of four months during the summer of 2015, right before the municipal elections in October that year. Paragraphs were chosen instead of documents, sentences or phrases, because we noticed that in political news articles the presence of a sentiment target is more likely to be found at the paragraph level.

This dataset was annotated by the first two authors of this paper, obtaining an agreement of ∼76% ($\kappa = 0.62$), which is well within the range of other studies of sentiment analysis in similar domains (Njølstad et al., In press). The 3016 remaining paragraphs were used for training and testing using 10-fold crossvalidation. The paragraphs are labeled with one of the following three categories: positive, negative and neutral. The subjective class consists of both positive and negative paragraphs.

The criteria used for the authors to annotate the paragraphs were adapted from (Balahur and Steinberger, 2009), and are summarized in the following points: (i) world knowledge can be used if it is not clearly biased towards an entity; (ii) factual information should not be annotated as subjective; (iii) if polarity shifting occurs, the text should be annotated according to the one bearing the strongest sentiment and the most important entity; (iv) in case of uncertainty, the text should be annotated as objective.

3.2 Features

As summarized in Table 1, the features used to train the classifier are positive, negative and neutral Co-Occurring Terms (COTs), and negation words.

COTs are words with importance to each other that co-occur in a sentence (Pang and Lee, 2008; Matsuo and Ishizuka, 2004). In languages with a lack of computational resources to perform domain-

---

Features	Description	Type
PositiveCots	Number positive COTs in a paragraph	Discrete
NeutralCots	Number of neutral COTs in a paragraph	Discrete
NegativeCots	Number of negative COTs in a paragraph	Discrete
Negations	Number of negation words in a paragraph	Discrete

Table 1: Input features for machine learning classifier.

specific sentiment analysis, the use of COTs as a sentiment lexicon has been shown to work effectively. Such domain-specific lexicon of sentiment-bearing COTs can be acquired with a reasonable amount of manual labour (Njølstad et al., In press).

The COTs used in the present work are two-word co-occurring terms, with a maximum of four words between, and belonging to one of the following morphosyntactic categories: adjectives, nouns, verbs and adverbs. The *term frequency - inverse document frequency* ranking function was used to limit the lexicon size to 4000 COTs. This ranking is important as we only want to include COTs that are relevant to our domain of investigation. The Oslo-Bergen-Tagger[4] for the Norwegian language was used to tokenize and part-of-speech tag paragraphs. Table 2 shows an excerpt from the lexicon obtained. Details on how this lexicon was acquired are outside the scope of this paper, but are described in detail in (Njølstad et al., In press).

Term 1	Term 2	Translation	Sentiment
ta	*ordet*	'take the word'	0
er	*allerede*	'is already'	0
stor	*glede*	'big happiness'	1
er	*misfornyd*	'is unsatisfied'	-1
skape	*arbeidsplasser*	'achieve job positions'	1
kan	*svekke*	'can weaken'	-1
skal	*i stedet*	'will instead'	0

Table 2: Excerpt from sentiment lexicon.

The second set of features used in this paper are negations. Negations are words that change the polarity of words, phrases or sentences and have been shown to work effectively to detect sentiment (Wilson et al., 2005). Table 3 shows the Norwegian negative words considered as features in this paper. As the dataset includes paragraphs written in both *bokmål* and *nynorsk* Norwegian, negation words from both varieties are included in this table.

Norwegian *bokmål*	Norwegian *nynorsk*	English
ikke	*ikkje*	'not'
ei	*ei*	'not'
nei	*nei*	'no'
aldri	*aldri*	'never'
neppe	*neppe*	'hardly'
ingen, inga, intet	*ingen, inga, inkje*	'none, any'

Table 3: Negation words in the Norwegian language.

We hypothesize that there will be more negations in negative sentiment-bearing paragraphs, which can positively contribute to classify a paragraph's sentiment. In order to observe this before obtaining the features, we analyzed the average number of negation words in each paragraph. As can be seen in

---

[4]http://www.tekstlab.uio.no/obt-ny/

Table 4, on average each paragraph annotated as negative has 0.53 negations, compared to only 0.23 per neutral paragraph. Besides, the subjective (positive+negative) class has 0.43 negations per paragraph, whereas the neutral class has only 0.23. These numbers suggest that negations can also be relevant to detect subjective paragraphs.

Annotated class	Paragraphs	Negations per paragraph
Positive	698	0.29
Neutral	1426	0.23
Negative	892	0.53
Positive+Negative	1590	0.43
All	3016	0.33

Table 4: Average number of negations per paragraphs.

4 Evaluation

The main goal of this paper was to experiment with simple feature combinations in a two-step binary classification process, in order to achieve state-of-the-art results within the domain of Norwegian political news. Tables 5 and 6 summarize the evaluation of this work. Three different classifiers were used: J48, Random Forest (RF), and Naïve Bayes (NB). These classifiers were selected because of their computational speed, as opposed to the more computational heavy algorithms such as SVM (Zhao and Zhang, 2008). In addition, current research has shown that J48 and RF yield higher precision results in the financial news domain (Njølstad et al., In press). All three classifiers in our system are set up using the WEKA framework.[5]

Table 5 presents the results of subjectivity classification. Highlighted in bold is the precision of the feature combination with the best result − 67.1%. This feature combination includes all three types of COTs. However, it does not include negations at all. We hypothesized that the difference in negations per paragraph between the neutral class and the subjective class could have a positive impact on the precision results for this step. Looking at these results, where there is a higher precision without the use of negations for both NB and J48, our hypothesis does not hold after all. RF is the only machine learning model which benefits from negations, though this model yielded unsatisfying results in general.

	PosCots	NeutCots	NegCots	Negations	Precision
NB	✓	✓	✓		**67.1**
	✓	✓	✓	✓	66.6
				✓	59.9
RF	✓	✓	✓		62.5
	✓	✓	✓	✓	63.8
				✓	59.8
J48	✓	✓	✓		67.2
	✓	✓	✓	✓	67.0
				✓	61.8

Table 5: Subjectivity classification precision results with various feature combinations.

As can be seen from Table 6, there are more feature combinations in the polarity classification step. The goal of this step is to differentiate between positive and negative paragraphs, and thus we could experiment with combinations that did not include neutral COTs. As can be seen from the highlighted precision score, this is in fact what yields the best result of 73.2%, which is in line with the current

---

[5]http://www.cs.waikato.ac.nz/~ml/index.html

state-of-the-art. Before objective paragraphs were removed, the system only yields 59.7% precision, which shows that a significant improvement is obtained by using a two-step classification system. It is interesting to observe that negations do not have a positive impact on these results either. Throwing out COTs yields very low precision scores in all cases.

	PosCots	NeutCots	NegCots	Negations	Precision
NB	✓	✓	✓		72.8
	✓	✓	✓	✓	69.5
	✓		✓		**73.2**
	✓		✓	✓	71.4
				✓	57.4
RF	✓	✓	✓		64.2
	✓	✓	✓	✓	63.5
	✓		✓		66.1
	✓		✓	✓	64.2
				✓	55.6
J48	✓	✓	✓		72.2
	✓	✓	✓	✓	71.1
	✓		✓		72.3
	✓		✓	✓	71.6
				✓	—

Table 6: Polarity classification precision results with various feature combinations.

For both steps, NB is the machine learning algorithm that performs best. However, J48 outperforms the other two when including negations. It is important to note that there is no score for J48 with only negations included. This is because this model was not able to build a decision tree based on this feature. Lastly, RF is the worst performer in all cases.

5 Discussion

The results in tables 5 and 6 only include the overall precision scores obtained during testing. However, in each classification step there are separate precision and recall scores for each class involved in the classification. These results can shed some light on which parts of the system perform better than others. Tables 7 and 8 summarize those results.

Class	Precision	Recall
Neutral	57.1%	**85%**
Subjective	**76%**	42.7%
Overall	67.1%	61.1%

Table 7: Precision and recall for subjectivity classification.

As can be seen in Table 7, the subjective class achieves a precision of 76%, which means that in a live system where the input of polarity classification are only subjective texts, 76% of these will be correctly classified. On the other hand, there will be many false negative cases, as can be seen from the low recall value. In contrast, the recall yields 85% in the objective class, while the precision is lower, 57.1%. This means that our engine does not have enough information to decide when a paragraph bears a sentiment. Instead, the features represented in this classification step are not suitable in order to detect sentiment, and thus it misclassifies too many subjective paragraphs as objective. As discussed in 3.2, the difference in the average of negations per paragraph seemed to indicate that this feature was suitable for subjectivity

classification. However, this is not the case. We believe that the complexity of the scope negation can be the reason for this. In order to include negations and observe a positive impact in the classifier, the syntactic structure of the sentence should probably be considered. This could, in turn, make the system considerably lower and make this solution unpractical from a real-system point of view.

Similar results are presented in Table 8. The positive class shows poor results in terms of recall, though in terms of precision ranks higher than the overall, achieving satisfactory results. For the negative class, precision is good enough, 70% whereas recall yields a much higher value of 88.2%.

Class	Precision	Recall
Positive	**77.4%**	51.6%
Negative	70%	**88.2%**
Overall	73.2%	72.1%

Table 8: Precision and recall for polarity classification.

6 Conclusion and future work

In this paper we have presented a two-step classification system to detect sentiment in the political news domain for an under-resourced language such as Norwegian. COTs and simple negation counts were included as features in this system. By performing a 10-fold crossvalidation, we obtain close to state-of-art results in this task, over 70%, which is an optimistic value given the inherent complexity of this domain. Interestingly, negation words do not contribute to either classification task. We believe that this is because of the semantics of the scope of negation, that goes beyond singular words. We intend to investigate further on how to deal with negation in real-time systems in future work. All experiments have been conducted in the Norwegian political newswire domain at the paragraph level. In future work, we want to continue the work in this domain. More specifically, we plan to experiment with sentiment targets, such as events or named entities, within text units to detect the sentiment around them.

References

Alexandra Balahur and Ralf Steinberger. 2009. Rethinking sentiment analysis in the news: from theory to practice and back. *Proceeding of WOMSA*, 9.

Evgenia Belyaeva and Erik van Der Goot. 2008. News bias of online headlines across languages. *The study of conflict between Russia and Georgia.*

Fortuna Blaz, Carolina Galleguillos, and Nello Cristianini. 2009. Detecting the bias in media with statistical learning methods text mining: Theory and applications.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

Gregory Grefenstette, Yan Qu, James G Shanahan, and David A. Evans. 2004. Coupling niche browsers and affect analysis for an opinion mining application. In *Coupling approaches, coupling media and coupling languages for information retrieval*, pages 186–194. Le Centre de Hautes Etudes Internationales d'Informatique Documentaire.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Mesut Kaya, Guven Fidan, and Ismail H. Toroslu. 2012. Sentiment analysis of turkish political news. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 174–180. IEEE Computer Society.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics.

William A. Ladusaw. 1992. Expressing negation. In Chris Barker and David Dowty, editors, *Proceedings of the second conference on semantics and linguistic theory*, pages 237–59, Columbus. The Ohio State University.

Howard Lasnik. 1972. *Analyses of Negation in English*. Ph.D. thesis, Massachusetts Institute of Technology.

Yutaka Matsuo and Mitsuru Ishizuka. 2004. Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, 13(01):157–169.

Rada Mihalcea, Carmen Banea, and Janyce M Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections.

Pål-Christian Salvesen Njølstad, Lars Smøras Høysæter, and Jon Atle Gulla. In press. Optimizing supervised sentiment lexicon acquisition: Selecting co-occurring terms to annotate for sentiment analysis of financial news. *Language Resources and Evaluation*.

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

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.

Barbara H. Partee. 1992. On the "scope of negation" and polarity sensitivity. In Eva Hajicova, editor, *Functional Description of Language: Proceedings of the Conference*, Faculty of Mathematics and Physics, Charles University, Prague, 24-27 November.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136. Association for Computational Linguistics.

Yongheng Zhao and Yanxia Zhang. 2008. Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12):1955–1959.