

Integrating Viewpoints into Newspaper Opinion Mining for a Media Response Analysis

Thomas Scholz

Heinrich-Heine University
Institute of Computer Science
D-40225 Düsseldorf, Germany
scholz@cs.uni-duesseldorf.de

Stefan Conrad

Heinrich-Heine University
Institute of Computer Science
D-40225 Düsseldorf, Germany
conrad@cs.uni-duesseldorf.de

Abstract

Opinion Mining in newspaper articles is very interesting for media monitoring, because the results can represent the foundation for many analyses such as a Media Response Analysis. But for these kinds of analyses we need to know the viewpoint for an extracted sentiment. In this paper we apply various commonly used approaches for sentiment analysis and expand research by analysing the specific point of view and viewpoint features. The evaluation on real world datasets of German Media Response Analyses in contrast to a state-of-the-art method illustrates that our approach provide the assignment of a point of view for a given statement. Furthermore, we show how the features can increase the performance of existing solutions and how the viewpoints influence the polarity of sentiment.

1 Introduction

With the growth of online portals of newspapers and news services, Opinion Mining in news has become a central issue for media monitoring. Companies, organisations such as political parties, associations or maybe even distinguished public figures would like to know the polarity of sentiment (we also call it the tonality) in texts which concern them. How does the media talk about the new product of company XY? Is the tonality in the news changing after the speech of the chairman/chairwoman of the party? These questions are answered through a Media Response Analysis (MRA) (Michaelson and Griffin, 2005; Watson and Noble, 2007).

1.1 Media Response Analysis

In order to create a report about the media attention, an initiator of a MRA has first to define key words of interest like the name of companies (subsidiary companies, competitors), products or important persons (chairmen/chairwomen, press agents). Thereby, all texts containing these words are automatically collected by a crawler system. In a MRA, several media analysts read all texts, extract relevant statements from the texts and determine a tonality and the belonging viewpoint for every statement. The following statement is positive for US president Barrack Obama and the Democratic Party, e.g.:

- President Obama made the tough, politically unpopular decision to rescue an industry in crisis, a decision that saved more than 1.4 million American jobs. (Code: positive, Democrats)

To put this into practice today, a big human effort is needed. At the same time, the tasks are getting more and more difficult. The number of potentially relevant articles is increasing and the clients want to obtain their results in real time, because they have to be able to react quickly to dramatic changes in tonality. As a consequence, more machine-aided methods are needed for the field of Opinion Mining in newspaper articles.

While much research has been carried out on the detection on the sentiment (Pang and Lee, 2008), we want to concentrate on the detection of the viewpoint of a statement.

1.2 Problem Definitions

In a MRA, the statements have a tonality and an assignment of the viewpoint (this could be the client’s organisation or a competitor).

Definition I: Viewpoint Tonality. *Given a statement s which consists of the words w_i with $i \in \{1, \dots, s_n\}$. The t_1 which determines a tonality y and a point of view assignment g for the statement s .*

$$t_1 : s = (w_1, \dots, w_{s_n}) \mapsto (y, g); \quad (1)$$
$$y \in \{pos, neg\}, g \in \{g_1, \dots, g_m\}$$

The m different views are known before the analysis. The following statement has a positive sentiment for the German chancellor Merkel and her political party CDU, e.g.:

- That itself illustrates how important it was that chancellor Merkel has implemented reforms in Europe. (Code: positive, CDU)

This is the way state-of-the-art companies in media monitoring code their MRA results. A harder task (even for humans and therefore not the common procedure) is the determination of the tonality from a certain point of view.

Definition II: View Modified Tonality. *Given a statement s and a point of view g . The function t_2 which determines the tonality y_g from this point of view.*

$$t_2 : (s, g) \mapsto y_g \in \{pos, neg\} \quad (2)$$

In this example, the tonality y_g is modified by a given point of view g . For example, a statement that is considered positive from a certain viewpoint A, can be negative for another viewpoint B.

- The government quarrels, the SPD acts: The time has come to establish legal rules for a quota of women in commercial enterprises. (Code A: positive, SPD; Code B: negative, CDU)

The SPD is the biggest opposing party in Germany and political competitor of the CDU. While the point of view should be extracted from the statement itself for the **Viewpoint Tonality**, this task needs more external or world knowledge (the SPD is a competitor of the CDU, e.g.).

In this paper, we examine these problems. We present a new ontology-based approach for the determination of viewpoints. In addition, we explain viewpoint features which improve current methods in sentiment analysis for statements which are modified through a viewpoint. We evaluate our approach against a state-of-the-art method on two German corpora: A corpus of a real MRA about the finance industry and the public available pressrelations dataset of the election campaigns which consists of press releases of the CDU and SPD (Scholz et al., 2012). Furthermore, we want to analyse the influence of the viewpoint to the tonality.

This paper contains the following: In section 2 we analyse the related work, before we give a short overview on the basic polarity determination in section 3. In section 4 we present our viewpoint assignment algorithm and viewpoint features. After that we evaluate our methods on real world datasets of a MRA in comparison with DASA (Qiu et al., 2010), before we give a conclusion and a brief outlook on our future work in the last section.

2 Related Work

Recent research in sentiment analysis is focused on Opinion Mining in customer reviews (Pang and Lee, 2008).

One major aspect of Opinion Mining in product reviews is the collection of sentiment bearing words (Harb et al., 2008; Kaji and Kitsuregawa, 2007) or the construction of complex patterns which represent not only the sentiment, but also extract the relationship between the sentiment and the features of the product (Kobayashi et al., 2004). The point of view aspect plays less important role, because a customer expresses only one view (his/her own view) and different viewpoints mostly occur by comparisons of different products (Liu, 2010). As shown in the examples, the viewpoints are almost essential in the newsarea domain and some statements do not have any sentiment without a viewpoint.

Many techniques rely on a sentiment dictionary such as SentiWordNet (Baccianella et al., 2010). In this work, we apply the SentiWS dictionary (Remus et al., 2010). The sources of SentiWS are an English lexicon (General Inquirer),

which is translated into German, a large collection of product reviews, and a special German dictionary. They compute and refine a sentiment score for their words based on the three sources by applying the Pointwise Mutual Information method (Church and Hanks, 1989) for the score. Hereby, they obtain 1686 positive and 1818 negative German words in lemma.

In social media, approaches (Harb et al., 2008; Pak and Paroubek, 2010) analyse tweets from Twitter or blog-entries to extract negative and positive opinions. Pak and Paroubek (2010) are looking for emoticons (smilies) to determine the tonality of the tweets. Likewise, they collect some statistics about frequent POS-tags: Personal pronouns appear more often in a subjective tweet and verbs in past tense are a small hint for a negative tonality. This is unfortunately inappropriate in our case, because news articles do not contain emoticons.

However, far too little attention has been paid to Opinion Mining in newspaper articles and especially the integration of viewpoints. Most of the approaches on this topic only work with single words/phrases (Wilson et al., 2009) or reported speech objects (Balahur et al., 2009; Balahur et al., 2010; Park et al., 2011), because quotations are often more subjective. The opinion holder (the speaker) can be deduced in the cases and sometimes even the object of the opinion (e.g. another person or an organisation which appears in the reported speech). But this technique can only capture opinions which are part of a quotation in the news. We have examined 4000 statements of our evaluation dataset, of which less than 22% of the statements contain quotation marks and only less than 5% have a proportion of quoted text bigger than 50% of the whole statement. As a result, this technique cannot analyse 78% to 95% of the statements. Another approach (Devitt and Ahmad, 2007) works with news articles about a company takeover. It computes graph-based features which require a sentiment dictionary as well as a common dictionary to create the graph. The nodes are concepts of words and the edges represent relations between the words, but this approach does not handle different viewpoints.

Park et al. (2011) extract groups for a certain topic. The groups have contrasting views about

this topic. To extract these groups, they identify the speaker of a reported speech object who agree or disagree with other speakers or organisations of the same group and the opposing group, respectively. Unfortunately, this approach does not fit in with the requirements of a MRA, because the determination of tonality and not the different opinion groups are interesting for such an analysis. The different groups are commonly known from the beginning, anyway. Thomas et al. (Thomas et al., 2006) headed in the same direction by using a graph-based approach with same speaker links and different speaker agreement links in congressional debates.

Qiu et al. (2010) propose an approach called DASA for an online advertising strategy. They analyse web forums for an ad selection based on the consumer attitudes. Although their intention is something different, the basic problem is the same: the extraction of a viewpoint. They use syntactic parsing and a rule base approach to extract topic words which are associated with negative sentiment to propose products from rivals. We also expand this approach to non negative sentiments to create topic words and use this approach for our comparison.

3 Determination of Sentiment Polarity

First, we determine four basic tonality features (**Basic Tonality Features** α) based on the four word categories adverbs, adjectives, nouns, and verbs, which are the most important word classes for tonality (Bollegala et al., 2011; Remus et al., 2010).

For the weighting of the polarity (our tonality score TS), we use existing methods such as chi-square (Kaji and Kitsuregawa, 2007), the Pointwise Mutual Information method (Church and Hanks, 1989; Kaji and Kitsuregawa, 2007; Turney, 2002) and the German sentiment lexicon SentiWS (Remus et al., 2010). The chi-square value (Kaji and Kitsuregawa, 2007) is given by the statistical measure in which the null hypothesis is assumed which says that one word appears in positive statements with the same probability as in negative statements.

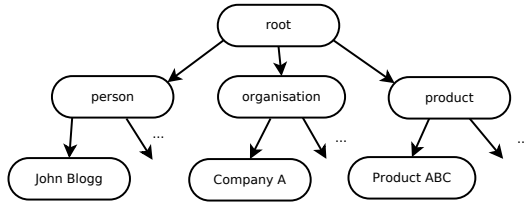


Figure 1: A sample ontology.

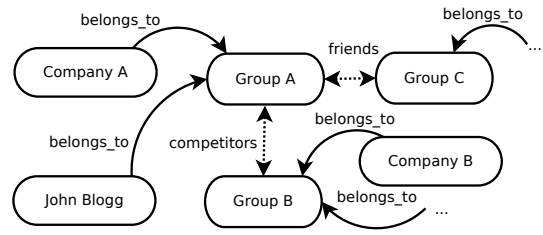


Figure 2: Sample ontology relationships.

$$TS_{\chi^2}(w) = \begin{cases} \chi^2(w) & \text{if } P(w|neg) < P(w|pos) \\ -\chi^2(w) & \text{otherwise} \end{cases} \quad (3)$$

The Pointwise Mutual Information (PMI) based tonality (Kaji and Kitsuregawa, 2007) uses the strength of the association between the word w and positive and negative statements, respectively.

$$TS_{PMI}(w) = \log_2 \frac{P(w|pos)}{P(w|neg)} \quad (4)$$

By using these methods, we construct dictionaries which contain words of the four categories weighted by the sentiment score TS .

We use the SentiWS (Remus et al., 2010) as another method. It contains a tonality value for 3504 words in lemma.

We compute four tonality features for one statement (**Basic Tonality Features** α). Every feature is the average of the tonality score in one category: The first feature is the average of the scores of all the statement's adjectives, the second of all nouns, and so on.

4 Viewpoint of Statements

For the viewpoint of a MRA it is very important to know which entities (persons, organisations, products) play a role in a given statement. To recognize viewpoints based on entities we propose an ontology based approach.

4.1 Ontology based Approach

Ontologies are very helpful in structuring knowledge. For our ontology based approach of viewpoint determination we organise our different entities in a hierarchical structure (shown in Figure 1). The entities are persons, organisations, and products. Persons can have a important role in an

organisation, e.g. a special function such as press agent or chairman. Organisations can be companies, political parties and so on. Products can be something the client's companies or the competitors are producing and/or selling.

All these entities have a group attribute. In other words, they belong to one group. Also, each group stands in relationship to every other group. These relationships can be neutral, friendly, or a competitive. Hence, every entity has one of these relationships to each of the other entities. Figure 2 shows an example.

It is not very time consuming to extract this information because media analysts keep this information in so-called code books. These pieces of information are used during the human analysis process (e.g. as tooltips) because it is very hard for humans to remember all the names and relationships between the entities.

4.2 Viewpoint Features

The approach considers how often the article mentions friendly entities and how often the competitors are mentioned. As a result, two features represent the persons and two features the organisations/products (**Viewpoint Features** β).

$$f_{\beta_1}(s) = \frac{pf(s, g)}{p(s)} \quad f_{\beta_2}(s) = \frac{of(s, g)}{o(s)} \quad (5)$$

In equation 5, $p(s)$ and $o(s)$ are the number of persons and organisations, respectively, in the statement s . $pf(s, g)$ and $of(s, g)$ are the friendly persons and friendly organisations/products, respectively, for group g . Friendly persons could be members of the initiator of the MRA (e.g. the managing director of the analysis customer) or a member of a cooperation organisation.

$$f_{\beta_3}(s) = \frac{pc(s, g)}{p(s)} \quad f_{\beta_4}(s) = \frac{oc(s, g)}{o(s)} \quad (6)$$

$pc(s, g)$ and $oc(s, g)$ are persons and organisations/products, respectively, of group g 's competitors. Friends and competitors are deduced by the relationships in the ontology.

Furthermore, the influenced tonality is stored by **Viewpoint Tonality Features** γ .

$$f_{\gamma_1}(s) = \sum_{w \in F_w} TS(w) \quad (7)$$

$$f_{\gamma_2}(s) = \sum_{w \in R_w} TS(w) \quad (8)$$

F_w and R_w are the sets of words which belongs to friend and competitor entities, respectively, and are determined in this way: Our method creates a scope around an entity. All words in the scope have a smaller distance to all other entities (in number of words between them). In the sentence "Merkel is acclaimed for her government's work, so the SPD is still performing relatively poorly in polls." the method would associate "acclaimed" with "Merkel" and "poorly" with "SPD".

4.3 Determination of the Assignment

To assign a viewpoint for the statements, our algorithm determines the probability of one entity belonging to one group (we use this algorithm also for the friend and competitor assignment). As a consequence, a statement belongs to one or more specific groups, if the probability is maximal under the assumption that all entities within this statement belong to this group (if the two probabilities are equal and maximal, the statement belongs to two groups and so on).

$$g = \arg \max_{g_i \in \{g_1, \dots, g_m\}} P(s|g_i) \quad (9)$$

$$P(s|g_i) = \sum_{e \in E_s} P(e|g_i) \quad (10)$$

The probability of one statement belonging to one specific group is the sum of all single probabilities with which entity e belongs to group g_i (E_s are all entities in statement s).

Algorithm 1: Entity Similarity

Data: statement entity e , group entity e_g
Result: similarity σ

```

1 if  $e$  and  $e_g$  are the same type then
2    $l_1 \leftarrow getListOfTokens(e)$ ;
3    $l_2 \leftarrow getListOfTokens(e_g)$ ;
4    $m = \max(getSize(l_1), getSize(l_2))$ ;
5   foreach token  $t_1 \in l_1$  do
6     foreach token  $t_2 \in l_2$  do
7       if  $t_1$  equals  $t_2$  then
8          $m = m - 1$ ;
9     end
10  end
11 end
12  $\sigma = 0.9^m$ ;
13 end
14 else
15    $\sigma = 0.0$ ;
16 end
```

Figure 3: Pseudocode of Entity Similarity.

For this purpose, an entity of statement s is compared to all entities E_{g_i} which belong to group g_i in the ontology.

$$P(e|g_i) = \max_{e_g \in E_{g_i}} (sim(e, e_g)) \quad (11)$$

The similarity function $sim(e, e_g)$ compares the name e of the entity in the text and the name of member e_g of group g . If they are the same, the value is 1.0. If e consists of the same tokens and only one token is different, the value is 0.9 (see the pseudocode in Figure 3). A token is one part of a name, e.g. the surname of a person. So, a name of a person could, for example, consist of two tokens: the first name and the surname. The method *equals* in Algorithm 1 checks, if two tokens are have the same string representation (e.g. both names start with "John").

This is useful, when persons are only mentioned with their surname or when a product's or organisation's name is not named in its entirety (company XY \rightarrow company XY Inc., product ABC \rightarrow product ABC international).

This also means that two persons, who share the same first name, could have a similarity of at least 0.9. This might sound unpleasant, but we have noticed that the persons are almost always mentioned by their full names in the complete articles, so we include a orthomatching module and a pronomial coreferencer (cf. section 5.2).

Group	Organis.	Persons	Products	all
Finance				
A	6	178	6	190
B	2	58	1	61
C	4	53	2	59
D	5	79	1	85
E	2	16	0	18
Politics				
CDU	9	237	0	246
SPD	9	149	0	158

Table 1: Size of the evaluation ontologies.

5 Evaluation

5.1 Experiment Design

We use two datasets for our experiments: The first test corpus called '**Finance**' (with 4,000 statements) represents a real MRA about a financial service provider and its four competitors. The corpus was created by up to ten media analysts (professional domain experts in the field of media response analyses). Their quality assurance ensures an inter-annotator agreement of at least 75% to 80% using Fleiss' kappa. For our second corpus which we call '**Politics**', we use the press-relations dataset¹ (Scholz et al., 2012) which consists of 1,521 statements from press releases and has two viewpoints (the two competitive parties: CDU and SPD). Here the agreement is 88.08% (Cohen's kappa, two annotators).

All statements (4,000 or 1,521, respectively) are annotated with a tonality and a viewpoint for this tonality (one of the five companies or the two parties, respectively). Small examples from the **Politics** dataset are mentioned in the first section.

For the **Finance** corpus, we create a RDF ontology by extracting all entities of the code book from customer group A (see Table 1; all companies' names are made anonymous for reasons of data protection). Group A has four competitors (group B to E) and the groups D and E have a friendly relationship, because D has taken over E in the first few days of the MRA. All other relationships are competitive. For the **Politics** corpus, we create an ontology in which CDU and SPD are competitors. We add all party members

¹<http://www.pressrelations.de/research/>

of the seventeenth German Bundestag² (the German parliament) and add some synonyms of the party and concepts such as 'government' or 'opposition' as organisations (see Table 1).

For evaluation of the **Viewpoint Features** β and **Viewpoint Tonality Features** γ , we use 30% of the statements to construct sentiment dictionaries which are weighted by the methods explained in section 3 and the remaining statements as training and test set (20% training and 80% test; we use this small training and big test set to guarantee that this approach will also work in practice). In addition, we use SentiWS (Remus et al., 2010) with 1686 positive and 1818 negative words as another baseline. We change the tonality according to viewpoint: the tonality is changed to the negative, if a statement is exclusively positive for a competitor and negative statements for a competitor become positive. For the classification, we use a SVM (Rapidminer³ standard implementation with default parameters) which performed better than other machine learning techniques (k-means, Naive Bayes) in this evaluation task. The evaluation shows the results in different combination of the features. So, $\alpha+\beta$ is the combination of set α and the feature set β and so on and all means the selection of all features.

5.2 Text Preprocessing

For better results, we analyse not only the statements, but also the whole text from which a statement is taken. The basic framework for our approach is GATE (General Architecture for Text Engineering)⁴, which provides a Named Entity Recognition (NER) for several languages. We tag POS with the TreeTagger (Schmid, 1995) whose results have been incorporated in the NER. We add the ontology entities and design new JAPE Rules⁵ to improve our NER: In order to guarantee that these entities are found in most case, the rules treat the ontology entities with the highest priority. Furthermore, we readjust the orthomatching module and edit some rules to avoid failures due to some of the names of organisations. In addi-

²collected from <http://www.bundestag.de>

³Rapid-I: <http://rapid-i.com/>

⁴GATE: <http://gate.ac.uk/>

⁵Developing Language Processing Components with GATE Version 6 (a User Guide): <http://gate.ac.uk/>

Method	$ s $	c	nc	w
DASA				
Viewpoint A	994	0.5613	0.2374	0.2012
Viewpoint B	1173	0.5303	0.2293	0.2404
Viewpoint C	812	0.5123	0.3067	0.1810
Viewpoint D	682	0.5967	0.2038	0.1994
Viewpoint E	339	0.6018	0.2655	0.1327
All	4000	0.5518	0.2458	0.2025
our approach				
Viewpoint A	994	0.7606	0.0986	0.1408
Viewpoint B	1173	0.7894	0.0648	0.1458
Viewpoint C	812	0.7906	0.0370	0.1724
Viewpoint D	682	0.8372	0.0440	0.1188
Viewpoint E	339	0.8348	0.0590	0.1062
All	4000	0.7945	0.0635	0.1420

Table 2: Results of the group assignment on **Finance** in comparison with DASA (Qiu et al., 2010).

tion, we implement a German pronominal coreferencer. Thus, we can resolve persons, organisations, and products in statements, even if they are only mentioned by he/she/it in the statement, but mentioned by name in the rest of the article.

5.3 Setup for DASA

For our comparison, we implement the DASA algorithm (Qiu et al., 2010). For the sentiment analysis step, we use the polarity annotation of the datasets to identify the opinion words with the same polarity, because we are more interested in the perspective component than in the sentiment analysis component. We calculate the dependencies between opinion and topic words with the Stanford Parser for German (Rafferty and Manning, 2008) and apply the rules R1 to R7 in descending order as described in Qiu (2010). Furthermore, we also expand their approach to non negative sentiments. For their task of selecting ads it is very reasonable to search for negative words and then to recommend rivals’ products, but we also need a viewpoint for non negative statements. We use the same RDF ontologies to create the input information for the DASA method, so that DASA knows the rivals and can use the entities as topic words for the assignment.

5.4 Experiment Results

Table 2 and 3 show the results of the viewpoints assignment. $|s|$ is the number of statements, c are the correctly assigned statements, w are the ones

Method	$ s $	c	nc	w
DASA				
CDU	992	0.5927	0.2258	0.1815
SPD	529	0.5350	0.2155	0.2495
All	1521	0.5726	0.2222	0.2051
our approach				
CDU	992	0.8700	0.0766	0.0534
SPD	529	0.6759	0.0964	0.2287
All	1521	0.8021	0.0835	0.1144

Table 3: Results of the group assignment on **Politics** in comparison with DASA (Qiu et al., 2010).

Method	α	$\alpha+\beta$	$\alpha+\gamma$	all
SentiWS	0.5066	0.5678	0.5200	0.5888
PMI	0.6094	0.6450	0.5909	0.6406
Chi-square	0.6275	0.6388	0.6272	0.6281

Table 4: Results of view based modified tonality on **Finance**.

incorrectly assigned, and nc could not be classified (the probability of all viewpoints is zero or DASA does not find a viewpoint, respectively). The average performance for statements are correctly classified in more than 79% or 80% of cases, less than 15% or 12% are classified incorrectly and over 6% or 8% are not classified at all. This is an improvement about 24% or 22% against the DASA algorithm.

We use the accuracy evaluation metric for the evaluation of the **View Modified Tonality**.

$$ACC = \frac{c}{n} \quad (12)$$

In equation 12, c is the number of correctly predicted statements and n is the number of all statements in the test set. In contrast to the view assignment, we only use the non neutral statements from the **Politics** corpus in this task (1,029 statements). But we use all 4,000 statements from **Finance**, because all statements are subjective (positive or negative).

Table 4 and 5 show the statement results of which the tonality is modified by a given point of view. The improvement expands from over 1% (chi-square, $\alpha+\beta$ combination on **Finance**) to over 8% (SentiWS, all features on both datasets). Almost all methods achieved the best results, if all features are combined.

In contrast, Table 6 shows the results using the **Basic Tonality Features** α , if the tonality is not

Method	α	$\alpha+\beta$	$\alpha+\gamma$	all
SentiWS	0.5342	0.5259	0.5259	0.6177
PMI	0.6294	0.6077	0.6260	0.6427
Chi-square	0.6694	0.6494	0.6678	0.6761

Table 5: Results of view based modified tonality on **Politics**.

Method	not mod. Finance	not mod. Politics
SentiWS	0.6455	0.5526
PMI	0.6393	0.6528
Chi-square	0.6848	0.6878

Table 6: Results of both datasets which are not modified through a viewpoint.

modified through a given view. The results are of course higher, because this task of a not modified tonality is simpler and the results do not provide any information about the viewpoint. But the combination of all viewpoint features approach these results (the PMI method or SentiWS with all features achieved even a better result on **Finance** or **Politics**, respectively).

5.5 Error Analysis

If we examine the statement set **Finance** for the group assignment task, the reason for the worst performance of the customer group is the type of the collected statements. The customer is given more attention during a MRA and so articles are collected which do not directly contain the known entities, but they include general messages: “Markets are suffering from the consequences of the economical crisis.” This is also the reason why this group has the highest rate of not classified statements.

One reason for the better performance of our approach against the DASA algorithm is the entity similarity and assignment of the most likely viewpoint which decreases the number wrong and especially not classified statements.

Both methods (DASA and our approach) perform only slightly better on **Politics**, although **Politics** has only two different viewpoints, because this area is more characterized by comparative statements. The press releases of CDU are talking much about the SPD and vice versa, what it makes more difficult to extract the correct viewpoint of a statement.

The results of the view modified tonality are not on such a high level. But this is a very hard task even for humans and not the state-of-art method to code statements in a real media analysis. Besides, the improvement of the viewpoint features approach the existing solutions of the not modified results which do not provide any information about the viewpoint.

6 Conclusion and Future Work

In conclusion, our ontology based approach provides a tonality based on a specific viewpoint. The group assignment algorithm and the viewpoint features allow the coding of statements into certain groups and tonality mutation based on viewpoint. Results of the evaluation suggest that both options (viewpoint assignment and viewpoint features) are possible for view based approach. The viewpoint features improve the accuracy (the results are not far away from the result, when the tonality is not modified through a view), while the assignment algorithm provide concrete viewpoint determination. This information is valuable and a two-process solution of calculating tonality and viewpoint fits the way of state-of-the-art companies coding their MRA.

But the evaluation also shows that the information itself about the viewpoint can increase the calculation of the tonality. So, our future work will cover the design of more techniques to improve the classification even further by using this approach for the viewpoint assignment first.

Acknowledgments.

This work is funded by the German Federal Ministry of Economics and Technology under the ZIM-program (Grant No. KF2846501ED1). The authors want to thank our cooperation partner pressrelations, especially Leszek Blacha for the creation of the dataset and Gudrun Karbe for helpful comments during the writing process.

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. of the 7th conference on International Language Resources and Evaluation (LREC’10)*.

- Alexandra Balahur, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, and Mijail Kabadjov. 2009. Opinion mining on newspaper quotations. In *Proc. of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 03*, pages 523–526.
- Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen, and Jenya Belyaeva. 2010. Sentiment analysis in the news. In *Proc. of the 7th conference on International Language Resources and Evaluation (LREC'10)*.
- Danushka Bollegala, David Weir, and John Carroll. 2011. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11*, pages 132–141.
- Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *Proc. of the 27th annual meeting on Association for Computational Linguistics, ACL '89*, pages 76–83.
- Ann Devitt and Khurshid Ahmad. 2007. Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*.
- Ali Harb, Michel Plantié, Gerard Dray, Mathieu Roche, François Troussel, and Pascal Poncelet. 2008. Web opinion mining: how to extract opinions from blogs? In *Proc. of the 5th international conference on Soft computing as transdisciplinary science and technology, CSTST '08*, pages 211–217.
- Nozohiro Kaji and Masaru Kitsuregawa. 2007. Building lexicon for sentiment analysis from massive collection of html documents. In *Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.
- Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, Kenji Tateishi, and Toshikazu Fukushima. 2004. Collecting evaluative expressions for opinion extraction. In *Proc. of the 1st International Joint Conference on Natural Language Processing (IJCNLP-04)*, pages 584–589.
- Bing Liu. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition*. CRC Press, Taylor and Francis Group.
- David Michaelson and Toni L. Griffin. 2005. A new model for media content analysis. Technical report, The Institute for Public Relations.
- Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. of the 7th conference on International Language Resources and Evaluation (LREC'10)*.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Souneil Park, KyungSoon Lee, and Junehwa Song. 2011. Contrasting opposing views of news articles on contentious issues. In *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11*, pages 340–349.
- Guang Qiu, Xiaofei He, Feng Zhang, Yuan Shi, Jiajun Bu, and Chun Chen. 2010. Dasa: Dissatisfaction-oriented advertising based on sentiment analysis. *Expert Systems with Applications*, 37(9):6182 – 6191.
- Anna N. Rafferty and Christopher D. Manning. 2008. Parsing three german treebanks: lexicalized and unlexicalized baselines. In *Proc. of the Workshop on Parsing German, PaGe '08*, pages 40–46.
- R. Remus, U. Quasthoff, and G. Heyer. 2010. SentiWS – a publicly available german-language resource for sentiment analysis. In *Proc. of the 7th International Language Resources and Evaluation (LREC'10)*.
- Helmut Schmid. 1995. Improvements in part-of-speech tagging with an application to german. In *In Proc. of the ACL SIGDAT-Workshop*, pages 47–50.
- Thomas Scholz, Stefan Conrad, and Lutz Hillekamps. 2012. Opinion mining on a german corpus of a media response analysis. In *Proc. of the 17th international conference on Natural language processing and information systems, TSD'12*.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of EMNLP*, pages 327–335.
- Peter D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proc. of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 417–424.
- T. Watson and P. Noble, 2007. *Evaluating public relations: a best practice guide to public relations planning, research & evaluation*, chapter 6, pages 107–138. PR in practice series. Kogan Page.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.