Research Group Social Computing
Department of Informatics
Technical University of Munich

Clustering-Based Sentiment Analysis for Media Agenda Setting

Wing Sheung Leung, Qiaoxi Liu
Department of Informatics, Technical University of Munich (TUM), Boltzmannstr. 3, 85748 Garching, Germany
July 31, 2020

Abstract — Most of the media agenda settings always are investigated in political events. However, it is also interesting to understand the effect from media to the public on other issues. In this paper, we investigate the agenda setting on an organic food domain. Although the salience on organic food domain is not as significant as the political domain, our results concluded from a clustering-based sentiment analysis approach turn out telling that media agenda settings exists in terms of discussion occurrences and sentiments polarity among different topics on organic food domain.

Acknowledgements

Through out the implementation of this project, there were many difficulties and unsatisfied results. Here we want to express our greatest thanks to Gerhard Hagerer for his guidance and Pan Yan for experience sharing and supports.

1 Introduction

An agenda setting study requires one to define a set of topics and to measure their salience. We propose a clustering-based sentiment analysis approach based on sentence level clustering and sentiment analysis for exploring a news corpus and measuring the media agenda by tagging sentences from news and its comments with topics. Eventually, in order to gain a insight towards sentiment distribution, four indicators are introduced to evaluate the statistical results as well as to construct a comprehensive analysis. Based on plenty of research on media agenda setting, four assumptions and questions are made as follows:

1. Coherent emotions: The emotion in the media can greatly transfer to the public.
2. Dominance/Attention of topics: The topic that media focuses on remains a leading topic for the readers.
3. Provoked fierce debate on topics: Whether/Which hot topic have the potential to arouse the similar topic(s) ?
4. What are the similarities and differences in the news media agendas across these two countries?

2 Methodology

A crucial part of an agenda setting study is the measuring of the media agenda. Since our approach based on sentence level, it carried out two parts for each individual sentence like following:

1. To define a set of agenda topics
   a) Topic discovery by unsupervised learning method k means
   b) Topic tagging by finding the top word list in each cluster
2. To assign a sentiment score between $[-1, 1]$

2.1 Data

In order to study the media agenda setting on the organic food domain, articles and their corresponding comments are crawled from two online news outlets, *nytimes.com (New York Times)*, *spiegel.de (Der Spiegel)* and a discussion forum, *quora.com (Quora)* if they consist of the search terms *organic food* and *organic farming* in English or *Bio-Lebensmittel* and *Bio-Landwirtschaft* in German. In this project, only the records labelled as 1.0 in 'relevant' field are used and Table 1 shows the statistics of these desired data.

As the media agenda setting is mainly described the influence of media towards its readers, com-

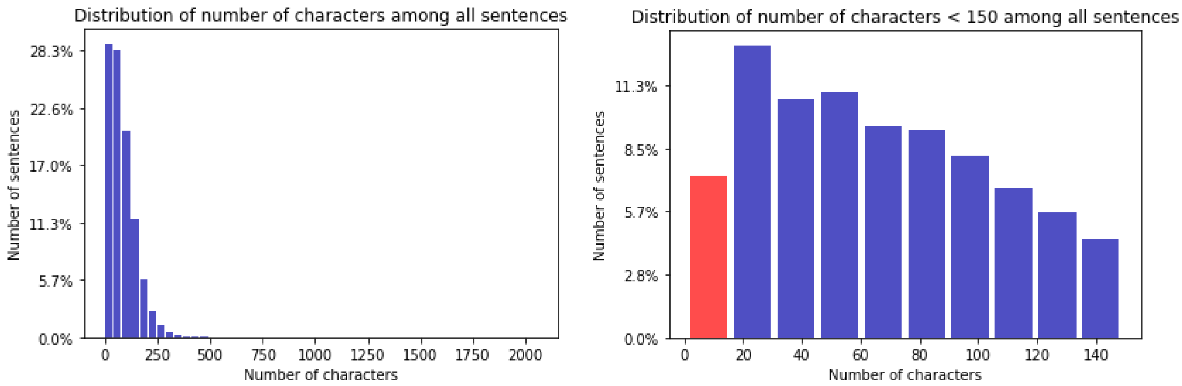


Figure 1 *Left*: 85.6% of sentences with smaller than 150 characters among all sentences. *Right*: 6.6% of sentences with character length smaller than 15 are ignored as they are inscrutable to be clustered into organic food related topics.

	Start	End	No. of articles
New York Times			
With comments	2006	2017	99
Without comments	1970	2017	228
Der Spiegel			
With comments	2007	2017	61
Without comments	2007	2017	91
Quora			
With comments	2009	2017	1304
Without comments	2010	2017	193

Table 1 Statistics for data crawled from *nytimes.com*, *spiegel.de* and *quora.com* with 'relevant' labelled as 1.0.

ments right under the news articles are treated as direct response while public discussions in *Quora* are counted as indirect response.

2.2 Text Preprocessing

In order to fit into the sentence-level-based analysis, plain text of articles and comments are first tokenized in sentences with *nltk.tokenize*. Both English and German text are well separated into sentences. Then, URLs, specially for those enclosed with HTML <a> tag, are replaced with string 'url'. After that, sentences with character length smaller than 15 are listed out. With manually observation, most of these sentences are determined as inscrutable to be clustered into organic food related topics. Since these sentences only take small portion, 6.6% as shown in Figure 1, they are ignored. In short, 127,464 English sentences and 200,627 German sentences, i.e. total 328,091 sentences are used for further analysis.

2.3 Sentence Embedding

With the list of sentences for each articles and comments obtained from text preprocessing stage, the Universal Sentence Encoder Cross-lingual (XLING) module is adopted for generating sentence embeddings. The module is built with the training setup based on the paper "Learning Cross-lingual Sentence Representations via a Multi-task Dual Encoder" [6] and dedicate for English and German sentences, phrases or short paragraphs which perfectly fit with our purpose. List of 512 dimensional vectors, i.e. sentence embeddings, are generated by passing variable length of English or German sentences into the module. Although the sentences are bilingual, XLING projects sentences with similar semantics closer together in the vector space. This property facilitates the next topic identification task.

2.4 Topic Modeling

K-means clustering algorithm is implemented on sentence embeddings. In order to get a promising and coherent clustering result, Elbow Method and Akaike Information Criterion (AIC) are first adopted to have a rough idea on what the potential number of clusters, k is. Then, *clarity* score calculation, which first introduced by Cronen-Townsend et al. [7] and enhanced by Angelidis and Lapata [3] on the determination of a ranked list of important terms that are the most characteristic for different aspects in the domain, is applied. As defined by Angelidis and Lapata [3], clarity measures how much more likely it is to observe word w in a subset of segments that discuss aspect a, compared to the corpus as a whole:

$$score_a(w) = t_a(w)log2\frac{t_a(w)}{t(w)}$$

where $t_a(w)$ and $t(w)$ are the l1-normalized tf-idf scores of w in the segments annotated with aspect a and in all annotated segments, respectively.

In another words, segments discussing aspect a are actually equivalent to sentences clustered into same cluster c in our settings. Besides, we try to get ranked list of symbolized terms for each cluster on each language l. Thus, the following clarity score calculation tells more concretely:

$$score_{l,c}(w) = t_{l,c}(w)log2\frac{t_{l,c}(w)}{t_l(w)}$$

where $t_{l,c}(w)$ and $t_l(w)$ are the l1-normalized tf-idf scores of w in the sentences labelled with cluster c and in all annotated sentences, respectively, for a certain language .

Listing the terms by their clarity score descending, the most characteristic terms for each cluster are found, i.e. top words. Finally, optimal k is determined over the potential ks obtained from Elbow Method and AIC with comparing the coherence of top words found by clarity scores among clusters. Last but not least, stopwords removal is also a concern when calculating the clarity score. In general, stopwords are the most common words in the documents and sometimes they are too dominant that interfere the result. Thus, terms appear the most frequently (TF-High) in documents among all the unique words for each language are shortlisted as stopwords according to the method suggested by Saif et al. [9] which is the inspired by Zipf's law. Figure 2 shows the rank-frequency distribution of unique words in both languages respectively with threshold frequency as 5000. It is believed that the TF-High stopwords defined are more specific to our organic food domain than those pre-defined stopwords corpus from famous online resources.

English
the, and, to, of, is, in, that, organic, it, are, not, for, you, food, as, have, be, on, with, they, or, but, this, from, more, we, do, can, there, if, by, at, all, foods, about, what, has, will, so, their, an, your, would, than, people, no, which, like, was, one, some, my
German
die, und, der, ist, das, nicht, von, in, es, sie, zu, den, ich, auch, mit, zitat, ein, sich, für, auf, man, sind, dass, aber, werden, wie, im, nur, oder, wenn, eine, so, bei, als, wird, aus, was, dem, noch, bio, an, dann, haben, kann, da, hat, mehr, wir, um, mal, doch, schon, ja, nach, sein, keine, immer, einen, des, gibt, hier, diese, durch

Table 2 TF-High stopwords with frequency larger than 5000.

2.5 Sentence Level Sentiment Assignment

Sentiment analysis is the process of computationally identifying and categorizing opinions expressed in text-based source materials [5]. Sentiment analysis can help agenda setting researchers discover the effective states and emotions expressed in messages. This analysis method recently has been commonly used for mining opinions on social media. By exploring the salience of positive and negative attributes associated to the agendas, researchers can investigate the second level agenda setting influence. In general, sentiment analysis has been investigated mainly at three levels: Document, Sentence and Entity. In our study, we determine whether each sentence expressed a positive, negative or neutral opinion.

As mentioned at the beginning, our goal is to transfer each free text sentence into a measurable object. Thus, we regard an sentence S(opinion) as a tuple which consists of two components: a topic t and a sentiment score s on the topic. i.e. (t, s). One Opinion can only have one topic as well as one sentiment score.

Not surprisingly, the most important indicators of sentiments are sentiment words. e.g. the adjective like *nice*, *interesting* are positive, or verb like *hate*, *discourage* are negative. Thus, our procedure for assigning sentiment score to a sentence S is based on aggregation of word sentiment:

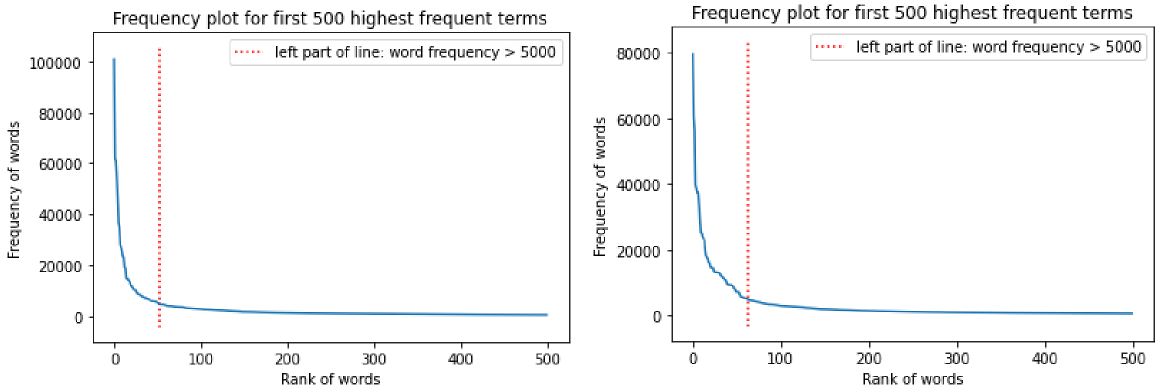


Figure 2 *Left*: Rank-frequency distribution of the top 500 English distinct words out of 43,255 words in the organic food corpus. All other words are removed from the plot for ease visualisation. 52 words, i.e. 0.12%, with frequency larger than threshold 5000. *Right*: Rank-frequency distribution of the top 500 German distinct words out of 73,081 words in the organic food corpus. 63 words, i.e. 0.09%, with frequency larger than threshold 5000.

1. Extract the words with part-of-speech tagging: ADJ, NOUN, ADV, VERB
2. For each tagged word $w \in S$, compute sentiment score $s_w \in [-1, 1]$ using pre-trained model
   - English: use Sentiwordnet [4] which based on synonym set to compute average sentiment score
   - German: use SentiWS [8] which based on semantic orientation calculated by Pointwise mutual information (PMI)
3. Assign the final score to S by taking average on the sentiment scores of all desired words in S, $s = avg(s_w)$

3 Results

3.1 Topic Modeling

3.1.1 K-Means Clustering on Multiple Languages Sentences Embedding

As XLING provides the ability to map both English and German sentences into vectors, k-means clustering is conducted with these sentence embeddings. From the perspective of Elbow Method, the change of sum of squared distances of samples to their closest cluster center along number of cluster k is shown in Figure 3. Since the change drops quite constantly, there is not indicative elbow helping to determine the suitable k.

While with AIC, a global minimum is found when k equals to 15 (Figure 4). Since we want to assign each cluster a topic, it is logical to also look into the topwords in each cluster before determine the optimal k. Thus, $k = \{13, 14, 15, 16\}$ are the potential k as there AIC values are so closed.

3.1.2 Word Cloud Generated Based on Clarity Scoring

We study the sample code for clarity score implementation from Angelidis and Lapata [3], they tried to use a global inverse document frequency (idf) generated by all segments (equivalent to sentences in our setting) for calculating all tf-idf scores. In our case, global idf is only for td-idf score calculation, $t_l(w)$, while idf for sentences labelled the same cluster is calculated separately for its own tf-idf, $t_{l,c}(w)$.

This clarity score calculation gives coherent results. For all the potential k, 11 reasonable topics related to organic food domain are distinguished from the word clouds (Figure 5, 6, 7 and 8). Besides these 11 topics, it is observed that other clusters are dominated by World Wide Web terms and very common words, like 'thanks', 'question', 'answer', 'mir' and 'ganz'. These clusters can be considered as garbage clusters as these top words cannot show descriptive view on the organic food domain. Indeed, as the k increase from 13 to 16, it is observed that only the garbage clusters differs. For example, it is found that a garbage cluster about 'references

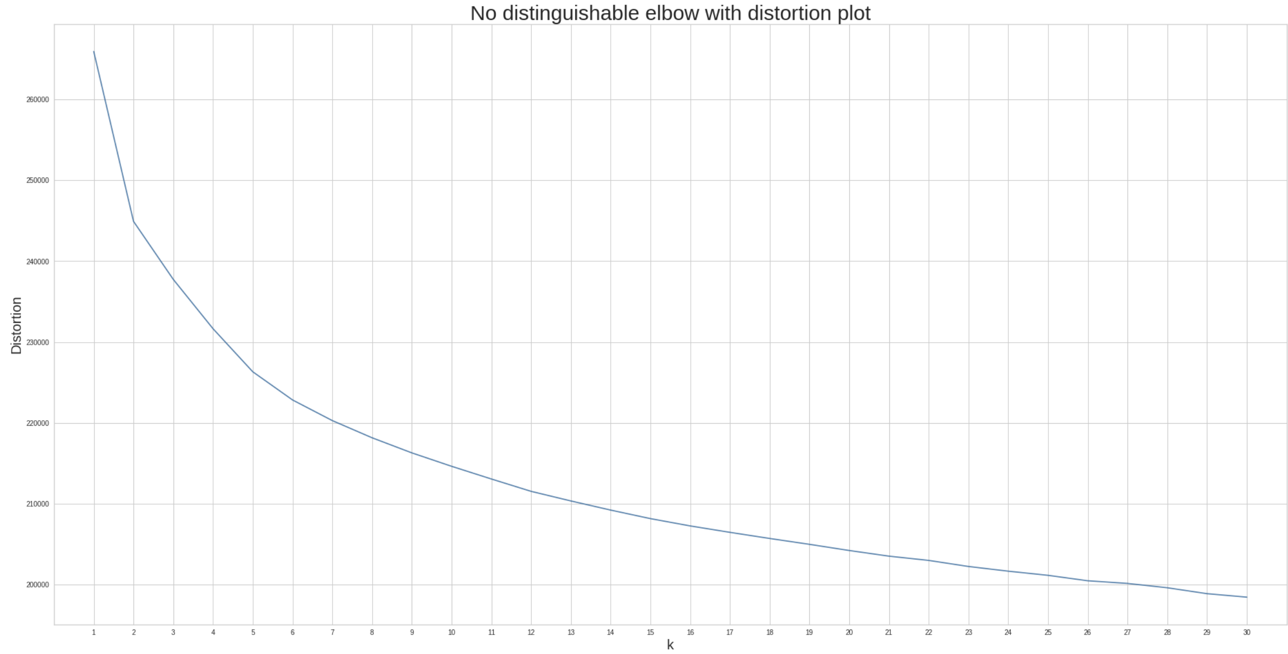


Figure 3 Elbow Method Plot. There is not indicative elbow helping to determine the suitable k in k-means clustering for topic modeling.

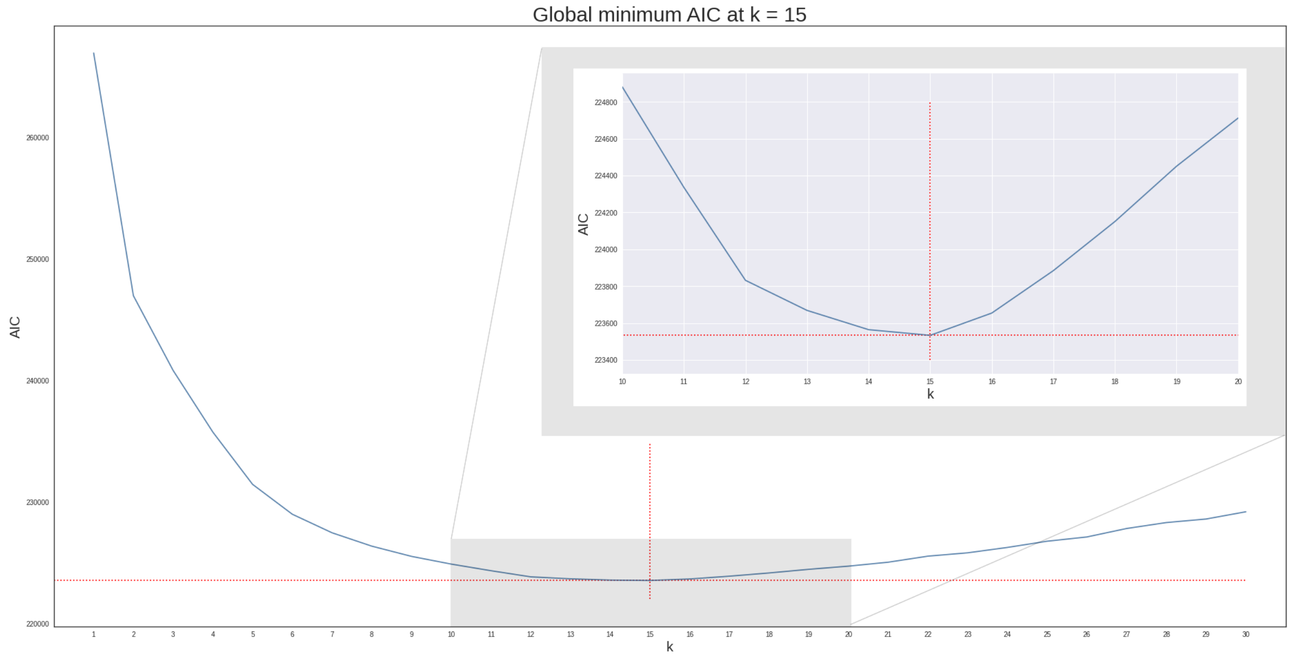


Figure 4 AIC Plot. $k = \{13, 14, 15, 16\}$ are the potential k as there AIC values are so closed while $k = 15$ is the global minimum.

Topics	Corresponding cluster when $k = 15$
Planting and gardening	0
Retail	1
GMO label and bio-products	3
Taste and food	5
Chemicals and cancer	6
Genetic research	7
Health and diet	8
Governance and public policy	10
Meat and animal feeding	11
Agriculture	12
Price and consumption	13

Table 3 Topics are identified from word cloud for $k = 15$

and quoting' is newly formed when comparing word clouds of $k = 15$ to that of $k = 16$.

3.1.3 Summary

In short summary, k-mean clustering on sentence embeddings for both English and German sentences gives a promising result with visualizing top words in word clouds. Coherent clustering results shows that there are 11 concrete organic food related topics (Table 3) mentioned by both English and German text from all potential ks. Since $k = 15$ scores a global minimum in AIC, we determined $k = 15$ as the optimal. Figures 9 shows the topic distribution in different sources and the sentences origins (article or comment).

3.2 Sentiment Distributions

Several systems have been built which attempt to quantify opinion for different application. In our use case, the media news and public opinion are much more complex than a product review which can only concentrate on one specific aspect. Thus, we do not assign for each news one topic but to analyse the news and comments through different dimensions. In order to measure the impact of our defined tasks, we performed global as well as individual views on the set of 99 *New York Times* news (14,202 sentences) with its comments (45,597 sentences) and 61 *Der Spiegel* news with its comments (121,369 sentences).

3.2.1 Basic Definitions

When characterizing sentiments we must make a clear distinction between different terms that may appear. Our data source are mainly from news medias and the comments below the news. Therefore, we declare following concepts for interpreting our results.

Document we claim that a document D is either a $news$ or the set of its $comments$. More clearly, we define D^{news} as a particular news and its corresponding set of comments as D^{comm}.

As a result, we can now combine the previous concept the Opinion S with document. For example, D is a set of $S \in S_1, S_2, ...$ we say the D has length $|S|$. Since every $S = (s, t)$ is assigned to one topic t, then D obtains all topics which are once assigned to the S.

Sentiment Indicators In order to articulate the sentiment for one topic t in D, we introduce four quantified indicators to measure the sentiment in four dimensions, respectively. Each of them also have a visualized interpretation for box plot.

Polarity It directly reflects the emotion of one topic. It is a score from $avg(s)$ where s are from the $S = (s, t)$ that have same topic t. For example, D has a topic *Retail*, we compute the mean of sentiment scores from all sentences given *Retail* and

Figure 5 *Top*: Word clouds for English sentences for $k = 13$ *Bottom*: Word clouds for German sentences for $k = 13$

Figure 6 *Top*: Word clouds for English sentences for $k = 14$ *Bottom*: Word clouds for German sentences for $k = 14$

Figure 7 *Top*: Word clouds for English sentences for $k = 15$ *Bottom*: Word clouds for German sentences for $k = 15$

Figure 8 *Top*: Word clouds for English sentences for $k = 16$ *Bottom*: Word clouds for German sentences for $k = 16$

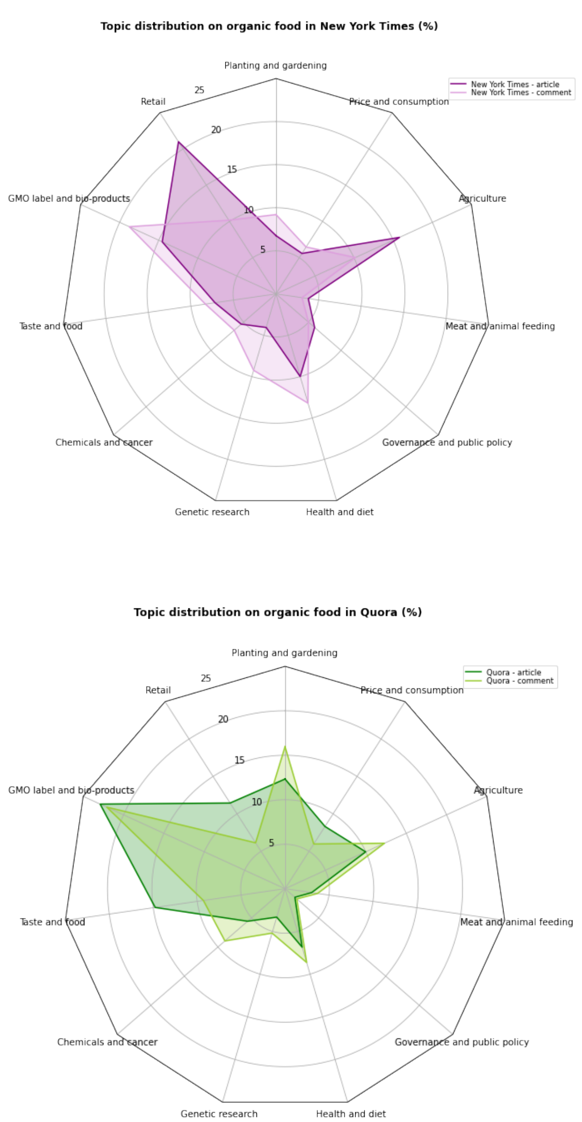


Figure 9 *Top left and right*: Topic distribution of comments for both *New York Times* and *Der Spiegel* close to their topic distribution of the news respectively; and due to the culture differences, the most discussed topics varies. *Bottom left*: Topic distribution for comments also almost matches that of the articles in discussion forum, *Quora*, alone. *Bottom right*: Compare to the other 3 radar plots, this one shows indirectly response between news articles from *New York Times* and *Der Spiegel* and comments from *Quora*. It shows that the *Quora* follows more to *New York Times* than *Der Spiegel* and it is probably due to the languages that English audiences are more affected by English media.

state topic *Retail* in D has polarity between -1 (negative) and 1 (positive). In the box plot, it is represented by gradient colors. (See Figure 11a mapping -1 to green and 1 to red.)

Dominance It shows how frequently a topic appears in one D. It defines as $\frac{|S \text{ assigned to } t|}{|S|}$ i.e. The length of D is $|S| = 100$ and the sentences assigned to topic *Retail* appears 40 times. Then the dominance of *Retail* in D is 0.4. Sum of the dominance value from all topics in one document should be 1. Its graphical representation in box plot is the width of the box.

Variance It illustrates how emotion varies inside one topic. Suppose $X = s_1, s_2, ...$ where s_i is the sentiment score from one topic. We claim the variance level is equal to the $Q_X(0.75) - Q_X(0.25)$ (where Q_X represent the quantile) which between 0 and 2. In box plot, the variance for each topic is visualized as the height of each box.

Hotness It is the multiplication of dominance and variance. Consequently in the box plot, it is translated to the area of the box. It intuitively describes how fierce this topic is. Namely, the hotter topic means the larger size of box, which means this topic not only is more frequent discussed and also has the stronger (opposite) emotions among public.

3.2.2 Global Sentiment Distribution Comparison

In favor of acquiring a global sentiment overview and to compare the similarities and difference across two countries, four distributions of sentiment scores are illustrated in Figure 10). In global analysis, the polarity and the variance indicators are first applied (See Table 4).

Sentiment Comparison Within-Country Compare a with b (or c with d) in Figure 10, the distributions between two corpus are similar. The polarity from comments tends to be more negative in both. The variance level in comments is also stronger than news (See Table 4). It implies that, in different cultural background, the public sentiment in a degree could be effected by media sentiment, whereas the public opinion tends to be less positive (more negative). As expected, the variance level which measures the fierceness of discussion in comments is much more intensive than news in both media.

Sentiment Comparison Across-Country Apparently, the sentiment polarity across two countries is different. Namely, in Germany, the sentiment overall is more negative than in USA. In addition, in Figure 10 we observe that in Germany the majority sentences are assigned to a neutral sentiment which remains an open question for future study. Turning now to the variance, although the emotion strength in comments are greater than in news in both medias, the German corpus indicates even stronger global variance in comments (more than twice much as in news).

3.2.3 Topic Sentiment Distribution

By visualizing the topic sentiment as box plot, the reader can quickly capture the overall sentiment features illustrated by the four indicators we mentioned earlier within one corpus, between news and comments as well as the difference across countries.

New York Times Part a in Figure 11 illustrates the topic sentiment distribution from *New York Times*. Each box represents a topic sentiment that can be interpreted by its polarity (color), dominance(width), variance (height) and hotness (area).

	comments	news	Δ
polarity	0.032	0.061	-0.029
variance	0.375	0.193	+0.182

(a) *New York Times*

	comments	news	Δ
polarity	-0.061	-0.050	-0.011
variance	0.402	0.144	+0.258

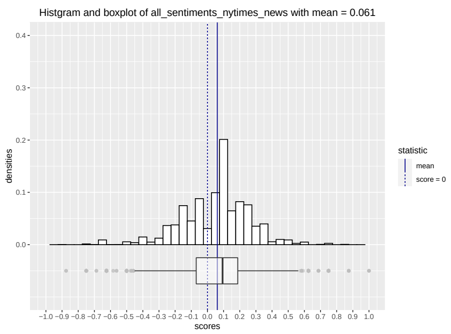
(b) *Der Spiegel*

Table 4 Global sentiment comparison between news and comments

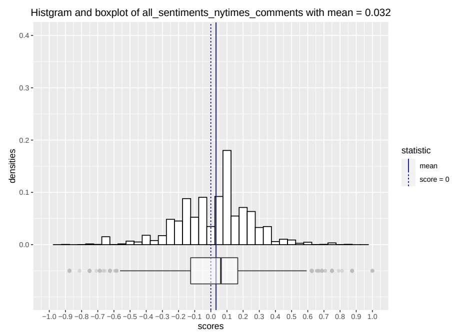
From the perspective of hotness, it is clearly to observe that the *hot* topics in news remains *hot*, like *Retail*, *GMO label and bio-products*, *Health and diet*. However, within these fierce topics, the dominance level (width of box) changes from news to comments. For instance, the dominance of *Retail* in news owns 35% which ranks the first, but turns to 9% in comments (See Table 5). It implies that the media in USA tends to set more agenda in *Retail* to arouse the public, whereas the concern about retail is though dominance, but might be weaker than in media. On the contrary, the topic *Health and diet* which is closely relevant to public life wins more dominance in the comments and unsurprisingly become less positive. Additionally, compare the polarity per topic of comments with news, we conclude that the polarity from each topic coherently goes slightly down to negative as we shown in the global polarity (See Table 4).

Der Spiegel Similarly to the *New York Times*, part b in Figure 11 reveals the features of sentiment in *Der Spiegel*. The global hot topics in news and comments are *Governance and public policy*, *Meat and animal feeding*, *Agriculture*, *Price and consumption*, which differ from in *New York Times*. On one hand, this significant bias indicates that in a local social environment, the media agenda setting shows its own cultural preference; On the other hand, it could be an interesting starting point to explore the media agenda setting from the intercultural and transcultural aspects.

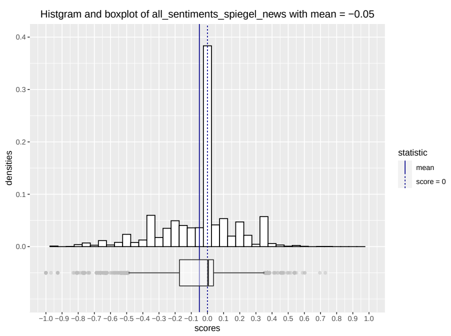
Regarding to dominance, interestingly, although the dominance from majority of topics are stay consistently in the similar level between comments and news, these topics are not perfect coherent: *Chemicals and cancer* (from 0.03 increases to 0.14,$\Delta = +0.11$), *Health and diet* ($\Delta = +0.08$) and *Agriculture* ($\Delta = -0.41$). These shifts which also happen in *New York Times* points towards the idea that the group of online readers who might be the majority internet user living in the city tends to concern more the topic affiliated their daily life (*Health and diet*, *Chemicals and cancer*) than the topic which not directly related to their manner of living *(*Agriculture).



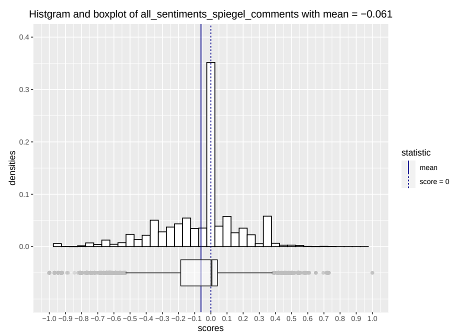
(a) *New York Times* News



(b) *New York Times* Comments

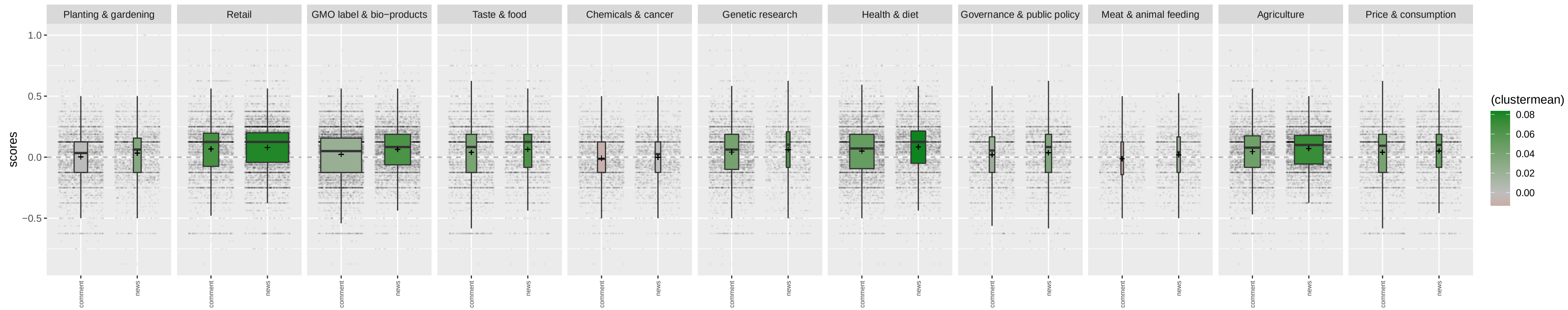


(c) *Der Spiegel* News

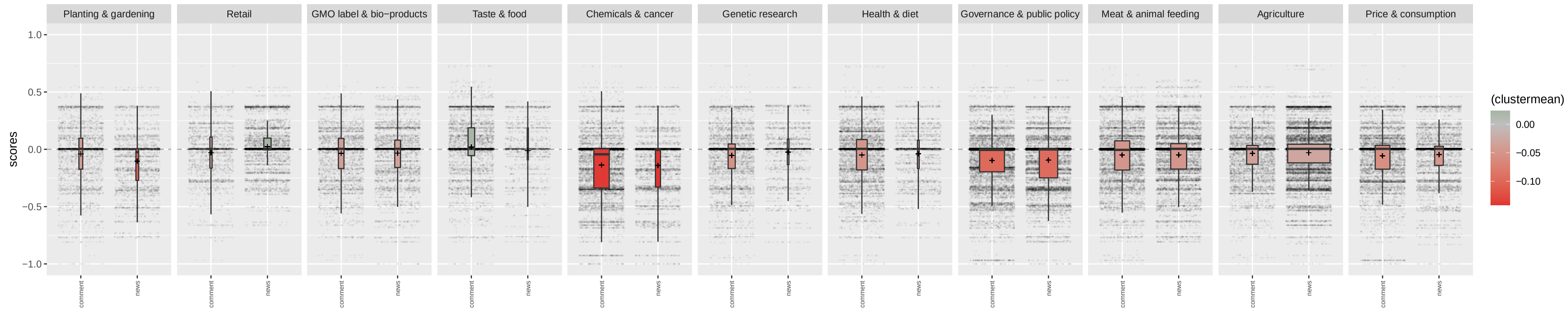


(d) *Der Spiegel* Comments

Figure 10 Score distributions over four corpus



(a) *New York Times*



(b) *Der Spiegel*

Figure 11 Topic sentiment box plot visualization

3.2.4 Relation Between Sentiment Variance and Size of Comments

Besides the global sentiment distribution analysis, we want to find the evidence to support the hypothesis that the variance level in comments related to the size of comments. We assume that the larger size of comments (number of sentences in comments) may have a tendency that the sentiment variance also rises. However, as illustrated in Figure 12, the variance fluctuates when the number of sentences is small (i.e. under 100). As the size grows, it slowly converges to a constant in both medias. As a result, regarding to our dataset, no signs of this assumption is found. The possible reason for this convergence could be that, the larger size the comments are, the more rational content (i.e. using more scientific facts) the readers expressed. Consequently, the polarity of these descriptive content might tend to be weaker than the simple emotional comments (i.e. I applaud Whole Foods for taking a stand!).

3.2.5 Sentiment Visualization on Individual Document

Apart from a global view of sentiment, we highlight the distribution from an individual perspective is also important to study. Hence, two news and its comments from two medias respectively are selected.

***New York Times*: Labeling GMO** The news [Lab] mainly talked about the requirement of labeling GMO from a grocery chain company. As results shown in Figure 13 a, it proves that our clustering satisfies the original news text: the two hot topics are *GMO label and bio-products* and *Re-*

ranking	dominance	topic
1	0.35	Retail
2	0.21	Agriculture
3	0.18	GMO label & bio-products
4	0.09	Health & Diet
5	0.04	Taste & Food
6	0.04	Planting & gardening
7	0.03	Governance & public policy
8	0.02	Price & consumption
9	0.02	Chemicals & Cancer
10	0.01	Genetic research
11	0.01	Meat & Animals

(a) News

ranking	dominance	topic
1	0.33	GMO label & bio-products
2	0.17	Health & Diet
3	0.09	Agriculture
4	0.09	Retail
5	0.08	Genetic research
6	0.08	Planting & gardening
7	0.05	Taste & Food
8	0.04	Chemicals & Cancer
9	0.04	Price & consumption
10	0.02	Governance & public policy
11	0.01	Meat & Animals

(b) Comments

Table 5 Dominance ranking in *New York Times*

ranking	dominance	topic
1	0.50	Agriculture
2	0.16	Governance & public policy
3	0.14	Meat & Animals
4	0.06	Price & consumption
5	0.05	Retail
6	0.03	GMO label & bio-products
7	0.03	Chemicals & Cancer
8	0.01	Planting & gardening
9	0.01	Health & Diet
10	0.01	Genetic research
11	0.00	Taste & Food

(a) News

ranking	dominance	topic
1	0.26	Governance & public policy
2	0.14	Chemicals & Cancer
3	0.13	Meat & Animals
4	0.13	Price & consumption
5	0.09	Agriculture
6	0.09	Health & Diet
7	0.05	Genetic research
8	0.04	Taste & Food
9	0.04	GMO label & bio-products
10	0.02	Planting & gardening
11	0.01	Retail

(b) Comments

Table 6 Dominance ranking in *Der Spiegel* *tail*. Whereas in the comments, the dominance of *GMO label and bio-products* stays but of *Retail* decreases. It provides an evidence which we observe from previously analysis in subsubsection 3.2.3 from an individual view that when the media proposes a commercial agenda to influence public's agenda, the audience are indeed affected has and it might spread out over more related topics. Remarkably, in this example the readers give more positive opinions for *Retail* than the sentiment polarity of *Retail* in news. Is it a coincident? This question leads to exploring the sentiment indicators' correlation (examined in subsubsection 3.2.6) between the comments and news.

In order to make a intuitive impression, some comments are listed in section 5.1 which articulates the comments from *New York Times* tend to be straightforward and simple emotional statement (like or dislike, comparative adjectives) and many of them use *I* as subject. That is to say, if some positive sentiment words (assuming already considered negation words) which are actually the POS words in pretrained model co-appear with *I* in a sentence, it has greater possibility to reason that this reader holds a positive opinion rather than a negative.

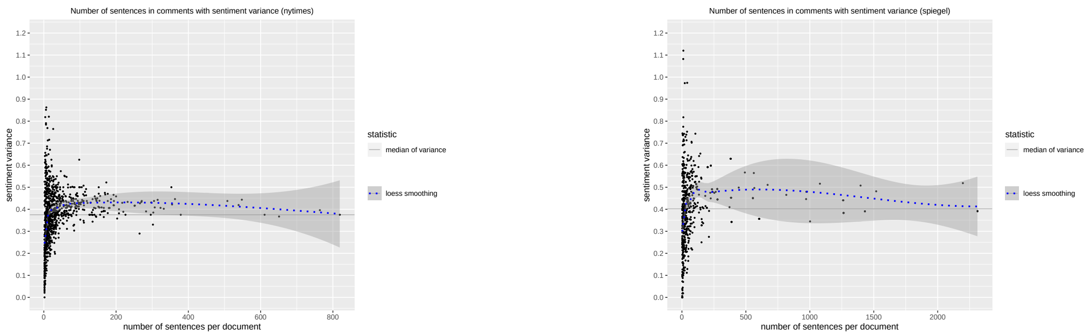
***Der Spiegel*: Der Skandal um Dioxin** In contradiction with above business-purpose oriented agenda, in Figure 13 b, the sentiment distributes totally different. The scandal of a harmful chemical is reported in this German article (See original news text [Gif]). Undoubtedly, the polarity in news and in comments are both negative. It is crucial to note that the topic *Chemicals and cancer* becomes a trigger which arouses several other topics (*Governance and public policy*, *Price and consumption* as well as *Health and diet*) which discussed fiercely by public but almost not mentioned in news. We believe it should be consider in further study that which topic under which condition could largely provoke wider debates in other area.

Regarding to the German comments, section 5.1 expresses different features compared with English comments. The sentences contain more sarcasms and irony emotion which are harder to detect and could make the sentiment analysis process error-prone. Because these figurative language can finally alter the meaning of the sentence.

3.2.6 Correlations Between Comments and News

According to the different distributions, We apply Pearson test for examining the correlation of polarity and Spearman test for computing correlation of dominance as well as variance because the distribution of last two sentiment indicators do not satisfy the normal distribution. The correlation result (See Figure 14) demonstrates in three dimensions for each topic:

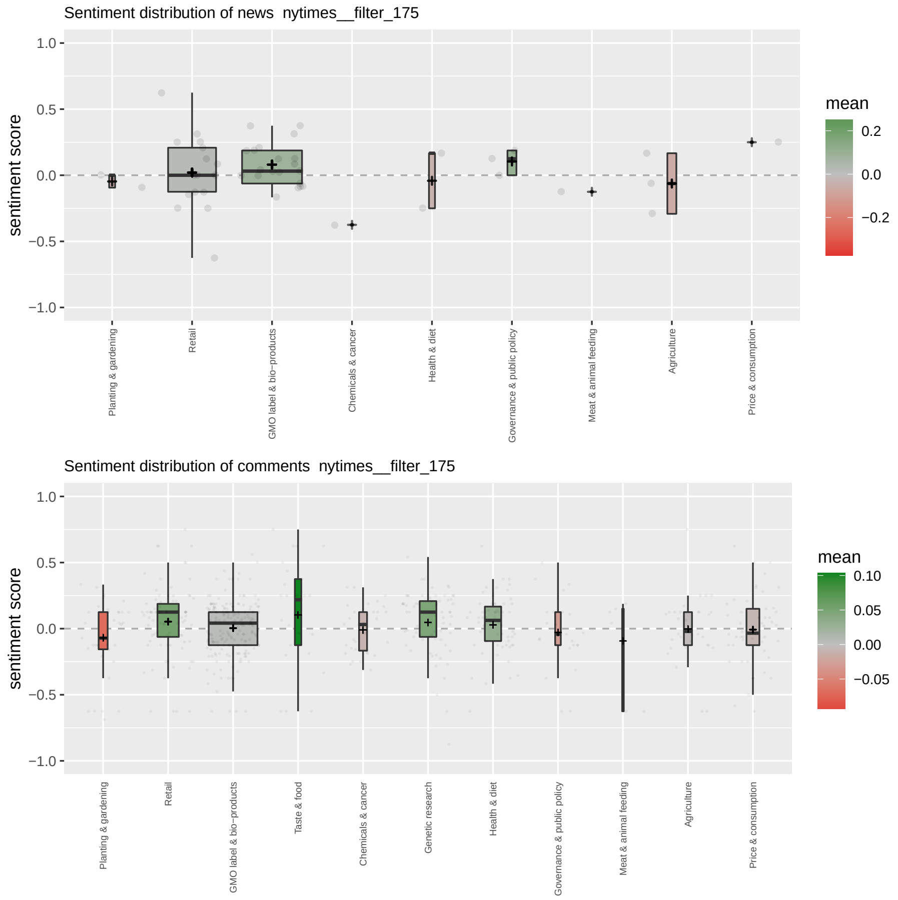
Polarity First let us review that polarity reflecting the positivenegative opinion is measured by mean of sentiment scores. With respect to correlation, we claim that the polarity of news and comments in one specific topic is positively correlated if as the polarity of news increases(decreases), the polarity of comments also rises (falls). Figure 14 a presents that only two topics in which it gets coefficient over 0.3 that implies a weak positive correlation between the new and comments whereas it shows a negative correlation between them in topic *Retail*. For



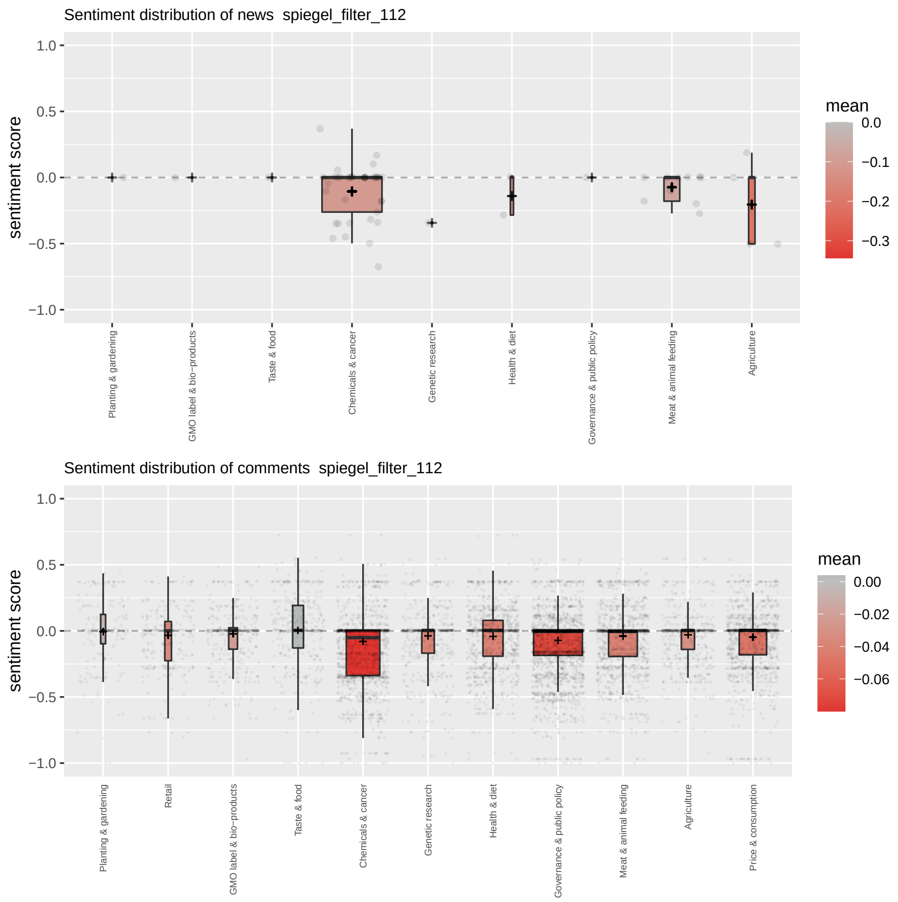
(a) *New York Times*

(b) *Der Spiegel*

Figure 12 Number of sentences in comments with sentiment variance



(a) Visualization of a news (top) and its comment (bottom) from *New York Times* (doc_id = 175)

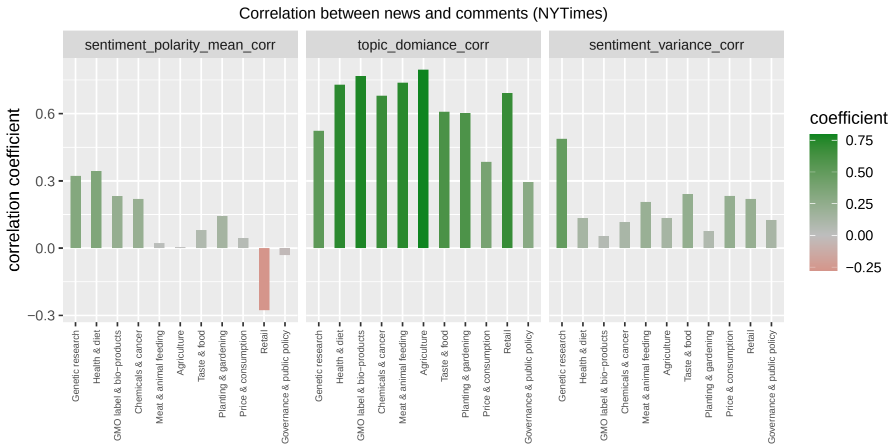


(b) Visualization of a news (top) and its comment (bottom) from *Der Spiegel* (doc_id = 112)

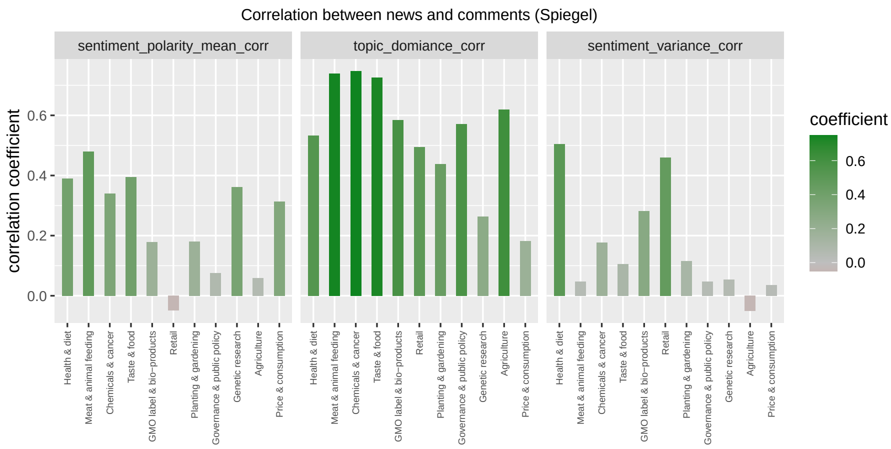
Figure 13 Individual sentiment visualization the rest topics there is no proof that they are significant related. In comparison to *New York Times*, the polarity correlation between news and comments in *Der Spiegel* refers generally stronger. Firstly, there are more topics in which the correlation are positive which implies intuitively that in *Der Spiegel* the opinion sentiment among the majority topics could be more coherent than in *New York Times*. Nevertheless, in the topic *Retail*, it slightly exhibits a negative coefficient that is similar to *New York Times*.

Dominance Recall the definition of dominance of a topic: it is calculated by the ratio of number of sentences assigned to this topic to the total number of sentences in the document. From Figure 14 a and b we can clearly observe that for most of topics, the news and comments have a strong positive correlation in both corpus. Despite the topics where there is only a weak correlation (the last two lowest coefficient: in *New York Times* are *Price and consumption* and *Governance and public policy* while in *Der Spiegel* are *Genetic research* and *Price and consumption*), it draws a small distinction between two medias.

Variance Turning now to the correlation of variance which interprets that as the sentiment of news within one topic varies more strongly, the emotion of corresponding comments also fluctuate intensively. Nevertheless, only in a few topics from both medias, there are clear positive correlations (coefficient greater than 0.3). Precisely, in *New York Times*, there is only one topic *Genetic research* while *in Der Spiegel*, it exists two: *Price and consumption* and *Retail*. The result infers that among these topics as the news presents a more debatable view, the public opinion also tends to be more controversial. In addition, the topic *Agriculture* in *Der Spiegel* shows a slightly negative correlation between news and comments which hints that when the media has



(a) *New York Times*



(b) *Der Spiegel*

Figure 14 Correlations between comments and news a wider range of statements about agriculture, the public does not tend to be talk more but rather ignore about that. This hint also supports aforementioned claim that these online readers are anticipated to be less expressive in such topic than other daily-life related topics.

4 Conclusions

Though out our methodologies, text from articles and comments are split into sentences for sentiment analysis for different topics around organic food domain. Our favourable results show that there are influences from media to their readers on public agenda. The influences are not limited to the discussion occurrences but also the sentiments polarity, dominance, variance and hotness. Regarding to the proposed questions mentioned in the introduction that we would like to dig into, it is found that:

- Coherent emotions: In a global perspective, we find that the media agenda could be a precursor for public sentiment due to the similar sentiment polarity distribution between news and comments.
  - The result of correlation of polarity in *New York Times* offers evidence for the limited influence (correlation coefficients are in general not higher than in *Der Spiegel*) of media agenda in terms of emotion contagion.
  - Combined with the results in section 3.2.2 and Figure 14, the distinction of sentiment trend between two medias infers that the media agenda setting effects will be more powerful (higher coherence level and stronger correlated) for the negative opinion than the positive in response to readers' expressions of emotion.
- Dominance/Attention of topics: In subsubsection 3.2.3 we concludes that the daily-life related organic food topics like health are more likely be affected by media agenda setting while the topic like *Agriculture* stimulates less concern among the online readers who mostly live in the cities.
- Provoked fierce debate on topics: Whether stronger media agenda setter is able to trigger the public at the same or even more intenser level depends on the specific topic according to our correlation analysis in Figure 14 combined with change of hotness degree visualized in Figure 11 . The public debate within *Agriculture* in *Der Spiegel* argued above is inclined to be less fierce by given a media context with greater sentiment variance.
- Similarities and Differences across countries: Due to the divergent cultural and language background, each has its own focus (See topic ranking from Table 5and Table 6). Therefore the public agenda partially effected by medias also varies. Although the public sentiment polarity turns to be less positive for both medias, due to more criticisms in German media agenda setting, it leans towards entrenching stronger emotional expressions in the public compared with the commercial-oriented agenda setting in *New York Times*.

5 Future Works

If it is possible, we would like to work more on the relationship between the sentiment influences between the news media and the discussions from public forum, *Quora* and the corresponding media agenda setting in temporal order. Besides, the evaluation on sentiment assignment between the two frameworks is hoped to be carried out in order to ensure the fairness on sentiment analysis comparison between two languages.

References

[Gif] Die wichtigsten fragen und antworten zum giftskandal. `https://www.spiegel.de/wissenschaft/medizin/dioxin-die-wichtigsten-fragen-und-antworten-zum-gift-skandal-a-737972.html`.

[Lab] Major grocer to label foods with gene-modified content. `https://www.nytimes.com/2013/03/09/business/grocery-chain-to-require-labels-for-genetically-modified-food.html`.

[3] Angelidis, S. and Lapata, M. (2018). Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. *ArXiv*, abs/1808.08858.

[4] Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204.

[5] Boyd-Graber, J. and Resnik, P. (2010). Holistic sentiment analysis across languages: Multilingual supervised latent dirichlet allocation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 45–55.

[6] Chidambaram, M., Yang, Y., Cer, D., Yuan, S., Sung, Y.-H., Strope, B., and Kurzweil, R. (2018). Learning cross-lingual sentence representations via a multi-task dual-encoder model. *arXiv preprint arXiv:1810.12836*.

[7] Cronen-Townsend, S., Zhou, Y., and Croft, W. B. (2002). Predicting query performance.

[8] Remus, R., Quasthoff, U., and Heyer, G. (2010). Sentiws-a publicly available german-language resource for sentiment analysis. In *LREC*. Citeseer.

[9] Saif, H., Fernández, M., He, Y., and Alani, H. (2014). On stopwords, filtering and data sparsity for sentiment analysis of twitter.

Appendix

5.1 Partial Content of News and Comments

***New York Times*: Labeling GMO[Lab]** Whole Foods Market, the grocery chain, on Friday became the first retailer in the United States to require labeling of all genetically modified foods sold in its stores, a move that some experts said could radically alter the food industry. ... Whole Foods, which specializes in organic products, tends to be favored by those types of consumers, and it enjoys strong sales of its private-label products, whose composition it controls. ... He said Whole Foods looked forward to working with suppliers on the labeling.

Comments

- *I have a cousin who works at Whole Foods. He is a happy employee and loves it. Thinks it is a great company.*
- *I am not sure if non-GMO foods are healthier to eat but they are certainly better for the environment.*
- *Ich kann und will niemandem verbieten Fleisch von deutschen Rindern zu essen.*
- *I applaud Whole Foods for at least taking a stand.*
- *Red meat consumption is associated with an increased risk of total, heart, and cancer mortality.*
- *Since apples are apparently the most pesticide-ridden fruit, I have gotten to like the more expensive but sweeter ones.*

***Der Spiegel*: Der Skandal um Dioxin[Gif]** Es ist einer der größten Giftskandale der vergangenen Jahre: Bis zu 3000 Tonnen dioxinverseuchtes Fett wurden laut Bundeslandwirtschaftsministerium an 25 Futtermittelhersteller in mindestens vier Bundesländern geliefert. Wo das Gift von dort aus hingelangte und welche Mengen an Nahrungsmitteln belastet sind, ist weitgehend unklar. Verbraucher reagieren zunehmend verunsichert: Der Verkauf von Hühnereiern ist "spürbar" gesunken, teilte die landwirtschaftliche Marktberichterstattungsstelle MEG mit. Welche Gefahren drohen durch die Einnahme von Dioxin? Welche Vorsichtsmaßnahmen können getroffen werden? SPIEGEL ONLINE gibt Antworten auf sieben Fragen.

Comments

- *Der Dioxin-Skandal war hoffentlich nicht der letzte. Es sollten so viele wie möglich vorkommen. Am besten aber wäre,wenn ein paar Konsumenten nachweislich an solchen oder anderen Giftstoffen in Lebensmitteln sterben.*
- *3000 Tonnen verseuchtes Tierfutter - das ist ein Terroranschlag.*
- *Ich kann und will niemandem verbieten Fleisch von deutschen Rindern zu essen.*
- *dengesetzlichen Vorgaben ist so eine Sache. Fahren Sie mal mit einem Fiat 500 mit 80 km/h frontal gegen eine S-Klasse!*
- *Wer sich mit Bio-Artikeln überfrisst, stirbt auch. Also sind Bio-Lebensmittel auch lebensgefährlich, wenn man falsch damit umgeht. Guten Appetit.*