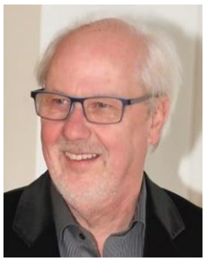
2019 Hawaii University International Conferences
Arts, Humanities, Social Sciences & Education January 3 - 5, 2019
Prince Waikiki Hotel, Honolulu, Hawaii

Country-specific Sentiment on Microblogs

Nikolic, Janina
Department of Information Science
Heinrich Heine University Düsseldorf
Düsseldorf, Germany



Dorsch, Isabelle
Scheibe, Katrin
Zimmer, Franziska
Stock, Wolfgang G.

Department of Information Science
Heinrich Heine University Düsseldorf
Düsseldorf, Germany

Janina Nikolic
Isabelle Dorsch
Katrin Scheibe
Franziska Zimmer
Wolfgang G. Stock
Department of Information Science
Heinrich Heine University Düsseldorf
Germany

Synopsis:

Can expressed sentiments (positive, neutral, negative) within text messages of microblogging services be differentiated by the users' country? Are there indeed country-specific emotional patterns of microblogging behavior? Using the example of the service Twitter, the work in hand intends to introduce a measure for the expressed sentiments of countries. This idea assumes that people of different regions and cultures have a specific emotionality, not only in daily life but also in writing microblogs.

Country-specific Sentiment on Microblogs

Can expressed sentiments (positive, neutral, negative) within text messages of microblogging services be differentiated by the users' country? Are there indeed country-specific emotional patterns of microblogging behavior? Using the example of the service Twitter, the work in hand intends to introduce a measure for the expressed sentiments of countries. This idea assumes that people of different regions and cultures have a specific emotionality, not only in daily life but also in writing microblogs. 14,192,182 tweets have been analyzed from different countries, languages, and topics to be able to get significant results in how distinct people represent their positive and negative feelings and opinions to the outside depending on their home and culture. We applied a lexicon-based approach and adjusted the term lists (words as well as emojis) intellectually. As indicators we introduced sentiment strength (general emotionality), sentiment climate (difference between positive and negative sentiment values) and the amount of positive as well as negative microblogs. Finally, we compared our findings with other indicators on country level (as Human Development Index or average temperature).

Keywords: Information behavior, Country, Microblogging, Emotion, Social media, Sentiment analysis, Twitter, Sentiment strength, Sentiment climate, Emoticons

1. Introduction

To find out what other people think has always been an important part of human information-gathering behavior. Nowadays, about 336 million people all over the world use Twitter to communicate, to publish their feelings and opinions [34]. It became a rich resource of short texts (commonly known as tweets) related to all kinds of topics. Therefore, and because of the easy and free access via API, there are numerous analyses with Twitter data on a large scale to predict trends, election results, stock market development, or to capture customer satisfaction; many of them applying sentiment analysis [2, 40].

Sentiment analysis aims to determine the attitude of a person regarding a topic. Given a natural language text, it identifies whether the expressed opinion is positive, negative, or neutral. Similar terms are sentiment detection or opinion mining. Considering several studies about sentiment analysis on a specific topic or methodology in Twitter [35, 37], there is a research gap regarding the general exposure of emotions: What about worldwide country-specific exposed emotionality on Twitter? From the angle of information science, the aim of this work is to investigate possible differences between countries according to their sentiment on this microblogging service. To combine these topics, theoretical foundations have to be acquired about different ways on how to measure emotions and approaches for sentiment analysis.

1.1 Approaches to Measure Emotion

In everyday language, emotion means a person's state of feeling in the sense of an affect. Although this term is used very frequently, there exist many different definitions, because emotions are complex and contain different components. Principally, there are two primary approaches to classify emotions: emotion as finite categories, and emotion as dimensions. Proponents of the theory of discrete categories assume that emotions consist of a finite number of existing basic emotions. These are emotions that cannot be traced back to other emotions. Although varying, according to different accounts and cultural contexts, these basic emotions generally include happiness, sadness, anger, disgust, and fear [10].

In contrast, the dimensional approach classifies emotions not as discrete categories but as multiple dimensions (mostly two or three) for emotional and affective states. Several dimension models of emotion have been developed with just a few remaining as dominant. Sentiment analysis often focuses on valence or the positive-negative dimension of emotion which results from the evidence in the psychology literature for valence having the greatest impact on cognitive processes in comparison to other emotion dimensions [23]. In general, cultural differences have been observed in the way how emotions are valued and expressed [29].

1.2 Sentiment Analysis Background

The two most common sentiment analysis tasks are subjectivity and polarity detection, where subjectivity detection predicts whether a given text is subjective or not and polarity detection predicts whether a subjective text is positive or negative. According to polarity detection, the sentiment strength detection predicts the strength of positive or negative sentiment [37]. One advantage of sentiment analysis in comparison to traditional opinion polls is that while most surveys only request what is already known because of predefined questions and partially slightly motivated participants, the focus in sentiment analysis is on listening instead of asking questions which leads to more realistic results [30].

In general, sentiment analysis can take place at three levels: document-level, sentence-level, or phrase-level [19]. A sentiment analysis on document-level is the grossest form of opinion detection because it is assumed that a document only contains one opinion about a topic while sentiment analysis on sentence-level takes into account that a topic can consist of several opinions. Indeed, even a sentence can contain multiple opinions to different aspects, e.g. in case of the attributes of a product [11] leading to the finest granularity level of single words or phrases.

Sentiment classification techniques can be roughly divided into machine learning (ML), lexicon-based, and hybrid approaches. Our work uses the lexicon-based approach which relies on a sentiment lexicon. Put simply, this is a collection of sentiment terms with positive and negative values in a given range. Our approach is based on the polarity detection and mainly of the sentiment strength detection and decides between positive, negative, and neutral sentiment for each tweet within the interval of [-5, 5].

In case of international comparisons of sentiment analysis, it is often not taken into consideration that there could be a fundamentally different distinct emotionality in countries. For example, people of one country could react more emotionally in general than people of another country, regardless of whether the topic is e.g. politics, sports, or entertainment.

Different studies about cross-country emotionality underline this statement. Gallup Poll, a division of the American company Gallup, annually measured daily emotions in more than 150 countries and areas by asking residents whether they experienced five positive and five negative emotions a lot the previous day. To measure the presence or absence of emotions, Gallup averaged together the percentage of residents in each country who said they experienced each of the 10 positive and negative emotions [6]. For example, in 2016, Ecuador was the most emotional country in the world according to Gallup [12].

Another study [3] measures the cultural relativity of emotional valence and arousal of the twelve most popular emotion keywords expressed on Twitter, distinguished between the three regions Europe, Asia, and North America. The results show that Europeans are, or at least present themselves as more positive and aroused, North Americans are more negative, and Asians appear to be more positive but less aroused when compared to global valence and arousal levels of the same emotion keywords.

Further investigations present cultural stereotypes of emotions and emotionality; for instance, East Asians are often identified as "inscrutable" by Westerners [32] and people from countries of warm climates are believed to be more passionate than people from countries of cold climates [25].

If these differences in emotionality turn out to be applicable for tweets, results in cross-border analyses can no longer be easily compared, and require a normalization factor to enable internationalization and bring opinions, for instance, on a product or a politician to the same level. In this connection, the major challenge might be the different languages in which tweets are written and how to measure the emotions in a comparable way.

The research of recognizing sentiments, especially in tweets, grew in recent years and produced lots of different approaches to gain the best results. One way to find emotional tweets is via hashtags because their usage is common and tweets which express sentiments often also represent this emotion through an according hashtag, e.g. #angry if the tweeter is angry about something. Qadir and Riloff [27] developed a bootstrapping algorithm to automatically learn hashtags that convey emotion, but only for tweets written in English. Since their approach does not rely on any language-specific techniques, they planned to learn emotion hashtags in other prominent languages such as Spanish, Portuguese, etc.

Cui et al. [8] present another approach to face the problem of multilingual Twitter sentiment analysis. They focus on the analysis of emotion tokens, including emotion symbols (e.g. emoticons), irregular forms of words and combined punctuations. The advantage of these emotion symbols is that they are independent of language; therefore, they are a useful signal for sentiment analysis on multilingual tweets. After these tokens are extracted automatically from tweets, a graph propagation algorithm labels their polarities. It is also mentioned that from a dataset of five million tweets 33% contain emotion tokens, so this approach only covers a part of the whole data and would have to be combined with other attempts. Likewise, Coats [7] and Wolny [43] examined emojis and emoticons in sentiment analysis on Twitter. Furthermore, a sentiment lexicon only for emojis was developed [17].

Another idea to implement international, multilingual sentiment analysis came from Bautin, Vijayarenu, and Skiena [5]. They performed a sentiment analysis by using existing translation programs and simply translate news texts in eight foreign languages to English before passing them to a sentiment analysis system. Mentioned difficulties include the loss of nuance incurred during the translation process, because "even state-of-the-art language translation programs fail to translate substantial amounts of text, make serious errors on what they do translate, and reduce well-formed texts to sentence fragments" [5, p. 19]. The authors also conclude that the success of their method is largely translator-independent since they compared the sentiment results of the same source text corpus across two distinct translators for Spanish. Another interesting fact is that they provide normalizing for cross-cultural language effects because they suggest that e.g. certain languages (e.g. Chinese and Korean) appear to produce significantly higher sentiment scores than others (e.g. Italian). Besides techniques to correct for such biases, they present a cross-cultural comparison of country sentiments by language.

There exist further multilingual sentiment studies on Twitter, considering a certain topic or analyze a limited number of different languages [1, 36, 33, 42]. However, none of them refer to a worldwide topic-independent sentiment analysis.

The approach of Roberts, Roach, Johnson, Guthrie, and Harabagiu [28] focuses on a corpus collected from Twitter with tweets annotated at the tweet-level with seven emotions: anger, disgust, fear, joy, love, sadness, and surprise. It was analyzed how emotions are distributed in the annotated data and afterwards compared to the distributions in other emotion-annotated corpora. They also used the annotated corpus to train a classifier that automatically discovers the emotions in tweets. Hence this is an option that focuses on more specific emotions which is not given in the present study. Besides, some studies lay a special focus on sentiment detection of happiness [9, 26, 31].

Concerning the distinction to other works, no papers could have been found which exactly address a multilingual worldwide country-specific sentiment analysis of Twitter short messages. Most existing sentiment analysis systems are designed to work in a single language, usually English, although many people also tweet in their mother tongue [20].

1.3 Country-specific Sentiment on Twitter

This work investigates—from an Information Science perspective—if inhabitants of different countries express themselves emotionally stronger or weaker in principle— independent of any topics. Foundations are tweets of a country whose sentiment is calculated by various sentiment features. This sentiment can have distinct orientations and is also dependent on the limitation of the considered languages. Coherences between the sentiment orientation and different indices are investigated to learn about the possible reasons for a specific sentiment. Besides the average yearly temperature which is a location-dependent factor, also economic, ecological, and cultural factors have been compared. According to the stated research aims, we want to tackle the following research questions:

RQ 1: How emotional are users from different countries around the world in general (independent of positive or negative sentiment)?

RQ 2: Do users in particular countries write more positive or negative Twitter messages in general?

RQ 3: Do features (e.g. emoticons or punctuation) have a distinct impact on the emotionality of different countries?

RQ 4: Are our results consistent with other country-level emotion studies?

RQ 5: Do the countries' microblogging sentiments correlate with other specific characteristics of the countries (Human Development Index, temperature, GDP per capita, Happy Planet Index, Six Dimensions of National Culture, and Global Emotions Report)?

2. Methods

In this work, four indices are presented to determine the specific sentiment for countries. The Index for Positive Sentiment (1) and the Index for Negative Sentiment (2) are raw indices; the Index for Sentiment Climate (3) and the Index for Sentiment Strength (4) are derived from raw data.

In the formulas, *pos* means the positive sentiment value of a tweet *i*, *neg* means the negative sentiment value of a tweet *j*. These summarized values are divided by the amount of tweets N_{pos} for the positive tweets or N_{neg} for the negative tweets.

$$\text{mean positive sentiment (ps)} = \frac{\sum_{i=1}^{N_{pos}} pos_i}{N_{pos}} \quad (1)$$

$$\text{mean negative sentiment (ns)} = \frac{\sum_{j=1}^{N_{neg}} neg_j}{N_{neg}} \quad (2)$$

The sentiment climate—similar to the Ifo Business Climate Index (which is an indicator for economic activity in Germany [18])—does not include neutral values and the positive and negative mean values are added.

$$\text{sentiment climate (sc)} = \frac{\sum_{i=1}^{N_{pos}} pos_i}{N_{pos}} + \frac{\sum_{j=1}^{N_{neg}} neg_j}{N_{neg}} \quad (3)$$

The higher the sentiment climate, the more positive the sentiment of a country is in general.

By contrast to the sentiment climate, to calculate the sentiment strength the absolute values of positive and negative sentiment *s* of a tweet *k* have to be added up and divided by the amount of positive and negative tweets per country.

$$\text{sentiment strength (ss)} = \frac{\sum_{k=1}^{N_{pos}+N_{neg}} |s|_k}{N_{pos}+N_{neg}} \quad (4)$$

$$where\ s \neq 0$$

The sentiment strength indicator is an expression of emotionality (or the emotions' arousal)—independent of positive or negative sentiment.

2.1 Data Collection

The data has been collected via the official Twitter Streaming API (Application Programming Interface). For this investigation, the Twitter stream filter "locations=[-180,-90,180,90]" have been used as a query to get all tweets worldwide containing a geolocation. Two datasets from different time periods have been collected to avoid that the data only show one snapshot of a country's emotionality but rather a country-specific emotionality regardless of a particular event. The first data set includes 6,961,676 tweets (time period December 15-17 2015), the second comprises 7,230,505 tweets (time period March 3, 6-8, 2016).

The number of unfiltered, raw tweets per dataset was around 7 million, still including countries not investigated because of the insufficient amount of tweets. Retweets—identified by “RT” inside the tweet—have not been processed because they do not reflect the opinion or emotions of the tweeter her- or himself.

2.2 Preprocessing

Before the gathered data could be analyzed, it had to be preprocessed. Usernames have been replaced with “@USERNAME” and links have been replaced with “URL” to ensure that possible sentiments containing names or symbols like “:/” do not change the meaning of the actual tweet. Also, the hashtag symbol has been removed from all tweets so that the words themselves can be identified. In addition, some special characters which have been represented in the wrong way have been fixed, e.g. “&lt” to “<.” The next step was the translation for all non-English tweets. From the whole data of 14,192,182 tweets, for 1,053,761 tweets the original language could not be identified, so they could not be translated into English. While in other investigations the preprocessing contains further steps as reducing words with repetitive characters or to replace abbreviations with their written-out word(s), it is not necessary for this study because the abbreviations have been included in an emotion lexicon and words with consecutively repeating characters were considered to be an expression of emotionality.

2.3 Sentiment Analysis

For the sentiment analysis the tweets have been analyzed on the granularity of a document level, so exactly one sentiment is assigned to a tweet. For the investigation of tweets in multiple languages, all non-English tweets have been translated to English (via the free Python library TextBlob)—because there are some evaluated anglophone emotion lexica and the most used language on Twitter is English as well. The difficulty with this option is the quality of the translation; so some tweets will not be translated or in a way that changes the original statement. But it still appeared as the most appropriate solution.

There are many different emotion lexicons and other possible features for conducting sentiment, so it had to be decided which one to use in this case. In a comparative study on explicit Twitter sentiment analysis, Koto and Adriani [16] have investigated which features are appropriate. By using four different datasets, the results reveal that AFINN lexicon and SentiStrength method are the best current approaches to perform Twitter sentiment analysis.

AFINN is a list of English words rated for valence with an integer between -5 (negative) and +5 (positive), specifically constructed for microblogs. The words were manually labeled by Finn Årup Nielsen in 2009-2011 [24]. SentiStrength is a lexicon-based classifier that uses additional linguistic information and rules to detect sentiment strength in short informal English texts. For each text, the SentiStrength output is expressed by two integers: 1 to 5 for positive sentiment strength and a separate score of 1 to 5 for negative sentiment strength. Two scales are used because even short texts can contain both positivity as well as negativity and the goal is to detect the sentiment expressed rather than its overall polarity [38].

In the end, to decide which emotion lexicon generates the best results for the gathered data, two sentiment analysis tools have been compared: AFINN and SentiStrength. For the comparison as well as an evaluation each lexicon has been used for the sentiment analysis of two evaluation datasets: 1,000 English tweets and 1,000 German tweets. These English and German tweets have been chosen manually from the gathered data to ensure that tweets with different kinds of content are contained.

All in all, seven people were involved in the evaluation of these tweets; two persons a time for different parts of data. The tweets have been labeled with 1 for a positive, 0 for a neutral or -1 for a negative sentiment. In the case that the two persons were divided over a tweets' sentiment, the evaluation group decided together how to assess the sentiment. In the next step, the automatically obtained results of the emotion lexica have been scaled from different values—according to the lexicon's range—to [1,0,-1] and aligned to the human evaluation. AFINN and SentiStrength both had an accordance of around 70% for the German tweets and a bit higher for the English ones. This result was not good enough to apply one method to the whole dataset, so the lexica had to be improved manually. That was the point when it came to a decision which lexicon would be used and enlarged for this investigation.

While SentiStrength is more complex and already includes an emoticon lexicon, a negation lexicon etc., it also is a complete program and less easy to adjust. For this reason, AFINN has been chosen but extended with further lexica and rules. So, all in all, the following lexica have been used: the original AFINN emotion lexicon—but extended with new words and deleted some original ones—, an emoticon lexicon, a negation lexicon as well as a lexicon for booster words —like "very," "totally," etc.—and a lexicon for phrases, using SentiStrength as a model.

After multiple repetitions of adaption of the lexica and rules, the final agreement for the evaluation datasets could be clearly improved up to 91.4% agreement for English tweets and 80.4% agreement for German tweets. Therefore, this approach has been used for the sentiment analysis of both datasets.

2.4 Preselection of Data Features for Further Processing

After the sentiment analysis, each tweet has a distinct value regarding its sentiment. These values have to be aggregated by country to get country-specific findings of emotionality.

To compare different countries worldwide, a selection of the most active countries had to be made to obtain significant results. If there are only a few tweets of one country in a big dataset, we are not able to infer meaningful conclusions from such a small empirical basis. While most of the sentiment analyses only focus on English tweets, many users also post in their mother tongue. In 2013, the most widely used language on Twitter was indeed English with an amount of 34%. The subsequent languages are Japanese (16%) and Spanish (12%) [4]. Therefore, English captures only one third. To gather as much content of a country as possible, for each country all tweets were collected. This proceeding also enables to recognize possible differences in emotionality between the use of mother tongue and English for countries with mainly non-native English speakers.

Tweets with an unrecognized language have been excluded from the analysis because they could not be translated correctly. Moreover, outliers have been excluded. The next step was the normalization of the sentiment within an interval [-5,5], so -5 is the value of the most negative tweets and 5 for the most positive ones. The normalized values finally can be used for the aggregation by country. Therefore, they are divided into positive, neutral, and negative. The aggregated values are the mean positive or negative sentiment of a country. 110 countries result from Dataset 1; 115 from Dataset 2. The mean values of all countries approach for both datasets, as well as the values of the different countries. For this reason, the two datasets have been combined for further computations. This raw dataset contains 14,192,182 tweets. If the requirements are a recognized tweet language, a number of tweets per country >= 1,000, and an absolute sentiment value <= 40, 13,109,881 tweets remain for analysis. All in all, 126 countries have been considered for this investigation.

The United States have the most tweets with over three million tweets, followed by Brazil with less than two million tweets, and Argentina with less than one million tweets. In this study, English is the most used language on Twitter by far with more than five million tweets, followed by Spanish and Portuguese tweets with around two million tweets. Although English is a world language and the dominant language on Twitter, the rate of English tweets varies from country to country. While the United States have an English rate of 94.74%, Brazil—which is the second-strongest country of the dataset—only has a rate of 4.88%.

3. Results

The results of this investigation are the outcome of the multilingual data since it gives a more extensive overview of the Twitter community of countries with a low rate of English tweets. It has to be kept in mind that the reliability of the sentiment is dependent on the translation and the lexica as well. Altogether, in the dataset (13,109,881 tweets) most tweets are neutral (6,219,971), followed by positive tweets (4,426,579). Negative tweets (2,463,331) are represented nearly half as often as the positive ones.

3.1 Sentiment Strength, Sentiment Climate, and the Amount of Positive and Negative Tweets (RQ 1 and RQ 2)

Regarding RQ 1, the sentiment strength shows the emotionality of people from different countries around the world (Figure 1). While at first glance no pattern can be identified, particular points are conspicuous. East Asia as a bigger cluster is less emotional than the rest of the world. Parts of Africa are very strongly emotional. The top 10 countries of the highest and lowest sentiment strength can be obtained from Table 1.

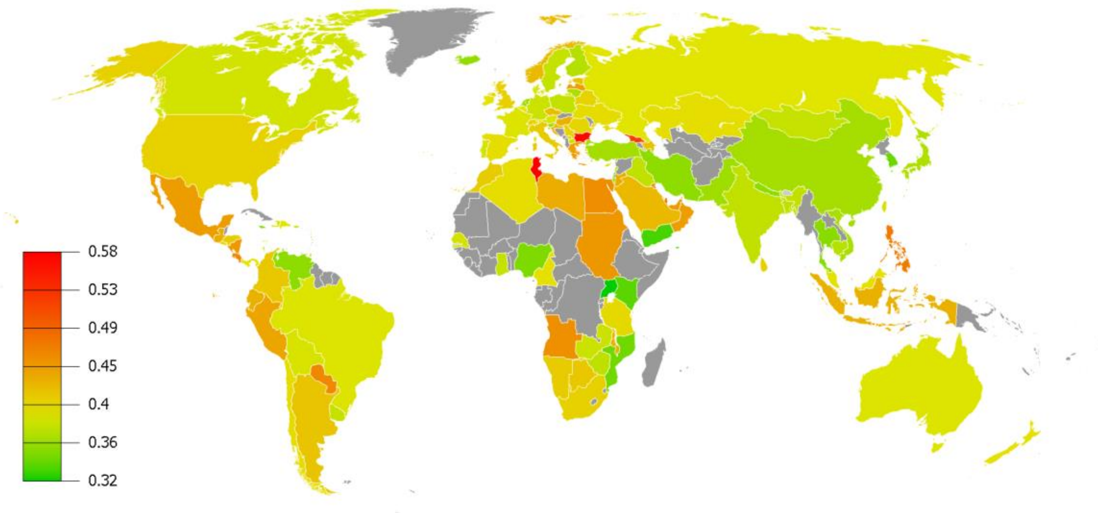


Figure 1 Sentiment strength map.

Table 1 Top 10 countries with the highest and lowest sentiment strength (n = 126 countries).

#	Top 10 Highest Sentiment Strength		Top 10 Lowest Sentiment Strength	
	Country	**Strength**	**Country**	**Strength**
1	Tunisia	0.578	Antarctica	0.312
2	Bulgaria	0.572	Uganda	0.316
3	Georgia	0.524	Yemen	0.328
4	Bahrain	0.476	Kenya	0.334
5	Philippines Qatar	0.473	South Korea	0.339
6	Guam	0.465	Mozambique	0.342
7	Costa Rica	0.464	Barbados Nigeria	0.347
8	Paraguay	0.463	Jamaica Thailand	0.351
9	Kuwait	0.462	Iran Nepal Venezuela	0.353
10	Kosovo	0.459	Iceland	0.356

According to the negative sentiment (Figure 2), countries like the United States or Russia as well as many countries in South America exhibit a rather negative sentiment. East Asia, in contrast, is still less negative. The highest negative sentiment received the countries Guam (-0.373), Kosovo (-0.369), and Nicaragua (-0.359).

Bulgaria and Tunisia are outliers regarding to the positive sentiment. For that reason, they have been examined more precisely. Bulgaria has a strong positive sentiment for Dataset 2 (0.787), so it has been searched for specific events in that timeframe. The top hashtag of this dataset was "KCA" with 160 mentions. KCA stands for the Nickelodeon Kids' Choice Awards. This is an annual award show, honoring the year's biggest television, movie, and music acts, as voted by Nickelodeon viewers and took place on 12th March 2016. Furthermore, the 3rd of March is a National Holiday (Bulgaria's Liberation from the Ottoman Empire) [21].

Tunisia has a strong positive sentiment for Dataset 1 (0.796). Therefore, the other timeframe has been investigated. The hashtag "NassimRaissi" has been used the most (262 times). Nassim Raissi is a Tunisian singer of the Star Academy 11 which is a reality talent show that brings together candidates from the Arab World. The hashtag "SanayaIrani" was second most (97 times) and stands for an Indian actress, while the third most hashtag was "KCA" (76 times) again. Furthermore, the 17th December 2015 was the fifth anniversary of the Tunisian Revolution and the wider Arab Spring.

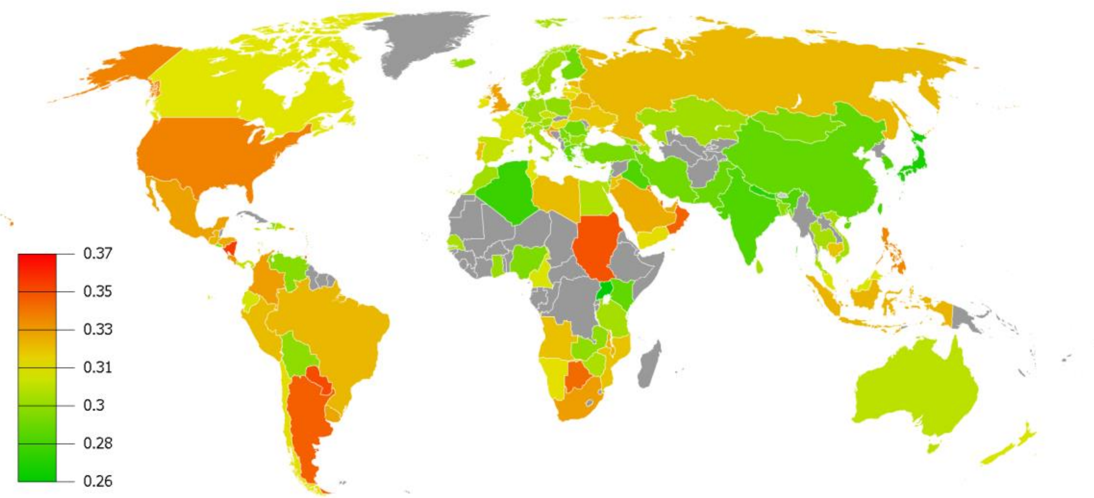


Figure 2 Negative sentiment map.

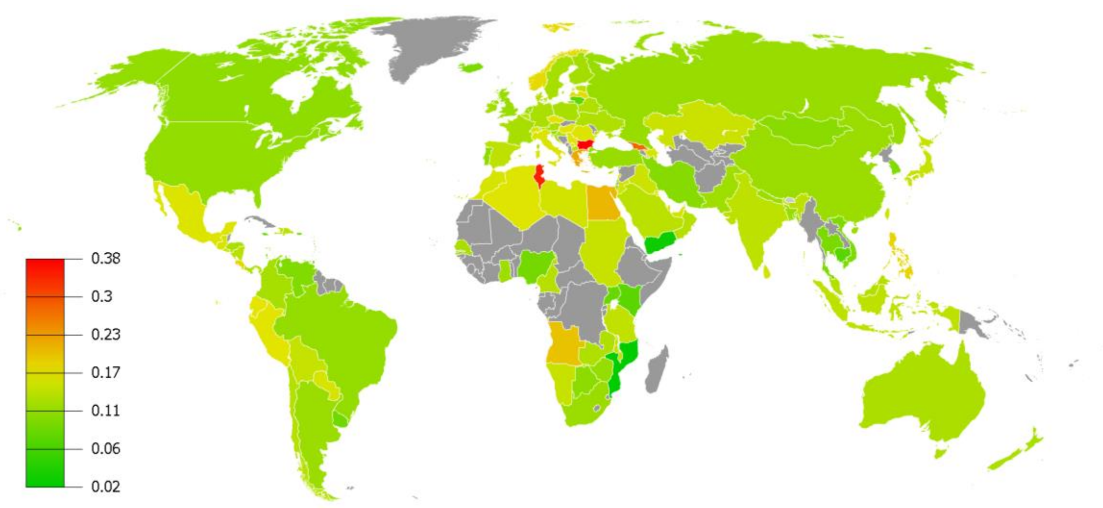


Figure 3 Sentiment climate map.

The sentiment climate (Figure 3) is consistently positive (except for Antarctica). While in most countries of the world a low positive sentiment predominates, in some African countries (and, of course, in Bulgaria) the sentiment climate is much higher. Table 2 displays the top 10 countries of the highest and lowest sentiment climate.

Table 2 Top 10 countries with the highest and lowest sentiment climate (n = 126 countries).

#	Top 10 Highest Sentiment Climate		Top 10 Lowest Sentiment Climate	
	Country	**Climate**	**Country**	**Climate**
1	Bulgaria	0.375	Antarctica	- 0.037
2	Tunisia	0.348	Mozambique	0.023
3	Georgia	0.284	Barbados	0.025
4	Greece	0.224	Yemen	0.027
5	Egypt	0.208	Cambodia	0.069
6	Angola	0.198	Antigua Barbuda	0.070
7	Brunei Darussalam	0.192	Kenya Lithuania	0.071
8	Costa Rica Philippines	0.182	Jamaica Malta	0.073
9	Bahrain	0.181	Uganda	0.080
10	Norway	0.176	Uruguay Puerto Rico	0.083

3.2 Features' Impact on Emotionality (RQ 3)

The following features were considered regarding their impact on the emotionality of different countries: positive sentiment, negative sentiment, sentiment climate, sentiment strength, sentiment word counter, emoticon counter, repeated punctuations counter, repeated characters counter, and uppercase counter.

All sentiment means of the stated features were correlated with each other (Table 3). It is not surprising that the positive sentiment correlates negatively with the negative sentiment (-0.363**). The positive sentiment is also strongly significant related to the sentiment climate (0.907**) and the sentiment strength (0.987**) which means that the positive emotions are definitely stronger than the negative ones. Furthermore, it is coherent with the number of emoticons (0.667**). We can conclude from this that more emoticons mean a higher positive sentiment. The positive emoticons prevail over the negative ones.

In contrast, the negative sentiment is negatively connected to the sentiment strength (-0.436**). It also correlates negatively with the number of emoticons (-0.480**) which supports the assumption that people rather use emoticons in a positive context. The repeated characters within words are negatively connected to the negative sentiment (-0.298**) which means that repeated characters often rather appear in a positive than in a negative context.

The sentiment climate strongly correlates with the sentiment strength (0.860**). This is a conclusion of the fact that the positive sentiment influences both features. Sentiment climate is also positively related to the number of sentiment words (0.229**) and emoticons (0.497**). Besides, the sentiment strength is connected to the number of emoticons (0.690**) as a consequence of the coherence between positive sentiment and emoticons. The number of sentiment words is related to the number of repeated punctuations (0.312**), accordingly punctuations are an expression of emotionality as they often appear with emotional words. The number of emoticons is negatively correlated with the number of uppercase sentences (-0.264**), so emoticons appear less in uppercase written sentences.

Table 3 Correlations between sentiment feature means by country.

	Positive Sentiment	Negative Sentiment	Sentiment Climate	Sentiment Strength	Sentiment Word Counter	Emoticon Counter	Repeated Punctuations Counter	Repeated Characters Counter	Uppercase Counter
Positive Sentiment	1								
Negative Sentiment	- 0.363**	1							
Sentiment Climate	0.907**	0.064	1						
Sentiment Strength	0.987**	- 0.436**	0.860**	1					
Sentiment Word Counter	0.133	0.192*	0.229**	0.090	1				
Emoticon Counter	0.667**	- 0.480**	0.497**	0.690**	- 0.030	1			
Repeated Punctuations Counter	0.085	0.037	0.108	0.064	0.312**	0.077	1		
Repeated Characters Counter	0.095	- 0.298**	0-.033	0.091	- 0.200*	0.122	- 0.005	1	
Uppercase Counter	0.012	- 0.127	-0.045	-0.008	0.173	-0.264**	0.192*	0.138	1

*: p <= 0.05, **: p <= 0.01, ***: p <= 0.001

3.3 Coherences with other Studies and Indices (RQ 4 and RQ 5)

To answer RQ 4 and RQ 5 the sentiment values have been compared with other studies and the following indexes: Human Development Index [41], the countries' average temperature, GDP per capita [39], Happy Planet Index [22], Six Dimensions of National Culture [14], and the Global Emotions Report [12].

The results do not show any significant correlations to the Human Development Index, GDP per capita, and the Happy Planet Index. A study says that people from countries of warm climates are believed to be more passionate than people from countries of cold climates [25]. Therefore, the average yearly temperature of each country (measured between 1961 and 1990) has been compared to the sentiment values for 118 countries. According to the Pearson correlation, only the negative sentiment mean of the countries has a slightly negative correlation to the annual average temperature (-0.277**). This would mean that negative sentiment slightly relates to a higher annual average temperature. However, there is no significant correlation between temperature and positive sentiment.

From Hofstede's six dimensions of national culture [14], only Indulgence versus Restraint (IND) correlates significantly with the sentiment (80 countries have been compared). "Indulgence stands for a society that allows relatively free gratification of basic and natural human drives related to enjoying life and having fun. Restraint stands for a society that suppresses gratification of needs and regulates it by means of strict social norms" [15]. North and South America both have high scores, as well as Oceania and some European and African countries meaning that these cultures are more indulgent. Hofstede explains that “people in societies classified by a high score in Indulgence generally exhibit a willingness to realize their impulses and desires with regard to enjoying life and having fun. They possess a positive attitude and have a tendency towards optimism. In addition, they place a higher degree of importance on leisure time, act as they please and spend money as they wish” [13]. On the other hand, countries with a low score have “a tendency to cynicism and pessimism” and “do not put much emphasis on leisure time and control the gratification of their desires” [13]. The slightly negative correlation is significant for the positive sentiment (-0.226*) as well as for the sentiment climate (-0.325**). Positive sentiment is rather strong for countries with less indulgent societies. This is not really consistent with the mentioned characteristics of an indulgent society.

4. Discussion

The aim of this work was to investigate differences between countries according to emotionality on Twitter. After data preparation, 13,109,881 tweets from 126 countries were analyzed. The results indicate that there is indeed a country-specific emotionality based on tweets.

There are countries that are generally more or less emotional (regardless of whether positive and negative). East Asia as a big cluster is less emotional in general, whereas parts of Africa are generally very emotional. The United States, Russia and many countries in South America are more negative in comparison to other countries in the world. Apart from Antarctica, all countries have a positive sentiment climate so their users rather write positive tweets than negative ones. Especially in Bulgaria and Tunisia, the positive sentiment is strong. However, this may be due to events that took place at the time of research.

Emoticons have a great power in this study as they strongly correlate with positive sentiment. In contrast to emojis, features as repeated punctuations or uppercases do not have a strong impact on the emotionality since they represent only a small part of tweets.

To evaluate if the data is consistent with other studies, different indices were compared with the given results. The average annual temperature and Hofstede’s IND (Indulgence vs. Restraint) dimension show slightly significant coherences. All of them measure different aspects which are only parts of multiple reasons for the given results.

Problematic within this investigation are the required translation, the small time frames and the fact that words can mean different things in different contexts. A sentiment lexicon is not adaptable in every single case. The emoticon lexicon also does not contain all existing emojis because they can differ in meaning and in the case of Japanese emojis they are just too many and it is hard to weight them consistently. Although two datasets from different times have been used, they cannot give a total reliable overview over the sentiment of countries on principle. Important factors are the events that took place in the analyzed timeframes. To understand the sentiment for the investigated timeframes, the top hashtags of each country could be considered, e.g. if there was a specific positive or negative emotion causing event. To get solid data, tweets from longer time periods should be collected.

Future works could try other variants of sentiment analysis, e.g. work with a machine learning approach or differentiate between more specific emotions than positive and negative ones, e.g. sadness, disgust, etc. Another option would be to develop a hashtag lexicon with sentiment words, e.g. #sad, #love, in different languages to collect tweets which definitely include sentiment and to compare these tweets by country.

In practice, our approach can be used for international comparisons to concrete topics as an election campaign, opinions to politicking, or other events. It could also be adapted to other digital communication platforms than Twitter in respect to different user groups (e.g. Weibo in China). All in all, this investigation answered the question if there is a country-specific sentiment—the answer being yes. Furthermore, and to our knowledge, it was a first idea to introduce sentiment strength and sentiment climate in the toolbox of sentiment analysis.

5. Acknowledgments

Computational support and infrastructure was provided by the “Centre for Information and Media Technology” (ZIM) at the University of Düsseldorf (Germany).

6. References

[1] F. Aisopos, D. Tzannetos, J. Violos, and T. Varvarigou, “Using n-gram graphs for sentiment analysis: An extended study on Twitter”, in *Proceedings of the 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 44-51), IEEE, Washington, DC, 2016.

[2] M. Arias, A. Arratia, and R. Xuriguera, “Forecasting with twitter data”, *ACM Transactions on Intelligent Systems and Technology, 5*(1), 2013, pp. 1-24.

[3] E.Y. Bann, and J.J. Bryson, “Measuring cultural relativity of emotional valence and arousal using semantic clustering and Twitter”, in *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1809-1814), Cognitive Science Society, Austin, TX, 2013.

[4] W. Baumann, *Twitter’s world,* 2013, Online: http://www.technologyreview.com/graphiti/522376/the-many-tongues-of-twitter/. (Retrieved November 16, 2015)

[5] M. Bautin, L. Vijayarenu, and S. Skiena, “International sentiment analysis for news and blogs”, in *Proceedings of the 2nd International Conference on Weblogs and Social Media – ICWSM,* The AAAI Press, Menlo Park, CA, 2008.

[6] J. Clifton, *Singapore Ranks as Least Emotional Country in the World*, 2012, Online: http://www.gallup.com/poll/158882/singapore-ranks-least-emotional-country-world.aspx#1. (Retrieved June 13, 2018)

[7] S. Coats, “Skin tone emoji and sentiment on twitter”, in E. Mäkelä, M. Tolonen, J. Tuominen (Eds.), *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference* (pp. 122-138), Helsinki, Finland, 2018.

[8] A. Cui, M. Zhang, Y. Liu, and S. Ma, “Emotion tokens: Bridging the gap among multilingual twitter sentiment analysis”, in *Proceedings of the 7th Asia Information Retrieval Societies Conference* (pp. 238-249), Springer, Berlin, Heidelberg, Dubai, UAE, 2011.

[9] A.O. Durahim, and M. Coşkun, “#iamhappybecause: Gross National Happiness through Twitter analysis and big data”, *Technological Forecasting & Social Change, 99,* 2015, pp. 92-105.

[10] P. Ekman, and W.V. Friesen, “Constants across cultures in the face and emotion”, *Journal of Personality and Social Psychology, 17*(2), 1971, pp. 124-129.

[11] R. Feldman, “Techniques and applications for sentiment analysis”, *Communications of the ACM, 4,* 2013, pp. 82-89.

[12] Gallup, Inc., *Gallup 2017 Global Emotions*. (Retrieved June 4, 2018)

[13] G. Hofstede, *National Culture*. n.d., Online: https://geert-hofstede.com/national-culture.html. (Retrieved July 10, 2016)

[14] G. Hofstede, “Dimensionalizing cultures: The Hofstede model in context”, *Online Readings in Psychology and Culture, 2*(1), 2011, pp. 1-26.

[15] Hofstede-Insights, *National Culture*, n.d., Online: https://www.hofstede-insights.com/models/national-culture/. (Retrieved July 11, 2018)

[16] F. Koto, and M. Adriani, “Comparative study on Twitter sentiment analysis: Which features are good?”, in C. Biemann, S. Handschuh, A. Freitas, F. Meziane, and E. Métais (Eds.), *Proceedings of the 20th International Conference on Applications of Natural Language to Information Systems, NLDB 2015* (pp. 453-457), Passau, Germany, 2015.

[17] P. Kralj-Novak, J. Smailović, B. Sluban, and I. Mozetič, “Sentiment of emojis”, *PLoS ONE, 10*(12), 2015. https://doi.org/10.1371/journal.pone.0144296

[18] J.D. Lindlbauer, “Konjunkturtest”, in K.H. Oppenländer, and G. Poser (Eds.), *Handbuch der ifo-Umfragen* (pp. 122-187), Duncker & Humblot, Berlin, Germany 1989.

[19] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool, San Rafael, CA, 2012.

[20] Mapbox, *Languages on Twitter*, 2015, Online: https://www.mapbox.com/labs/twitter-gnip/languages/. (Retrieved November 6, 2015)

[21] National Assembly of the Republic of Bulgaria, *Bulgarian Official Holidays*, n.d., Online: http://www.parliament.bg/en/24. (Retrieved July 21, 2016)

[22] New Economics Foundation, *Happy Planet Index*, 2018, Online: http://www.happyplanetindex.org/. (Retrieved June 6, 2018)

[23] P.M. Niedenthal, J.B. Halberstadt, J. Margolin, and Å.H. Innes-Ker, “Emotional state and the detection of change in facial expression of emotion”, *European Journal of Social Psychology, 30*(2), 2000, pp. 211-223.

[24] F. Å. Nielsen, “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs”, in *Proceedings of the ESWC 2011 Workshop on 'Making Sense of Microposts': Big Things Come in Small Packages* (pp. 93-98), Heraklion, Greece, 2011.

[25] J.W. Pennebaker, B. Rimé, and V. E. Blankenship, “Stereotypes of emotional expressiveness of northerners and southerners: A cross-cultural test of Montesquieu’s hypotheses”, *Journal of Personality and Social Psychology, 70*(2), 1996, pp. 372-380.

[26] B. Poblete, R. Garcia, M. Mendoza, and A. Jaimes, “Do all birds tweet the same? Characterizing twitter around the world”, in *CIKM '11 Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 1025-1030), ACM, New York, NY, 2011.

[27] A. Qadir, and E. Riloff, “Bootstrapped learning of emotion hashtags #hashtags4you”, in *4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 2-11), Association for Computational Linguistics, Atlanta, Georgia, USA, 2013.

[28] K. Roberts, M.A. Roach, J. Johnson, J. Guthrie, and S.M. Harabagiu, “EmpaTweet: Annotating and detecting emotions on Twitter”, in *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3806-3813), European Language Resources Association (ELRA), Istanbul, Turkey, 2012.

[29] J. Russell, “Culture and categorization of emotions”, *Psychological Bulletin, 110*(3), 1991, pp. 426-450.

[30] N.M. Shelke, S. Deshpande, and V. Thakre, “Survey of techniques for opinion mining”, *International Journal of Computer Applications, 57*(13), 2012, pp. 30-35.

[31] K. Singh, H.K. Shakya, and B. Biswas, “Happiness index in social network”, in D. Singh, B. Raman, A. Luhach, and P. Lingras (Eds.), *Advanced Informatics for Computing Research. ICAICR 2017. Communications in Computer and Information Science* (pp. 261-270), Springer, Singapore, 2017.

[32] P.B. Smith, and M.H. Bond, *Social psychology across cultures,* 2nd Ed., Allyn & Bacon, Boston, MA, 1998.

[33] K.S. Smith, R. McCreadie, C. Macdonald, and I. Ounis, “Regional sentiment bias in social media reporting during crises”, *Information Systems Frontiers*, 20(5), 2018, pp. 1013-1025.

[34] Statista, *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2018 (in millions),* 2018, Online: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/. (Retrieved June 3, 2018)

[35] W.G. Stock, and M. Stock, *Handbook of Information Science,* De Gruyter Saur, Berlin, Germany, and Boston, MA, 2013.

[36] E.S. Tellez, S. Miranda-Jiménez, M. Graff, D. Moctezuma, R.R. Suárez, and O.S. Siordia, “A simple approach to multilingual polarity classification in Twitter”, *Pattern Recognition Letters, 94*, 2017, pp. 68-74.

[37] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment strength detection for the social web”, *Journal of the American Society for Information Science and Technology, 63*(1), 2012, pp. 163-173.

[38] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, “Sentiment strength detection in short informal text”, *Journal of the American Society for Informational Science and Technology, 61*(12), 2010, pp. 2544-2558.

[39] The World Bank Group, *GDP per Capita, PPP*, Online: http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD. (Retrieved June 11, 2018)

[40] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, “Predicting elections with Twitter: What 140 characters reveal about political sentiment”, in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media* (pp. 178-185), The AAAI Press, Menlo Park, CA, 2010.

[41] United Nations Development Programme, *Human Development Index (HDI)*, n.d. Online: http://hdr.undp.org/en/content/human-development-index-hdi. (Retrieved June 11, 2018)

[42] Y. Wang, M.S.M. Pozi, Y. Kawai, A. Jatowt, and T. Akiyama, "Exploring cross-cultural crowd sentiments on Twitter", in *HT 2017 - Proceedings of the 28th ACM Conference on Hypertext and Social Media* (pp. 321-322), ACM, New York, NY, 2017.

[43] W. Wolny, "Emotion analysis of Twitter data that use emoticons and emoji ideograms", in J. Gołuchowski, M. Pańkowska, C. Barry, M. Lang, H. Linger, and C. Schneider (Eds.), *Information Systems Development: Complexity in Information Systems Development (ISD2016 Proceedings)* (pp. 163-171), University of Economics in Katowice, Katowice, Poland, 2016.