

Sentiment Analysis of National Tourism Organizations on Social Media

Jan HRUŠKA

University of Hradec Králové, Hradec Králové, Czech Republic; jan.hruska.3@uhk.cz

Abstract: Social media is probably currently the largest source of human-generated text content. User opinions, feedback, comments, and criticism points to their mood and sentiment towards different topics, especially destinations, products or services. The rapid rise in amount of data and constantly generated content require the need to automate both data acquisition and processing to identify important information and knowledge. Sentiment analysis provides the opportunity to detect opinion, feeling and sentiment from unstructured texts on social media. To analyze the sentiment Machine Learning with Google Natural Language API Client Libraries and Google Cloud SDK (Software development kit) was used. NTOs (National Tourism Organizations) social media have been chosen for analysis in which emotional messages can be expected to stimulate potential visitors to the destination. It was found that all selected NTOs add mostly positive posts and in the sample of two hundred contributions there are only seven with negative polarity of sentiment. There was a moderate correlation between customer growth and positive polarity in the contribution. The results show that creating stable positive descriptions for posts can be one of the key variables for the growth of the fan base and stimulation of potential visitors.

Keywords: sentiment analysis; national tourism organizations; social media

JEL Classification: O35; M31

1. Introduction

The huge increase in popularity of big data on social media and social media in general makes it possible for the general public to express their opinions on a wide range of areas such as the state of the economy, enthusiasm or disappointment with a particular product, or to show pleasure in making a purchase (Shayaa et al. 2017).

Internet content has changed over the last few years, and so has the distribution of content. The shift was made possible by new technologies, new media and new communication tools. This change has led to a massive expansion of social media and a huge increase in short informal messages that are accessible to the public (Nakov et al. 2016). With the increasing number of different blogs, groups, forums, and social media as a source of online communication, it is now possible to analyze a huge amount of data that shows mood and feelings (Lyu and Kim 2016). Properly analyzing this data can significantly improve business efficiency (Hedvičáková and Král 2019).

Content on all possible social media is spreading very quickly just because users can freely share whatever they want. This is one of the reasons why the social media posts may have millions of reactions in a few days (Yoo, Song, and Jeong 2018). One of the ways marketers are currently trying to exploit this content and the response of users is sentiment analysis.

Sentiment analysis is the use of natural language processing and text mining to identify and extract subjective information from source materials, most often texts (Xiang et al. 2017; Alaei, Becken, and Stantic 2017). Finding these opinions can play a key role in understanding user, consumer or voter behavior.

This fact has recently brought great interest in analyzing sentiment and gathering opinions mainly from the text. Most current methods and approaches to sentiment analysis mainly use a number of positive and negative words in the text to reveal the polarity of sentiment (Dridi and Reforgiato Recupero 2019).

Sentiment analysis is used to gather opinions from different types of data (text, video or audio). It is most often text analysis that is frequently used for different areas of marketing, customer needs and customer service, including tourism. The mood or relationship of the customer or users in the examined articles is mainly analyzed by texts such as contributions, ratings and comments.

There are countless opinions on social media that express users' response to events, friend activities, products, or services. While these messages are short, such as on Twitter, up to 140 characters, the short messages can be used to identify a person's mood and feelings through classification, sentiment analysis, or machine learning. (Gaspar et al. 2016).

Most of the data available on social media is unstructured (Gaspar et al. 2016). According to them, approximately 80% of the total data in the world is unstructured, which makes this data difficult to analyze and extract valuable information from it.

Two very important techniques that can help detect emotions and opinions from data available on social media are:

- Sentiment analysis
- Opinion mining

If the mood and feelings of users are correctly detected, these findings can help solve problems in many areas such as elections, public opinion, advertising, marketing, healthcare, public satisfaction and, of course, tourism (Ahmed, Tazi, and Hossny 2015). They highlight, that one of the problems with sentiment analysis is that sometimes it is difficult to tell from the text what emotion the user is trying to convey. They further state, that sentiment analysis can also be used to define trust on social media for a brand or service.

Traditional statistical analyses and methods are not always suitable for big data analysis, precisely because the data is often unstructured. Methods associated with Big Data analysis unify specific tools to find similarities and patterns in large volumes of data. These methods include natural language processing, artificial intelligence, data mining or predictive analysis (Kirilenko et al. 2018).

Social media such as Twitter, Facebook and Chinese Weibo are the ultimate platforms for users to share comments, experience about a product, service, or even the entire business. Businesses are constantly trying to expand their offline and online network (Pochobradská and Marešová 2018). These social media are priceless for those who focus and understand the sentiment of the public (Wang et al. 2016). The main goal in case of sentiment analysis is to correctly classify user generated content (usually text) into either positive or negative polarity (Dhaoui, Webster, and Tan 2017). This was the main impulse for the creation of the sentiment analysis and the subsequent use of this analysis on social media. Sentiment analysis is an important research field today, which still faces many challenges due to the typical structure of social media and microblogs. These contributions are typically very short and full of noise unlike conventional texts such as newspapers (Dridi and Reforgiato Recupero 2019).

It is a bit surprising that even though sentiment analysis is increasingly being used for various purposes, it does not have as much attention among scientists for the overall use of sentiment analysis as an online marketing tool (Rambocas and Pacheco 2018). Over the past few years, sentiment analysis or online feedback have been increasingly used posts. Sentiment analysis uses the principles of natural language processing to identify attitudes and opinions about a specific product, description, or value. Thanks to the huge amount of information and data on the Internet, manual evaluation of sentiment is not a suitable option. Automating the process of collecting and evaluating data is the only practical solution to determine usable opinion from data available on the Internet. These evaluated data can then be used to improve decision making. Improved decision-making on sentiment analysis can be beneficial in many areas including financial market, marketing, e-commerce, politics, law, public decision-making and tourism.

Sentiment analysis has become a standard component of the social media analysis toolkit of marketers and customer relation managers in large organizations (Thelwall 2019). For its use from 2014 to 2019, a review study by Drus and Khalid (2019) demonstrates the prevailing lexicon-based approach. In tourism, the analysis of sentiment in social media is used to research the perception and evaluation of the quality of tourist services (Airport Service Quality – Martin-Domingo, Martín, and Mandsberg 2019) and analysis of tourist relations to destinations (Liu et al. 2019). Sentiment analysis

is also part of research into various features of review sites (e.g. Xiang et al. 2017), where sentiment analysis could also be used as a tool to detect false reviews (Kauffmann et al. 2019).

2. Methodology

The purpose of this research was to find out which sentiment NTOs use in their posts in social media and compare this sentiment with followers for the past two years. Facebook was chosen as the focus social medium because it is the dominant social network and contains the largest number of companies trying to promote their products and services. All analyzed posts in this work are from the social network Facebook. NTOs from the countries with the highest visitor numbers of international tourists were selected. According to UNWTO (2019), the top ten most visited countries in the world in 2018 were France, USA, Spain, China, Italy, Great Britain, Germany, Mexico, Thailand and Turkey.

Of the ten countries selected, Thailand is the only country that does not post on Facebook in English, and for this reason, another country has been added, namely Australia.

A Google product and library package called Artificial Intelligence and Machine Learning with Google Natural Language API (Application Programming Interface) Client Libraries and Google Cloud SDK (Software development kit) was used to analyze sentiment.

Algorithm for sentiment analysis is written in language Python in software PyCharm. The sentiment analysis goes through the inserted text and identifies the emotional opinion in the text, to reveal the author's opinion as either positive, negative or neutral. The sentiment analysis determines what polarity prevails in the text. Output variables from this method are score and magnitude.

The variable score points to the overall emotional state of the inserted text. Magnitude then shows how much emotional content is present in the text. The higher the magnitude, the usually higher the number of positive and negative words in the text. The Natural Language API shows the differences between positive and negative emotions in the selected text, but does not show specific emotions. For example, "annoyed", "sad" or "disappointed" are considered negative emotions. If the score of the pasted text is around zero (neutral), it may mean text with a low number of words expressing emotions, or it may mean mixed emotions with both positive and negative words in the text. Authors used magnitude values to uncover these cases, as truly neutral documents will have a low magnitude value, while mixed documents will have higher magnitude values. The example below shows some sample values and how to interpret them:

- Score 0.9 and magnitude 4.2 (Clearly Positive)
- Score -0.7 and magnitude 3.8 (Clearly Negative)
- Score 0.2 and magnitude 0.1 (Neutral)
- Score 0.0 and magnitude 5.4 (Mixed)

The original research plan consisted of analyzing the posts that are a clickable article (that is, not to analyze posts that are a video or image) and the text on the social media for the post, and then re-analyzing the sentiment in the article after clicking the post. After examining hundreds of contributions from selected NTOs, this procedure has been changed to study text on social media only, since most NTOs did not add ten posts in 2019 as an article. They focus mainly on photos and videos.

The text used to analyze sentiment will be twenty posts from each Facebook social network for each national tourism organization (these posts are usually shared in the same form on Twitter). The text used on social network will be analyzed. Total of 200 posts from 10 selected NTOs will be tested. Every character from post will be analyzed. Only links will be excluded from the sentiment analysis as they have no value for this test.

The following research question will be asked: Does the polarity of sentiment analysis affect subscriber growth on social media?

Pearson correlation in statistical software IBM SPSS Statistics will be used for statistical analysis.

3. Results

To determine the sentiment used in selected national tourism organizations, a total of 200 posts were analyzed. Analyzed was text attached to post on social media. Posts were selected using the

random selection method from all posts inserted in 2019 and are mostly image or video posts. These 200 posts were selected from a total of almost 2,500 uploaded posts.

One very positive, very negative and neutral short text will be presented to demonstrate the functionality of the algorithm.

Clearly Positive

Text: "The product is amazing! It solved my problem and I highly recommend it!"

Output: Sentiment: 0.9, magnitude 1.9

Clearly Negative

Text: "The movie was awful, the performances were terrible, and even the music was not very good."

Output: Sentiment: -0.9, magnitude 1.8

Neutral

Text: "The book is well written even though it has weaker passages, the main character is awesome too."

Output: Sentiment: 0.0, magnitude 0.0

The table below shows the selected countries with the largest number of international tourists and their number of fans as of 31st of December 2017 (column titled "Subs. in 2017"), followed by the number of their fans as of 31st of December 2019 (column titled "Subs. in 2019). The column labeled "Increase" shows the increase in the number of fans over the last two years. "Score" and "Magnitude" show the average sentiment of the twenty posts on Facebook in 2019 and the column "Char." shows the average number of characters in the post, including spaces of those twenty posts.

Table 1. Number of subscribers and sentiment of the posts for selected NTOs.

Country	Subs. in 2017	Subs. in 2019	Increase	Score	Magnitude	Char.
France	1,526,227	1,989,406	30%	0.58	1.12	105
Turkey	5,299	6,562	24%	0.75	0.88	227
China	23,524	27,869	18%	0.38	0.98	170
Germany	2,339,192	2,749,899	18%	0.48	1.22	154
Australia	7,612,922	8,343,652	10%	0.27	0.67	139
Italy	448,994	491,709	10%	0.62	1.30	208
Mexico	5,161,340	5,495,632	6%	0.47	0.69	108
USA	6,457,274	6,763,487	5%	0.44	1.74	233
Spain	1,745,456	1,816,522	4%	0.43	0.92	166
Great Britain	3,316,366	3,383,393	2 %	0.49	0.80	108

The table shows that all selected NTOs add positive polarity contributions on average. Of the two hundred contributions analyzed, only seven had negative polarity. Specifically, one weak negative contribution for Spain with a polarity of -0.1, two negative contributions for China with a polarity of -0.2, one negative contribution for Mexico with a polarity of -0.4, one negative contribution for Turkey with a polarity of -0.1, and two negative contributions for Australia with polarity -0.3 and -0.1.

Average of sentiment analysis score is 0.49 and average of magnitude is 1.03 across all analyzed 200 posts and all NTOs.

France had the highest fan growth over the last two years, with an increase of 30% and a sentiment value of 0.58, which is above average. In the second place in the growth of the number of fans is Turkey with an increase of 24% over the last two years and an average sentiment value of 0.75, which is also above average.

Other findings when examining posts include that each NTO has its own style of posting and is more or less adhering to it. For instance, France will never forget to put a positive word in its

contributions, and because of this, the value of sentiment polarity is one of the highest. Social contributions in the case of France are usually shorter in two lines. In contrast, the US has contributions longer, often around four lines. Spain posts are in two languages and contain a higher number of hashtags. China and Australia often make use of social Instagram posts from other users. Italy is very active on social media and adds several posts per day. On the other hand, Turkey is not so active and will only add a few posts a month. Great Britain almost always puts emoticons in the post.

Furthermore, the Pearson correlation coefficient for these variables was found, including the increase in the number of fans over the last two years, sentiment analysis score, magnitude, and the number of characters in the text, including spaces.

Table 2. Correlation coefficients for increase in fans over the last two years, sentiment analysis score, magnitude, and the number of characters in the text.

Correlations				
	Increase	Score	Magnitude	Characters
Increase	1	.451**	.025	-.006
Score		1	.166	.333
Magnitude			1	.593
Characters				1

** . Correlation is significant at the 0.01 level (2-tailed).

The table shows the correlation between the increase in a fan base and the sentiment polarity value of 0.451. This correlation is referred to as weak to moderate. Sentiment polarity of posts can therefore be one of the key variables for the growth in the number of social media fans, as this is the first thing that should interest users. It can also be seen from the correlations that magnitude and the number of characters in such short texts do not affect the growth of a fan base.

4. Discussion

Social media are, as confirmed by the findings of a number of authors (e.g. Wang et al. 2016; Dhaoui, Webster, and Tan 2017), intensively used not only in the NTO's marketing but in majority of industries. The dominant social medium in the NTOs' marketing of the ten most visited countries is Facebook (Hruška and Pásková 2018).

This paper focused on the analysis of sentiment on social media posts and linking the polarity of sentiment to the growth of the fan base. According to the correlation, it was found that the polarity of sentiment has moderate influence on the growth in the number of subscribers. Of course, a number of other factors, such as the quality of posts, the quality of shared images and videos, could influence subscriber growth. The political situation, nationality, the size of the country, marketing and investment in tourism are also important (Ahmed, Tazi, and Hossny, 2015; Drus and Khalid 2019). Nonetheless, this research highlights the importance of the polarity of sentiment in contributions. The main social network where this research was conducted is Facebook, but NTOs often share posts from Facebook in the same wording on Twitter and sometimes Instagram (this is consistent with the results published by Hruška and Pásková 2018). It is therefore possible that the same conclusions will apply to other social media.

All countries add mostly positive contributions and out of 200 tested contributions only seven had negative sentiment. Another finding was a moderate correlation between the sentiment value in the post and the increase in the number of subscribers over the last two years. These findings suggest that, although the number of fans is certainly influenced by many variables, the sentiment of posts could play a very important role in the growth of social network accounts.

The method of analysis performed in this work is not suitable for detecting sentiment polarity from videos, but this is likely to be a development in the future. Social media users usually do not want to read anything, so there could be another focus on analyzing video sentiment, on YouTube, for example.

The limits of this work are that it targets only one factor (text added to social media post and its sentiment) and does not take into account other factors such as video quality or shared image quality. Moreover, the analysis was carried out on a relatively small sample of two hundred contributions. The further research could consist of analyzing a larger amount of posts, including the text inside the post, after clicking, and thus analyzing sentiment on longer texts in areas other than tourism. It would be advisable to use a grounded theory approach (Lai, 2015) for further research methodology with consistent definition of research objectives, type of social media, sample size and results interpretation.

5. Conclusions

This work analyzed the sentiment of social media posts on Facebook for countries (NTOs) with the most international tourists. The result of this work is finding that there is a moderate correlation between sentiment and growth in the number of fans of NTOs social media.

Concerning the most successful NTOs as for the increase in subscribers over the last two years on Facebook, France has the highest subscriber growth (30%) in the last two years, a sentiment score of 0.58, and its average number of characters in the post description is 105. In comparison, the average sentiment of all selected countries and posts is 0.49. Turkey came second with 24% subscriber growth, an even higher sentiment value of 0.75, and an average post length of 227 characters. China ranked third in the number of subscriber growth, with a lower sentiment value of 0.38, and an average post length of 170 characters, but also a smaller subscriber growth (18%). The research also found that France has added most of the posts that contain an article among all the selected countries. The other selected NTOs usually only add a combination of video and pictures. Another finding was that the last four countries in subscriber increases have less than the average sentiment value. Observation of accounts of selected NTOs has also revealed that they share user generated content (typically an Instagram photo or a YouTube fan video) on its social media, especially on Facebook. Selected NTOs usually add their own text to such posts, and so this factor should not have a big impact on the analysis in this paper.

Acknowledgments: This work was supported by the FIM UHK under Grant of Specific Research Project "Information and knowledge management and cognitive science in tourism".

References

- Ahmed Khaled, Tazi Neamat E., and Hossny Ahmad H. 2015. Sentiment Analysis over Social Networks: An Overview. *2015 IEEE International Conference on Systems, Man, and Cybernetics*: 2174–2179. <https://doi.org/10.1109/SMC.2015.380>.
- Alaei Ali R., Becken Susanne, and Stantic Bela. 2019. Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research*: 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>.
- Dhaoui Chedia, Webster Cynthia M., and Tan Lay P. 2017. Social media sentiment analysis: Lexicon versus machine learning. *Journal of Consumer Marketing*: 34(6), 480–488. <https://doi.org/10.1108/JCM-03-2017-2141>.
- Dridi Amna and Reforgiato Recupero D. 2019. Leveraging semantics for sentiment polarity detection in social media. *International Journal of Machine Learning and Cybernetics*: 10(8), 2045–2055. <https://doi.org/10.1007/s13042-017-0727-z>.
- Drus Zulfadzli and Khalid Haliyana. 2019. Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*: 161, 707–714. <https://doi.org/10.1016/j.procs.2019.11.174>.
- Gaspar Rui, Pedro Cláudia, Panagiotopoulos Panos, and Seibt Beate. 2016. Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*: 56, 179–191. <https://doi.org/10.1016/j.chb.2015.11.040>.
- Hedvičáková Martina and Král Martin. 2019. Benefits of KPIs for Industry Sector Evaluation: The Case Study from the Czech Republic. *E&M Ekonomie a Management*: 22(2), 97–113. <https://doi.org/10.15240/tul/001/2019-2-007>.
- Hruška Jan and Pásková Martina. 2018. How do the National Tourism Organizations use the social media? *E&M Ekonomie a Management*: 21(4), 226–240. <https://doi.org/10.15240/tul/001/2018-4-015>.
- Kauffmann Erick, Peral Jesús, Gil David, Ferrández Antonio, Sellers Ricardo, and Mora Higinio. 2019. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*: in print. <https://doi.org/10.1016/j.indmarman.2019.08.003>.

- Kirilenko Andrei P., Stepchenkova Svetlana O., Kim Hany, and Li Xiang (Robert). 2018. Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*: 57(8), 1012–1025. <https://doi.org/10.1177/0047287517729757>.
- Lai Linda S. L. 2015. Content analysis of social media: a grounded theory approach. *Social Media Content Analysis*: 16(2), 16.
- Liu Yi, Huang Kaixuan, Bao Jigang, and Chen Kaiqi. 2019. Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations. *Tourism Management*: 71, 337–347. <https://doi.org/10.1016/j.tourman.2018.10.004>.
- Lyu Kigon and Kim Hyeoncheol. 2016. Sentiment Analysis Using Word Polarity of Social Media. *Wireless Personal Communications*: 89(3), 941–958. <https://doi.org/10.1007/s11277-016-3346-1>.
- Martin-Domingo Luis, Martín Juan Carlos, and Mandsberg Glen. 2019. Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*: 78, 106–115. <https://doi.org/10.1016/j.jairtraman.2019.01.004>.
- Nakov Preslav, Rosenthal Sara, Kiritchenko Svetlana, Mohammad Saif M., Kozareva Zornitsa, Ritter Alan, Stoyanov Veselin, and Zhu Xiaodan. 2016. Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources and Evaluation*: 50(1), 35–65. <https://doi.org/10.1007/s10579-015-9328-1>.
- Pathak Xema and Pathak-Shelat Manisha. 2017. Sentiment analysis of virtual brand communities for effective tribal marketing. *Journal of Research in Interactive Marketing*: 11(1), 16–38. <https://doi.org/10.1108/JRIM-09-2015-0069>.
- Pochobradská Lucie, and Marešová Petra. 2018. Supporting the economic pillar of sustainable development through effective offline networking of companies. In *5th SGEM International Multidisciplinary Scientific Conferences on SOCIAL SCIENCES and ARTS SGEM2018*. Vol. 18, pp. 983–988. <https://doi.org/10.5593/sgemsocial2018/1.5/S05.122>.
- Rambocas Meena, and Pacheco Barney G. 2018. Online sentiment analysis in marketing research: A review. *Journal of Research in Interactive Marketing*: 12(2), 146–163. <https://doi.org/10.1108/JRIM-05-2017-0030>.
- Shayaa Shahid, Al-Garadi Mohammed Ali, Piprani Arsalan Z., Ashraf Muhammad, and Sulaiman Ainin. 2017. Social Media Sentiment Analysis of Consumer Purchasing Behavior vs Consumer Confidence Index. *Proceedings of the International Conference on Big Data and Internet of Thing - BDIOT2017*. Pp. 32–35. <https://doi.org/10.1145/3175684.3175712>.
- Thelwall Mike. 2019. Sentiment Analysis for Tourism. In *Big Data and Innovation in Tourism, Travel, and Hospitality*, pp. 87–104. https://doi.org/10.1007/978-981-13-6339-9_6.
- Wang Zhaoxia, Chong Chee Seng, Lan Landy, Yang Yinping, Beng Ho Seng, and Tong Joo Chuan. 2016. Fine-grained sentiment analysis of social media with emotion sensing. *2016 Future Technologies Conference (FTC)*: 1361–1364. <https://doi.org/10.1109/FTC.2016.7821783>.
- Xiang Zheng, Du Qianzhou, Ma Yufeng, and Fan Weiguo. 2017. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*: 58, 51–65. <https://doi.org/10.1016/j.tourman.2016.10.001>.
- Yoo SoYeop, Song JeIn, and Jeong OkRan. 2018. Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications*, 105, 102–111. <https://doi.org/10.1016/j.eswa.2018.03.055>.