Sentiment analysis of national tourism organizations on social media

Jan HRUŠKA\_1 1,\* and Josef ZELENKA\_2 2

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

2 University of Hradec Králové, Hradec Králové, Czech Republic; josef.zelenka@uhk.cz

**\*** Correspondence: 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, products or services. The rapid rise in amount of data and constantly generated content requires the need to automate both data acquisition and processing to identify important information and knowledge. This data is often unstructured and requires special procedures to process it. Sentiment analysis provides the opportunity to detect opinion, feeling and sentiment from texts on social media. To analyze sentiment is used 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). It was found that all selected NTOs (National Tourism Organizations) add mostly positive posts and in the sample of two hundred contributions there are only seven of them with negative polarity of sentiment. There was a moderate correlation between customer growth and positive polarity in the contribution. It shows that creating stable positive descriptions for posts can be one of the key variables for the growth of the fan base.

**Keywords:** sentiment analysis; national tourism organizations; social media

JEL Classification: O35; M31

1. Introduction

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

The immense popularity of big data of social media opens the door for the consumer to express their opinions and views on a wide range of topics like state of economy, disappointment over certain products or services, or express their happiness of making purchases (Shayaa, Al-Garadi, Piprani, Ashraf, & Sulaiman, 2017). Social media contents spread swiftly while users communicate and share them freely within the site (Yoo, Song, & Jeong, 2018).

The Internet has democratized content creation enabling a number of new technologies, media and tools of communication, and ultimately leading to the rise of social media and an explosion in the availability of short informal text messages that are publicly available (Nakov et al., 2016).

With the increase use of microblogs and social media platforms as forms of on-line communication, we now have a huge amount of opinionated data reflecting people’s opinions and attitudes in form of reviews, forum discussions, blogs and tweets (Lyu & Kim, 2016). This has recently brought great interest to sentiment analysis and opinion mining field that analyzes people’s feelings and attitudes from written language. Most of the existing approaches on sentiment analysis rely mainly on the presence of affect words that explicitly reflect sentiment (Dridi & Reforgiato Recupero, 2019).

Sentiment analysis is used to gather opinions from large amounts of data. This data can be either text, video or audio. Most often it is text analysis. This analysis is often used for different areas of marketing, customer needs and customer service. The mood or relationship of the customer or users in the examined articles is mainly analyzed by texts such as contributions, ratings and comments.

Social media applications host a large volume of opinions that reflect people's reaction to events. Even as brief as Twitter's 140 characters, social media reactions function as user-driven data that can be for example automatically classified in terms of their sentiment using opinion mining or machine learning techniques (Gaspar, Pedro, Panagiotopoulos, & Seibt, 2016).

Most of the data that exist in social networks is unstructured (Gaspar et al., 2016). Such unstructured data is approximately 80% of the data all over the world. This makes it difficult to analyze and gain valuable insights from such data. Sentiment analysis or opinion mining are two important techniques, which help in detecting emotions and opinions on social media data. This can help in solving many problems and provide many indicators in election, public opinion, and advertisement, health care and public satisfaction (Ahmed, Tazi, & Hossny, 2015).

Spam or fake sentiment detection in reviews or posts is an important application of the the sentiment analysis. Sentiment analysis also can be used to define trust over social network for a brand or a service to build recommendation systems, which recommend a service, a place or a product for user (Ahmed et al., 2015).

Because of the unstructured nature of Big Data, traditional sampling and statistical analysis methods are not always suitable, and Big Data analytics unifies a diverse set of methods specifically targeted at finding patterns in the large volumes of data. These methods include predictive analytics, data mining, artificial intelligence, natural language processing (Kirilenko et al., 2018).

Social media, such as Twitter, Facebook and Chinese Weibo, is overwhelmingly the go-to platform for internet users to share their comments or experiences towards certain products, services or policies. It is a gold mine for those who appreciate the value of understanding public sentiment (Wang et al., 2016). The objective is to classify positive and negative consumer generated content, typically text-based, according to some manual or automated classification methods (Dhaoui, Webster, & Tan, 2017). This led to the emergence of sentiment analysis over social media as an important research field nowadays, facing new challenges due to the typical nature and structure of microblogging data, considering the shortness and the noisiness of messages differently from the conventional long text (e.g., newspapers) (Dridi & 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 & Pacheco, 2018). Over the past few years, sentiment analysis or online feedback has 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 (Visvizi & Lytras, 2019). 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 (Ghosh et al., 2018). 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 organisations (Thelwall, 2019).

2. Methodology

The purpose of this research was to find out which sentiment use NTOs on social media in their posts and compare this sentiment with followers in past two years. NTOs was selected from the countries with the highest visitor numbers of international tourists. According to UNWTO (2019), the top ten most visited countries in the world in 2016 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 in English, and for this reason, another country has been added, namely Australia.

To analyze sentiment was used 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).

Algorithm for sentiment analysis is written in language Python in software PyCharm. Sentiment Analysis inspects the given text and identifies the prevailing emotional opinion within the text, especially to determine a writer's attitude as positive, negative, or neutral. The algorithm output is *score* and *magnitude*.

The score of a document's sentiment indicates the overall emotion of a document. The magnitude of a document's sentiment indicates how much emotional content is present within the document, and this value is often proportional to the length of the document. It is important to note that the Natural Language API indicates differences between positive and negative emotion in a document, but does not identify specific positive and negative emotions. For example, "angry" and "sad" are both considered negative emotions. However, when the Natural Language API analyzes text that is considered "angry", or text that is considered "sad", the response only indicates that the sentiment in the text is negative, not "sad" or "angry".

A document with a neutral score (around 0.0) may indicate a low-emotion document, or may indicate mixed emotions, with both high positive and negative values which cancel each out. Then authors used magnitude values to disambiguate 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 0.8 and magnitude 3.0 (Clearly Positive)
* Score -0.6 and magnitude 4.0 (Clearly Negative)
* Score 0.1 and magnitude 0.0 (Neutral)
* Score 0.0 and magnitude 4.0 (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 tests.

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

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

3. Results

To determine the sentiment used in selected national tourism organizations, a total of 200 posts were analyzed and the algorithm was started a total of 200 times. Launches were for the text provided by the NTOs social media account manager on text attached to post on social media. Posts were randomly selected from 2019 and are mostly image or video 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 tho.”

Output: Sentiment: 0.0, magnitude 0.0

The table below shows the selected countries with the largest number of international tourists and their fan numbers in December 2017 (column titled "Subs. in 2017"), followed by the number of their fans in December 2019 (column titled "Subs. in 2019). The column labeled "Increase" shows the increase in 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 length 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, | 130% | 0,58 | 1,12 | 105 |
| Turkey | 5 299 | 6 562 | 124% | 0,75 | 0,88 | 227 |
| China | 23 524 | 27 869 | 118% | 0,38 | 0,98 | 170 |
| Germany | 2 339 192 | 2 749 899 | 118% | 0,48 | 1,22 | 154 |
| Australia | 7 612 922 | 8 343 652 | 110% | 0,27 | 0,67 | 139 |
| Italy | 448 994 | 491 709 | 110% | 0,62 | 1,30 | 208 |
| Mexico | 5 161 340 | 5 495 632 | 106% | 0,47 | 0,69 | 108 |
| USA | 6 457 274 | 6 763 487 | 105% | 0,44 | 1,74 | 233 |
| Spain | 1 745 456 | 1 816 522 | 104% | 0,43 | 0,92 | 166 |
| Great Britain | 3 316 366 | 3 383 393 | 102 % | 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 has had the highest fan growth in the last two years, with an increase of 130% and a sentiment value of 0.58, which is above average. In second place in the growth of fans is Turkey with an increase of 124% in 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. So 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 adds a greater number of hastags. 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 in 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 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 | ,451 | 1 | ,166 | ,333 |
| Magnitude | ,025 | ,166 | 1 | ,593 |
| Characters | -,006 | ,333 | ,593 | 1 |

The table shows the correlation between fan increase 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 of social media fans, as this is the first thing that should interest users.

4. Discussion

Social media are, as confirmed by the findings of a number of authors (Wang et al., 2016), intensively used not only in the NTO’s marketing but used in majority of industries. The dominant social media in the NTOs’ marketing of the ten most visited countries is Facebook (Hruška & 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. Two hundred posts were analyzed. According to the correlation, it was found that the polarity of sentiment has moderate influence on growth of customers. Of course, a number of other factors, such as the quality of posts, the quality of shared images and videos, could influence subscriber growth. Also, the political situation, the size of the country, marketing and investment in tourism, but even so, 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 in Hruška & Pásková, 2018). It is therefore possible that the same conclusions will apply to other social media.

The limits of this work are that it is very specifically targeted to one narrow area and does not take into account other factors such as video quality or shared image quality.

5. Conclusions

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 are than tourism. Result then compare to the study made in this paper. It would be advisable to use a grounded theory approach (Lai, To, 2015) for further research methodology with consistent definition of research objectives, type of social media, sample size and results interpretation.

The results of this work show that France with the highest subscriber growth (130%) in the last two years has a sentiment score of 0.58 and an average length of characters in the post description of 105. The average sentiment of all selected countries and posts is 0.49. Turkey came second with 124% subscriber growth and 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 (118%). Research also found that France, as the most successful in the increase in subscribers over the last two years, 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. It was also found that the last four countries in subscriber increases have less than the average sentiment value. All countries add mostly positive contributions and out of 200 tested contributions only seven of them had negative sentiment. There was found also a moderate correlation between the sentiment value in the post and the increase in subscribers over the last two years. From these findings, it can be concluded 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.

**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, K., Tazi, N. E., & Hossny, A. 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

Dhaoui, C., Webster, C. M., & Tan, L. 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, A., & 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

Gaspar, R., Pedro, C., Panagiotopoulos, P., & Seibt, B. (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

Hruška, J., & Pásková, M. (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

Lyu, K., & Kim, H. (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

Nakov, P., Rosenthal, S., Kiritchenko, S., Mohammad, S. M., Kozareva, Z., Ritter, A., … Zhu, X. (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

Shayaa, S., Al-Garadi, M. A., Piprani, A. Z., Ashraf, M., & Sulaiman, A. (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*, 32–35. https://doi.org/10.1145/3175684.3175712

Wang, Z., Chong, C. S., Lan, L., Yang, Y., Beng Ho, S., & Tong, J. C. (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

Yoo, S., Song, J., & Jeong, O. (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

|  |  |
| --- | --- |
| copyRight | © 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). |

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. Journal of Travel Research, 58(2), 175–191. <https://doi.org/10.1177/0047287517747753>

Ghosh, A. K., Kotekal, P., & Chakraborty, G. (2018). Attribute-based sentiment analysis in marketing: Application and strategic implications. Model Assisted Statistics and Applications, 13(4), 311–318. <https://doi.org/10.3233/MAS-180441>

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (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, L. S. L., To, W. M. (2015). Content analysis of social media: A grounded theory approach. Journal of Electronic Commerce Research, 16(2), 138-152.

Meena Rambocas, Barney G. Pacheco, (2018) "Online sentiment analysis in marketing research: areview", Journal of Research in Interactive Marketing, <https://doi.org/10.1108/JRIM-05-2017-0030>

Thelwall, M. (2019). Sentiment Analysis for Tourism. In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), Big Data and Innovation in Tourism, Travel, and Hospitality (pp. 87–104). Springer Singapore. https://doi.org/10.1007/978-981-13-6339-9\_6

Visvizi, A., & Lytras, M. D. (Eds.). (2019). Research & Innovation Forum 2019: Technology, Innovation, Education, and their Social Impact. Springer International Publishing. <https://doi.org/10.1007/978-3-030-30809-4>

Xema Pathak, Manisha Pathak-Shelat, (2017) "Sentiment analysis of virtual brand communities for effective tribal marketing", Journal of Research in Interactive Marketing, Vol. 11 Issue: 1, pp.16-38, doi: 10.1108/JRIM-09-2015-0069

Xiang, Z., Du, Q., Ma, Y., Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism, Tourism Management, 58, 51-65, doi: j.tourman.2016.10.001.