A Proposed Method for Analysing a Facebook User's Profile Based on Text Sentiment Analysis

Tayf Tariq Abdul Sahib

Muhanad Abdul Elah Al-khalisy

University Information Technology and Communication

Abstract

To increase the efficiency of sales and reduce the cost of advertising products, many companies are interested in identifying and analyzing customer preferences. With the increase in number of users on each day on a social media platform that generates a huge amount of data; today data analysis plays a big role. We focus on Facebook's mining role in extracting useful information that provides user data such as user interest, account name, some behaviors, etc. In this paper, the method of forming the profile of users of the social network Facebook based on the analysis of the sensitivity of the text and user reactions is described. The method is based on forming a list of keywords and determining their emotional assessment: positive or negative. A lexico-grammatical analysis based StanfordPOSTagger and a sensitivity analysis based AylienText library were used. In addition, the user's activity time in the context of the day of the week and time is determined. A software solution has been developed that implements the method. Studies were conducted on a sample of 300 users for whom the corresponding profiles were formed.

Keywords: Social media, Facebook profile, sentiment analysis, emotional, user activity, customer preferences.

Introduction

Many companies are very interested in identifying and analyzing users' preferences in the Internet as these data are used to reduce the cost of advertising products and increase the sales efficiency of those companies. Usually, such information can be collected in various ways, including with the use of methods of explicit collection in the use of which, a direct interaction with users. Often, such approaches are ineffective for a number of reasons: in particular, due to the different attitudes of people to the scales of assessment. Consequently, it is necessary to look for new approaches to collection data on user preferences. Currently, social networks such as Facebook, Twitter and Google+ have become the main platforms for self-expression. Facebook the most popular social network, which has (1.6) billion users (Barnard, Bothma, and Cant 2017). For example, structured data from Facebook can be used to classify users by areas of interest. Unlike Twitter, in which the post is limited to 140 characters, Facebook supports the creation of messages to 63,206 characters (Sandim et al. 2017). The classification is based on the concept of a user profile. The profile contains information about the user: a unique identifier, a name, interests, some behaviors, etc. The most important information from the point of view of companies is the interests of users and their behavior properties, which can be expressed in terms of keywords, information about the user's activity time. Understanding this information in terms of marketing makes it possible to create targeted advertising more effective. This work is devoted to the problem of forming Facebook user profiles based on analysis of the text of posts, comments, "likes" of the user.

Related Works

In the course of the work existing studies of the sentimental analysis of text messages based on user data on social networks Twitter, Facebook; selection of keywords of users using the tool "Linguistic Inquiry and Word Count" and patterns of behavior of users of social networks descried bellow:

(Awrahman and Alatas 1843) Emotion analysis and mining research have been well illustrated. Many aspects of emotion analysis have been explained and various steps have been introduced in the process of emotion analysis, techniques for analyzing different emotions have been extensively discussed with the introduction of research challenges. (Chen et al. 2017) A unified method was used to construct the user profile of the social network depending on the social media language in Facebook status updates where emotion analysis was applied to create a range of user effects plus training random forest model to configure SWL using user effects and features language to update status. (Settanni and Marengo 2015) The study examined the relationship between the textual content of the social network user and the emotional expression of the feelings. The study focused on feelings of depression, anxiety and stress, which was the result of the collection and analysis of 201 users of Facebook where the extraction of textual stimuli and expression symbols associated with emotion through automated text analysis. The results showed that the expression of anxiety was high in the elderly. (Liu and Jansen 2017) This study is based on the evaluation of the practical performance of the services of questioning the social context in the social networks by analyzing the behavior of the exchange of knowledge among users in the process of questions and answers in terms of their contributions and their concerns and interests where a large number of questions and received answers, which revealed the roles of participants. Where the analysis of 3006 Q & A results showed that the prediction of the user's feelings depending on that method is accurate by 70%. (Markovikj et al. 2013) was searched for the feasibility of modeling the personality of the user of social networks based on the characteristics extracted from the Facebook user data based on a pre-proposed set of features where the results were positive and encouraging. Numerous classification techniques were used. (Shen et al. 2017) Aims to detect depression by analyzing the amount of data on Twitter social media. This method is used for depression and non-depressive speech groups. Six groups related to depression with clinical criteria were used, as well as other groups related to depression behaviors used in social media. The Multiple Depression Dictionary model was proposed to determine the signs of depression for Twitter users. Data were analyzed to reveal the basic behaviors that distinguish between a depressed person or not. (Fournier-viger and Nkambou 2015) An algorithm was suggested to deduce the profile of the Facebook user accurately. The algorithm allows the user to control and trade between the amount of information accessed through the social graph and the accuracy of predictions. The algorithm was applied to more than 11.247 Facebook user profiles. The study showed that the profile of the user was more accurately determined by using a small part of the social graph data compared to the algorithm "state-of-Arts". (Fernandez, Levinson, and Rodebaugh 2012) Objective criteria, such as user concerns, were tested for a profile of Facebook users to assess whether they discriminated between higher and lower social anxiety.

After a comparison between the method proposed in this paper and the methods used previously found that the previous methods did not address the use of such types of procedures and tools in the same application where the methods used in this paper are characterized by speed and high accuracy of analysis.

Proposed Methods

This work proposed method of forming Facebook's user profiles by analysis of text messages on user data (posts, comments, like). The analysis is base of sentimental analysis and the selection of keywords of users using the tool "Linguistic Inquiry and Word Count" In addition to extraction a user's patterns of behavior on those social networks. To implement the proposed approaches and integrate them, a proposed software solution was used. Fig. 1 shows the architecture of the proposed software solution.

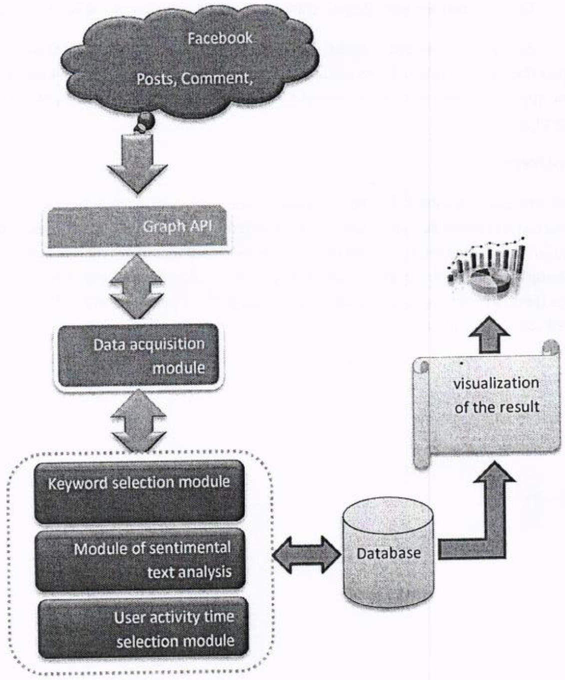

Fig. 1 the architecture of the proposed software solution for the formation of the Facebook user profile

Data collection

To automate the collection of data, a script was created in php language. Collection information on posts, likes and comments is carried out through a request with using the Facebook APIs through php script by using (Graph API SDK) (Mahmood et al. 2016). The login to Facebook app done with "Access tokens", this process conform to the (OAuth 2.0) protocol used in Facebook app. OAuth 2.0 allows entities such as a User, app or a Page to authorize selected tokens. When authorized, apps can use those tokens to access specific information like user profile, posts, and comments. The results come in a file in JSON format (Krishnan 2017). Fig. 2 (a) shows the get access token and authorizes (login) process scripts code used in the proposed software solution while Fig. 2 (b) shows the script code for get user profile information process.

```
$fb = new Facebook\Facebook([
  'app_id' => '123458',
  'app_secret' => '454544',
  'default_graph_version' => 'v2.2',]);

$helper = $fb->getRedirectLoginHelper();
$accessToken = $helper->getAccessToken();

if (isset($accessToken)) {
echo '<h3>Access Token</h3>';
var_dump($accessToken->getValue());}
$oAuth2Client = $fb->getOAuth2Client();
$tokenMetadata = $oAuth2Client->debugToken($accessToken);
echo '<h3>Metadata</h3>';
$_SESSION['fb_access_token'] = (string) $accessToken;
```

Fig. 2 (a) script code for get access token and authorizes process:

```

$response = $fb->get('/me?fields=id,name', '$accesstoken');
$user = $response->getGraphUser();
echo 'Name: ' . $user['name'];

```

Fig. 2 (b) script code for get user profile information process:

Selecting keywords

"From the text of the post (posts/comments) selected keywords using lexico-grammatical analysis. AT This work used the (StanfordPOSTagger) library to parse text messages. For each word, a part of the speech was defined and a corresponding POST tag was marked: a verb, a noun, an adjective, and so on. StanfordPOSTagger allows you to operate with 36 tags (Ritter, Clark, and Etzioni 2011), but the work used 12 tags to select keywords. Selected tags are presented in Table 1."

Table 1.Tags used in the work

Tag group	Tags (part of speech)
Noun	NNS (Common Nouns Plural), NNPS (Proper Nouns Plural), NN (Common Nouns), NNP (Proper Nouns Singular)
"Verbs"	"VB (Verbs (base form)) ", (Verb, (3rd person singular present)"", "VBG (verb, gerund) ", VBN (verb, past participle) ", VBP (Verb, non 3rd person

	singular present)" ,VBD (Verbs (past tense))", VBZ
Adjective	JJ (adjective)
Symbol	SYM (symbol)

Consider the following example of analyzing a text comment and generating keywords. Let there be a phrase.

Ali pay more concern to SPIDERMAN than to POKEMON

After lexico-grammatical analysis of this text, the following result is obtained.

Ali/NN pay/VBP more/JJR concern/NN to/TO SPIDERMAN /NNP than/IN to/TO POKEMON /NNP

Note that the result is the keywords from this sentence, such as: {concern // NN, SPIDERMAN // NNP, POKEMON // NNP}

Algorithm for determining sentimentality posts and user comments

Data received as a result of a query include a lot of overhead, so the service information is separated from the desired. Thus, after the selection of basic information for words from texts in English, parts of speech are determined. Let's denote the selected word as wi.

Post consists of the following fields: identifier (ID), user name, text messages, reactions, comments (Van Dam and Van De Velden 2015). The analysis uses all of the listed fields. The reactions are divided into 5 kinds: LIKE (like), LOVE (like), HAHA (ridiculous), SAD (sad), ANGRY (evil). Reactions can be divided into positive and negative reactions. We refer to the positive ones: LIKE, LOVE, HAHA. To negative: SAD, ANGRY. The emotional coloring of the selected words is determined, for which a part of the speech Si (wi , D) is defined, where D is the dictionary.

$$\mathrm{S}_i(\mathrm{w}_i, \mathrm{D}) = \begin{cases} NEGATIVE\,, & If\ the\ word\ w_i\ is\ marked\ as\ negative \\ POSITIVE\,, & If\ the\ word\ w_i\ is\ marked\ as\ positave \end{cases}$$

To determine the emotional coloring of user's text posts or comments (sentimentality), the algorithm shown in Fig. 3 was proposed:

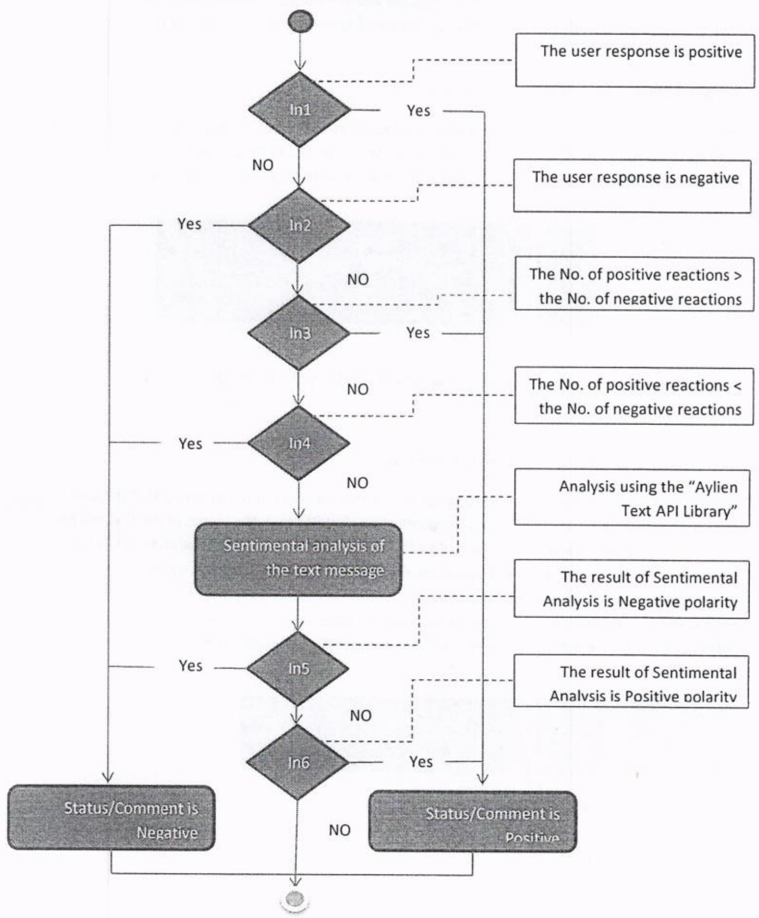

Fig. 3. Algorithm for determining the sensitivity of the post text or user's comments

Next, a list of keywords w is formed for each of the words with the following set of attributes: part of speech, emotional coloring (NEG / POS) and number of keywords and user ID that uses the keyword (w).

Retrieving Time Information for User Activity

Information about the user's activity in time is obtained as a result of a request using the Facebook API. Fig. 4 presents the response to the request in JSON format where "created_time" tag is a timestamp for the Facebook's user e_action; id is the action ID (Mahmood et al. 2016).


```
"'data'": [
{
"Created_time": "2017-12-15T15:11:25+0000"
```


Fig. 4 JSON responses to the request for the time of compiling the message

Analysis of the Sensitivity of Text Using AylienText

According to the algorithm for evaluating the text of the post or comment (Figure 3), if the user presses a positive reaction button and the publication contains text, the words of the text will be assigned to the list of positive keywords. Consider the case if there is a publication text, but the user did not click the response button. To evaluate the text of a message, a sensitivity analysis of the text is required. This analysis is performed using the AylienText library. If the user used an emoji sign within the text the emoji classified based on polarized table assign before. Fig. 5 presents a sample of code for assign polarity value for emoji. Table 2 presents examples of sensitivity analysis of the text.

```
[ { name: '100', emoji: '💯', polarity: 3 },
  { name: 'angry', emoji: '😠', polarity: -3 },
  { name: 'anguished', emoji: '😧', polarity: -3 },
  { name: 'astonished', emoji: '😲', polarity: 2 },
  { name: 'black_heart', emoji: '🖤', polarity: 3 } ]
```

Fig. 5 sample code for assign polarity value for emoji

Table 2 Examples of analysis result text sensitivity

Message	Positive / Negative	Confidence
Today it's very hot ☺	Positive (POS)	0.785
Today it's very hot!	Positive (POS)	0.787
Today it's very hot ☹	Negative (NEG)	0.365
Today it's very hot ☹ ☹	Negative (NEG)	0.955

Thus, the assignment of the text message to positive or negative is determined by the signs and symbols that are used by the user.

Result and Discussion

By using the proposed software solution that illustrated in Fig. 1, information was collected on a number of users. The most popular positive keywords were: happy, love, life, world, good, thank and the most popular negative keywords were: "hate", "cry", "sad". After processing the data on the activity time of users, it was determined the distribution of the use of a particular keyword in Facebook with a discretization by the hour from 1:00 Am to 12.00 Am and for each day of the week. Fig. 6 presents the user activity by days of week.

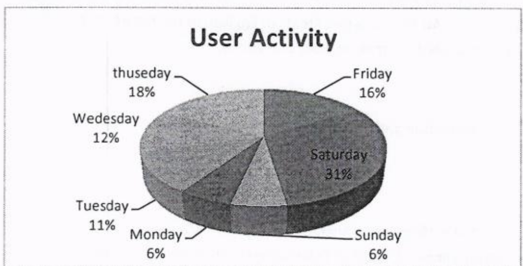

Fig. 6 user activity time in Facebook social network

Next, we searched for users whose keywords partially overlap. After receiving the records, a list of users is formed, the number of their keywords.

In this study, the test of the proposed method was performed on a sample of 300 users (223 men and 77 women). A total of 141148 status updates were analyzed. Table 3 shows the statistics of popular users' keywords by usage shares (the ratio of the number of used keywords to the total number of keywords).

Table 3 Usage statistics popular keywords

Keyword	Percentage of use	Man (%)	Woman (%)
Good	4%	15%	65%
Love	1.5%	47%	53%
Happy	8.5%	56%	64%
Believe	4.4%	50%	55%
Thank	9.18%	47.5%	66.5

Conclusion

The using of different analysis techniques in Facebook data helps to structure data and better understand it. This data will be beneficial for company in sales and advertising products. In this paper, the method of forming the profile of users pf social network Facebook is considered. The method is based on forming a list of keywords and determining their emotional assessment: positive or negative. A lexico-grammatical analysis based StanfordPOSTagger and a sensitivity analysis based AylienText library were used In addition, the user's activity time in the context of the days of the week and time is determined. A proposed software solution has been design for this method. Studies were conducted on a sample of 300 users for whom the corresponding profiles were formed. Currently, the development is performed for the analysis of texts in English. In the future, it is planned to refine the program for the analysis of texts in Arabic language.

ACKNOWLEDGMENT

Special thanks to the University of Information and Communication Technology for their support in accomplishing this research.

References

Awrahman, Banan, and Bilal Alatas. 1843. "Sentiment Analysis and Opinion Mining within Social Networks Using Konstanz Information Miner." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 9(1): 15–22.

Barnard, S. M., C. H. Bothma, and M. C. Cant. 2017. "The Identification of Criteria for the Optimal Use of Facebook Pages for Marketing Purposes: A South African Perspective." Banach Journal of Mathematical Analysis 11(2): 65–78.

Chen, L. et al. 2017. "Building a Profile of Subjective Well-Being for Social Media Users." PLoS ONE 12(11): 1–15.

Van Dam, Jan Willem, and Michel Van De Velden. 2015. "Online Profiling and Clustering of Facebook Users." Decision Support Systems 70: 60–72.

Fernandez, Katya C., Cheri A. Levinson, and Thomas L. Rodebaugh. 2012. "Profiling: Predicting Social Anxiety from Facebook Profiles." Social Psychological and Personality Science 3(6): 706–13.

Fournier-viger, Philippe, and Roger Nkambou. 2015. "Inferring User Profiles in Online Social Networks Using a Partial Social Graph." Advances in Artificial Intelligence 9091: 84–99. http://link.springer.com/10.1007/978-3-319-18356-5.

Krishnan, Hema. 2017. "Emotion Detection of Tweets Using Naïve Bayes Classifier." 4(11): 457–62.

Liu, Zhe, and Bernard J. Jansen. 2017. "Identifying and Predicting the Desire to Help in Social Question and Answering." Information Processing and Management 53(2): 490–504. http://dx.doi.org/10.1016/j.ipm.2016.05.001.

Mahmood, Rana Arsal, Hassan Ali Khan, Amna Yasin, and Ammar Rafique. 2016. "Twibook – A Systematic Approach towards Twitter & Facebook." NFC-IEFR Journal of Engineering and Scientific Research 4(1): 45–48. http://nijesr.iefr.edu.pk/topics_detail.php?getPaper=70.

Markovikj, Dejan, Sonja Gievska, Michal Kosinski, and David Stillwell. 2013. "Mining Facebook Data for Predictive Personality Modeling." Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA: 23–26.

Ritter, Alan, Sam Clark, and Oren Etzioni. 2011. "Named Entity Recognition in Tweets: An Experimental Study." ... of the Conference on Empirical Methods ...: 1524–34. http://dl.acm.org/citation.cfm?id=2145595.

Sandim, Miguel, Paula Fortuna, Alvaro Figueira, and Luciana Oliveira. 2017. "Journalistic Relevance Classification in Social Network Messages: An Exploratory Approach." Studies in Computational Intelligence 693: 631–42.

Settanni, Michele, and Davide Marengo. 2015. "Sharing Feelings Online: Studying Emotional Well-Being via Automated Text Analysis of Facebook Posts." Frontiers in Psychology 6(JUL): 1–7.

Shen, Guangyao et al. 2017. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." IJCAI International Joint Conference on Artificial Intelligence: 3838–44.