

The Weighting Factors to Improve Predictability on Twitter

Jorge Arroba Rimassa¹, Rafael Muñoz Guillena², Fernando Llopis²

¹Facultad de Ingeniería Ciencias Físicas y Matemáticas, Universidad Central del Ecuador, Quito, Ecuador

²Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Alicante, España

Email: jarroba@uce.edu.ec, rafael@dlsi.ua.es, llopis@dlsi.ua.es

How to cite this paper: Arroba Rimassa, J., Muñoz Guillena, R. and Llopis, F. (2018) The Weighting Factors to Improve Predictability on Twitter. *Technology and Investment*, 9, 68-79.

<https://doi.org/10.4236/ti.2018.91005>

Received: December 12, 2017

Accepted: February 24, 2018

Published: February 27, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The result of the analysis of a thematic in a social network is to find a measure that allows the principal actors to know their performance, that is, they can define or maintain strategies and courses of action in order to optimize their communication. It is necessary to formally define the principles of analysis in Social Networks in order to use their characteristics better and to be able to contextualize the concept and use of weighting factors to improve their predictability. When Social Networks are going to be used as a mechanism to predict social behavior, for example, to predict the outcome of a political election, weighting factors must be introduced to try to match the data collected from the Social Network with those of a sample. In this article we have defined the methodology to incorporate the geographic weighting factors and several formulas have been created that allow reprocessing the data downloaded from Twitter in which its polarity has been determined by classical NLP methods to increase the predictive power.

Keywords

Twitter, Weighting Factors, Formalization of Social Networks

1. Introduction

Today, all with 280 characters we can say about everything; these opinions are added in a massive way to build a thematic analysis that is a theoretical construct on various aspects.

In this context, opinion leaders, product brands, thematic positions, official or private organizations, entities, etc. are considered as principal actors that interact with the different users who in turn express their opinion and take some position with respect to these actors.

The realization of an analysis of this information that flows in a bidirectional way, between principal and active actors, the users, is what gives more power to the Social Networks; because these are becoming a tool to “influence” the user. Then the analysis of a topic in a social network becomes a measure that allows in many cases to the principal actors to know their performance; if what they express is well positioned, their measure of acceptance; that is, we have the measure to define or maintain strategies and courses of action in order to optimize communication and performance.

All people have access to a social network, think, criticize, warn and they make value judgments on any topic and from social networks. One of those that has more followers and allows the download for further analysis is Twitter to use the messages as a tool to evaluate the performance of the principal actors. A conversion process must be carried out so that the opinion given in the twitters is equivalent to those of the general population, since Twitter users are a subset of these.

Before beginning the analysis of Social Networks, we must begin to contextualize their concept; we must formally define their principles in order to better exploit their characteristics.

We must also formalize the conversion process so that its use is more widespread and that the a-posteriori results are closer to reality and reflect the true feelings of the users.

2. Methods and Procedures for Processing Information

Performing analysis in order to predict certain behaviors using social networks has positions. From the most extreme ones that state that it definitely does not make sense to make any attempt to analyze the information of social networks to those that tend to use and abuse to try to predict the different behaviors of the actors.

Different authors have used different techniques of polarity analysis and even more in electoral processes, which become a thermometer to the extent that their results can be contrasted with those issued by the predictions that are made, besides the simple counting of the mentions. **Table 1** shows a taxonomy of antinomic positions in terms of whether social networks can be used or not to

Table 1. Positions on the use of twitters for predictive analysis.

	If they predict	If they not predict
Do not perform polarity analysis	Tumasjan <i>et al.</i> , 2010	
	Fernández Crespo, 2013	
	Zarella, 2010	
	http://tweetminster.co.uk , 2010	Gayo-Avello, Metaxas, and Mustafaraj, 2011
Do perform polarity analysis	Montesinos, 2013	
	Ramadhan, Nurhadryani, and Hermadi, 2014	
	Tsakalidis <i>et al.</i> 2014	

predict results and the use or not of methods to determine the polarity of messages; in which the main authors have been annotated in relation to political prediction.

In the first quadrant, those who state that the analysis of the twitters serves to predict results but without using methods to detect the polarity, that is, working with the total number of mentions that one or another principal actor can have. The works of [1] in the elections of Germany, [2] in the general elections of Spain and [3] in several legislative elections in the United States in the political field are iconic.

In the second quadrant, there are those that use and perform the analysis of the twitter to predict, but incorporate the procedures for determining the polarity of the messages. They then count the mentions of those who have positive polarity or acceptance to a certain principal actor. The analyzes published by tweet minster for elections in the United Kingdom, [4] in primary elections in Chile, [5] in the elections in Jakarta and [6] in various elections in Europe also in the political field, they are referents as much by the results as by the diverse in the geographic regions that apply them.

In the third quadrant, however, we mention [7] as an advocate of those who state that the twitter analysis cannot predict any result by the erratic results and values of the average absolute error, MAE, which are obtained high.

3. Contextualization of the Analysis of Social Networks

We can consider that a message is a dimensional $p + 1$ vector:

$$t\$ = (v_1, v_2, \dots, v_p, x)$$

where I saw them v_i are variables that identify the user and also determine the context in which the message is made, x is the content that can be: text, photo, image, video, audio, etc.

Clearly an x can contain a text plus an image or the meeting of all them.

If we define the set:

$$RS = \{t\$ \mid \text{it is a message}\}$$

how the whole Social Network.

We affirm that a Social Network contains all the messages transmitted to the present.

The set A of all the users who send, receive or visualize messages in RS is called the set of Actors. The actors are those entities that emit any message, either to publicize a particular event, an opinion or simply wish to retransmit the content x of a message $t\$$; there are also those who only observe in the social network. We must distinguish some types of actors: those we want to analyze will be the principal actors, here are the candidates, the brands, etc.; those who send and receive messages will be the active actors, in this group are all the users or the general public and those who only observe the activity in the social network, without participating will be the passive actors; since they do not interact

with anyone, they cannot be followed or analyzed, so we will exclude them from the analysis. The active actors of the principal actor $a \in A$ will be all the u that “interact with” a .

We can define in the Social Network, the relationship “interact with” as:

$$R: A \times A \rightarrow \{0,1\}$$

$$(u, a) \mapsto R(u, a)$$

where:

$$R(u, a) = \begin{cases} 1 & \text{if } u \text{ follow to } a \\ 0 & \text{otherwise} \end{cases}$$

The set of all the active actors of the principal actor a will be noted by $U_a = \{u_{1,a}, u_{2,a}, \dots, u_{q,a}\}$.

The set of all the messages of the set of actors A in the temporality $T = [t_i, t_f]$ where t_i and t_f represent the time or the start and end date of reception of the messages; it will be noted by: $RS_{A,T}$ and we will call it the Thematic of actors A in temporality T in the social network RS .

The messages of the active actor $u_{j,a} \in U_a$ will be: $t\$_{j,1,a}, t\$_{j,2,a}, \dots, t\$_{j,k,a}, \dots, t\$_{j,n_j,a}$ since an active actor can send n_j messages to the principal actor a , the matrix representation of the messages for the principal actor a will be:

$$Mt\$_a = \begin{pmatrix} t\$_{1,1,a} \\ t\$_{1,2,a} \\ \vdots \\ t\$_{1,n_1,a} \\ t\$_{2,1,a} \\ t\$_{2,2,a} \\ \vdots \\ t\$_{2,n_2,a} \\ \vdots \\ \vdots \\ t\$_{q,1,a} \\ t\$_{q,2,a} \\ \vdots \\ t\$_{q,n_q,a} \end{pmatrix} = \begin{pmatrix} v_{1,1,a,1} & v_{1,1,a,2} & \dots & v_{1,1,a,p} & x_{1,1,a} \\ v_{1,2,a,1} & v_{1,2,a,2} & \dots & v_{1,2,a,p} & x_{1,2,a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{1,n_1,a,1} & v_{1,n_1,a,2} & \dots & v_{1,n_1,a,p} & x_{1,n_1,a} \\ v_{2,1,a,1} & v_{2,1,a,2} & \dots & v_{2,1,a,p} & x_{2,1,a} \\ v_{2,2,a,1} & v_{2,2,a,2} & \dots & v_{2,2,a,p} & x_{2,2,a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{2,n_2,a,1} & v_{2,n_2,a,2} & \dots & v_{2,n_2,a,p} & x_{2,n_2,a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{q,1,a,1} & v_{q,1,a,2} & \dots & v_{q,1,a,p} & x_{q,1,a} \\ v_{q,2,a,1} & v_{q,2,a,2} & \dots & v_{q,2,a,p} & x_{q,2,a} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{q,n_q,a,1} & v_{q,n_q,a,2} & \dots & v_{q,n_q,a,p} & x_{q,n_q,a} \end{pmatrix}$$

Be the matrix $Mt\$_a$ and $Mt\$_b$:

$$Mt\$_a = \begin{pmatrix} v_{1,1,a,1} & \dots & v_{1,1,a,p} & x_{1,1,a} \\ \vdots & \ddots & \vdots & \vdots \\ v_{q,n_q,a,1} & \dots & v_{q,n_q,a,p} & x_{q,n_q,a} \end{pmatrix},$$

$$Mt\$_b = \begin{pmatrix} v_{1,1,b,1} & \dots & v_{1,1,b,p} & x_{1,1,b} \\ \vdots & \ddots & \vdots & \vdots \\ v_{r,n_r,b,1} & \dots & v_{r,n_r,b,p} & x_{r,n_r,b} \end{pmatrix}$$

The matrix operation \otimes is defined as:

$$Mt\$_a \otimes Mt\$_b = \begin{pmatrix} v_{1,1,a,1} & \cdots & v_{1,1,a,p} & x_{1,1,a} \\ \vdots & \ddots & \vdots & \vdots \\ v_{q,n_q,a,1} & \cdots & v_{q,n_q,a,p} & x_{q,n_q,a} \\ v_{1,1,b,1} & \cdots & v_{1,1,b,p} & x_{1,1,b} \\ \vdots & \ddots & \vdots & \vdots \\ v_{r,n_r,b,1} & \cdots & v_{r,n_r,b,p} & x_{r,n_r,b} \end{pmatrix}$$

And we will have the matrix representation of the Thematic $RS_{A,T}$:

$$Mt\$ = \otimes_{a \in A} Mt\$_a$$

Facebook, Twitter, Whatsapp, Linkedin, etc. they are examples of Social Networks, which contain messages from actors, and a theme is for example “an advertising campaign” launched a month ago; where the principal actors are those who want to “sell” the product and also those who want us not to “buy” it, surely their competitors, the active actors would be all the users who receive the messages from these principal actors that show us the virtues and defects of the product. Temporality will be the period from one month ago to the present.

Without loss of generality, to the extent that one knows and has perfectly defined which are the principal actors to be analyzed, while the active actors are rather unknown and that is sometimes one of the objectives of performing the analysis in a Social Network, is that the set of actors A , will be defined based on the principal actors.

And then what does it mean to perform an analysis of $RS_{A,T}$?

It is to answer a series of questions that can be asked, basically, is to know certain performance measures. In the preceding example it can be: the percentage of users that will accept the “product”.

The present investigation wants to give a contribution so that the measures of performance are more reliable making use only of the $RS_{A,T}$ and we will be restricted in the social network Twitter treating only those textual contents of the messages.

The processes to be followed in the analysis procedures of an $RS_{A,T}$ theme are basically the same: defining the actors and temporality, extracting the data, debugging the data, measuring the polarity of the messages and obtaining the performance measures.

Be the principal actors $A = \{a_1, a_2, \dots, a_r\}$, the determination of the active actors u of each principal actor a , that is to say U_a as mentioned is not fundamental. What we should observe in data cleansing is that some active actor are not “bots”, which today have become the biggest problem that social networks have since they distort the performance measures of any actor; often increasing its presence and temporality $T = [t_p, t_f]$ where t_i and t_f represent the start and end date of reception of the messages. This is how the Thematic $RS_{A,T}$ of a certain determined investigation is defined.

Then we define a mechanism to download $RS_{A,T}$ for twitter there are several

free and commercial applications to obtain the matrices $M_{j,a}$ for all the principal actors to which they were defined. The debugging of the data has to do basically to avoid the “bots” mentioned, being an elementary procedure to eliminate those messages $t_{j,i,a}$ and $t_{j,k,a}$ from an active actor that contain the same content x .

We also eliminate the spaces, external links and only keep the text that represents the topic of twitter.

The next step is to define the polarity of the content x of the messages t . The goal of polarity is to define a measure of acceptance, rejection or neutrality of each content x made by the active actor u on the principal actor a . That is, we have to find a function f such that:

$$p_{u,a} = f(t) = \begin{cases} 1 & \text{if } x \text{ is accepted} \\ 0 & \text{if } x \text{ is neutral} \\ -1 & \text{if } x \text{ is rejected} \end{cases}$$

This problem can be addressed by several methods, the most used:

3.1. Supervised Methods

It is required to “train” through an “agent”, depending on the subject of the analysis, the agent can be a function of discrimination, in other cases the agent must be a person.

The number of elements to train can be calculated by the size of a simple random sampling, SRS, predefining α and e (reliability and error). For small sizes, one third of the population of the x can be trained.

Independent of the algorithm, the values of the Confusion Matrix must be checked, specifically the CM trace will be added and divided by the trained size. The technique will be accepted if this value is greater than the value that the researcher has set as the admissible limit of acceptability; otherwise, you should look for another method to find the function f .

3.2. Unsupervised Methods

Through, for example, the use of dictionaries. For which the x must be prepared. Each content x can be decomposed in a sequence $\langle x \rangle$ in which each element of the sequence is a “debugged” word, which is the one in which special reference characters, stopwords, lemmatizations, etc. have been eliminated.

Thus: if $x = \langle x_1, x_2, \dots, x_\beta \rangle$ and we are using the dictionaries $DP = \{p/p \text{ is a positive word}\}$ and $DN = \{n/n \text{ is a negative word}\}$ using the counting method we would have the following polarity function:

$$p_{u,a} = \text{sgn} \left(\sum_{i=1}^{\beta} (|\{x_i\} \cap DP| - |\{x_i\} \cap DN|) \right)$$

(The symbol $|C|$ represents the cardinal of set C).

For applications where you want to have only two types of polarity, 1 if you are in favor and 0 otherwise we will use the following function that is evaluated by the Equation (1):

$$p_{u,a} = \text{sgn} \left(\text{sgn} \left(\sum_{i=1}^{\beta} (|\{x_i\} \cap DP| - |\{x_i\} \cap DN|) \right) \right) \quad (1)$$

The following procedure, subject of the present investigation, is as follows.

4. The Weighting Factors and Their Application in the Analysis of Social Network

The weighting factors is the mechanism by which a conversion process should be carried out so that the opinion given in the twitters equals those of the general population, given that Twitter users are a subset of these.

As the processes of analysis of a thematic is carried out in a specific geographical location, it will be necessary to determine the administrative divisions of a lower level.

In view of the fact that the only “geographical” data available are those of the geo-coordinates, which is generally not available because most users prefer not to enable it, and the other is the locality from which the messages are made twitter, we will try to identify the location and parameterize it with the province of origin where the message was issued and for those that do not have a location in any of these categories will be defined as others.

Define the set of provinces and their locations as follows:

$$P = \{P_i \mid P_i \text{ it is a province, } i = 1, 2, \dots, N\}$$

$$P_i = \{loc_{i,1}, loc_{i,2}, \dots, loc_{i,m_i}\}, \forall i, i = 1, 2, \dots, N$$

First we must determine, for each observation j the number of mentions (with positive polarity) in total that exist in the province of observation j . The only data we have is that of the loc_j locality and it may be the case that it is not defined or is simply expressed through some colloquialism. In this case this locality will pass to the group of others or those not located. According to [8] getting a real place from these data is a very complex problem.

That is why in the P_i set of localities of the province i , it should be the most extensive, in the sense of putting with the highest level of detail the localities or administrative areas of lower level. The number of mentions of province P_i in the matrix will be determined by Equation (2):

$$N_j = \left(\left(pol_j \sum_{i=1}^N |\{loc_j\} \cap P_i| \left[\sum_{k=1}^{NOBS} |\{loc_k\} \cap P_i| pol_k \right] - 1 \right) \right) |\{loc_j\} \cap \bigcup_{i=1}^N P_i| pol_{j+1} \quad (2)$$

where the variable $NOBS$ is the total number of rows of the matrix Mt , pol_k is a polarity the message k , loc_k is a location of each observation k .

Next we determine the total of localities that belong to any province of the considered ones. This value is calculated using the Equation (3):

$$D = \sum_{k=1}^{NOBS} pol_k \left| \{loc_k\} \cap \bigcup_{i=1}^N P_i \right| \quad (3)$$

Then the sample proportion for observation j will be calculated using the Eq-

uation (4):

$$M_j = \frac{N_j}{D} \tag{4}$$

Similar process must be done to obtain the proportion of the universe, for each observation j , as for each P_i we have the data of the Population (electoral) proportion pP_i , but this percentage is not valid for those other category locations, so a reconversion must be carried out, by the Equation (5):

$$U_j = pol_j \sum_{i=1}^N \left| \{loc_j\} \cap P_i \right| pP_i \tag{5}$$

Then the weighting factor that will be applied to each observation j , both for those that have localities located, and other type localities will be estimated by the Equation (6):

$$fp_j = \left(\left(\frac{U_j}{M_j} - 1 \right) \left| \{loc_j\} \cap \bigcup_{i=1}^N P_i \right| + 1 \right) pol_j \tag{6}$$

In order to obtain the final assessment for each of the actors, we will evaluate their mentions with positive polarity, affecting it with the weighting factor.

Through the following example we will test the indicated methodology.

Let $P = \{P_1, P_2, P_3\}$ be a geographical entity with $N = 3$ provinces, where each of these have the following localities (municipalities, cantons, town, etc.):

$$P_1 = \{town1, town5\}$$

$$P_2 = \{town2, town3, town4\}$$

$$P_3 = \{town6\}$$

And the principal actors are: $A = \{a1, a2, a3, a4\}$.

Table 2 shows the $Mt\$$ matrix representation of $RS_{A,T}$ (we assume that the temporality has been collected on the scheduled dates). As we mentioned, it does not matter to meet the principal actors.

The number of messages to be analyzed is $NOBS = 16$.

Let us also suppose that these messages are already debugged.

For the determination of polarity we will use the dictionary of positive and negative terms, being:

$$DP = \{a, b, c\}$$

$$DN = \{r, s, t\}$$

For purposes of the analysis that is required, we will use Equation (1) so that the polarity takes only two values: 1 if it is positive and 0 otherwise.

In this example, we determine the polarity determination, *ceteris paribus*, because what we want to demonstrate is the use of the weighting factors that should be used to improve predictability. The count of positive words in each content, as well as that of negative words has been evaluated for the contents x of each one of the observations considered in the matrix $Mt\$$, in order to calculate the polarity, which we will notice by pol_j is shown in **Table 3**.

Table 2. Matrix representation of $RS_{A,T}$

Observation	Location	Active Actors	Principal Actors	Content
obs	loc	u	act	x
1	town1	u1	a1	abc def
2	town1	u2	a1	ace g pq
3	town1	u3	a1	aprst
4	town2	u4	a1	cd
5	town3	u5	a1	rst
6	town3	u6	a1	bpq
7	town4	u7	a2	cdc
8	town5	u8	a2	
9	town5	u9	a2	des
10	town6	u10	a3	abet
11	town1	u11	a4	btb
12	town1	u12	a4	ars
13	town3	u13	a4	rsr
14	town7	u14	a4	abcs
15	town7	u15	a4	dptes
16	town7	u16	a4	rrr

Table 3. Polarity evaluation.

obs	Positive count	Negative count	pol_j
1	3	0	1
2	2	0	1
3	1	3	0
4	0	0	0
5	0	3	0
6	1	0	1
7	2	0	1
8	0	0	0
9	0	1	0
10	2	1	1
11	2	1	1
12	1	2	0
13	0	3	0
14	3	1	1
15	0	2	0
16	0	3	0

The next step is to evaluate the weighting factors for each of the messages. The population proportions of each province are shown in **Table 4**, with province P_3 being the one that concentrates the greatest amount of population, this concentrates 40% of the total population of the geographic entity considered.

Using the Equations (2)-(3) to evaluate the terms of N_j and D , with these values we calculate, using the Equation (4) the sample proportion given by M_j . The proportion of the terms of the population universe is evaluated using the Equation (5) and finally we have that the weighting fp_j given by the Equation (6), for each observation are shown in **Table 5**.

Finally, it is necessary to calculate the valuation for each one of the actors, use the Equation (7).

$$V_a = \sum_{k=1}^{NOBS} pol_k \cdot fp_k \left| \{a\} \cap \{act_k\} \right|. \tag{7}$$

In **Table 6**, the second column shows the assessment made for each the actors

Table 4. Population proportion of the provinces.

Provinces	Population proportion pP_i
P_1	0.35
P_2	0.25
P_3	0.40
Total	1.00

Table 5. Determination of the weighting factors.

obs	N_j	D	M_j	U_j	fp_j
1	3	6	0.50	0.35	0.70
2	3	6	0.50	0.35	0.70
3	1	6	0.17	0.35	0.00
4	1	6	0.17	0.25	0.00
5	1	6	0.17	0.25	0.00
6	2	6	0.33	0.25	0.75
7	2	6	0.33	0.25	0.75
8	1	6	0.17	0.35	0.00
9	1	6	0.17	0.35	0.00
10	1	6	0.17	0.40	2.40
11	3	6	0.50	0.35	0.70
12	1	6	0.17	0.35	0.00
13	1	6	0.17	0.25	0.00
14	1	6	0.17	0.00	1.00
15	1	6	0.17	0.00	0.00
16	1	6	0.17	0.00	0.00

Table 6. Comparison of the valuation by three calculation methods.

Principal actor	Polarity and weighting factor	% polarity and weighting factor	Polarity	% polarity	mentions	% mentions
a1	2.15	30.7	3	42.9	6	37.5
a2	0.75	10.7	1	14.3	3	18.8
a3	2.40	34.3	1	14.3	1	6.3
a4	1.70	24.3	2	28.6	6	37.5
total	7	100.0	7	100.0	16	100.0

Table 7. Order of location for each one of the calculation methods.

Method		
Polarity and weighting factor	Polarity	Mentions
a3	a1	a1 - a4
a1	a4	a2
a4	a2 - a3	a3
a2		

using the Equation (7), this is the assessment using the procedure described in this article, in the third column is the percentage evaluation with the aim of normalizing this data; next, we have the assessment, as traditionally done, using only the polarity and without using weighting factors, which would correspond to the procedures described in quadrant two given in the state of the art and six column the valuation calculated only with the mentions of each actor, in which neither the calculation of the polarities or the calculation of the weighting factors, this method corresponds instead to the first quadrant described in the state of the art; for each of these, the percentage value is incorporated with respect to the total, for comparative purposes.

The results are contradictory. The best valued by our method, actor 3, a3 with 34.3% is the worst rated with the other two methods. It must be noted that the messages issued from province 3, P₃, which concentrates the largest number of habitants, must be resized because we have to give this province its true proportion in the population universe of study.

Table 7 shows the order of assessment and the location of each actor according to each of the methods used. In this it is observed that the methods of polarity and the mentions almost give similar results.

5. Conclusions

In the present article the aspects referring to the analysis of the Social Networks have been formalized and the formulas for the incorporation of the weighting factors have been developed in order to increase the efficiency in the assessment of the actors.

Through the comparative analysis, using an idealized data set, the variation of

the results in the final assessment is demonstrated.

This methodology has already been put to the test in the presidential election of February 19, 2017 in the Republic of Ecuador; an MAE = 1.1 was obtained that demonstrated the effectiveness of this incorporation of the weighting factors in the analysis of a Thematic in a Social Network.

Acknowledgements

The authors declare that no conflict of interest exists with the results and conclusions presented in this paper. Publication ethics have been observed.

References

- [1] Tumasjan, A., Sprenger, T., Sandner, P. and Welpe, I. (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington DC, 23-26 May 2010, 178-185.
- [2] Fernández Crespo, M. (2013) Predicción electoral mediante análisis de redes sociales. Ph.D. Dissertation, Universidad Complutense de Madrid, Madrid.
- [3] Zarrela, D. (2009) ReTweets Change Everything. <https://www.asgroupinc.com/new-data-can-twitter-predict-elections/>
- [4] Montesinos García, L. (2014) Análisis de sentimientos y predicción de eventos en twitter. Ing. Dissertation, Universidad de Chile, Santiago.
- [5] Ramadhan, D., Nurhadryani, Y. and Hermadi, I. (2014) Campaign 2.0: Analysis of Social Media Utilization in 2014 Jakarta Legislative Election. *ICACISIS 2014*, Jakarta, 18-19 October 2014, 102-107. <https://doi.org/10.1109/ICACISIS.2014.7065881>
- [6] Tsakalidis, A., Papadopoulos, S., Cristea, A. and Kompatsiaris, Y. (2015) Predicting Elections for Multiple Countries Using Twitter and Polls. *Predictive Analytics. IEEE Intelligent Systems*, **30**, 10-17. <https://doi.org/10.1109/MIS.2015.17>
- [7] Gayo-Avello, D., Metaxas, P. and Mustafaraj, E. (2011) Limits of Electoral Predictions Using Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social*, Barcelona, Catalonia, 17-21 July 2011, 490-493.
- [8] Peregrino, F.S., Tomás, D. and Llopis, F. (2013) Every Move You Make I'll Be Watching You: Geographical Focus Detection on Twitter. *Proceedings of the 7th Workshop on Geographic Information Retrieval ACM*, Orlando, 5 November 2013, 1-8. <https://doi.org/10.1145/2533888.2533928>