Microblog Sentiment Prediction based on User Past Content

Yassin Belhareth[1] and Chiraz Latiri[2]
[1]*LIPAH, ENSI, University of Manouba, Tunis, Tunisia*
[2]*University of Tunis El Manar, Tunis, Tunisia*

Keywords: Opinion Mining, Sentiment Analysis, User Past Content.

Abstract: Analyzing massive, noisy and short microblogs is a very challenging task where traditional sentiment analysis and classification methods are not easily applicable due to inherent characteristics such social media content. Sentiment analysis, also known as opinion mining, is a mechanism for understanding the natural disposition that people possess towards a specific topic. Therefore, it is very important to consider the user context that usually indicates that microblogs posted by the same person tend to have the same sentiment label. One of the main research issue is how to predict twitter sentiment as regards a topic on social media? In this paper, we propose a sentiment mining approach based on sentiment analysis and supervised machine learning principles to the tweets extracted from Twitter. The originality of the suggested approach is that classification does not rely on tweet text to detect polarity, but it depends on users' past text content. Experimental validation is conducted on a tweet corpus taken from data of SemEval 2016. These tweets talk about several topics, and are annotated in advance at the level of sentiment polarity. We have collected the past tweets of each author of the collection tweets. As an initial experiment in the prediction of user sentiment on a topic, based on his past, the results obtained seem acceptable, and could be improved in future work.

1 INTRODUCTION

Nowadays the social networks form an integral part of our daily life. The last statistics compiled by the agency *"We Are Social Singapore"*[1] assure us the close relationship between the public and the social networks, where 51% of the population of the territory are Net surfers, 40% are users of the social networks and 37% are users of the social networks on mobile. Social networks are ideal tools for sending messages, giving advice or sharing opinions on social issues, so businesses, political parties, sociologists or other organizations rely heavily on social networks. Sentiment analysis is regarded as the key to those who want to exploit the feedbacks and the public opinions. In this domain, specifically in the web, there are two basic tasks as Cambria et al. pointed in their book (Cambria et al., 2017): emotion recognition(extracting emotion labels) and polarity detection(input classification as positive or negative). In this work, we focus on the polarity detection task in the social media. There are a lot of studies which have been done on this task (Tang et al., 2014; Barbosa and Feng, 2010; Cliche, 2017). These studies are based on the direct analysis of a message or a sentence in order to extract its polarity. Our proposal in this paper tackles the sentiment mining issue and relies on sentiment analysis and supervised machine learning principles, but in an original manner where polarity detection is based on the one hand on the users past content and on the published topic, on the other hand. In that case, we will predict the future message polarity of a user according to a defined topic. In our approach, we utilize the social networking site Twitter, which is the most popular microblog, with more than 300 million monthly active users. The famous limitation of the messages to 140 characters(which has recently been upgraded to 280 characters to encourage users to write more but we will use a corpus of tweets that are tweeted before the update) and the availability of the tweets collection allows us to choose it. We test our approach on the tweets collection of SemEval 2016. They are annotated tweets(positive or negative), and are related to definite topics. We have collected the past tweets of each user ourselves.

[1]https://wearesocial.com/fr/blog/2018/01/global-digital-report-2018

Belhareth, Y. and Latiri, C.
Microblog Sentiment Prediction based on User Past Content.
DOI: 10.5220/0008073102500256
In *Proceedings of the 15th International Conference on Web Information Systems and Technologies (WEBIST 2019)*, pages 250-256
ISBN: 978-989-758-386-5

2 RELATED WORK

In the abstract, the research field is called ”Affective forecasting”, and it can be divided into four tasks: predicting valence (i.e. positive or negative), predicting specific emotions, predicting intensity and duration(Wilson and Gilbert, 2003). These tasks have been used in different domains like economies, health and law. In this paper, we focus on the first task ”predicting valence” in social media. Actually, there are many studies, but we recall some of them :
(Asur and Huberman, 2010) did a study entitled ”Predicting the Future With Social Media”. They made a model that predict the movies box office revenues using Twitter data. They used the model machine learning ”linear regression”, and their principal parameters were: rate of attention and the polarity of sentiments. There is another study focused on election prediction using Twitter and specifically it was on the Germany election in 2009 (Tumasjan et al., 2010). The study proved the importance and rich of Twitter data and it reflected the political sentiment in a meaningful way. Therefore it could use it to predict the popularity of parties or coalitions in the real word.
(Nguyen et al., 2012) created a model for predicting the dynamics of collective sentiment in Twitter, it depends on three main parameters: the time of tweet history, the time to demonstrate the response of Twitter and its duration. They utilized automatic learning models such as ”SVM”, ”Logistic Regression” and Decision Tree”.
It may be observed that the aforementioned works used a random data collected within a stipulated period of time. In our work , we are going to use a past content for each user in order to predict the polarity sentiment according to a topic.

3 METHOD

3.1 General Layout of Proposed Model

Our approach tackles the prediction of Twitter user sentiment on specific topics. Predicted sentiment is simply the orientation of sentiment that can be positive, negative or neutral, but in this work, the prediction is positive or negative, since the corpus used contains tweets annotated on these two poles. As regards the prediction, it depends essentially on past tweets of each user. Our approach is based on supervised classification. Figure 1 provides us with an overall view on it. Firstly, we start by creating a classification model, so, we need a collection that contains a set of tweets grouped by topics. Each tweet is accompanied by its polarity and the name of its author (tweeter). Then we collect past tweets of each tweeter to extract his features. On the one hand, they were based on a semantic comparison. On the other hand, they are based on a sentiment polarity detection of past tweets. Finally, we obtain a classification model based on a supervised classifier. To test our model, we repeat the steps of creating the model up to the step of extracting features on other users and apply our model on the extracted features to get their predicted polarity.

Table 1: Summary of notations.

Notation	Description
U	Set of Twitter users.
T	Set of topics represented by terms.
E	Set of combinations: tweet, polarity.
C	Set of past contents of the users U
N	Number of tweeters
M	Number of topics
P	Number of tweets or size of E

3.2 Notations and Problem Definition

Before the problem definition, we point out that the table 1 provides the an overview of the notations used in this section. We consider a set of twitter users $\mathrm{U} = \{u_1,...,u_N\}$ and a set of topics $\mathrm{T} = \{t_1,...,t_M\}$, where t_j represents a set of terms. We also consider set $\mathrm{E} = \{(e_1, p_1),...,(e_P, p_P)\}$, where e_k represents the text of the tweet k on a topic belonging to T, and its author belongs to U, $p_k \in \{-1,1\}$ with -1 indicating negative sentiment and 1 indicates positive sentiment. The P size of the E set must be greater than or equal to N and M.
$C = \{c_1,...,c_p\}$ is the set of past contents of users U, where c_k is the past content of the user who wrote tweet e_k, which is a set of tweets of maximum and fixed size N_{max} submitted before date T_j which is the date of the first tweet belonging to T on topic t_j.
Our objective is to create a model that depends on supervised learning. Its role is to classify the past content of a user according to a specific topic. The classification can be either positive or negative. To do this, we consider a training set $\mathrm{D}_P = \{(x_1,y_1),...(x_P,y_P)\}$ and $F(X) = Y$ where F is a function modelling the relation between $X = \{x_1,...,x_P\}$ and $Y = \{y_1,...y_P\}$, when x_k represents the features vector (see equation 1) and y_k is the sentiment polarity with $y_k = e_k$.
Let consider 3 vectors V, W and Z, such that:

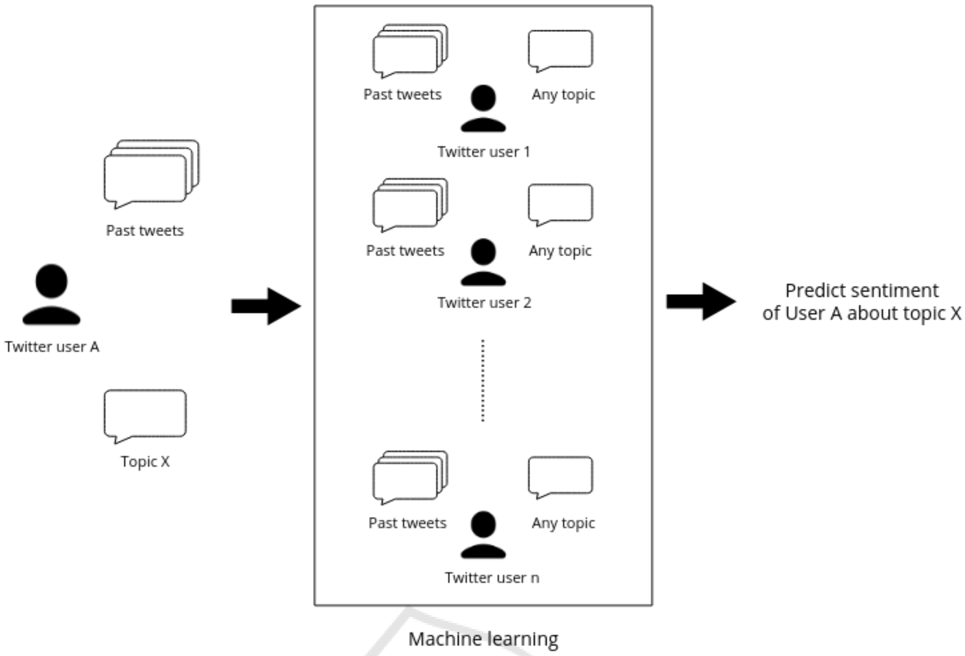

Figure 1: Approach Overview.

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_r \\ \vdots \\ v_{N_{max}} \end{pmatrix}, W = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_r \\ \vdots \\ w_{N_{max}} \end{pmatrix}, Z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \\ \vdots \\ z_{N_{max}} \end{pmatrix}$$

We obtain :

$$x_k = \left[v_1, w_1, z_1, \ldots, v_r, w_r, z_r, \ldots, v_{N_{max}}, w_{N_{max}}, z_{N_{max}}\right] \quad (1)$$

- r: r^{th} most recent tweet of c_k.
- v_r: It is the sentiment polarity of the r^{th} tweet (-1 or 1). If the r^{th} tweet does not exist, v_r is equal to 0.
- w_r: It is the semantic measure between the r^{th} tweet and the corresponding topic. If the r^{th} tweet does not exist, w_r is equal to 0.
- z_r: It is a score between 0 and 1 where the score tends towards 1 when the r^{th} tweet is close to the present (in our case it is close to the T_j). If the r^{th} tweet does not exist, z_r is equal to 0 (z_r = (date of r^{th} - date of first tweet (it was in 2006)) / (T_j - date of first tweet (it was in 2006))).

We notice that the features are divided into three types. The first type agrees with the past tweets sentiment, the second type agrees with their semantic relation to the corresponding topics and the last type is interested in the time factor. We also notice that the features of each tweet are kept, and they are not combined with each other. This is because any tweet is applied for on a specific date with such a sentiment (its absence is possible) on a specific topic. This leads us to treat tweets independently.

4 EXPERIMENTS AND RESULTS

4.1 Data Collection

We choose the data shared by the SemEval2016 team. It is a continuous series of evaluations of computational semantic analysis systems. It includes several tasks, but we focus on the data of the sentiment analysis task (Nakov et al., 2016), and specifically the subtask data that aims to classify tweets according to the sentiment polarity. In fact, It is a corpus of English tweets that are collected from July to December 2015. Tweets cover several topic categories such as books, movies, artists, social phenomena, etc. The corpus is composed of two sets: test and training. Each set represents instances that are composed of four attributes: the tweet identifier (code), the topic (as a term), the polarity and the text of the tweet (Table 3 shows an extract from the collection).

For each instance, we take its identifier in order to retrieve all the tweet information (retrieval is done by a python library[2] that uses the Twitter API[3]) including the tweet author's name. In this step we lost some instances due to the unavailability of the tweet. Then we retrieve each user's past tweets using a python program . It uses the same principle as if a user's twitter scrolls down to get past tweets. We set a maximum number of 300 tweets passed for each user. Finally, we delete users who have a number of past tweets less than 30. We also delete instances of users who applied more than once on the same topic. We keep that first tweet applied for each user (Table 2 shows the statistics before and after the recovery of past tweets). Figure 2 shows the distribution of the number of users according to their numbers of past tweets in the two sets (test and training). We notice that the two histograms have the same look and that the majority of the tweeters are very active(a Chi-Squared test is done and it showed a high dependence between the two sets with *p-value* $= 2.14e^{-11}$).

Table 2: Statistic collection.

	Topic	Positive	Negative	Total	User
Train	60	3,591	755	4,643	—
Test	100	8,212	2,339	10,551	—
After past tweets collection					
Train	60	2,144	440	2,581	2,565
Test	100	4,433	1,161	5,594	5,563

4.2 Sentiment Measure

To predict the sentiment of a Twitter user on a topic, we need to analyze his past tweets at the level of sentiment polarity. On a practical level, we cannot give to each user (tweeter) his past tweets to a group of experts in order to obtain their polarities. For that we need a tool that automatically does this task.
We have chosen the Vader-sentiment tool (Gilbert, 2014) and its advantage is that it does not depend on the learning approach, so it does not need training data. It is a lexicon and its lexemes are collected from three important lexicons in the field of sentiment analysis (LIWC(Pennebaker et al., 2007),ANEW(Nielsen, 2011), and GI(Stone et al., 1966)), similarly, it has been added by lexicons used in social networks (list of emoticons[4], list of slang terms[5] and list of acronyms[6]). The lexicon was evaluated by a group of experts to assign a real type value to each lexicon that represents a positive or negative intensity (between -4 and 4).

[2]https://github.com/bear/python-twitter
[3]https://dev.twitter.com/
[4]http://en.wikipedia.org/wiki/List_of_emoticons#Western
[5]https://www.internetslang.com/
[6]http://en.wikipedia.org/wiki/List_of_acronyms

On the other hand, Vader-Sentiment can evaluate a sentence, text or micro blog expressed in social networks through its lexicon and grammatical and syntactical rules. Table 4 illustrates a sentiment evaluation of a tweet. The positive, negative and neutral measures represent the rates of each category in the input text, and for the compound measure, it is the sum of intensity of each lexicon with a normalization between -1 and 1. This measure is used for the input classification (greater than 0.05 gives a positive input, less than -0.05 gives a negative input, otherwise it gives a neutral input). We did a Vader-Sentiment experiment on the SemEval 2016 collection (test and training). We have had f1-score=0.60 and Accuracy=0.70. The evaluation results are a little low, so we have decided to take these three intensities as features instead of just one feature (1: positive, -1: negative).

4.3 Enriching Topics

The tweets presented in the corpus have been postulated on well-defined topics. Each topic is expressed by a single term or token. Each group of tweets is postulated on a specific aspect of topic over a certain period of time, so we need the topics expressed by several specific terms that express the correct aspect of the topic in order to make a semantic comparison between the topics and their past tweets. To enrich the topics with appropriate terms, we have grouped the tweets belonging to the same topic into a single text, and from each text we have extracted the keywords through a keyword extractor (De Sousa Webber, 2015). This tool does not just depend on statistical measures, it also depends on semantic measures. We see in Table 5 some examples of topics with their enriched terms. In absolute terms, if we want to apply the approach in the real world, we can introduce our own terms.

4.4 Evaluation Protocol and Metrics

For the evaluation metrics, we have kept those were determined by the SemEval 2016 team, they used the following evaluation measures:

- ρ^{PN} is the macroaveraged recall:

$$\begin{aligned}\rho^{PN} &= \tfrac{1}{2}(\rho^{P} + \rho^{N}) \\ &= \tfrac{1}{2}\left(\tfrac{TP}{TP+FP} + \tfrac{TN}{TN+FN}\right)\end{aligned} \tag{2}$$

  where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives, respectively.

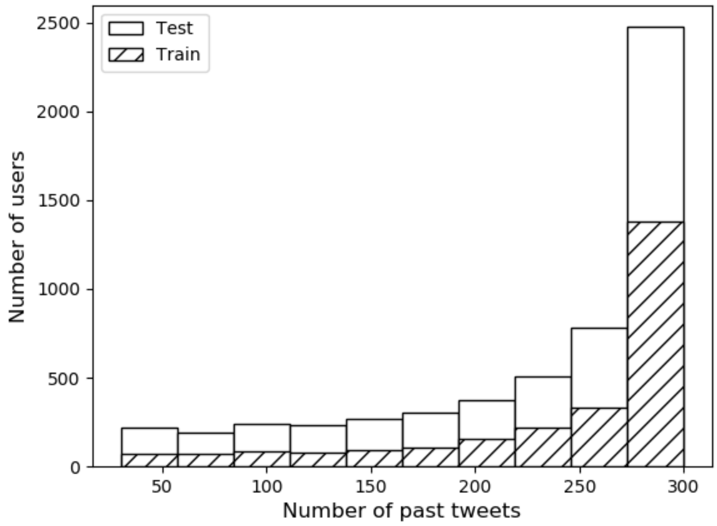

Figure 2: Number of users according to their number of past tweets.

Table 3: Some instances of the collection SemEval2016.

631676347362897920	big brother	positive	BIG Brother tomorrow is going to be so good
639659729480908804	big brother	negative	This may be the worst Big Brother episode in history.
639820535917092864	@microsoft	positive	@festivaluprise @Microsoft Good luck to all Pitch Battlers. May the best pitch win!
628949369883000832	@microsoft	negative	dear @Microsoft the newOoffice for Mac is great and all, but no Lync update? C'mon.
635014939581784064	angela merkel	positive	'Angela Merkel is right: the migration crisis will define this decade'. http://t.co/xFxH6tR7g8
634098002328743936	angela merkel	negative	But I thought Angela Merkel was the most evil woman in Europe? https://t.co/sqUYES8QjI

Table 4: Sentiment analysis of a tweet example by Vader-Sentiment.

BIG Brother tomorrow is going to be so good			
compound	negative	neutral	positive
0.578	0.0	0.681	0.319

- F_1^{PN} is the $F_1 scores$ average of positive and negative classes:

$$F_1^{PN} = \frac{F_1^P + F_1^N}{2} \tag{3}$$

$$F_1^P = \frac{2\pi^P \rho^P}{\pi^P + \rho^P} \tag{4}$$

with π^P and ρ^P designate precision and recall for the positive class, respectively:

$$\pi^P = \frac{TP}{TP+FP} \tag{5}$$

$$\rho^P = \frac{TP}{TP+FN} \tag{6}$$

- Classifier accuracy:

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

The adoption of measures 1 and 2 resides in the high sensitivity of the latter to class imbalances, and this imbalance exists approximately in the collection , where 20% of instances are negative and 80% are positive in the two sets. They also determine the functioning of the baseline classifier where it assigns to each instance the positive class, it is a method or any trivial classifier able to achieve it.

4.5 Results and Discussion

In this part, we will see the results of the experiments of our approach on the collection defined above. Before starting, it is necessary to point out the method of semantic measurement between topics and past tweets. We have chosen the approach of (Mnasri et al., 2015) that has shown better performances with the speed of execution. It is a measure of cosine similarity between the word vectors of the tweets and those of the topics. The vectors are generated from a word embedding model according to the Word2vec (Mikolov et al., 2013) algorithm.

Table 5: Example of extended topics.

Topic	**Terms**
christians	sunday, christ, christians, muslims, god, jews, worship, tomorrow, persecution, bible
google+	page, google, twitter, workshop, tomorrow, friday, youtube, google+, facebook, sunday
disneyland	love, halloween, guys, bunch, kid, someone, tomorrow, friday, disneyland, gonna
iphone	announcement, upgrade, iphone, tonight, os, apple, watch, tomorrow, plus, phone

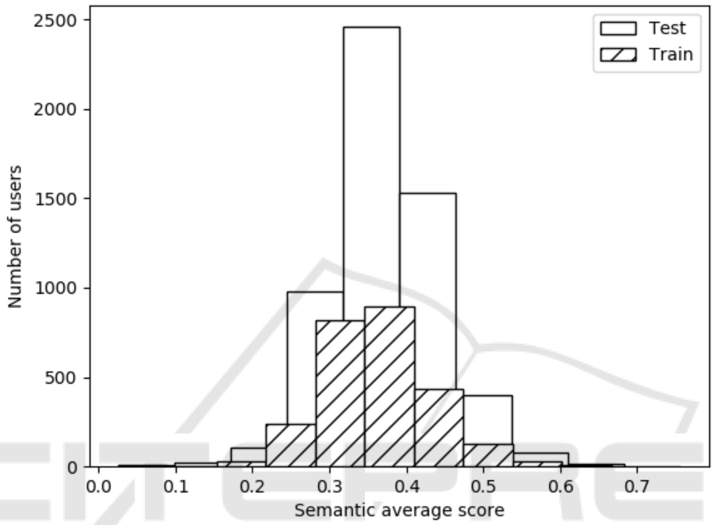

Figure 3: Number of users according to the semantic average between their past tweet and the corresponding topics.

Table 6: Classifier results.

Classifier	K	ρ^{PN}	F_1^{PN}	*Acc*
Naive_Bayes	250	0.592	0.55	0.620
Logistic_Regression	150	0.533	0.53	0.760
Decision_Tree	250	0.527	0.52	0.700
SVM	300	0.521	0.51	0.765
Baseline	—	0.500	0.44	0792

The model[7] is built from a google-news dataset. Histogram 3 shows a distribution of the number of users in relation to the semantic average between their past tweets and the corresponding topics. We observe that most users have an average between 0.25 and 0.5, which allows us to have a collection adequate to our approach where the past content of each user resembles relatively to the corresponding topic which makes the sentiment prediction more logical.

We have tried several classifiers but we have indicated in Table 6 only the results of the classifiers that are remarkable. For each classifier, we have estimated parameter k. It is the number of past tweets for all users. The estimation of k and classifier parameters is done by the 3-fold cross-validation procedure on the training set. The results show a slight improvement according to the SVM, RF and LR classifiers, and an acceptable improvement for the NB classifiers. The performance of the classifiers could be improved at several levels. First, we need to adopt a more effective sentiment tool that takes into account sarcasm expressions in social networks. Second, we need to improve the performance of semantic measurement especially since we have measured its quality according to the mean absolute error measure between a vector filled with 1 (ideal score) and the vector that contains the semantic scores between the tweets of the Semeval2016 collection (test and train) and its corresponding topics, and we had a relatively large error value of 0.34. This error can be improved by changing the corpus used to extract the word embedding model with another one that fits with the topics and with the short text. Finally, we can apply models based on neural networks in order to improve accuracy.

[7] https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit

5 CONCLUSION AND FUTURE WORK

In this paper, we presented an approach to predicting sentiment polarity in Twitter. The prediction depends on the past content of users according to identified topics. The utilized collection is that of the sentiment classification task of the SemEval 2016 edition. Added to that, we collected past tweets of each author in the collection. Our approach depends on supervised learning. The used features are sentiment measures at the tweets level, semantic measures between tweets and topics, and time scores. As a first experiment, the results obtained are acceptable, and for this reason we will try to improve the performance in future work by adopting deep learning as well as testing the approach on a larger collection.

REFERENCES

Asur, S. and Huberman, B. A. (2010). Predicting the future with social media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 492–499. IEEE Computer Society.

Barbosa, L. and Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd international conference on computational linguistics: posters*, pages 36–44. Association for Computational Linguistics.

Cambria, E., Das, D., Bandyopadhyay, S., and Feraco, A. (2017). *A practical guide to sentiment analysis*, volume 5. Springer.

Cliche, M. (2017). Bb_twtr at semeval-2017 task 4: Twitter sentiment analysis with cnns and lstms. *arXiv preprint arXiv:1704.06125*.

De Sousa Webber, F. (2015). Semantic folding theory and its application in semantic fingerprinting. *arXiv preprint arXiv:1511.08855*, page 38.

Gilbert, C. H. E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Mnasri, M., De Chalendar, G., and Ferret, O. (2015). Intégration de la similarité entre phrases comme critère pour le résumé multi-document. In *23ème Conférence sur le Traitement Automatique des Langues Naturelles (JEP-TALN-RECITAL 2016), session articles courts*, pages 482–489.

Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., and Stoyanov, V. (2016). Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*, pages 1–18.

Nguyen, L. T., Wu, P., Chan, W., Peng, W., and Zhang, Y. (2012). Predicting collective sentiment dynamics from time-series social media. In *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining*, page 6. ACM.

Nielsen, F. Å. (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv:1103.2903*.

Pennebaker, J., Chung, C., Ireland, M., Gonzales, A., and Booth, R. (2007). The development and psychometric properties of liwc2007: Liwc. net. *Google Scholar*.

Stone, P. J., Dunphy, D. C., and Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1555–1565.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Fourth international AAAI conference on weblogs and social media*.

Wilson, T. D. and Gilbert, D. T. (2003). Affective forecasting. *Advances in experimental social psychology*, 35(35):345–411.