J. Health Inform. 2018 Janeiro-Março; 10(1): 17-24 **Artigo Original**

Sentiment Analysis of Twitter's Health Messages in Brazilian Portuguese

Análise de Sentimento de Mensagens de Saúde no Twitter em Português Brasileiro

Análisis de sentimientos de los mensajes de salud de Twitter en portugués brasileño

Gabriela Denise de Araujo[1], Fabio Oliveira Teixeira[2], Felipe Mancini[3], Marcelo de Paiva Guimarães[3], Ivan Torres Pisa[4]

ABSTRACT

Keywords: Social Media; Cancer; Delivery of Health Care

Objective: To present results of a sentiment classification methodology, here denominated Sentiment Descriptor Indexing (SDI), to be applied in Brazilian Portuguese Twitter's messages related to health topics. **Methods**: The first step considered the construction of an algorithm that is based on the co-occurrence of Twitter terms with sentiment descriptor vocabulary known as ANEW-BR. In the second stage, an evaluation of SDI algorithm performance for messages about "cancer" of a period of three weeks was performed. The ratings were paired, to generate a performance appraisal.**Results**: The precision and recall values were 0.68 and 0.67, respectively. A total of 25,230 messages on the topic "cancer" with a positive feeling classification (71%) were collected. **Conclusion**: The contributions of this work aim to fill the lack of methods of analysis of feelings for the Portuguese Portuguese language.

RESUMO

Descritores: Mídias sociais; Câncer; Assistência à Saúde

Objetivo: Apresentar os resultados de uma metodologia de classificação de sentimento, aqui denominada Sentiment Descriptor Indexing (SDI), para aplicar em mensagens do Twitter em português brasileiro relacionadas a temas de saúde. **Métodos**: A primeira etapa considerou a construção do algoritmo SDI que se baseia na coocorrência de termos do Twitter com descritores do vocabulário ANEW-BR. Na segunda etapa foi realizada uma avaliação do desempenho do algoritmo SDI para mensagens sobre o tema "câncer" de um período de três semanas. As mensagens foram classificadas por voluntários e em paralelo pelo SDI. As classificações foram pareadas gerando uma avaliação de desempenho. **Resultados**: Os valores de precisão e recuperação resultaram 0,68 e 0,67 respectivamente. Coletou-se um total de 25.230 mensagens sobre o tema "câncer" com classificação de sentimento positiva (71%). **Conclusão**: As contribuições deste trabalho visam suprir a falta de métodos de análise de sentimentos para a língua portuguesa brasileira.

RESUMEN

Descriptores: Medios de comunicación sociales; Cáncer; Prestación de Atención de Salud

Objetivo: Presentar los resultados de una sensación de metodología de clasificación, aquí se llama Sentiment Descriptor Indexing (SDI), para aplicar en los mensajes de Twitter relacionados con temas de salud por el idioma portugués de Brasil. **Métodos**: El primer paso considera la construcción del algoritmo de SDI que se basa en la co-ocurrencia de términos do Twitter con el vocabulario de sentimiento conocido como ANEW-BR. En la segunda etapa se llevó a cabo una evaluación del rendimiento de los algoritmos SDI para los mensajes sobre "cáncer" de un período de tres semanas. Los mensajes se ordenan por voluntarios y en paralelo por SDI. Las clasificaciones fueron emparejados generar una evaluación de desempeño. **Resultados**: Los valores de precisión y recuperación fueron 0,68 y 0,67, respectivamente. Se recogió un total de 25.230 mensajes sobre "cáncer" con la clasificación de sentimiento positivo (71%). **Conclusión**: Las contribuciones de este trabajo tienen como objetivo hacer frente a la falta de métodos de análisis de sentimientos por el idioma portugués de Brasil.

[1] *Mestre em Ciências. Programa de Pós-graduação em Gestão e Informática em Saúde – PPGIS, Universidade Federal de São Paulo - UNIFESP, São Paulo, (SP), Brasil.*

[2] *Doutor em Ciências. Programa de Pós-graduação em Gestão e Informática em Saúde – PPGIS, Universidade Federal de São Paulo - UNIFESP, São Paulo, (SP), Brasil.*

[3] *Professor Doutor. Secretaria de Educação a Distância, Universidade Federal de São Paulo -UNIFESP, São Paulo, (SP), Brasil.*

[4] *Professor Livre-Docente do Departamento de Informática em Saúde, Universidade Federal de São Paulo -UNIFESP, São Paulo, (SP), Brasil.*

Autor Correspondente: **Gabriela Denise de Araujo**
e-mail: **gabriela.denise@unifesp.br**

Artigo recebido: **16/04/2017**
Aprovado: **24/08/2017**

INTRODUCTION

In recent years, the growth in the volume of data available on many areas of the web, such as blogs, corporative portals and social media, has modified the behavior of people and organizations in search of information. Nowadays, the search of information is usually initially concentrated on the web and is only later expanded to other media when necessary (printed media for example). This has been occurring because the web has become so widespread that it covers user's particular experiences, even including government prevention campaigns. In this sense the available data naturally becomes a rich source of information and evidence on various subjects. As a result, investigations of new concepts and methodologies for the analysis of such data have been compelled[1]. So much so that virtual environments that provide space for publication of opinions and news, like Twitter (twitter.com), already offer tools that enable access to data for analysis and studies for those interested in it[2].

Analyzing web data and identifying relevant information is a current and significant computational challenge, which is controlled by the area of artificial intelligence in particular by the subarea of natural language processing (NLP).

Examining texts written by people who express their opinions and feelings in one of their specific tasks. Sentiment analysis (SA) is a technique that has been gaining ground in the treatment of recognizing the affection inferred from written excerpts[1].

In this context, new opportunities arise in research, for example, to understand how health consumers use social networks to find information about treatments, and to understand opinions about diseases and aggregate knowledge about health. Studies focus on monitoring and identifying human sentiment in messages shared on social media, based on the assumption that a significant percentage of online data is related to emotions and opinions expressed by users[3-6].

According to Bing Liu[7], opinions are important because people and organizations are influenced by them when making a decision. Moreover, they can contribute to improvements in service. Chew and Eysenbach conducted a content analysis study of messages posted on Twitter during the H1N1 outbreak in 2009[8]. The study showed that the messages were used primarily to disseminate information of credible sources, but also opinions and personal experiences. This study found that the messages can be used for real-time content analysis and different sources of opinions and experiences, providing information to public health authorities so they may more easily capture public concerns.

In 2011, Yu[9] published a study on the classification of emotion types expressed by members of online health communities. This study revealed patterns in the expressions of users of different profiles, such as: inquirers, respondents, men, women, doctors, patients, caregivers, and others. Furthermore, it allowed for the classification and comparison of user's emotional communication. In order to understand how words express happiness in the United States, Lewis Mitchell et al[10] conducted a survey on the use of words in urban areas, considering states and cities, which enabled mapping and comparing areas of high and low happiness. This research used Twitter messages and correlated demographic, geographic, health and emotional characteristics. Among the many findings was a correlation between positive terms and the topic of obesity. This study demonstrated this technique's potential for the identification of areas of high obesity based on the use of significant words. Brazilian researchers conducted a study in which they presented an observation on dengue fever[11]. This study considers Twitter as a continuous source of epidemiological information. A tool that tracks the opinion of the population about dengue fever through the flow of messages has been developed, thus enabling real-time monitoring by health agencies and facilitating that intervention measures can be defined in advance. In each study cited, different methods of classification were used, adopting SA classification approaches, and with satisfactory results achieved, however the most of the studies deal with a specific topic and the sentiment analysis approaches end up becoming very restricted to the topic and consequently the language.

Along with a vast source of available data on the web, and the possible application of computational techniques to identify relevant data, we realized the necessity of exploring SA techniques and extracting relevant information and new concepts from the population regarding health topics in general. Furthermore, because users commonly express opinions and feelings on social media, these networks become ideal sources for research.

Considering the difficulties in finding sentiment analysis methods for Brazilian Portuguese and the complexity of applying the same approach to more than one topic, our study considered the construction of a sentiment classification method, here on called Sentiment Descriptor Indexing (SDI), to be applied in Twitter's messages related to health topics in Brazilian Portuguese. This method provides a polarity (positive or negative) of the user's opinion on published content, enabling sentiment analysis to be used to characterize aspects of the popularity and impact of the subject in focus. The engine developed is associated to a word selection approach, which involves evaluating content words and identifying those which are positive from those which are negative, based on a pre-sorted list of descriptors, along with a statistical approach to calculate the co-occurrence between words and sentiment descriptors. The list of sentiment descriptors used in this study was based on the affective vocabulary known as Affective Norms for English Words - BR (ANEW-BR)[12]. This set of words was put together from a study that translated and adapted the Affective Norms for English Words (ANEW) into Brazilian Portuguese, resulting in two emotional dimensions (valence and arousal) for a set of 1,046 words. The method proposed in this paper uses only the first dimension (valence) to sort messages. The contributions of this work aim to address the lack of SA methods for the Brazilian Portuguese language and to encourage its application in improving other activities in NLP.

METHODS

Three major steps were completed to fulfill the proposed objective of this study: collection and storage of data, construction of a SDI sentiment classifier, and testing and assessment of the SDI Classifier.

Collection and storage of data

Regarding the collection and storage of messages (tweets) on Twitter, we created an automatic mechanism using Java language and the Twitter4J (http://twitter4j.org) open code library. Firstly, it was created a database with one million tweets in Brazilian Portuguese containing the 1,046 sentiment's descriptors of the affective vocabulary ANEW-BR. This database was used in the construction of the sentiment classifier.

Construction of a sentiment classifier

To build the SDI classifier, we used the approach of the Journal Descriptor Indexing (JDI)[13], developed by the National Library of Medicine (NLM). The JDI is an automatic categorization tool of texts that present significant results for classification and indexing of scientific articles. There are studies that also show its application for general texts in health[14-15]. In short, the SDI uses co-occurrence between terms and categories, generating a score of relevance (degree of proximity) between them. In SDI, the categories are the sentiment descriptors of the affective vocabulary ANEW-BR that make up a vector of descriptors for each Twitter message analyzed. As an intermediate step, each term found in a message generates a vector with relevance scores relating this term to the sentiment descriptors. Subsequently, a final message vector is constructed by calculating an arithmetic average of the scores found in the relevant intermediate vectors. The final vector is ordered in the relevance of descending order. The five highest average relevant descriptors are then used to define the polarity of the message investigated.

First of all, to get a relevance score of a term about the sentiment descriptors is used the Equation (1).

$$\mathbf{SDI}_{k,i} = \frac{\mathbf{n}_{k,i}}{|\{j : t_i \in c_j\}|} \quad (1)$$

In this equation, the parameter $n_{k,i}$ represents the number of messages that the term t_i occurs with a sentiment descriptor. The denominator of the equation represented by the parameter $|\{j: t_i \in c_i\}||\{j: t_i \in c_j\}|$ expresses the sum of the number of messages in which the term occurs. The value of the relevance score can range from 0 to 1, with values closest to 1 being the most

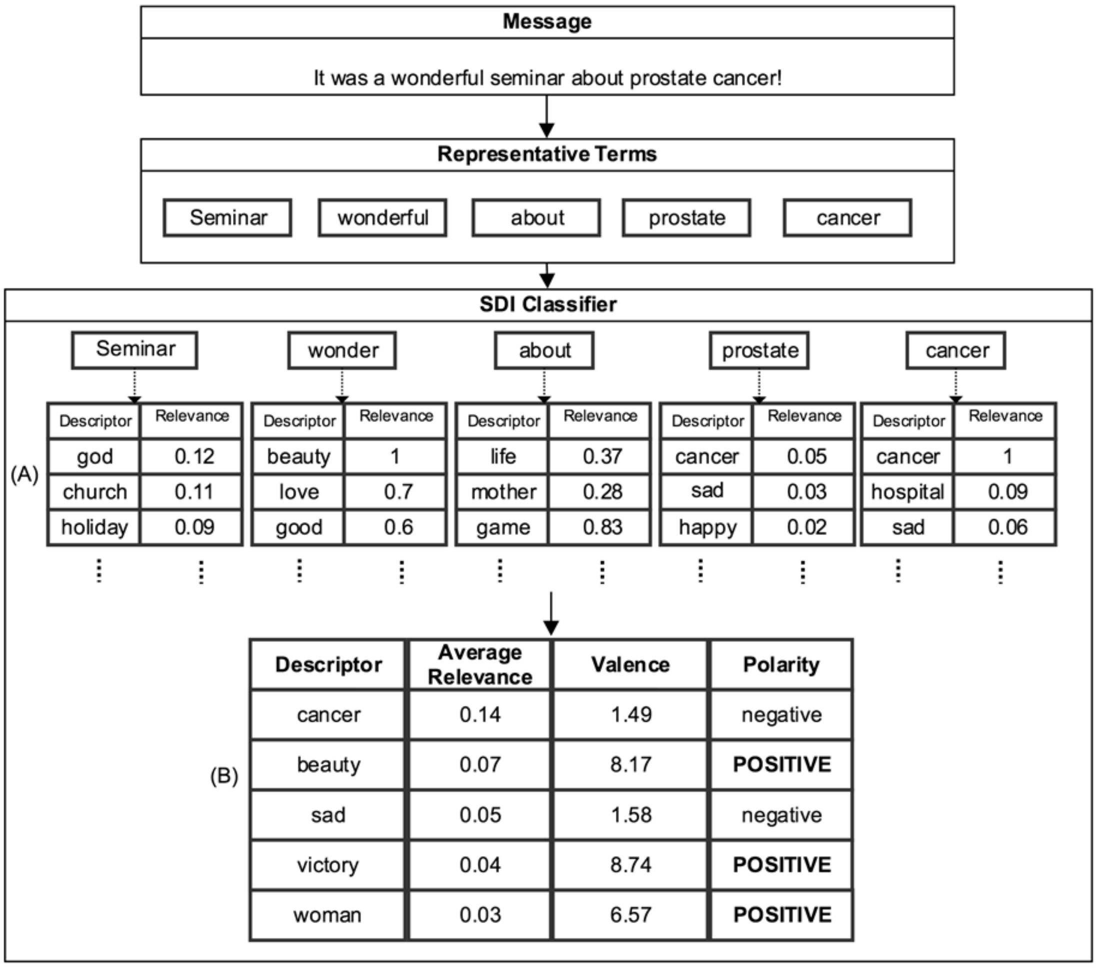

Figure 1 - Message classification example. Each representative term will have a vector of sentiment descriptors (A) with their respective relevance values. The average of the relevance's values of each descriptor is calculated to generate the final vector (B). The final vector is sorted of the highest to lowest relevance and the five descriptors of the top positions provide a classification. In this example, the message has been classified as positive.

relevant (higher co-occurrence).

Each term is associated with a vector whose size was defined by the total number of 1,046 descriptors of the ANEW-BR. The elements of this vector contain the relevance value in relation to its descriptor. A matrix is calculated based on Equation (1) for the group of 192,113 terms extracted from a million collected tweets, generating an occurrence matrix of dimensions 1,046 x 192,113.

From this matrix, the next step is the creation of a method to design a vector of sentiment descriptors for each database message to be investigated in this study. In this method, for the construction of sentiment descriptors vector we calculated the arithmetic average of the relevance of terms in relation to sentiment descriptors. The final vector generated for the message contains the 1,046 sentiment descriptors and their average relevance values. These values form a ranking when ordered from highest to lowest, placing the most appropriate descriptors in the top positions for labeling the message as negative or positive from the valence value of these descriptors.

To provide a final classification, the valence value of the first 5 descriptors of this vector is considered. The message is classified as negative if 3 or more descriptors have negative valence, and otherwise are classified as positive. Figure 1 shows a classification example of a message that was classified as positive by SDI.

In order to obtain a more accurate result on the sentimental classification of the messages and increase the scope of classification, two additions were combined to the developed mechanism. The first evaluates the presence of 93 complement emoticons to generate the classification adapted from the Wikipedia[16], with positive or negative polarity. The second addition evaluates the presence of the word "not" as a denial message[17] by reversing the polarity originally obtained in these cases.

The 3,797 collected messages must be pre-processed before being inserted into the SDI classifier. Among the collected messages on Twitter are those that are not directly related to the cancer topic as a health issue because the word cancer, in Brazilian culture, has two very popular meanings. The first one is a term related to a disease or diagnosis, and the second is related to a horoscope sign. Then, we exclude from the database messages containing ambiguous terms resulting in 3,119 messages for analysis.

With the remaining messages was applied these processes: (i) removing the user from the message when it is a retweet message or when the user sends a message to another specific user; (ii) removing a shortened link if any; (iii) removing hashtags, which are specific markers of events, people or things; (iv) converting abbreviated words to their no abbreviated format; (v) removing special text characters such as punctuation and symbols. For the conversion of abbreviated words into their no abbreviated form, we adapted a list of 56 abbreviation words. This list was constructed with the aid of some studies[18] and also by analyzing the Twitter messages collected for this study.

Testing and assessment of the classifier

An experiment with human classification was used to detect human perception in sentimental message classification on the topic of health, following the classification proposed by the SDI classifier. The 3,119 posts on cancer were classified by 299 student volunteers from a Specialized Health Informatics course by further education, provided by the Open University of Brazil (UAB) in partnership with Federal University of Sao Paulo (UNIFESP). The volunteers had training in medicine, nursing, psychology, teaching, computer science, physics, engineering, law and journalism. In order for volunteers to label messages, they accessed a website built specifically for this purpose. On this page, the participants could sort messages into positive, negative, or neutral categories, and also rated as "health" or "not health".

A message was considered positive, negative or neutral if one of these categories received 60% or more of the votes. It was considered as a "dispersed" category when there was no agreement among raters. The classification into "health" or "not health" considered a minimum of 50% of the votes; in the event of a tie, it was considered "dispersed". Only messages classified as health were compared to the SDI classification.

As a stage of evaluating the SDI classifier performance, we used 3 standard metrics: precision, recall and F-measure[19]. These metrics were applied to the results of the SDI classifier. The ratings carried out by volunteers for the same set of data were considered the gold standard.

RESULTS

A total of 3,119 tweets on cancer collected in May 2013 were analyzed and classified by 299 volunteers. Table 1 showed the results of the classification of tweets in the categories of health, not health and dispersed, as performed by the volunteers.

Table 1 - The amount of messages classified by humans into the classes: health-related, non-health related and dispersed.

Class	Amount	%
Health-related	1,146	36.7
Non-health related	1,868	59.9
Dispersed	105	3.4
Total	3,119	100

It is noted that over 50% of the tweets were considered non-health related and a small percentage of the total was labeled dispersed. Therefore, we disregarded tweets labeled as non-health-related and dispersed for the establishment of gold standard classification. In Table 2, we can observe the sentiment classification results performed by volunteers. It is noted that most of the classified tweets were categorized as dispersed sentiment, that is, most of the positive, negative and neutral categories did not reach 60% or more of the votes, resulting in poor agreement between them.

The neutral category includes tweets with no clear polarity trend, meaning tweets that may contain facts or news, so it is not possible to identify a feeling or an opinion.

Thus messages classified as neutral and dispersed were excluded from the gold standard. In the end, the topic of health-related cancer was obtained, discussed on Twitter in May 2013, with a generally positive classification according to the volunteers of this research. The SDI classifier was applied to 311 tweets classified as health-related, and labeled as positive and negative by the volunteers. The results are shown in Table 3.

Table 2 - The amount of messages classified by humans in positive, negative, neutral and disperse classes.

Class	Amount	%
Positive	194	17
Negative	117	10
Neutral	36	3
Dispersed	799	70
Total	1,146	100

Table 3 - The amount of messages labeled by the SDI classifier into positive and negative.

Class	Amount	%
Positive	255	82
Negative	56	18
Total	311	100

The overall result of the classification of sentiments by the SDI classifier also found a positive opinion for the cancer topic during the study period. The confusion matrix and the SDI classifier performance compared to the classification made by humans are shown in Table 4 and Table 5, respectively.

Table 4 - Confusion matrix obtained when the SDI method was used to classify messages about cancer.

		Human classification	
		Positive	Negative
SDI Classifier	Positive	131 (42%)	61 (20%)
	Negative	63 (20%)	56 (18%)

In Table 5, the performance of the SDI classifier is computed using the following standard metrics of Precision, Recall and F-measure based on the measures True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) shown in Table 4. The traditional F-measure or F_1 score is a balancing measure between the recall and the precision that calculates the harmonic mean between the two metrics. Besides, there are two other frequently used F-measures, F_2 which weighs recall higher than precision, and $F_{0.5}$ which weighs precision higher than recall.From analyzing precision, recall and F-measure, we observed that all achieved values were similar. Thus, the performance of the SDI classifier for the classification of posts in the positive and negative categories was approximately 68%.We carried out popularity and impact analyzes with the aim of identifying which subjects were associated with the research topic and how they impacted and influenced users to share the same positive or negative sentiment. A cloud illustration of words was generated to view the terms that obtained the highest frequency scores. Terms relating to matters that occurred more frequently in the tweets appear in greater prominence in the cloud (Figure 2), from May 2013, in which the font size is adjusted to the frequency of words from the collection of texts. Consequently, the most common terms show that they had a higher impact during the study period. The term cancer, which was also an attractor term, had the highest frequency score compared to the other terms. Then, this term was excluded from the cloud as the intention of the cloud was to represent issues related to cancer.

Table 5 - Precision, recall and F-measure values obtained when the SDI classifier was used to classify messages about cancer.

			F-measure		
	Precision	Recall	0.5	1	2
SDI Classifier	0.68	0.67	0.68	0.67	0.67

Issues concerning famous people (who had the disease during the study period and were widely reported in the media), cancer treatment hospitals, health campaigns, and types of cancer can be extracted from the words in the cloud, among others. Although cancer is a negative theme as it relates to a health condition and it consequently suggests negative comments and discussions on social networks, we found that the users' opinions on these issues

Figure 2 - Word cloud of the most frequent words about cancer in May 2013 excluding the term cancer.

identified as the most discussed on Twitter were positive, during the study period. The whole process of data capture in this study was performed in a period of six months, from May to October 2013. The total of collected messages on the subject during that period was 25,230.

There remained 18,668 messages that had been classified by SDI after the data pre-processing. The general sentiment classification of the messages was positive, reaching 73%. Therefore, it can be considered that the subject of cancer between May and October was positively commented on by Twitter users. Figure 3 presents a sentiment classification of cancer on messages in all months of study, showing a positive labeling with the highest score in each month.

DISCUSSION

As is an arduous task that is used in processing a large volume of opinion text, with abbreviated terms, misspellings, the high subjectivity of content, among other limiting factors. Nevertheless, it is a technique that presents significant scientific findings and has been little explored in the health area[20]. The messages shared on social networks are the most inputs used in analyzes and studies applying SA techniques.

The first difficulty faced in this study was related to the chosen theme. The topic of Cancer was first selected for being one of the topics most discussed by volunteers who assisted in the investigation of the themes for this study and because it is a recurring and important issue in the area of health. However, the word cancer has two commonly known meanings. The first is the name of a group of more than 100 diseases that have uncontrolled growth of cells that invade tissues and organs and may spread to other regions of the body[21]. The second concept is related to the horoscope, and is the fourth astrological sign in the Zodiac[22]. Thus, many collected messages were not directly related to the topic as a health issue, and despite the pre-processing prepared to discard messages with content related to the cancer sign and horoscope, there still remained some messages that were deleted only after human ratings of being health and non-health related.

Regarding the sentiment classification performed by student volunteer evaluators, agreement was expected between the judges at first so that messages could be readily labeled into positive, negative and neutral classes by adding the amounts of votes. However, the low agreement of the evaluators was remarkable. The dispersed class presented a total of 70%, meaning less than half of the investigated messages were safely labeled positive or negative by humans. On the other hand, the disagreement was much lower for the classification of health-related and non-health related posts. Only 3% of messages were identified as dispersed. Even with the pre-processing applied to eliminate the messages related to the cancer horoscope sign, 60% of messages were still labeled as non-health related. Because of this, and from analyzing these messages labeled in the non-health-related category, we found that users post messages with attractors under the topic of cancer, inducing other meanings for the attractors in message content. Some messages indicate that the word cancer can be freely used as an adjective in a context that may hinder a polarity analysis. Therefore, because it is not being used in the well-defined semantic meaning, it causes subjectivity and ambiguity in the text, a common phenomenon in the Brazilian Portuguese language. Nevertheless, the human intelligence in tasks such as categorization of text is notable, and the capable to realize irony, ambiguity and subjectivity in texts is very relevant. Therefore, in general it is more reasonable to use human classification as a gold standard[23], even the agreement among the humans is not completely accurate. In fact, the agreement can be improved, for example, by increasing the number of classification by tweet, in other words, more humans classifying the same tweet, or making a preprocessing of tweets to separate messages that really contain opinions, becoming more accurate the sentiment classification, facilitating the reproducibility.

CONCLUSION

The classifier constructed in this study used the word selection approach, along with a statistical approach for the sentiment classification of messages, and considering the use of ANEW-BR vocabulary. This research highlighted the use of ANEW-BR vocabulary as the main

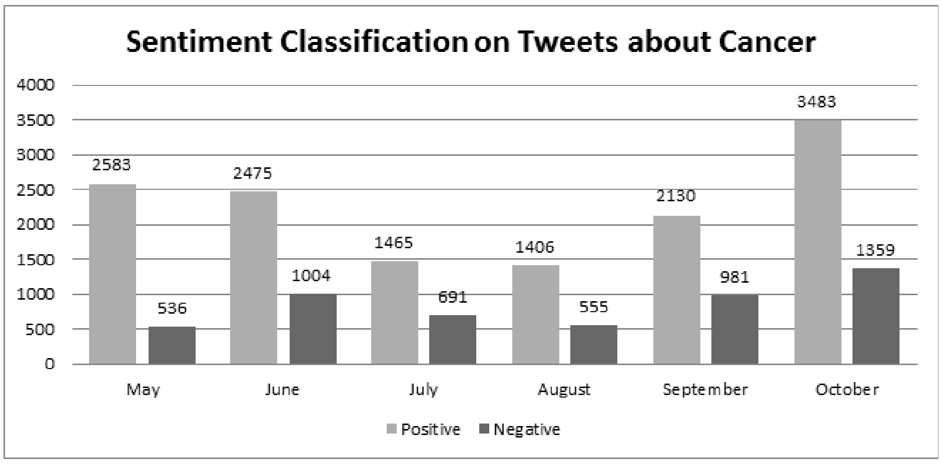

Figure 3 - Sentiment classification on cancer messages in all months of the study.

instrument of sentiment classification of words, and investigated the suitability of the ANEW-BR to classify messages with health content. One of the strengths of using the word selection approach, which makes use of sentiment vocabulary, is the fact that the system is domain-free. This is because the analysis is performed on the words that make up the investigated text, which consequently must be in the ANEW-BR vocabulary. The statistical approach helps even more, for fact that the system is less susceptible to context dependency, complementing the word selection approach. Applications that use supervised machine learning algorithms may have a greater success rate compared with other techniques, such as the selection of words, which is unsupervised. However, in these cases these applications are compiled from training conducted in their classifier in the construction phase, usually focusing on a specific domain. This training is usually done from a set of data with pre-rated content by people from that particular domain. That is, the data train a classification algorithm that learns to classify other non-classified data. In the case of constructing a SDI algorithm like in this work, this pre-assessment by people does not exist. The first reason is due to the large volume of Twitter messages, which makes it unfeasible to obtain such prior classification. The second reason is partly due to a lot of disagreement about the sentiment rating (positive or negative) of this type of message (70% of disagreement in tweets about cancer). In the case of humans, it is not possible to suggest without further specific studies on the systematic (regular) nature of these disagreements. Thus, the proposed SDI algorithm has satisfactory efficiency results, even more so when considering that their regular classification errors are systemic. Therefore, it presents itself as a reliable tweets classification instrument, even with its limitations.

ACKNOWLEDGEMENTS

The authors would like to thank the health professional consultants for their valuable and constructive suggestions about the health topics to be collected in this research and the 299 student volunteers for their help to classify the tweets.

This research received financial support from Coordination for the Improvement of Higher Level Personnel (CAPES) Brazil.

REFERÊNCIAS

1. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inf Retr. 2008 Jan;2(1-2):1-135.
2. Júnior CAS, Albuquerque JP, Mulatinho N, Nunes AM, Queiroz FS, Lima SR, et al. Difusão da informação no Twitter: fluxo da informação da audiência invisível. ENANCIB 2013 [Internet]. 2013 [citado 2014 jul 14]. Disponível em: http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/468
3. Ohana B, Tierney B. Sentiment classification of reviews using SentiWordNet. Proceeding of the 9th IT & T Conference [Internet]; 2009 Oct 22-23; Dublin, Ireland. Available from: http://arrow.dit.ie/ittpapnin/13
4. Zhang W, Xu H, Wan W. Weakness Finder: find product weakness from chinese reviews by using aspects based sentiment analysis. Expert Syst Appl. 2012 Sep;39(11):10283-91.
5. Dodds PS, Danforth CM. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. J Happiness Stud. 2009 Jul;11(4):441-56.
6. Paul MJ, Dredze M. You are what you tweet/ : analyzing twitter for public health. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence; 2011 Jul 17-21; Barcelona ES [cited 2015 aug 20];38:265-72. Available from: http:// www. aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880/ 3264
7. Liu B. Sentiment analysis: a multi-faceted problem. IEEE Intell Syst. 2010;25(3):76-80.
8. Chew C, Eysenbach G. Pandemics in the age of twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS ONE. 2010 Nov;5(11):e14118.
9. Yu B. The emotional world of health online communities. Proceedings of the 2011 iConference; 2011 Feb 8-11, Seattle, WA, USA: ACM; 2011 [cited 2012 feb 15]. p. 806-7. Available from: http://doi.acm.org/10.1145/1940761.1940914
10. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS ONE. 2013 May;8(5):e64417.
11. Silva IS, Gomide J, Barbosa GAR, Santos W, Veloso A, Meria W, et al. Observatório da dengue: surveillance based on twitter sentiment stream analysis. [Internet]. Scribd. [cited 2015 jun 3]. Available from: https://www.scribd.com/doc/266238638/Observatorio-da-Dengue-Surveillance-based-on-Twitter-Sentiment-Stream-Analysis
12. Kristensen CH, Gomes CF de A, Justo AR, Vieira K. Brazilian norms for the affective norms for English words. Trends Psychiatry Psychother. 2011 Jan;33(3):135-46.
13. Humphrey SM, Névéol A, Gobeil J, Ruch P, Darmoni SJ, Browne A. Comparing a rule based vs statistical system for automatic categorization of MEDLINE Documents According to Biomedical Specialty. J Am Soc Inf Sci Technol. 2009 Dec;60(12):2530-9.
14. Teixeira F, Falcão AJ, Sousa FS, Hummel AD, Costa TM, Mancini F, et al. Similarity-based scoring method for classification of Health Informatics content. J. Health Inform [Internet]. 2011 Jun [cited 2015 jun 3];3(2):35-42. Available from: http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/137
15. Bangalore AK, Divita G, Humphrey S, Browne A, Thorn KE. Automatic categorization of google search results for medical queries using JDI. In: Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; 2011 Jan; Brisbane, QLD. Amsterdam: IOS Press; 2007. p. 2253-4.
16. List of emoticons [Internet]. Wikipedia, the free encyclopedia. 2015 [cited 2015 Jun 3]. Available from:http://en.wikipedia.org/w/index.php?title=List_of_emoticons&oldid=663735680
17. Liu B. Sentiment analysis and subjectivity. Handbook of natural language processing. 2a. Ed. Chicago:Taylor and Francis Group; 2010.
18. Silva M. Língua portuguesa na internet: o caso das abreviações em salas de bate-papo [Internet]. 2014 Jul [citado 2014 jul 14]. Disponível em: http://www.filologia.org.br/anais/anais%20iv/civ06_6.htm
19. Sebastiani F. Machine learning in automated text categorization. ACM Comput Surv. 2002 Mar;34(1):1-47.
20. Araujo GD, Sousa FS, Teixeira F, Mancini F, Domenico EBLD, Guimarães M de P, et al. Análise de sentimentos sobre temas de saúde em mídia social. J. Health Inform [Internet]. 2012 Sep [citado 2015 jun 8];4(3):95-9. Disponível em: http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/195
21. Instituto Nacional do Câncer. Inca. O que é o câncer? [Internet]. [citado 2014 jul 14]. Disponível em: http://www1.inca.gov.br/conteudo_view.asp?id=322

22. Wikipedia: a encyclopedia livre [Internet]. Câncer (astrologia). 2015 [citado 2015 Jun 8]. Disponível em: http://pt.wikipedia.org/w/index.php?title=C%C3%A2ncer_(astrologia)&oldid=41644871
23. Nishioka C, Scherp A, Dellschaft K. Comparing tweet classification by author's hashtags, machine learning, and human annotators. Proceeding of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT); 2015 Dec 6-9; Singapore; 2015. pp. 67-74.