Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

© 2020 European Federation for Medical Informatics (EFMI) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI200294

Analysing Sentiment and Topics Related to Multiple Sclerosis on Twitter

Guido GIUNTI[a,1], Maëlick CLAES[a], Enrique DORRONZORO ZUBIETE[b], Octavio RIVERA-ROMERO[b], Elia GABARRON[c]
[a] *University of Oulu, Oulu, Finland*
[b] *Universidad de Sevilla, Seville, Spain*
[c] *Norwegian Centre for E-health Research, University Hospital North Norway, Tromsø, Norway*

Abstract. *Background* and objective: Social media could be valuable tools to support people with multiple sclerosis (MS). There is little evidence on the MS-related topics that are discussed on social media, and the sentiment linked to these topics. The objective of this work is to identify the MS-related main topics discussed on Twitter, and the sentiment linked to them. *Methods*: Tweets dealing with MS in the English language were extracted. Latent-Dirilecht Allocation (LDA) was used to identify the main topics discussed in these tweets. Iterative inductive process was used to group the tweets into recurrent topics. The sentiment analysis of these tweets was performed using SentiStrength. *Results:* LDA' identified topics were grouped into 4 categories, tweets dealing with: related chronic conditions; condition burden; disease-modifying drugs; and awareness-raising. Tweets on condition burden and related chronic conditions were the most negative ($p<0.001$). A significant lower positive sentiment was found for both tweets dealing with disease-modifying drugs, condition burden, and related chronic conditions ($p<0.001$). Only tweets on awareness-raising were most positive than the average ($p<0.001$). *Discussion*: The use of both tools to identify the main discussed topics on social media and to analyse the sentiment of these topics, increases the knowledge of the themes that could represent the bigger burden for persons affected with MS. This knowledge can help to improve support and therapeutic approaches addressed to them.

Keywords. Multiple sclerosis, Twitter, natural language processing, sentiment analysis, topic modelling

1. Introduction

Multiple Sclerosis (MS) is the leading cause of non-traumatic neurological disability in young adults [1]. In order to successfully self-manage such a chronic condition, patients require to learn about and manage their symptoms and problems [2-4]. Moreover, they also need social support that enhances the awareness that they are cared for.

Social support is known to be an important aspect for persons with MS (pwMS) [5-7]. However, problems in societal participation, such as reduced interactions with family or friends, are common among people affected by this condition due to physical

[1] Corresponding Author: Guido Giunti E-mail: Guido.Giunti@oulu.fi.

disabilities and MS fatigue [8]. In this sense, the ubiquity of social media and their great potential to engage and communicate with others could make these channels valuable tools to support pwMS [9-10]. PwMS using social media could benefit from communities that share content linked to positive sentiment, as positive emotion motivates cooperation [11]. However, there is not much evidence on what are the topics being discussed on social media related to MS, and if these topics are linked to positive sentiment, and therefore could provide the needed social support.

The objective of this work is to identify the main topics related to multiple sclerosis that are discussed on Twitter, and to analyse the sentiment polarity linked to these topics.

2. Methods

We collected tweets from Twitter using the Twitter API matching the query: #ms OR #multiplesclerosis OR "multiple sclerosis" that were posted between February 9th 2019 and June 26th 2019, and were in English language.

In order to identify the topics discussed in these tweets, we utilized Latent-Dirilecht Allocation (LDA) [12]. LDA is a generative statistical model that assigns to each document (e.g., a single tweet) a probability to belong to each topic of a list of topics which number is fixed and usually small[13]. We pre-processed each tweet with the following steps: 1) All URLs were removed using the following regular expression: https?://[[:alnum:][:punct:]]*; 2) All tweets were tokenized[2]; 3) All tokens were lemmatized using the R packages *textstem*[3] and *lexicon*[4]; and 4) Tokens were filtered to remove stop words using stopword lists from R packages *lexicon* and *tm* [14] (English stopwords).

We built a document-term matrix from the lemmatized and filtered tokens of all tweets collected. Then we trained an LDA model using the 47,852 original English tweets posted between February 9th and May 2nd 2019 and 20 topics. We fine-tuned LDA hyper-parameters using a differential evolution algorithm [15] with R package *DEoptim* [16] with 200 iterations and a population size of 30. For each population member (a pair of values for hyperparameters) considered, an LDA model was generated using a random sample of 80% of the training data. The model was then evaluated computing the perplexity of the other 20% of the training data. At the end, we obtained hyperparameter values 0.292217 and 0.026492 that were used for the final model.

The sentiment analysis of the tweets was performed on the whole dataset (74,076 tweets) using SentiStrength[17]. This software is specially designed to analyse the sentiment in short texts by assigning two scores: the intensity of the positive sentiment (1 to 5) and the intensity of the negative sentiment (-1 to -5). To apply SentiStrength, all tweets were allocated in plain text file, which means that unicode special characters were removed from the original text. As a result of the analysis every tweet was assigned a positive and negative sentiment score.

[2] https://cran.r-project.org/web/packages/text2vec/index.html

[3] https://cran.r-project.org/web/packages/textstem/index.html

[4] https://cran.r-project.org/web/packages/lexicon/index.html

The treatment of personal information for this study was approved by the data-protection officer at the University Hospital of North Norway (Ref:02275).

3. Results

The API extracted a total of 74,076 original tweets (i.e., not retweeted) on MS written in English. The sentiment analysis showed an average positive strength of 1.6 (SD=0.8); and a negative strength of -1.78 (SD=1.0) on the extracted MS tweets.

The 20 topics identified with LDA were analysed and recurring themes were identified and grouped into 4 main topics through an iterative inductive process. These 4 topics dealt with: Related chronic conditions; Condition burden; Disease modifying drugs; and Awareness raising. Table 1 summarizes the main keywords identified with LDA and included in each of the topics.

Table 1. Recurring MS' topics identified with LDA and grouped through iterative inductive process

Topic	Keywords included
Related chronic conditions	Autoimmune; autoimmune_disease; parkinsons; alzheimers; rheumatoid; rheumatoid_arthritis; chronic_illness; crwriter; neurodegenerative; mental; Gofundme; injury; nervous; nervous_system; spinal; central; spinal_cord; central_nervous; cord; brain_injury
Condition burden	Sign; ms_symptom; symptom_ms; quality_life; quality; warn_sign; sign_ms; sign_symptom; early_warn; risk_multiple; Cbd; cancer; marijuana; pain; dying; chronic_pain; bill; spasm; plea; cancer_multiple; Diagnose_multiple; diagnose; letsbeatms; intezaarnakaro_letsbeatms; inzaarnakaro; visit_intezaarnakaro; suffer_multiple; income; disability_progression
Disease modifying drugs	Cell; immune; sclerosis_patient; therapy; stem; trial; stem_cell; immune_system; disease_modify; phase; Drug; relapse; fda; approve; fda_approve; remit; treat_multiple; relapse_remit; oral; novartis
Awareness raising	Goal; goal_multiple; blog; count_donation; sclerosis_count; donation_donation; raise_goal; donation; count; energy; Month; awareness_month; march; ms_awareness; sclerosis_awareness; awareness; awareness_week; hob; ov; march_multiple

When comparing the sentiment polarity of tweets, we find that those with a lower sentiment polarity are the ones with a probability >0.6 to belong to the topics "disease modifying drugs" and "condition burden" (1.40 and 1.44 respectively, vs 1.60 for all other tweets, t-test $p<0.01$). Only tweets on awareness raising had a higher positive strength than the average (1.80 vs 1.60, t-test $p<0.01$). On the other hand, the tweets with a more negative strength were the ones dealing with condition burden and related chronic conditions (-2.53 and -2.14 respectively, vs -1.78, t-test $p<0.001$). Only tweets on awareness raising and disease modifying drugs had lower negative strength than the average (-1.23 and -1.61, vs -1.78, t-test $p<0.05$). Table 2 shows the average sentiment positive and negative sentiment values of the tweets with a probability >0.60 to belong to one of the 4 identified LDA topics, (t-tests compare the sentiment of these tweets with the tweets within the same topic with a probability <0.60).

Table 2. Average sentiment of tweets with a probability >0.60 to belong to each topic

LDA topic	Number of tweets	Positive Mean (SD)	Negative Mean (SD)
Related chronic conditions	1106	1.58 (0.6)*	-2.14 (1.3)*
Condition burden	2256	1.44 (0.6)*	-2.53 (1.3)*
Disease modifying drugs	1749	1.40 (0.6)*	-1.61 (0.7)*
Awareness raising	1171	1.80 (0.7)*	-1.23 (0.6)*
ALL TWEETS	**74076**	**1.60 (0.8)**	**-1.78 (1.0)**

* Students t-test, $p<0.01$

4. Discussion

The MS topics discussed on Twitter that were identified with LDA, and that we grouped into four main categories, are consistent with the topics discussed in a recent publication on MS Facebook groups [18]. The importance of information and awareness, treatment related issues, and other MS-related symptoms was also key in that study [18].

It is well-known that people affected with chronic conditions, such as MS, carry painful emotions in need of sharing [19], and that expressing these emotions is linked to feelings of relief and alleviation [20]. In our study we found that Tweets dealing with disease-modifying drugs, condition burden, and related chronic conditions had a significant lower positive sentiment than the average of Tweets on MS; tweets on condition burden and related chronic conditions were the most negative ones. These could represent the topics with more emotional burden for those affected with MS, and therefore the topics for which they should receive greater social support. This is in line with other literature on reported feelings of pwMS[5].

Strengths and limitations: Our study has several limitations. We extracted tweets using the standard Twitter API, that did not index all tweets published in the considered dates, and therefore, the data extraction might not be representative, and should be considered as a random sample of tweets on MS. Relevant content may not have been included due to our search strategy. Additional research could consider including other keywords related to MS, expand the search by including other languages, include other social media channels, analyse emojis and metadata such as users' profiles. We used SentiStrength software to conduct the sentiment analysis of the collected tweets on MS. This software has not yet been validated for use in the health domain. However, it was developed using posts on Myspace and shows a good performance with informal text, such the ones that can be found on Twitter.

Conclusions: The use of tools to identify main discussed topics on social media, and to analyse the sentiment of these topics increases the knowledge of the themes that could be representing the bigger burden for people with MS. This knowledge can help to improve support and therapeutic approaches addressed to people with multiple sclerosis.

Acknowledgments

EDZ receives funding and is supported by the V Plan Propio de Investigación de la Universidad de Sevilla, Spain.

References

[1] A. Compston, A. Coles, *Multiple sclerosis*. Lancet 2002;359(9313):1221-1231.

[2] Schulman-Green D, Jaser S, Martin F, et al. *Processes of Self-Management in Chronic Illness*. J. Nurs. Scholarsh. 2012;44(2):136-144.

[3] M. Marziniak, G. Brichetto, P. Feys, et al. *The Use of Digital and Remote Communication Technologies as a Tool for Multiple Sclerosis Management: Narrative Review*. JMIR Rehabil Assist Technol 2018;5(1):e5.

[4] D. Kantor, J.R Bright, J. Burtchell. *Perspectives from the Patient and the Healthcare Professional in Multiple Sclerosis: Social Media and Patient Education.* Neurol Ther 2018;7(1), 23–36.

[5] G. Giunti, J. Kool, O. Rivera-Romero, et al. *Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: a mixed-methods study*. JMIR mHealth uHealth 2018;6(2):e37.

[6] D. Kantor, J.R Bright, J. Burtchell. *Perspectives from the Patient and the Healthcare Professional in Multiple Sclerosis: Social Media and Participatory Medicine*. Neurol Ther 2018;7(1), 37–49.

[7] L. Lavorgna, D. Ippolito, S. Esposito, et al. *A disease in the age of the web: How to help people with Multiple Sclerosis in social media interaction*. Mult Scler Relat Disord 2017;17:238-239.

[8] A.L Katz, T.J Braley, E. Foxen-Craft, et al. *How Do Pain, Fatigue, Depressive, and Cognitive Symptoms Relate to Well-Being and Social and Physical Functioning in the Daily Lives*. Arch Phys Med Rehabil 2017;98(11):2160-2166.

[9] A. Arguel, O. Perez-Concha, L. SYW, et al. *Theoretical approaches of online social network interventions and implications for behavioral change: a systematic review*. J Eval Clin Pract 2018;24(1):212-221.

[10] M.S.C Lim, C.J.C Wright, E.R Carrotte, et al. *Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites*. Health Promot J Austr 2016;27(3):187-197.

[11] D.G Rand, G. Kraft-Todd, J. Gruber. *The Collective Benefits of Feeling Good and Letting Go: Positive Emotion and (dis)Inhibition Interact to Predict Cooperative Behavior*. PLoS One 2015;10(1):e0117426.

[12] D.M Blei, A.Y Ng, M.I Jordan. *Latent dirichlet allocation.* Journal of machine Learning research 2003;3(Jan):993-1022.

[13] C. Montenegro, C. Ligutom, III, J. Vincent Orio, and D.A.M Ramacho. 2018. Using Latent Dirichlet Allocation for Topic Modeling and Document Clustering of Dumaguete City Twitter Dataset. In Proceedings of the 2018 International Conference on Computing and Data Engineering (ICCDE 2018). ACM, New York, NY, USA, 1-5. DOI: https://doi.org/10.1145/3219788.3219799

[14] I. Feinerer, K. Hornik, D. Meyer. *Text Mining Infrastructure in R.* Journal of Statistical Software 2008;25(5):1-54.

[15] K.V Price, R.M Storn, J.A Lampinen (2006). Differential Evolution - A Practical Approach to Global Optimization. Berlin Heidelberg: Springer-Verlag.

[16] K. Mullen, D. Ardia, D. Gil, et al., 'DEoptim': An R Package for Global Optimization by Differential Evolution. Journal of Statistical Software 2011;40(6):1-26.

[17] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas. Sentiment strength detection in short informal text. J Am Soc Inf Sci 2010 Dec 15;61(12):2544-2558. [doi: 10.1002/asi.21416]

[18] S. D Rosa, F Sen. *Health Topics on Facebook Groups: Content Analysis of Posts in Multiple Sclerosis Communities.* Interact J Med Res 2019;8(1):e10146.

[19] B. Rimé. *Comment: Social Integration and Health: Contributions of the Social Sharing of Emotion at the Individual, the Interpersonal, and the Collective Level*. Emotion Review 2018;10(1):67-70.

[20] J.A Holyst (Ed.) Cyberemotions: Collective emotions in cyberspace. Springer 2016.