







Chapter 16

A Sentiment Analysis Method for Analyzing Users Opinions About Drugs for Chronic Diseases



María del Pilar Salas-Zárate , Giner Alor-Hernández ,
Jorge Luis García-Alcaraz , Luis Omar Colombo-Mendoza ,
Mario Andrés Paredes-Valverde , and José Luis Sánchez-Cervantes 

Contents

16.1 Introduction	217
16.2 Related Work	219
16.3 Sentiment Analysis Method for Drug Opinions Analysis	220
16.4 Evaluation	222
16.5 Conclusions and Future Work	226
References	227

16.1 Introduction

The information on the Web is growing dramatically, which is leading to a series of challenges that range from security and privacy issues to technical and technological issues. At the same time, it is turning the Web into a vast source of information

M. del Pilar Salas-Zárate (✉) · G. Alor-Hernández · L. O. Colombo-Mendoza
M. A. Paredes-Valverde
Tecnológico Nacional de México/I. T. Orizaba, Orizaba, Veracruz, México
e-mail: msalasz@ito-depi.edu.mx; galar@ito-depi.edu.mx; lcolombom@ito-depi.edu.mx;
mparedesv@ito-depi.edu.mx

J. L. García-Alcaraz
Department of Industrial Engineering, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez,
Chihuahua, Mexico
e-mail: jorge.garcia@uacj.mx

J. L. Sánchez-Cervantes
CONACYT-Tecnológico Nacional de México/I. T. Orizaba, Orizaba, Veracruz, México
e-mail: jlsanchez@conacyt.mx

that can be exploited to derive actionable knowledge about virtually anything. In this context, as the years go by, more and more people use the Web, e.g., blogs, forums, wikis, social networks, and review websites, as the primary medium for answering their queries and making decisions in multiple domains, not being the exception the healthcare context. Considering that chronic diseases are prevalent medical conditions in low and middle,¹ it makes sense to think that most of the Web queries about medical conditions are related to chronic diseases. In fact, according to the National Center for Chronic Disease Prevention and Health Promotion of the USA, 6 in 10 adults in that country have one chronic disease.² Similarly, according to the National Institute of Geography, Statistics, and Informatics of Mexico, by 2015, diabetes was the leading cause of death among Mexican adults above 60 years.³ Hence the importance of developing computational methods and tools to automatically derive actionable knowledge about chronic diseases that is valuable not only for patients but also for doctors, health professionals, and health authorities from Web resources.

Sentiment analysis is an on-going and fast-growing subfield of computer science that analyzes users' opinions about different topics such as individuals, organizations, products, services, events, as well as their attributes [1]. It has been successfully applied to analyze user-generated content on the Web in several different domains. Sentiment analysis can benefit the healthcare domain for developing multiple methods and tools such as methods for detecting adverse drug reactions mentions [2, 3], opinion summarization and extraction systems for drugs and doctors [4, 5], and recommender systems for health self-management [6, 7].

Several authors have introduced methods to analyze and process opinions from social media data. Most of these efforts are based on machine learning (ML) and semantic orientation (SO) approaches. The ML approach requires a set of data to train an algorithm and build a predictive model; and a set of data for evaluate the built model. On the other hand, the SO approach is based on lexicons (SentiWordNet, ML-Senticon, WordNet-Affect, etc.). However, there are several linguistic features expressed in opinions that they are not considered; for example, features related to cultural, social, psychologic aspects can have an important impact on sentiment analysis.

This work proposes a sentiment analysis method, which combines machine learning with psycholinguistic features to analyze opinions of users about drugs and determine their polarity. For this purpose, this method uses the LIWC tool [8], which already has been used to analyze the emotional well-being of people by analyzing their social media posts [9, 10] and to identify language features that distinguish demographic and psychological attributes from language in blogs [11]. An experiment was specifically conducted over a corpus of Mexican patient

¹https://www.who.int/nutrition/topics/2_background/en/.

²<https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm>.

³https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2017/muertos2017_Nal.pdf.

opinions about drugs for diabetes and hypertension that we compiled through an ad-hoc drug information website.

This work is structured as follows: Sect. 16.2 presents the relevant literature on sentiment analysis in the health self-care domain. Then, Sect. 16.3 describes the sentiment analysis method, whereas Sect. 16.4 describes a case study that shows its application for analyzing opinions about drugs. Finally, we present our conclusions and future directions in Sect. 16.5.

16.2 Related Work

Diverse sentiment analysis approaches in the form of methods, tools, and platforms, ranging from machine learning-based approaches to semantic orientation-based approaches, have been reported in the literature recently. Machine learning-based approaches mostly employ supervised machine learning algorithms; these algorithms initially fit a model from the features of the documents in a corpus using a training dataset and require a second dataset called test dataset to evaluate a final model built from the training dataset. In regard to semantic orientation-based approaches, sentiment lexicons such as SentiWordNet, iSOL, and eSOL [12] are employed. A basic sentiment analysis task is sentiment polarity detection, often called sentiment polarity classification, which involves identifying the polarity of an opinion expressed at the level of a document, sentence, entity, or aspect using a scale consisting of three degrees: positive, negative, or neutral [13].

With the aim of identifying the opportunity niches of our research, we analyzed and contrasted some works on sentiment analysis, specifically, on sentiment polarity detection, in the domain of chronic diseases and associated drugs.

Jiménez-Zafra et al. [14] studied how people express their opinion about drugs and doctors in medical forums in Spanish. They used both a machine learning approach and a semantic orientation-based approach to polarity detection over two different corpora extracted from the forums [mimedicamento.es](#) and [masquemedicos.com](#) for that purpose. In particular, the support vector machines algorithm and the iSOL sentiment lexicon were, respectively, used as the foundation of these approaches. A probabilistic aspect mining approach for drug reviews is presented in [15]. This approach differs from most approaches to aspect level-based opinion mining in the sense that it is not aimed at extracting the aspects and their sentiments from opinions but at identifying the aspects related to class labels or categorical meta-information in opinions. In fact, it is more related to topic modeling than to aspect level-based opinion mining. Gopalakrishnan et al. [16] studied patient satisfaction with drugs using online drug reviews. A machine learning-based approach to polarity detection that relies on artificial neural networks was proposed for that purpose. According to the results of an experiment performed on a corpus of drugs opinions obtained from the [askapatient.com](#) website, this approach performs better than other related approaches when the Radial Basis Function Neural Network model is specifically used. A study of the impact of

sentiment analysis features in detecting adverse drug reaction (ADR) mentions in tweets and forum posts was presented in [3]. In this study, a semantic orientation-based approach to polarity detection was integrated into an algorithm for ADR mentions extraction, namely a supervised conditional random field classifier for sequence labeling that is called ADRMine. A corpus of tweets and posts from support groups or forums of the DailyStrength social network was used by the authors for validation. Biyani et al. [17] analyzed the sentiment of users posts about cancer in English. In particular, the co-training machine learning algorithm was used to conduct polarity detection as a classification task based on both domain-dependent sentiment features and domain-independent sentiment features. A corpus of posts from the Cancer Survivors' Social Network (CSN) of the American Cancer Society was used by the authors for validation. Similarly, in [18] a sentiment analysis method and a tool to detect the polarity of the posts written by cancer patients in Brazilian online cancer communities were presented. This method corresponds to a semantic orientation-based approach and uses the dictionary of Portuguese terms used by the SentiStrength sentiment analysis tool. A series of experiments were conducted by the authors for validation purposes using a corpus of posts collected from the Facebook social network. The authors also proposed in a previous work [12] an aspect-level method for sentiment analysis of tweets about diabetes written in English. In particular, this method performs polarity detection at the aspect level by (1) semantically annotating tweets to identify aspects using an ontology, namely the diabetes diagnosis ontology and (2) detecting the polarity of each of the identified aspects using the SentiWordNet lexicon.

Based on the related work discussion above presented, next conclusions can be derived: (1) there is a lack of works on sentiment analysis methods for analyzing patient opinions about chronic disease medicines written in Spanish; (2) most of the works proposed in this domain that use hybrid approaches to polarity detection do not study the effect of psychological features but only the effect of linguistic features; and (3) no evaluation or validation on real-world datasets of Mexican patient opinions about drugs for diabetes and hypertension is reported in the literature on sentiment analysis. We intend to address these issues as will be seen later in this paper.

16.3 Sentiment Analysis Method for Drug Opinions Analysis

The sentiment analysis method proposed in this work relies on machine learning and psycholinguistic analysis technologies to analyze the user's opinions on drugs. As depicted in Fig. 16.1, before any analysis can proceed, a corpus of opinions is necessary. Then, corpus processing, psycholinguistic features extraction, and learning process are performed. Next sections thoroughly describe these phases.

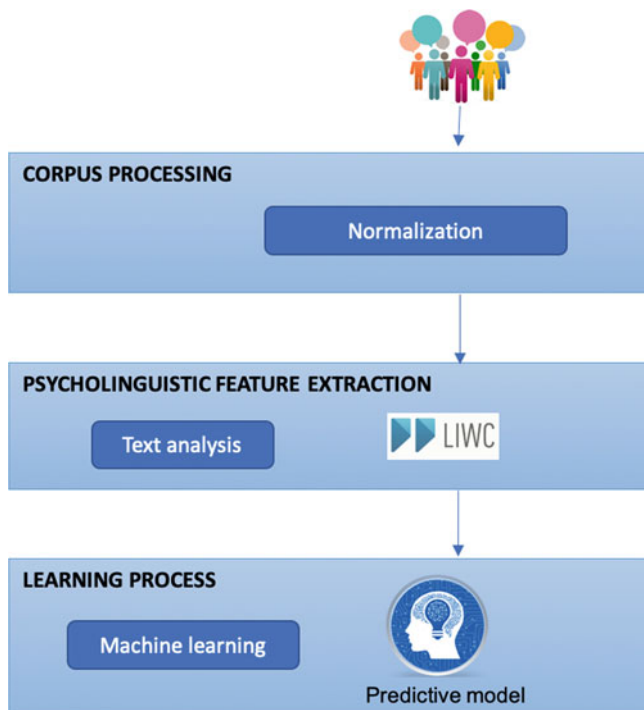


Fig. 16.1 Sentiment analysis method

Corpus Processing

Corpus processing. In the health self-care domain, most opinions about drugs are written by non-expert people. Therefore, the language used in these opinions is characterized by informal writing with spelling errors. Aiming to make the corpus collected more useful for linguistic research purposes, it is subjected to a process where spelling errors are corrected. Hunspell tool is used for finding spelling errors within an opinion. In this way, words such as “daibetes” were corrected to “diabetes” (diabetes). Hunspell was selected because it has been successfully applied for text preprocessing purposes in the sentiment analysis domain [19–21].

Psycholinguistic Feature Extraction

In this paper, we used the LIWC tool to obtain features such as psychologic linguistic of the text. LIWC carries out an analysis based on five main macro-categories (linguistic, psychological, personal concerns, spoken, and punctuation marks).

LIWC has been used in sentiment analysis only as a lexicon based on semantic orientation approach, which is constructed from two subcategories (positive emotions and negative emotions) of the 72 subcategories of LIWC [22] and [23]. However, only a few works have studied the importance of this type of features in the sentiment analysis.

Learning Process

In this phase, machine learning algorithms are trained. To this aim, WEKA tool [24] was used. WEKA provides several classification algorithms that allow the generation of models depending on the data and objective. The classification algorithms are categorized into seven groups: (1) Bayesian algorithms: BayesNet, Naive Bayes, etc.; (2) functions algorithms: logistic, linear regression, SMO, etc.; (3) lazy algorithms: LWL, IBk, etc.; (4) meta classifiers: Vote, Bagging, etc.; (5) miscellaneous: InputMappedClassifier, SerializedClassifier, etc.; (6) rules algorithms OneR, DecisionTable, etc.; (7) trees algorithms: RandomTree, J48, etc. Specifically, we select two algorithms for the present study, BayesNet and SMO.

16.4 Evaluation


In this section, the evaluation process is described. The objective is to measure the effectiveness of the proposed sentiment analysis method regarding users' opinions classification in the context of drugs for diabetes and hypertension. The below sections describe the people involved and corpus used in this work and discuss the obtained results.

Subjects

For the purposes of this case study, 20 people with diabetes and/or hypertension were asked to interact with the Web application aiming to collect opinions about drugs they use for managing the aforementioned diseases (see Fig. 16.2). All participants were native Spanish speakers, ranged from 35 to 60 years old. Furthermore, these people have lived with diabetes and/or hypertension for many years which make them suitable for this work because of their experience with the use of drugs for managing these chronic diseases.

MEDICAMENTOS
APP MOVIL
FARMACIAS
CONTACTO
PERFIL
CERRAR SESION

METFORMINA



PROMEDIO

★ 1.0

CALIFICA ESTE MEDICAMENTO

☆☆☆☆☆

👁️ 1 💬 0

FAVORITO

📌

DESCRIPCIÓN DEL MEDICAMENTO:

La metformina se utiliza sola o con otros medicamentos, incluyendo insulina, para tratar diabetes tipo 2 (condición en la que el cuerpo no usa la insulina normalmente y, por lo tanto, no puede controlar la cantidad de azúcar en la sangre). La metformina es una clase de medicamentos llamados biguanidas

MECANISMOS DE ACCION

Reduce la glucosa en plasma postprandial y basal. Actúa por 3 mecanismos. 1: Reduce la producción hepática de glucosa por inhibición de gluconeogénesis y glucogenolisis. 2: En el músculo incrementa la sensibilidad a insulina y mejora de captación de glucosa periférica y su utilización. 3: Retrasa la absorción intestinal de glucosa. No estimula la secreción de insulina por lo que no provoca hipoglucemia.

INDICACIONES

Diabetes mellitus tipo 2 en especial en pacientes con sobrepeso, cuando no logran control glucémico adecuado solo con dieta y ejercicio.

Agregar comentario

Agregar comentario

Fig. 16.2 Web interface

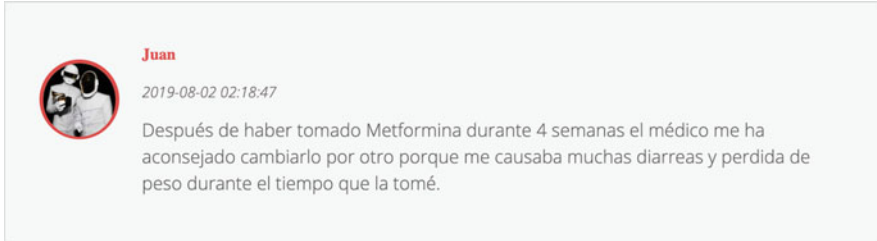


Fig. 16.3 Example of user's opinion about drugs

Corpus

The case study described in this work was performed using a corpus of reviews written in Spanish that describe users view or judgment about drugs for diabetes and hypertension. This corpus was collected through a web platform of chronic diseases (see Fig. 16.2). It must be mentioned that a manual review of all collected opinions was carried out aiming to discard reviews that seem out of the scope of this work or do not reflect an opinion about a drug. In this way, the resulting corpus consists of 740 opinions about 24 drugs. Figure 16.3 depicts an example of a user's opinion about "metformin" drug, which is used for the treatment of Diabetes type 2 in adults that helps control blood sugar levels.

Results

In the present work, recall (R), precision (P), and F-measure (F1) metrics were selected to measure the performance of the proposed method. Recall (see Eq. (16.1)) represents the number of correctly classified opinions divided by the number of data that should be identified as positive, negative, or neutral; Precision (see Eq. (16.2)) represents the number of correctly classified opinions divided by the number of all opinions returned by the classifier as correctly classified; and the F-measure (see Eq. (16.3)) is the harmonic mean of precision and recall.

$$R = \frac{TP}{TP + FN} \quad (16.1)$$

$$P = \frac{TP}{TP + FP} \quad (16.2)$$

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16.3)$$

where

Table 16.1 Evaluation results with SMO algorithm

LIWC categories	2 classes			3 classes		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
Linguistic process	0.713	0.712	0.712	0.657	0.657	0.657
Psychological process	0.736	0.736	0.736	0.649	0.768	0.704
Personal concerns	0.581	0.566	0.573	0.556	0.527	0.541
Spoken categories	0.528	0.513	0.520	0.514	0.514	0.514
Punctuation marks	0.633	0.595	0.613	0.546	0.604	0.574
Combination of all categories	0.794	0.793	0.793	0.754	0.81	0.781

Table 16.2 Evaluation results with BayesNet algorithm

LIWC categories	2 classes			3 classes		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>R</i>	<i>F1</i>
Linguistic process	0.698	0.698	0.698	0.653	0.715	0.683
Psychological process	0.708	0.708	0.708	0.675	0.738	0.705
Personal concerns	0.593	0.591	0.592	0.552	0.552	0.552
Spoken categories	0.553	0.553	0.553	0.52	0.52	0.520
Punctuation marks	0.588	0.569	0.578	0.502	0.578	0.537
Combination of all categories	0.726	0.725	0.725	0.697	0.738	0.717

True Positive (TP) are those positive, negative, or neutral opinions that the system classified as such; False Negatives (FN) are those positive, negative, or neutral opinions that the classifier not detected as such; and False Positives (FP) are the opinions that the system classified into another different class than the one it belongs to.

Tables 16.1 and 16.2 present the results obtained with SMO and BayesNet algorithms for the sentiment analysis based on psycholinguistic features.

As can be observed, the general results show that the best features for classification of drug opinions are psychological and linguistic features. On the other hand, the features that do not contribute to better classification are personal concerns and spoken features. We ascribe this to two reasons: (1) opinions contain a great number of grammatical words that are part of the linguistic category; (2) opinions contain words related to the feelings of the users. It is important to mention that people with a chronic illness such as diabetes usually suffer from a mental illness such as anxiety or depression, which can be reflected in their opinions.

On the other hand, the best results are obtained with the combination of all LIWC categories with an F-measure of 79.3%.

The results also show that SMO algorithm is more accurate for this classification problem than BayesNet. This result can be due to the study presented in [25], where authors conclude that SMO is more robust and accurate than other algorithms. Finally, the classification with 2 classes (positive opinions and negative opinions) provides better results than the classification with three classes (positive opinions, negative opinions, and neutral opinions). Therefore, the results demonstrate that with fewer categories the classification algorithm performs better.

16.5 Conclusions and Future Work

In the health self-care domain, most opinions about drugs are written on the Web by consumers themselves, i.e., by people who are not necessarily medical experts. Therefore, informal media such as social networks and user review websites become prominent sources of consumer opinions about drugs. Presumably, one of the hottest topics in this context are drugs for chronic diseases.

Sentiment analysis or opinion mining can be exploited to automatically derive actionable knowledge about drugs for chronic diseases from consumer opinions on the Web that is valuable not only for the patients themselves but also for doctors, health professionals, and health authorities.

Chronic diseases are among the most prevalent medical conditions around the world, and the Americas is not the exception. Nonetheless, there is a lack of works on sentiment analysis methods for analyzing patient opinions about chronic disease medicines written in Spanish. In this piece of research, we intended to address this issue by proposing a hybrid sentiment analysis method that combines machine learning features with psycholinguistic features to determine the polarity of patient opinions about drugs written in Spanish.

We validated our method using a corpus of Mexican patient opinions about drugs for diabetes and hypertension that was compiled through a drug information website that we developed for the purposes of this study. We obtained hopeful results with a F1-measure score of 0.793, which support our assumption that using a combination of linguistic and psychological features yields better results than using linguistic features solely. In fact, the combination of all the five categories of LIWC, namely linguistic, psychological, and personal concerns, spoken categories, and punctuation marks yielded better results relation to F1-measure values in comparison with each one of the individual categories.

Furthermore, regarding the number of polarity classes employed in our experiment, the use of two classes, namely positive and negative, yielded better results. This can be interpreted as evidence that opinions about chronic diseases drugs on the Web are highly polarized among Mexican patients.

As regards future work, we have planned to integrate new techniques to improve the sentiment analysis method. To this end, we will analyze the possibility to use an ontology that represents semantic knowledge about chronic diseases and related drugs to identify aspects in patient opinions and conduct sentiment analysis at the aspect level. In this context, calculating the polarity of consumer opinions about drugs at the sentence level does not necessarily results in identifying what a consumer likes or dislikes about a drug, because the same sentence can actually contain multiple opinions about multiple features of one single drug, e.g., cost and side effects. We will also study the use of topic models, e.g., Latent Dirichlet Allocation, to conduct the aspect mining process as a topic (or rather, sub-topic) discovery process.

Likewise, considering that we can find consumer opinions about drugs written in Spanish on multiple sources on the Web, we plan to integrate new data sources,

specifically, social networks, e.g., Twitter and Facebook, to validate our proposal using diverse opinion corpora.

Finally, the proposed sentiment analysis method may be integrated into complex applications or systems that can be especially valuable in the healthcare domain, specifically, for health self-management purposes; recommender systems and question answering systems are great examples of these complex systems.

Acknowledgments This work was supported in part by the National Council of Science and Technology of Mexico (CONACYT) and the Secretariat of Public Education of Mexico (SEP) through the Professional Development Program for University Professors (PRODEP). Additionally, Tecnológico Nacional de Mexico (TecNM) supported this research.

References

1. B. Liu, Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **5**, 1–167 (2012)
2. A. Sarker, G. Gonzalez, Portable automatic text classification for adverse drug reaction detection via multi-corpus training. *J. Biomed. Inform.* **53**, 196–207 (2015)
3. I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, G.H. Gonzalez, Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *J. Biomed. Inform.* **62**, 148–158 (2016)
4. R. Rivas, N. Montazeri, N.X. Le, V. Hristidis, Automatic classification of online doctor reviews: Evaluation of text classifier algorithms. *J. Med. Internet Res.* **20**, e11141 (2018)
5. D. Cavalcanti, R. Prudêncio, Aspect-Based Opinion Mining in Drug Reviews, in ed. by E. Oliveira, J. Gama, Z. Vale, H. Lopes Cardoso, *Progress in Artificial Intelligence, EPIA 2017, Lecture Notes in Computer Science*, vol 10423 (Springer, Cham, 2017)
6. R.L. Rosa, G.M. Schwartz, W.V. Ruggiero, D.Z. Rodriguez, A knowledge-based recommendation system that includes sentiment analysis and deep learning. *IEEE Trans. Ind. Inf.* **15**, 2124–2135 (2019)
7. D. Yang, C. Huang, M. Wang, A social recommender system by combining social network and sentiment similarity: A case study of healthcare. *J. Inf. Sci.* **43**, 635–648 (2017)
8. J.W. Pennebaker, M.E. Francis, R.J. Booth, Linguistic inquiry and word count: LIWC 2001. Mahw. Lawrence Erlbaum Assoc. **71**, 2001 (2001)
9. L. Zhang, X. Huang, T. Liu, Z. Chen, T. Zhu, Using Linguistic Features to Estimate Suicide Probability of Chinese Microblog Users, in ed. by Q. Zu, B. Hu, N. Gu, S. Seng, *Human Centered Computing, HCC 2014, Lecture Notes in Computer Science*, vol 8944 (Springer, Cham, 2014)
10. M. Settanni, D. Marengo, Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. *Front. Psychol.* **6**, 1045 (2015)
11. T. Yarkoni, Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *J. Res. Pers.* **44**, 363–373 (2010)
12. M. Del Pilar Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M.Á. Rodríguez-García, R. Valencia-García, Sentiment analysis on tweets about diabetes: An aspect-level approach. *Comput. Math. Methods Med* **2017**, 1–9 (2017)
13. E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis. *IEEE Intell. Syst.* **28**, 15–21 (2013)
14. S.M. Jiménez-Zafra, M.T. Martín-Valdivia, M.D. Molina-González, L.A. Ureña-López, How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. *Artif. Intell. Med.* **93**, 50–57 (2019)

15. V.C. Cheng, C.H.C. Leung, J. Liu, A. Milani, Probabilistic aspect mining model for drug reviews. *IEEE Trans. Knowl. Data Eng.* **26**, 2002–2013 (2014)
16. V. Gopalakrishnan, C. Ramaswamy, Patient opinion mining to analyze drugs satisfaction using supervised learning. *J. Appl. Res. Technol.* **15**, 311–319 (2017)
17. P. Biyani, C. Caragea, P. Mitra, C. Zhou, J. Yen, G.E. Greer, K. Portier, Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community, in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM'13* (ACM Press, New York, 2013), pp. 413–417
18. R.G. Rodrigues, R.M. Das Dores, C.G. Camilo-Junior, T.C. Rosa, SentiHealth-cancer: A sentiment analysis tool to help detecting mood of patients in online social networks. *Int. J. Med. Inform.* **85**, 80–95 (2016)
19. J.M. Coteló, F.L. Cruz, J.A. Troyano, F.J. Ortega, A modular approach for lexical normalization applied to Spanish tweets. *Expert Syst. Appl.* **42**, 4743–4754 (2015)
20. F.H. Khan, S. Bashir, U. Qamar, TOM: Twitter opinion mining framework using hybrid classification scheme. *Decis. Support Syst.* **57**, 245–257 (2014)
21. M.T. Wiley, C. Jin, V. Hristidis, K.M. Esterling, Pharmaceutical drugs chatter on Online Social Networks. *J. Biomed. Inform.* **49**, 245–254 (2014)
22. P. Gonçalves, M. Araújo, F. Benevenuto, M. Cha, Comparing and combining sentiment analysis methods, in *Proceedings of the First ACM Conference on Online Social Networks - COSN'13* (ACM Press, New York, 2013), pp. 27–38
23. C.J. Hutto, E. Gilbert, VADER: A parsimonious rule-based model for sentiment analysis of social media text, in *Eighth International AAAI Conference on Weblogs and Social Media* (Georgia Institute of Technology, Atlanta, GA, 2014)
24. R.R. Bouckaert, E. Frank, M.A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, *WEKA—Experiences with a Java Open-Source Project* (MIT Press, Cambridge, 2001)
25. H. Bhavsar, A. Ganatra, A comparative study of training algorithms for supervised machine learning. *Int. J. Soft Comput. Eng.* **2**, 2231–2307 (2012)