ANÁLISE DE SENTIMENTO PARA AVALIAR A ADERÊNCIA: UMA REVISÃO SISTEMÁTICA

SENTIMENT ANALYSIS TO ASSESS THE PATIENT'S ADHERENCE: A SYSTEMATIC REVIEW

EL ANÁLISIS DE SENTIMIENTOS PARA EVALUAR LA ADHERENCIA DEL PACIENTE: UNA REVISIÓN SISTEMÁTICA

Gabriel do Nascimento Ribeiro[1], Marcia Ito[2]

Master of Computing Engineering , Instituto de Pesquisas Tecnológica – IPT, São Paulo, SP, Brasil.

Research Staff Member, IBM Research Brasil – IBM-BRL, São Paulo, SP, Brasil

Abstract: Objective To conduct a systematic review of the use in sentiment analysis on social media to identify or assess patient's treatment adherence, and evaluate its application, benefits and future research. **Methods** A systematic review of the literature was carried out by identifying published articles on the main databases of computing and healthcare. Search strings were built by combining keywords related to adherence, social media, data analysis and sentiment analysis. **Results** From a total of 709 articles screened, it wasn't possible to identify any study related to the objective. However, we could select 15 which presented some similarity degree and yet very heterogeneous, they were analyzed according to six dimensions: Adherence, Data Source, Psychology, Methods, Tools and Sentiment Analysis. **Conclusions** A strong agreement and trend can be observed on the potential use and importance of automatic techniques to collect and analyzed online patient data, especially related to assessment of adherence with sentiment analysis.

Descriptors: adherence; sentiment analysis; systematic review.

Introduction

Patient's treatment adherence continues to be a common problem in the healthcare today, leading to high rates of deaths and costs to providers. Despite decades of studies, the literature is still growing every year with many papers published trying to understand why a patient abandon its treatment, what can be done to prevent and assess possible non-adherence behavior and how to intervene when necessary.[1,2] The strategies to address adherence remained the same until now. Theories and models from the health psychology fields were developed between 50's and 80's to explain human behavior,[3] but today people lives and interacts in a new way. The assessments based on manual methods, such as interviews, questionnaires and surveys now can rely on new technologies. These changes are a unique opportunity to reformulate the problem and obtain better outcomes. Since adherence is strictly related to human behavior, and behavior, in turn, is directly reflected on how people uses online social medias,[4] a huge source of raw data is ready to be collected and analyzed, seeking for traces of personality, sentiments and opinions. Many of the innovative studies of adherences are based on sensors or mobile.[6,7] These approaches can automatically collect data from patients in their daily livings, monitor objective and subjective[8] variables and help to identify non-adherence behavior. Despite initial studies shows good results, problems related to costs, accuracy, ethic and privacy still need to be resolved.

Artificial intelligence can help in this task with sentiment analysis, a technique which can extract subjective information from textual data by utilizing text mining, machine learning and natural language processing. Analysis of social media and sentiment analysis have been studied for different purposes. So the goal of this work is to conduct a systematic review of the use in sentiment analysis on social media to identify or assess patient's treatment adherence, and evaluate its application, benefits and future research.

Method

This systematic review followed the methods outlined in Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA)[9]. The search strategy version of this systematic review was generated by combining keywords for patient's treatment adherence and for sentiment analysis. The search string *("patient adherence" OR "treatment adherence") AND "sentiment analysis"* and others variations to include plurals and synonyms were performed in all selected data sources, and brought zero results in all of them. This naturally led to a questioning about: (1) effectiveness of the search strategy used: are the keywords not suitable to gather all the published related literature? And (2) feasibility of the objective of this systematic review: why there are not yet published material related to sentiment analysis and adherence? Is this research field not reasonable, and existing studies have failed? Or are the research community not aware of its potential?

Although such results initially suggests, probably as a misconception, that studies in the area doesn't exist, the authors believe that articles with any degree of similarity were already published in the literature, motivating this systematic review. Since additional efforts and deeper investigation are necessary to confirm this hypothesis, the search definition had to be reviewed, expanding the scope and selection criteria and aggregating similar fields and keywords, in the attempt to collect as many related paper as possible (and probably with lower precision and recall). The final search definition of this systematic review was elaborated by dividing the objective into four main areas: (1) patient's treatment adherence, (2) social media, source of most online patient's data analysis and sentiment analysis applications, (3) data analysis, because many researchers have a similar approach to sentiment analysis without explicitly using this term, and (4) sentiment analysis. The four main areas described earlier were used to generated several keyword (see Table 1), which then were combined into logical expressions with "OR" and "AND" operators to build different search strings. This approach, with several combinations of the areas, its keywords and search strings, allows the researcher to manually control the search granularity, adjusting it dynamically according to the results obtained and the scope, precision and recall desired. A total of 24 different search strings combining all the areas and keywords were built for this systematic review, and will not be described here due to space limitations.

Table 1 The four search definition areas and keywords

Area	Keyword
(1) patient's treatment adherence	"patient"
	"treatment"
	"adherence"
	"patient adherence"
	"treatment adherence"
(2) social medias	"social media"
	"social network"
(3) data analysis	"data"
	"analysis"
	"data analysis"
	"mining"
	"data mining"
	"text mining"
	"opinion mining"
(4) sentiment analysis	"sentiment analysis"

The searches were performed manually between April and May 2015 on the main articles databases for computing and healthcare, including ACM, IEEE, PubMed, PMC and ScienceDirect. Each article obtained were evaluated by its title, abstract and superficial content analysis in case of any doubt. To be selected in this systematic review, articles have ideally to apply sentiment analysis in social media to identify or assess adherences. Since the scope had to be expanded to aggregate many similar studies as possible, we changed the selection criteria to select articles that belong to at least two of the four search definition areas described above. For example, articles which apply sentiment analysis in patient data but not related to social media and articles which used any other technique on online patient data in social media were selected to this systematic review. There is no publication date constraint, articles not in English were excluded and grey literature was not considered. The metadata extraction consisting of titles, abstracts, sources, publication dates and URLs was carried out automatically by using a Google Chrome's extension called Web Scraper.

Papers were selected by the selection criteria were full reviewed and analyzed. A standardized form were used to extract the relevant data, including reviewer's commentaries, category, tags, similarity degree and six dimensions: Adherence, Data Sources, Psychology, Methods, Tools and Sentiment Analysis. Due to the results of the first search definition in which the union of areas (1) and (4) brought zero results, the authors thought it may be interesting to analyze each area separately and then its unions. This task was carried out by performing the search strings on the selected data sources and collecting the results and publication dates. All the values were normalized according to the following formula:

$$normalized\ data = \frac{data - minimum}{maximun - minimum}$$

In the Figure 1, it's possible to analyze the rapid and constant growing of each area separately in the last 15 years, showing clearly that the research interest continues to increase over the years.

Some areas are a recent phenomena, like (2)/(3) data analysis in social media, due to Web 2.0 and (4) sentiment analysis, with the first study published in 2003 and less than 1000 articles published until 2010.

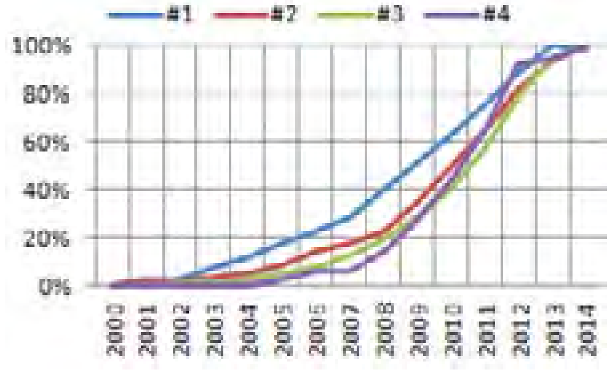

Figure 1 Distribution of articles by area from 2000 to 2014.

Results and Discussion

As a result of all searches performed, 709 articles were found and screened by its title, abstract and a superficial content analysis in case of doubt. According to the selection criteria, 672 articles were excluded and 37 were left to full text review, which revealed 4 duplicated articles, 12 articles which didn't met the selection criteria, 4 articles with access restriction and a total of 15 selected studies. A summary of the study selection flow can be observed in Figure 3:

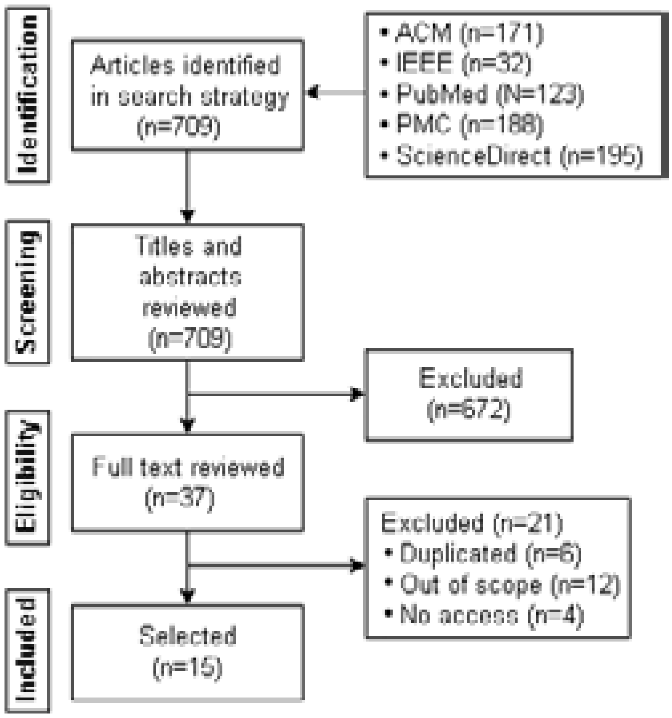

Figure 3 Study selection flow diagram.

The Figure 4 shows the trend of study selection publication date, with a high number of studies published in the last few years and growing exponentially.

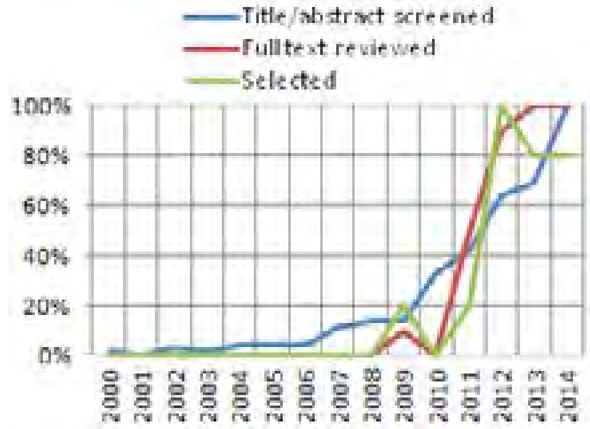

Figure 4 Distribution of study selection publication date

The full text review of the 15 articles selected showed a high heterogeneous content, hard to compare and analyze with traditional models and even to classify into categories One suitable approach found was to analyze each article in different dimensions, as follows: (1) Adherence, (2), Data Sources, where the analyzed data comes from, (3) Psychology, because many adherence studies are based on models and frameworks from psychology, (4) Methods, the manual or automatic technique used, , (5) Tools, the software or strategy used to carry out the method and (6) Sentiment Analysis. A summary is presented in Table 2.

Adherence and Data Source

We could observe that the strategy used to identify or assess adherence is strictly related to where the patient data comes from. Thus, these dimensions were merged and classified into 5 subgroups: (1) Patient experience, (2) Health Big Data, (3) Information health in Social Medias, (4) Internet heath information influence and (5) Doctor-patient interaction.

The (1) Patient experience subgroup contains all studies related to online commentaries, complains or opinions from patients about the treatment received from hospitals, including its infrastructure, medical staff or any other factor that matters to the patient perspective. The aim of the studies are to help in hospital management by automatically analyzing manual surveys and identifying common problems and opportunities. Alemi *et al*[11] and Greaves *et al*[12] used automatic techniques to identify the content and sentiment of online messages from patients about its experience in hospitals, collecting data from a doctor's rates and reviews website called Rate My MD (http://ratemymd.com) and data from the English National Health Service, respectively.

Subgroup (2) Health Big Data proposes a conceptual model of Cloud Computing and a Big Data for Health, containing big volume of patient heterogeneous data from different data sources, such as mobile, sensors, devices and internet. All this information then can be analyzed with any technique, algorithm or statistical model to gain insights about patient's life, behavior and adherence. Hsueh, Ramakrishnan and Chang[13]leading to wastes of resources and increasing patient churn rates. In current practice, mitigating the risk of non-adherence cases is a labor-intensive task that requires additional efforts from healthcare professionals to handle on a case-by-case basis. Previous work has investigated into the possibility of modeling patient adherence behavior, but left questions about the accountability of such models in services. With the prevalence of mobile devices and maturing cloud-based service models, more patient data are fed to cloud services from a variety of sources (e.g., health records, surveys, sensors, embedded GPS modules proposed a risk mitigation cloud service fed with heterogeneous patient data from a variety of data sources that can construct statistical models of adherence and intervention need. Vicente *et al*[14] proposes a conceptual reverse logistic model of Health Big Data with a cognitive computing in the backend responsible to data analysis. Yang and Lai[15] presented the concept of “peers” - patients facing the same self-care management problem in dealing with the same disease -, and proposed a conceptual system capable of identify personal adherence problems, analyze cluster peers and recommend suitable cures, all through a keyword-based analysis of information in Big Data.

The subgroup (3) Information health in Social Medias is related to the collect and analysis of online data about health in Social Medias, some popular and general-purpose such as Facebook and Twitter, and others specific to medical information change. The analysis of such data can be carried out with a variety of techniques and algorithms, and we are specially interested in those related to artificial intelligence and sentiment analysis, that can be used to automatically identify or asses, among others things, adherence and non-adherence behavior. Benton *et al*[16] analyzed online medical message boards of several health websites with the objective to identify potential adverse drug effects related to breast cancer discussed by users and could find several documented and undocumented items. Jamison-Powell *et al*[17] have collected and analyzed messages on Twitter about insomnia, and could concluded that people uses this social media to describe its coping and experience related to this disease. Mishra *et al*[18] have collected and analyzed a large sample of anonymous internet conversation in several health websites regarding treatment of prostate cancer with active surveillance. Murnane and Counts[19] had a similar approach from the first study, with the analysis of messages on Twitter but related to smoking cessation process. Park *et al*[20] focused on a celiac disease community on Facebook to gather information about the patient’s perspective of this disease. Ramagopalan, Wasiak and Cox[21] used Twitter to investigate opinions about multiple sclerosis treatments.

The subgroup (4) Internet heath information influence is related to the patient behavior change that internet users demonstrate after seeking online information about health. Cobb, Mays and Graham[22] didn’t aimed adherence, but analyzed how exposure to messages about the cessation drug varenicline affects smoker’s decision making around its use, on an online social media dedicated to smoking cessation and relapse prevention called QuitNet (http://www.quitnet.com). Weaver *et al*[23] compared personal and environmental determinants of internet users who report non-adherence behavior from their counterparts, using healthcare information obtained online.

The subgroup (5) Doctor-patient interaction analyzes how a good relationship and communication in medical consultations between doctor and patient can affect positively the treatment adherence. Howes, Purver and McCabe[24] have demonstrated in their study the feasibility of automatic prediction of adherence and therapy outcomes in schizophrenia through analysis of conversation topic between doctor-patient. Patterson *et al*[25] stated that patient education and counseling is a key factor in achieving and maintaining adherence, and thus developed an automatic system to detect, in clinical notes, text documentation of dietary counseling by physicians to gout patients.

Psychology

Adherence is strictly related to behavior, and many attempts to understand factors of non-adherence are based on conceptual frameworks and models found in the Psychology field. From the 15 articles selected to this systematic review, we could identify 3 which used human behavior’s theories to help in the task of automatically identify or assess adherence.

Hsueh, Ramakrishnan and Chang[13]leading to wastes of resources and increasing patient churn rates. In current practice, mitigating the risk of non-adherence cases is a labor-intensive task that requires additional efforts from healthcare professionals to handle on a case-by-case basis. Previous work has investigated into the possibility of modeling patient adherence behavior, but left questions about the accountability of such models in services. With the prevalence of mobile devices and maturing cloud-based service models, more patient data are fed to cloud services from a variety of sources (e.g., health records, surveys, sensors, embedded GPS modules used three theories to model the patient behavior, divided in two categories. The first one views the user behavior as an outcome, with the assumption that the degree to which patients follow the recommended care is determined by a set of predetermined factors. The Health Belief Model have been proposed to explain the relationship between beliefs and behavior, and the Decomposed Theory of Planned Behavior (DTPB) explains how patient behavior and the likelihood of adhering to recommended care plans are influenced by a set of behavioral intention-related attributes, such as attitude, subjective norm, and perceived behavior control. In the second category, the user behavior is view as a triggered action, and the Action Trigger Model can help providers to identify appropriate triggers to engage different types of users (e.g., users of different ability and motivation) who are at the different phases (e.g., awareness, motivation, and action) of behavioral change.

According to Murnane and Counts,[19] recent studies establishes that a combination of personal, behavioral, and environmental factors influence why an individual sets a goal, performs positively or negatively during its pursuit, and ultimately reaches success or failure. Thus, the Transtheoretical Model (TTM) provides a conceptual framework to evaluate an individual's readiness to embark on such goal and to monitor progress through stages of behavior change.

Weaver *et al*[23] uses the Social Cognitive Theory (SCT), which posits a model of reciprocal causation in which environmental factors, personal factors, and behavior all operate as interacting determinants of each other.

Methods, Tools and Sentiment Analysis

Here we are concentrated on what methods were chosen to analyze the data and which tools were used to carry out this analyze, with a special attention to sentiment analysis. We could identify 5 common approaches: (1) Manual, (2) Statistical, (3) Supervised Learning, (4) Natural Language Processing and (5) Sentiment Analysis.

Alemi *et al*[11] and Greaves *et al*[12] adjusted, applied and compared the supervised learning algorithms Decision Trees, Bagging, Support Vector Machines and Naive Bayes from the WEKA software package to conduct sentiment analysis on patient's experience.

Benton *et al*[16] aimed to identify potential adverse drug effects on a online medical message board, and analyzed all the data by applying statistical co-occurrence with the help of Natural Language Toolkit.

Cobb, Mays and Graham[22] were trying to determine the impact of online messages on smoker's choice to use varenicline, and used the software Salience Engine 4.1 to perform sentiment analysis with a keyword approach and a precoded dictionary.

Howes, Purver and McCabe,[24] to predict therapy outcomes using doctor-patient conversation topics, applied an unsupervised learning algorithm called Latent Dirichlet Allocation (LTA) present in MALLET toolkit and classified the results with Decision Trees and Support Vector Machines presents in WEKA.

Hsueh, Ramakrishnan and Chang[13]leading to wastes of resources and increasing patient churn rates. In current practice, mitigating the risk of non-adherence cases is a labor-intensive task that requires additional efforts from healthcare professionals to handle on a case-by-case basis. Previous work has investigated into the possibility of modeling patient adherence behavior, but left questions about the accountability of such models in services. With the prevalence of mobile devices and maturing cloud-based service models, more patient data are fed to cloud services from a variety of sources (e.g., health records, surveys, sensors, embedded GPS modules proposes a theoretical model of risk mediation cloud service which constructs models of patient adherence through statistics modeling.

Jamison-Powell *et al*[17] investigated insomnia on Twitter using with LIWC software to carry out a content analysis with character count, word class classification and word root type and a basic sentiment analysis.

Mishra *et al*[18]aggregate, and analyze content from the world-wide-web for ICs centered on AS. Collection of ICs was not restricted to any specific geographic region of origin. NLP was used to evaluate content and perform a sentiment analysis. Conversations were scored as positive, negative, or neutral. A sentiment index (SI used the Wool Labs software to perform a sentiment analysis on online conversation about active surveillance treatment. A machine learning approach with a internal data dictionary were used, and a sentiment index were calculated according to the numbers of positive and negative terms presents.

Murnane and Counts[19] used a manual method to analyze Twitter messages about smoking cessation events with Amazon Mechanical Turks, classifying survival or relapse, online activity, social media and behavior change process.

Park *et al*[20] collected surveys from a celiac disease community on Facebook, and analyzed all the data manually looking for patient-reported outcomes.

Patterson *et al*[25] applied a Natural Language Processing pipeline based on UIMA software to collect provider counseling practices in clinical notes.

Ramagopalan, Wasiak and Cox[21] collected data from Twitter regarding multiple sclerosis treatment's opinions, and performed sentiment analysis with Jeffrey Breen's code, calculating the difference of positive and negative terms in each message.

Vicente *et al*[14] proposes a theoretical reverse logistic model of Health Big Data, where all the data analysis are carried out by a cognitive computing system in the background which should have understanding of natural language, and ability to learn through iterative feedback and to personalize its response according to each patient.

Weaver *et al*[23] used surveys to gather data about internet users who change its behavior after seeking online health information and analyzed the results without any automatic technique.

Yang and Lai[15] developed a custom algorithm called iKeyword, where each keyword mimics a brain neuron, creating a media which characterizes patients and its similarity, according to the concept of "peers".

Table 2 Dimensions analysis of the selected articles

Article	Adherence	Data Sources	Psychology	Methods	Tools	Sentiment Analysis
ALEMI *et al.*[11]	Patient experience	Rate My MD	-	Decision Tree, Bagging, Support Vector Machine, Naive Bayes and Natural Language Processing	WEKA	Yes
BENTON *et al.*[16]	Information health in Social Medias	Several health websites	-	Statistical co-occurrence	Natural Language Toolkit	No
COBB, MAYS and GRAHAM[22]	Internet heath information influence	QuitNet	-	Keyword, pre-coded dictionary	Salience Engine 4.1	Yes
GREAVES *et al.*[12]	Patient experience	English National Health Service	-	Decision Tree, Bagging, Support Vector Machine, Naive Bayes and Natural Language Processing	WEKA	Yes

HOWES, PURVER and MCCABE[24]	Doctor-patient interaction	Audio-visual consultation records	-	Latent Dirichlet Allocation, Decision Trees and Support Vector Machine	WEKA and MALLET	No
HSUEH, RAMAKRISHNAN and CHANG[13]	Health Big Data	Cloud services	Health Belief Model, Decomposed Theory of Planned Behavior and Action Trigger Model	Statistical modeling	-	No
JAMISON-POWELL *et al.*[17]	Information health in Social Medias	Twitter	-	-	LIWC	Yes
MISHRA *et al.*[18]	Information health in Social Medias	Several health websites	-	Machine learning with internal dictionary	Wool Labs	Yes
MURNANE and COUNTS[19]	Information health in Social Medias	Twitter	Transtheoretical Model	Manual	Amazon Mechanical Turk	No
PARK *et al.*[20]	Information health in Social Medias	Facebook	-	Survey	-	No
PATTERSON *et al.*[25]	Doctor-patient interaction	Clinical notes	-	Natural Language Processing	UIMA	No
RAMAGOPALAN, WASIAK and COX[21]	Information health in Social Medias	Twitter	-	-	Jeffrey Breen	Yes
VICENTE *et al.*[14]	Health Big Data	Health Big Data	-	Cognitive Computing	-	No
WEAVER *et al.*[23]	Internet heath information influence	Survey	Social Cognitive Theory	Survey	-	No
YANG and LAI[15]	Health Big Data	Big Data and mobile	-	Keyword analysis	iKeyword	No

Conclusion

The analysis of the selected studies showed many articles similar to an automatic system capable of crawl online patient data on social media, analyze such information and its sentiment associated and finally identify and assess non-adherence behavior for management and intervention purposes.

However, several notable differences exists that avoids the aforementioned idea to be a reality today and, which makes this topic totally new in the literature to the best of our knowledge. Despite the adherence approach and its five subgroups presented, most of the studies[11,12,14,16–22,25]real-time surveys must be radically short. The shortest possible survey is a comment card. Patients' comments can be found online at sites organized for rating clinical care, within e-mails, in hospital complaint registries, or through simplified satisfaction surveys such as \"Minute Survey.\" Sentiment analysis uses patterns among words to classify a comment into a complaint, or praise. It further classifies complaints into specific reasons for dissatisfaction, similar to broad categories found in longer surveys such as Consumer Assessment of Healthcare Providers and Systems. In this manner, sentiment analysis allows one to re-create responses to longer satisfaction surveys from a list of comments. To demonstrate, this article provides an analysis of sentiments expressed in 995 online comments made at the RateMDs.com Web site. We focused on pediatrician and obstetrician/gynecologist physicians in District of Columbia, Maryland, and Virginia. We were able to classify patients' reasons for dissatisfaction and the analysis provided information on how practices can improve their care. This article reports the accuracy of classifications of comments. Accuracy will improve as the number of comments received increases. In addition, we ranked physicians using the concept of time-to-next complaint. A time-between control chart was used to assess whether time-to-next complaint exceeded historical patterns and therefore suggested a departure from norms. These findings suggest that (1 don't aim directly the automatic identification of patient behavior for adherence management or intervention, yet clearly is possible to aggregate and adapt each of them in a adherence-driven way. Patients who are complaining publicity on internet about its bad past experience and patients who are discussing adverse effects of a certain drug or treatment option in a medical message board or Twitter are in the risk group of treatment abandon, and we can expand this group by analyzing medical-patient poor communications and investigation why users are seeking for online health information if they already have a treatment plan. All this model of adherence management fits in the model of Health Big Data, which contains important patient data from a variety of data sources to be analyzed. A common problem when applying any automatic technique, especially in the medicine field, is the validity of the results obtained. Studies struggles to confirm that the outputs reflect the reality and are not being affected by any noise or statistical issue. The use of models and frameworks from the psychology shows a driven way to apply such techniques, serving as a starting point and as a base for comparing and validating its efficacy. We can name, for instance, the study which collected several adverse drug effects from online patients discussions, most of them already documented in previous researches and a few with a high potential to be new undiscovered adverse effects. The methods to carry out the studies are another heterogeneous dimension. Some surveys analysis[20,23] and manual methods[19] leave a good opportunity to be substituted by automatic techniques related to artificial intelligence, especially sentiment analysis. We could notice most of the articles[11,12,14,16,18,20,22,23]real-time surveys must be radically short. The shortest possible survey is a comment card. Patients' comments can be found online at sites organized for rating clinical care, within e-mails, in hospital complaint registries, or through simplified satisfaction surveys such as \"Minute Survey.\" Sentiment analysis uses patterns among words to classify a comment into a complaint, or praise. It further classifies complaints into specific reasons for dissatisfaction, similar to broad categories found in longer surveys such as Consumer Assessment of Healthcare Providers and Systems. In this manner, sentiment analysis allows one to re-create responses to longer satisfaction surveys from a list of comments. To demonstrate, this article provides an analysis of sentiments expressed in 995 online comments made at the RateMDs.com Web site. We focused on pediatrician and obstetrician/gynecologist physicians in District of Columbia, Maryland, and Virginia. We were able to classify patients' reasons for dissatisfaction and the analysis provided information on how practices can improve their care. This article reports the accuracy of classifications of comments. Accuracy will improve as the number of comments received increases. In addition, we ranked physicians using the concept of time-to-next complaint. A time-between control chart was used to assess whether time-to-next complaint exceeded historical patterns and therefore suggested a departure from norms. These findings suggest that (1 are more related to the healthcare field than computing science, where the authors are essentially medical doctors applying computing tools. It suggests that the computing community is not yet aware of all the practical application opportunities in the healthcare, while medical community are by themselves learning, applying and benefiting from computing. As consequence, we could observe the superficial level of computing in those studies. Outcomes could be improved if more accurate data pre-processing, parameters adjusting and training-sets were applied.

References

[1] Mathes T, Jaschinski T, Pieper D. Adherence influencing factors – a systematic review of systematic reviews. *Arch Public Heal*. 2014;72(1):37.

[2] Sapkota S, Brien J, Greenfield J, Aslani P. A Systematic Review of Interventions Addressing Adherence to Anti-Diabetic Medications in Patients with Type 2 Diabetes—Impact on Adherence. *PLoS One*. 2015;10(2):1-17. doi:10.1371/journal.pone.0118296.

[3] Glanz K, Rimer BK, Viswanath K. *Health Behavior and Health Education: Theory, Research and Practice*. John Wiley & Sons; 2008. doi:10.1016/S0014-1606(49)81524-1.

[4] Stoughton JW, Thompson LF, Meade AW. Big five personality traits reflected in job applicants' social media postings. *Cyberpsychol Behav Soc Netw*. 2013;16(11):800-806. doi:10.1089/cyber.2012.0163.

[5] Bachrach Y, Kosinski M, Graepel T, Kohli P, Stillwell D. Personality and patterns of Facebook usage. *Proc 3rd Annu ACM Web Sci Conf*. 2012:24-32.

[6] Mistry N, Keepanasseril a., Wilczynski NL, Nieuwlaat R, Ravall M, Haynes RB. Technology-mediated interventions for enhancing medication adherence. *J Am Med Informatics Assoc*. 2015. doi:10.1093/jamia/ocu047.

[7] Varshney U. Mobile health: Four emerging themes of research. *Decis Support Syst*. 2014;66:20-16. doi:10.1016/j.dss.2014.06.001.

[8] Glenn T, Monteith S. New Measures of Mental State and Behavior Based on Data Collected From Sensors, Smartphones, and the Internet. *Curr Psychiatry Rep*. 2014;16(12):1-10.

[9] Prisma. [internet] [acesso em 2015 nov 25] Disponível em: http://www.prisma-statement.org/

[10] Ji X, Chun SA, Geller J. Monitoring public health concerns using twitter sentiment classifications. *Proc - 2013 IEEE Int Conf Healthc Informatics, ICHI 2013*. 2013:145-344.

[11] Alemi F, Torii M, Clementz L, Aron DC. Feasibility of real-time satisfaction surveys through automated analysis of patients' unstructured comments and sentiments. *Qual Manag Health Care*. 2012;21(1):9-19.

[12] Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. *J Med Internet Res*. 2013;15(11):1-11.

[13] Hsueh PY, Ramakrishnan S, Chang H. Risk mediation cloud service: Constructing statistical models of patient adherence for sustainable case management. *Annu SRII Glob Conf*. 2012:237-242.

[14] Vicente S, Martin D, Vicente O, Doblas M, Maynar M. Reverse logistic on health. *Proc 2014 Summer Simul Multiconference*. 2014;(59):1-7.

[15] Yang YP, Lai TY. The development of a Patient Adherence Management System Applying Intelligent Keyword (PAMSIK) empowered by peer learning mechanism. *IEEE 14th Int Conf e-Health Networking, Appl Serv*. 2012:105-110.

[16] Benton A, Ungar L, Hill S, et al. Identifying potential adverse effects using the web: A new approach to medical hypothesis generation. *J Biomed Inform*. 2011;44(6):989-996.

[17] Jamison-Powell S, Linehan C, Daley L, Garbett A, Lawson S. "I can't get no sleep": discussing #insomnia on Twitter. *Proc SIGCHI Conf Hum Factors Comput Syst*. 2012:1501-1510.

[18] Mishra M V, Bennett M, Vincent A, et al. Identifying barriers to patient acceptance of active surveillance: content analysis of online patient communications. *PLoS One*. 2013;8(9):1-7.

[19] Murnane EL, Counts S. Unraveling abstinence and relapse: smoking cessation reflected in social media. *Proc 32Nd Annu ACM Conf Hum Factors Comput Syst*. 2014:1345-1164.

[20] Park K, Harris M, Khavari N, Khosla C. Rationale for Using Social Media to Collect Patient-Reported Outcomes in Patients with Celiac Disease. *J Gastrointest Dig Syst*. 2014;4(1):1-5. Secreted.

[21] Ramagopalan S, Wasiak R, Cox AP. Using Twitter to investigate opinions about multiple sclerosis treatments: a descriptive, exploratory study. *F1000Research*. 2014:1-9.

[22] Cobb NK, Mays D, Graham AL. Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. *J Natl Cancer Inst Monogr*. 2013;(47):224-230.

[23] Weaver JB, Thompson NJ, Weaver SS, Hopkins GL. Healthcare non-adherence decisions and internet health information. *Comput Human Behav*. 2009;25(6):1373-1380.

[24] Howes C, Purver M, McCabe R. Using Conversation Topics for Predicting Therapy Outcomes in Schizophrenia. *Biomed Inform Insights*. 2013;6:39-50.

[25] Patterson O V., Kerr GS, Richards JS, et al. Identifying provider counseling practices using natural language processing: Gout example. *IEEE Second Conf Healthc Informatics, Imaging Syst Biol*. 2012:127.

Contato

Autor correspondente:
Gabriel Ribeiro
email: gribeiro@br.ibm.com