2018

IT for diabetes self-management - What are the patients expectations?

Omar El-Gayar
Dakota State University

Tareq Nasralah
Dakota State University

Ahmed Elnoshokaty
Dakota State University

IT for diabetes self-management - What are the patients' expectations?

Omar El-Gayar
Dakota State University

Tareq Nasralah
Dakota State University

Ahmed Elnoshokaty
Dakota State University

IT for diabetes self-management-What are the patients' expectations? IT for diabetes self-management-What are the patients' expectations?

3 authors:

Omar F. El-Gayar
Dakota State University
156 PUBLICATIONS   **1,501** CITATIONS

Tareq Nasralah
Northeastern University
15 PUBLICATIONS   **13** CITATIONS

Ahmed El Noshokaty
Northern Michigan University
6 PUBLICATIONS   **17** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Online petition Success Factors - A Linguistic Cue-based Approach   View project

Project   Smart Agriculture   View project

IT for diabetes self-management - What are the patients' expectations?

Completed Research

Omar El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Tareq Nasralah
Dakota State University
tareq.nasralah@trojans.dsu.edu

Ahmed El Noshokaty
Dakota State University
ahmed.elnoshokaty@trojans.dsu.edu

Abstract

Diabetes is a debilitating chronic condition that is touted to inflict 9% of adults. Despite the information technology (IT) solutions for the self-management of diabetes, there is a gap between user expectations, and what the IT solutions are offering. To understand these expectations, this research investigates the topics shared by a diabetic community on Twitter as an instant day-to-day micro-blogging platform and TuDiabetes as a social media forum for the community. We focus on studying personal experiences related to diabetes through adapt and apply text mining techniques to tweets of self-identified diabetic patients on Twitter and TuDiabetes. Through the analysis 75,864 tweets and 1,424 TuDiabetes discussions, we identify dominating talks on continuous glucose monitoring on both Twitter and TuDiabetes, we also spotted insightful discussions on diet, wearables, and technology for Type two diabetes patients on Twitter. The results could guide future IT interventions into the self-management of the condition.

Keywords

Diabetes; social media; text mining; topic modeling; unsupervised learning.

Introduction

Diabetes is a widespread, costly condition related with major illness and death rate. Diabetes is one of the most important chronic diseases in the healthcare area, and this is due to the increase in the number of patients with type 1 as well as type 2 (Renders et al., 2001). Diabetes is expected to be the seventh leading cause of death by 2013 and is touted to inflict nine percent of adults 18 years or older (Chulis, 2015). The prevalence can also vary drastically by country reaching 20% in some countries. Diabetes comes in different types, where Type 2 (which is preventable and manageable by adjusting one's dietary, exercise, and other behavioral choices) represent 90% of worldwide diabetes cases. The increasing number of diabetes cases and associated side effects is a key driver for medical research and IT-based innovations in an attempt to stem this impending epidemic.

As with most chronic conditions, self-management defined as "Self-management relates to the tasks that an individual must under- take to live well with one or more chronic conditions. These tasks include gaining confidence to deal with medical management, role management, and emotional management." (Corrigan, Greiner, & Adams, 2004) is critical to mitigate the progression and serious effects of this condition. A distinguishing element of this definition is the description of self-management as a set of behaviors by the patient. In diabetes, examples of such behavior include self-monitoring of blood glucose, physical activity or exercise, diet, and medications.

Advances in technology such smartphones, mobile applications and wireless networks have resulted in enhanced capability and increased interest by the medical community and patients alike for the self-management of chronic conditions. A systematic review by El-Gayar et al. (El-Gayar, Timsina, Nawar, & Eid, 2013b) focusing on information technology (IT) the self- management of diabetes indicates further

attest to the viability of the IT solutions. In this context, IT included the Internet/web-based systems, mobile devices, decision support systems, and telemedicine. The review identified a number of issues with IT use in diabetes self-management such as value-added, satisfaction, and adoption. The latter is particularly important where some studies indicates that patients showed less interest in continuation of technological interventions after the pilot studies (Bujnowska-Fedak, Puchała, & Steciwko, 2006; Mollon et al., 2008).

While El-Gayar et al. (El-Gayar et al., 2013b) highlight the importance of theoretically grounded IT interventions, user-centeredness, and socio-technical design to improve adoption and acceptance, we argue that that there may be additional issues at play that impedes the large-scale diffusion of IT in the self-management of diabetes. In essence, despite the proliferation of apps and other technical solutions for the self-management of diabetes, there is still a gap / mismatch between user expectations/needs and what the technology solution is offering. There exist a need to better understand the patients' expectations, needs and priorities with respect to the self-management of their conditions.

In that regard, social media is gradually showing promise as a viable source of data for gaining insights into various disease condition, patients' attitudes and behaviors, medications, etc. Social media channels can also serve as a conduit to health behavioral change through messaging (Korda & Itani, 2013). A key advantage of social media as a data source is that while the confidentiality and privacy of patient data is protected by the Health Insurance Portability and Accountability Act (HIPAA), many patients would willingly discuss and share health-related information about their condition. Coupled with text mining and machine learning, social media such as Twitter can serve as a rich resource of health (Dredze, 2012; Dredze, Cheng, Paul, & Broniatowski, 2014)

The objectives of this research are 1) Explore the use of social media as a source for eliciting patients' needs and expectations with respect to the use of information technology (IT) for the self-management of diabetes, and 2) Highlight challenges pertaining to mining social media for insights, particularly in the healthcare domain.

The key implication for practice is the importance of focusing on building IT solutions for the self-management of diabetes that are centered around continuous glucose monitoring. The research and development emphasis on elements such as usability, reliability, etc. will continue to be important and play a role. However, only after, and around the continuous measurement of glucose levels.

The theoretical implications of this research are two folds: 1) It highlights the importance of further developing and adapting text mining techniques to social media. Such media represent inherent challenges for text mining given the amount of noise and distortion in the data. 2) It brings the possibility of organically inferring users' interests and needs to light. While this research demonstrated that we can infer high-level needs of potential users, future research can continue to explore the idea of mining features and requirements beyond needs.

In essence, the research demonstrates that patients' expectations can vary from that of researchers and developers, and highlights the potential of social media to provide insight into patients' expectations. To be able to leverage the richness of social media, further research is warranted. Such research need to capture the salient characteristics of social media in pre-processing, feature extraction and selection, and the selection of analytic techniques.

The remainder of the paper is organized as follows: the next section provides a brief synopsis of the use of IT in diabetes self-management and social media mining in healthcare. This is followed by a description of the methodology (data collection, pre-processing, and analysis), results, and discussion. The last section concludes the paper with a summary, key findings, and recommendations for future work.

Literature review

Social media such as online forums and twitter are used by patients to exchange information and discuss different health related topics. (Tapi Nzali, Bringay, Lavergne, Mollevi, & Opitz, 2017). Discussion forums are kind of online communities' application, which hold incredible educational perspective (Helic, Maurer, Scerbakov, Supported, & Media, 2004). Online discussion forums are one of the most common social media provider people use. Studies have indicated that members in online community such as discussion forums can advance health results (Abrahamson & Rubin, 2012). Discussion forums offer a revenues for both of

patients and health care providers to discuss subjects associated to the Diabetes domain (de Clercq, Hasman, & Wolffenbuttel, 2001). Offering discussion forum in the online patient education of diabetes sites is a good assistance for patients by having online based support intermediations and assisting them in managing diabetes (Online & Than Win, 2010). Several studies have used the online health communities to help better understand the patients' needs. Lu et al. (2013) have conducted a study to detect the health-related hot topic in online communities using text clustering.

With respect to Twitter as a microblogging service, users tweet short text messages that often contain links to news stories and comment (Lerman & Ghosh, 2010). Several studies have used Twitter as a source of input data to identify the public's reactions to the opioid epidemic by detecting the most popular topics tweeted by users (Glowacki, Glowacki, & Wilcox, 2017), for marijuana content analysis (Cortés, Velásquez, & Ibáñez, 2017). Keywords have been used to filter Marijuana related tweets (Daniulaityte et al., 2015; Tian, Lagisetty, & Li, 2016) or tweets related to potential drug effects (Jiang & Zheng, 2013). In Twitter diabetes (Batool, Khattak, Maqbool, & Lee, 2013) related keywords ("diabetes," "Food," "Diet," and "Blood pressure") are proposed to precisely filter in diabetes-related tweets.

Studying contributions by diabetes patients will help us to discover the dominant diabetes management themes within the community, as well as the most important topics within each identified theme. By uncovering these topics and themes, healthcare providers will have more information about what diabetes patients are more concerned about, and by doing so, they can help improve the services provided, which in turn will be reflected positively on patient's health status.

Methodology

Figure 1 depicts the data processing and analysis process. The following subsection describes various elements of the process.

Data Collection and Preprocessing

Tudiabetes data collection and preprocessing

We create a crawler to collect data from Tudiabetes.org which is an online community of people touched by diabetes as one of online health forums that aims to improve the lives of people impacted by diabetes (322,577 posts collected from March 10, 2007, to Oct 13, 2017). These collected posts belong to users who self-identified themselves as diabetes patients with Type 1 and Type 2. Out of all posts, we extract the users' technology discussions posts (26776 posts belong to different discussions about diabetes technology in general). In practice, there was significant noise in the posts due to semantic ambiguity. We developed several steps to prepare the data for the analysis. The preprocessing steps for the posts as the following:

- Removal of all whitespaces and cleaning punctuation, HTML tags, hyperlinks and URLs paths.
- Splitting attached words: after removal of punctuation or white spaces, words can be attached. This happens especially when deleting the periods at the end of the sentences. The corpus might look like: "it will get better that was my experienceGood luck". So, there is a need to split "experienceGood" into two separate words.
- Standardize words (remove multiple letters): we applied some correction for the words' spelling. Where users' emotions of online social media sometimes can play a role while they make the posts, such as "gooood" instead of "good" or "annnd" instead of "and" and so on.
- Convert the text to lowercase and remove stop words: stop words are basically a set of commonly used words in any language. By removing the words that are very commonly used in each language, we can focus only on the important words instead, and improve the accuracy of the text processing.
- Replacement of hyperlinks and email addresses. All the hypertext links are replaced by the term "link" and all the email addresses are replaced by the term "mail."
- Appling stemming and lemmatization for all words to reduce inflectional word forms to linguistically valid lemmas.

After preprocessing the posts text, we create discussions where each discussion is the aggregated users' posts that belong to same discussion title. This gives us 1424 discussions in total.

Twitter data collection and preprocessing

From 32,653 followers on TuDiabetes Twitter account, we aimed to spot self-identified Diabetic patients from others. We filtered out users that did not have diabetes keywords (Diabetes, Type 1, Type 2...etc.) in their Twitter account biography. Keywords are referenced in (Batool et al., 2013; "Common Diabetes Terms," 2013), only 1,221 self-identified Twitter users were retrieved. Then we manually labeled these users through manually going through their bio description and pointed 792 self-identified diabetic patients. We collected a total of 2,268,657 tweets for the self-identified diabetes patients over the period from 11/7/2017 to 11/15/2017. We then filtered in on any of the keywords diabetes-related terms as mentioned earlier, filtered out retweets, tweets containing links non-English, and duplicate tweets resulting in 75,864 of personal experience tweets by self-identified diabetes patients.

Data Analysis

We applied unsupervised text modeling process by using Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) to extract the different topics that the users have in their discussions and tweets. The LDA model is one of the most common topic model currently in use, this because its conceptual advantage over other latent topic models (Blei et al., 2003). The model generates automatic summaries of topics in terms of a discrete probability distribution over words for each topic. The interaction between the observed documents and hidden topic structure is manifested in the probabilistic generative process associated with LDA. This generative process can be thought of as a random process that is assumed to have produced the observed document (Bao & Datta, 2014). We use a metric called perplexity that is conventional in language modeling. Perplexity can be understood as the predicted number of equally likely words for a word position on average and is a monotonically decreasing function of the log-likelihood. Thus, a lower perplexity over a held-out document is equivalent to a higher log-likelihood, which indicates better predictive performance (Blei et al., 2003). We calculated perplexity scores for various number of topics to deduce a suitable number of topics to use by the LDA algorithm (Azzopardi, Girolami, & van Risjbergen, 2003).
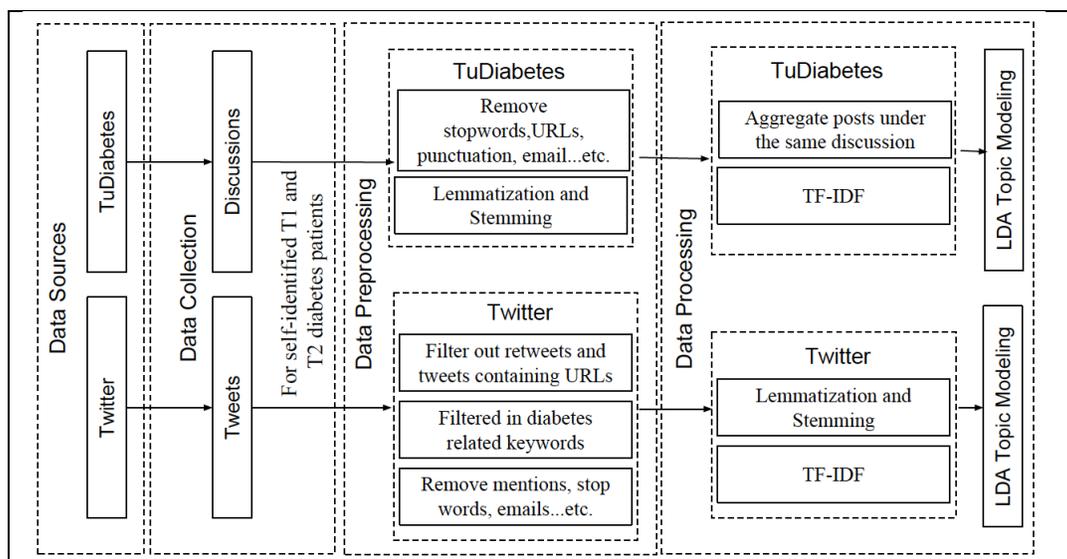
Figure 1. Data Processing and Analysis Process

Results and discussion

Tudiabetes data analysis

After applying the analysis on the discussions documents, the output of the LDA topic modeling for Tudiabetes discussions was a set of topics and the keywords related with each topic. The 1428 preprocessed discussions were treated as the input for the LDA. After studying the perplexity score as shown in Figure 2, to define the best number of topics to set the LDA model on, we decide to set the number of topics to fifty.

Out of the fifty topics, we took top seven topics that have most coherence keywords. Table 1 showing the seven topics with associated keywords. The topics where mostly related to continues glucose monitoring (CGM) technologies and related issues, e.g., usage, calibration, insurance coverage, software support, etc.
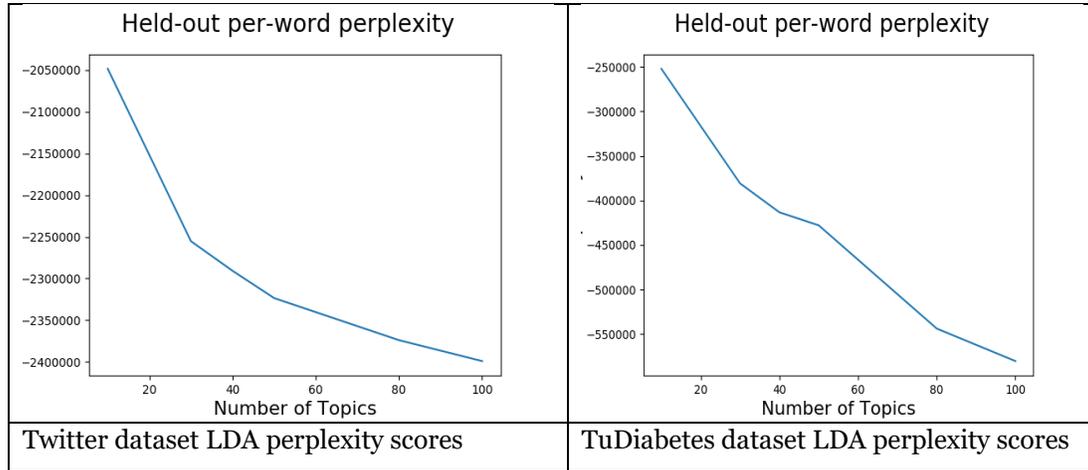
Figure 2. perplexity scores

Dexcom Calibration	Dexcom Usage	BG Meters	Insurance	Mobile Apps	System Support	Security/Risk
sensor	sensor	meter	dexcom	phone	software	Device
dexcom	skin	strip	insurance	receiver	window	Data
reading	tape	glucose	medicare	app	mac	Driver
accurate	dexcom	finger	company	iPhone	dexcom	Security
calibration	adhesive	blood	supply	battery	pump	Health
accurate	transmitter	lancing	coverage	device	computer	Information
cgm	arm	reading	cost	data	usb	Risk
range	work	change	pay	android	file	Patient
work	area	test	medical	export	program	Record
time	problem	data	sensor	exercise	cable	Command

Table 1. Tudiabetes Seven Topics with Associated Keywords.

Overall, the results seem to indicate that despite the extensive research in the use of IT for diabetes self-management, e.g., using the Internet (web), text messaging, mobile apps, etc. (El-Gayar et al, 2013), the interest and priorities of the patients appear to be different. Specifically, there is a significant attention to continuous glucose monitoring via latest innovations in the field such as Dexcom's G5 CGM system. This particular system is comprised of a small sensor that measures glucose levels just underneath the skin, a transmitter that fits onto the sensor and sends data wirelessly to a receiver or a compatible smart device that displays real-time glucose data. These results are particularly interesting as clinical guidelines do not generally prescribe CGM except in certain situations (usually type 1 and extreme cases of uncontrolled type 2 diabetes). This is also confirmed by physicians where the general position is to optimize the measurement of glucose levels to minimize undue stress on the patient due to finger pricking (the more prevalent approach for measuring glucose levels).

Other than the significant interest in continuous glucose monitoring, there were references (albeit sparse) to mobile apps and information security/risk. The interest in apps references data export echoing aspects of integration with medical devices (Al-Ramahi, Liu, & El-Gayar, 2017). However, contrary to expectations, we did not find evidence for discussion topics related to usability and ease of use, adoption, communication with physicians, induced behavioral changes, etc. that are commonly referred to in the literature as drivers to the use of IT in diabetes self-management.

Twitter data analysis

Twitter dataset is classified as shown in Table 2, also the data preprocessing is captured in Figure 1.

Item	Count
Self-identified diabetes patients	792
Number of Tweets for Self-identified diabetes patients	75,864
Number of T1 Self-identified diabetes patients	592
Number of Tweets for T1 Self-identified diabetes patients	52,523
Number of T2 Self-identified diabetes patients	58
Number of Tweets for T2 Self-identified diabetes patients	23,341

Table 2. Twitter Dataset

We ran topic modeling LDA (Blei, Ng, & Jordan, 2003) on all 75,864 tweets for all self-identified diabetes and fine-tuned the number of topics through computing the perplexity for some topics 10, 30, 50, 80, and 100. According to perplexity scores as shown in Figure 2. We chose the number of topics to be fifty as the difference in perplexity from hundred topics was minor, however, it was easier to label topics with fifty topics. The discussion topics are labeled as shown in Table 3.

Overall, while continuous glucose monitoring was a dominant topic (represented under one topic below), the results reflect a slightly broader set of topics with reference to diet and nutrition, social support, insurance/cost, medications and wearables. Only wearables contained direct reference to technology.

Diet and Nutrition	Diabetes Community Networking	Continuous Glucose monitoring	medical insurance	mental state impact	diabetes drugs	wearables and Mobile applications
weight	awar	dex	insur	life	drug	Check
Loss	talk	easi	take	happi	afrezza	Myfitnessp
nutrit	rais	track	onlin	health	doc	exercise
food	challeng	smart	digit	camp	a1c	Goal
calori	stori	Access	pancrea	need	team	Call
Fat	wddchat15	wear	edicar	person	meet	Phon
Eat	Dsma	cgm	ask	impact	fda	Hate
healthi	pwd	pump	manag	fun	treatment	Agre
drink	world	insulin	artifici	read	dsma	Diagnos
snack	Gbdoc	glucos	option	save	diabet	Idea

Table 3. Twitter Seven Topics with Associated Keywords.

Also, we ran LDA separately on 3,259 tweets of the minority T2 Self-identified diabetes patients and we achieved similar results with the majority of tweets talking about Continuous Glucose Monitoring. However, more discussion is also focused on Diet and Exercise as well as mobile applications as shown in Table 4.

Tweet	Topic
@KellyRawlings I keep wanting to test that out but forget to take my meter to town. 😕	Continuous Glucose Monitoring
I earned a Fitbit Adjustment of 584 calories. #LoseIt	Diet and Exercise
Completed his food and exercise diary for 11/18/2014 #myfitnesspal	Mobile Applications
@amydiabetesx2 I'll try after I eat. I've been taking it then eating. Morning has been the worst until now. Concerned about fluid loss too.	Diet and nutrition

Table 4. Examples of Tweets with Personal Experiences for T2 self-identified patients

Overall, the topics analysis results of both datasets (the forum's discussions and tweets) show there is a common interest for the diabetes patients in needs to continue monitoring their glucose level. As well, we found common topics talking about mobile apps and medical insurance. Similar to that, in the forum discussions, some topics related to technical support and security / risk of using the technology were discussed through Twitter too. However, some topics are discussed in the twitter datasets were not found in the forum discussions, such as: diet and nutrition and social support.

For both sources, there were significant challenges in pre-processing the data. Examples include the proliferation of spelling errors, abbreviations, slang, etc. While, the pre-processing techniques applied addressed some of these problems, further research into social media / source specific techniques is warranted. For twitter, the length of the tweets also presented issues aligning with appropriate topics. More importantly, feature extraction and selection that is solely based on a bag-of-words (BoW) approach can prove limiting.

Conclusion

In this research we leveraged text mining, machine learning to explore diabetes patients' needs and expectations with respect to the use of IT in diabetes self-management. The underlying motivation and driver for the research is explaining what appears to be a paradox or inconsistency between what is observed as a great interest and potential by the medical community (see El-Gayar (2013) for a systematic review) and the questionable adoption and retention by diabetes patients for IT in the self-management of the condition (Bujnowska-Fedak et al., 2006; Mollon et al., 2008). With this motivation, we aimed to 1) Explore the use of social media as a source for eliciting patients' needs and expectations with respect to the use of information technology (IT) for the self-management of diabetes, and 2) Highlight challenges pertaining to mining social media for insights, particularly in the healthcare domain. The initial results highlight the potential of using social media to elicit patients' needs as well as the possible challenges encountered. With respect to the underlying motivation, the results indicate that a potential explanation lies in a mismatch between the expectations of the medical community and the patients (actual users/receipts of the technology). While the research focuses on patient-provided communication, design of IT-enabled systems, etc., patients appear interested in innovations related to continuous glucose monitoring (even if their condition) does not need or require such continuous monitoring.

In summary, the results demonstrate the utility of social media mining in identifying users' (patients in this context) needs that may very well be different from domain experts' (physicians and medical researchers in this context). The results also shed light into explaining the diabetes app landscape that is characterized by a large number of offerings, yet fairly low level of adoption and retention - less than six months. Despite significant contributions into the design of mobile app (El-Gayar, Timsina, Nawar, & Eid, 2013a), it appears the more prevalent issue and patient need is the importance of continuously monitoring their glucose levels.

Also despite previous research emphasizing diet management, physical activity, medication, and blood glucose monitoring as the primary features related to the self-management of diabetes (El-Gayar et al., 2013a), this study highlights the importance of blood glucose monitoring features since it is a widely discussed concerns on both Tudiabetes and Twitter. Most prominently, is the importance of efficiently integrating blood glucose calibration in mobile applications. In addition, new features for linking the online diabetes community should be incorporated in IT solutions as found on Twitter.

From a practitioner's' perspective, the results emphasize the importance of focusing on building IT solutions for the self-management of diabetes that are centered around continuous glucose monitoring regardless of the type or status of the disease. Such emphasis can complement ongoing research and development that focuses on elements such as usability, reliability, etc. will continue to be important and play a role.

From a theoretical perspective, the research highlights the importance and potential for organically inferring users' interests and needs from unstructured social media data. This is in contrast with structured interviews and surveys such as those encountered in requirement elicitation and engineering or research endeavors. It also highlights the notion of relying on computational research (relying on large volumes of data to automatically detect patterns in data) versus 'traditional' research (e.g., in social sciences that relies on the design of experiments and field studies to test hypotheses based on pertinent theories)

Naturally the research has a number of limitations: 1) Reliance on limited data sources (TuDiabetes and Twitter). Other alternatives include other discussion forums, Facebook groups, etc., 2) Limitations with the data set, e.g., time and technical (limitations on the Twitter API) constraints the ability to extract sets of related tweets which we believe would have improved our results, and 3) Establishing a 'ground truth' for a more systematic validation of the results.

In future research we plan to address the aforementioned limitations. Further refining the data collection process, we seek to explore the use of vocabularies such as the Open-Access Collaborative Consumer Health Vocabulary ("OAC CHV," "CHV" for short) which was developed as a collection of expressions and concepts that are commonly used by ordinary health information users (Zeng & Tse, 2006) and UMLS (Park et al. (2016). Other than LDA, we also plan to apply other text analysis methods, such as document clustering by using the K-means algorithm and its variants (Andrews & Fox, 2007) as well as word2vec (Goldberg & Levy, 2014). Last but not least, we aim to work towards a framework for the standardization of data management, preparation, and analysis that is specifically customized to the problem under consideration as noted in Chulis (2015).

In conclusion, despite the gradual proliferation of research at the intersection of social media and text mining and machine learning, we contend that we are only witnessing the tip of an iceberg. The rich information embedded in social media is yet waiting to be discovered. It is paramount upon the information and computing communities to work hand-in-hand with healthcare researchers and experts to realize the true potential of this data and adapt and advance state-of-the-art technologies towards this goal.

REFERENCES

Abrahamson, J. A., & Rubin, V. L. 2012. "Discourse structure differences in lay and professional health communication," *Journal of Documentation*, *68*(6), 826–851.

Al-Ramahi, M. A., Liu, J., & El-Gayar, O. F. 2017. "Discovering Design Principles for Health Behavioral Change Support Systems," *ACM Transactions on Management Information Systems*, *8*(2–3), 1–24.

Andrews, Nicholas, O., Fox, Edward A., 2007. "Recent developments in document clustering," Technical report TR-07-35. Deparment of Computer Science, Virginia

Tech Azzopardi, L., Girolami, M., & van Risjbergen, K. 2003. "Investigating the relationship between language model perplexity and IR precision-recall measures," In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR '03* (p. 369). New York, New York, USA: ACM Press.

Bao, Y., & Datta, A. 2014. "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures," *Management Science*, *60*(6), 1371–1391.

Batool, R., Khattak, A. M., Maqbool, J., & Lee, S. 2013. "Precise tweet classification and sentiment analysis,"

2013 IEEE/ACIS 12th International Conference on Computer and Information Science, ICIS 2013 - Proceedings, 461–466.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. "Latent Dirichlet Allocation. Journal of Machine Learning Research".

Bujnowska-Fedak, M. M., Puchała, E., & Steciwko, A. 2006. "Telemedicine for diabetes support in family doctors' practices: A pilot project," *Journal of Telemedicine and Telecare*, *12*(1_suppl), 8–10.

Chulis, K. 2015. "*Data mining Twitter for cancer, diabetes, and asthma insights*," Purdue University.

Common Diabetes Terms. 2013.

Corrigan, J., Greiner, A., & Adams, K. 2004. "*The 1st Annual Crossing the Quality Chasm Summit: A Focus on Communities: Report of a Summit,"* Washington, DC: National Academies Press,.

Cortés, V. D., Velásquez, J. D., & Ibáñez, C. F. 2017. "Twitter for marijuana infodemiology," In *WI '17 Proceedings of the International Conference on Web Intelligence*.

Daniulaityte, R., Nahhas, R. W., Wijeratne, S., Carlson, R. G., Lamy, F. R., Martins, S. S., … Sheth, A. 2015. "Time for dabs": Analyzing Twitter data on marijuana concentrates across the U.S. *Drug and Alcohol Dependence*, *155*, 307–311.

de Clercq, P. A., Hasman, A., & Wolffenbuttel, B. H. 2001. "Design of a consumer health record for supporting the patient-centered management of chronic diseases," *Studies in Health Technology and Informatics*, *84*(Pt 2), 1445–9.

Dredze, M. 2012. "How Social Media Will Change Public Health," *IEEE Intelligent Systems*, *27*(4), 81–84.

Dredze, M., Cheng, R., Paul, M. J., & Broniatowski, D. 2014. "HealthTweets.org: A Platform for Public Health Surveillance using Twitter."

El-Gayar, O., Timsina, P., Nawar, N., & Eid, W. 2013a, January. "Mobile applications for diabetes self-management: Status and potential," *Journal of Diabetes Science and Technology*.

El-Gayar, O., Timsina, P., Nawar, N., & Eid, W. 2013b, August. "A systematic review of IT for diabetes self-management: Are we there yet?" *International Journal of Medical Informatics*. Elsevier.

Glowacki, E. M., Glowacki, J. B., & Wilcox, G. B. 2017. "A text-mining analysis of the public's reactions to the opioid crisis," *Substance Abuse*, 1–5.

Goldberg, Y., & Levy, O. 2014. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, (2), 1–5.

Helic, D., Maurer, H., Scerbakov, N., Supported, C., & Media, N. 2004. "Discussion Forums As Learning Resources in Web- Based Education," *Advanced Technology for Learning*.

Jiang, K., & Zheng, Y. 2013. "Mining Twitter Data for Potential Drug Effects," In *9th International Conference, ADMA 2013* (Vol. 8346, pp. 434–443). Hangzhou, China: Springer, Berlin, Heidelberg.

Korda, H., & Itani, Z. 2013. "Harnessing Social Media for Health Promotion and Behavior Change," *Health Promotion Practice*, *14*(1), 15–23.

Lerman, K., & Ghosh, R. 2010. "Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 90–97.

Lu, Y., Zhang, P., Liu, J., Li, J., & Deng, S. 2013. "Health-Related Hot Topic Detection in Online Communities Using Text Clustering," *PLoS ONE*, *8*(2), e56221.

Mollon, B., Holbrook, A. M., Keshavjee, K., Troyan, S., Gaebel, K., Thabane, L., & Perera, G. 2008. "Automated telephone reminder messages can assist electronic diabetes care," *Journal of Telemedicine and Telecare*, *14*(1), 32–36.

Online, R., & Than Win, K. 2010. Implementing patient accessible health information site for diabetes management, 1790–1797.

Renders, C. M., Valk, G. D., Griffin, S. J., Wagner, E. H., Van Eijk, J. T. M., & Assendelft, W. J. J. 2001. "Interventions to improve the management of diabetes in primary care, outpatient, and community settings: A systematic review," *Diabetes Care*, *24*(10), 1821–1833.

Tapi Nzali, M. D., Bringay, S., Lavergne, C., Mollevi, C., & Opitz, T. 2017. "What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer," *JMIR Medical Informatics*, *5*(3), e23.

Tian, Q., Lagisetty, J., & Li, B. 2016. "Finding needles of interested tweets in the haystack of Twitter network," *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 447–452.

Zeng, Q. T., & Tse, T. 2006, January. "Exploring and developing consumer health vocabularies," *Journal of the American Medical Informatics Association*. Oxford University Press.