Social Media Analytics for Behavioral Health

Rose Yesha, Aryya Gangopadhyay

Department of Information Systems University of Maryland Baltimore, County Baltimore, Maryland, USA

MENTAL HEALTH AND SOCIAL MEDIA

Mental health conditions affect a large percentage of individuals each year. Traditional mental health studies have relied on information collected through contact with the mental health practitioner. There has been research on the utility of social media for depression, but there have been limited evaluations of other mental health conditions (Jan-Are, Jan & Deede, 2002). First, we will examine specific techniques that have previously been used to analyze forum data, de ne behavioral health and public health issues, and lastly, we will explore the implications that this research has for big data analytics.

Analysis of Social Media

In this part of the paper, we explore the various techniques that have been previously used to analyze the data found in social media sites. The rise of social media sites, forums, blogs, and other communications tools has created an online community of individuals who are able to socialize and express their thoughts through various applications (Georgios & Mike, 2012). Microblogging has become a very popular tool for communication among users. The individuals who write these messages blog about their lives share opinions, and discuss current events. As more individuals participate in these micro blogging services, more information about their messages becomes available. The massive amount of data in user updates creates the need for accurate and efficient clustering of short messages on a larger scale (Chen & Liu, 2014). Certain research areas have chosen to focus on the opinions and sentiments of these messages (Si et al., 2014), community detection (Newman, 2004), politics (Tumasjan, Sprenger, Sandner, & Welpe, 2010), and user interests (Li et al., 2014). Techniques for clustering this data have included document clustering, topic modeling sentiment analysis, and text mining.

Topic modeling

Recent years have seen a surge in information that is both digitized and stored. As this trend continues, it has become increasingly di cult for users to find what they are looking for. Novel computational tools are needed to help organize, search, and comprehend these large amounts of data (Chen & Liu, 2014). Currently, we are able to type keywords into a search and find documents that are related to them. However, there is a crucial element that is missing from this process. Specifically, it is important to utilize themes to explore specific topics. A thematic structure could serve as a portal through which users could explore and obtain knowledge about various topics. Topic modeling algorithms are statistical methods that analyze the words of the original documents and discover themes that occur. Furthermore, topic modeling analyzes how these themes relate to one another, and how they differ over time (Blei, 2012). These algorithms do not need any previous annotations or labeling of the documents, these topics surface automatically form the analysis of the original texts. Blei (2012) describes latent Dirichlet allocation (LDA), which is the simplest type of topic model. LDA is a statistical model of a collection of documents that tries to validate the intuition that documents exhibit multiple topics. The simple LDA model provides an effective and powerful way to discover and exploit the hidden thematic structures

*Correspondence regarding this article should be directed to: gangopad@umbc.edu

found in large amounts of text data.

Sentiment analysis

Microblogging websites have developed into a source for varied types of information. Individuals post messages about their opinions, current events, complaints, and sentiments about products they use in their daily lives (Liu, 2012). It is very often that companies study these user reactions on microblogging sites. The challenge then becomes how to build a technology that can detect and summarize an overall sentiment. A large amount of social media contains sentences that are sentiment-based. Sentiment is defined as a personal belief or judgment that is not founded on proof or certainty (Davidov, Tsur & Rappoport, 2010). Sentiment involves the use of Natural Language Processing (NLP), statistics, or machine learning methods to extract, identify, or characterize the sentiment content of a text source (Liu, 2012). The automated identification of sentiment types can be beneficial for many NLP systems.

Text mining

Text mining is the discovery of new information by automatically extracting information from a large amount of various unstructured textual resources (Aggarwal & Zhai, 2012). Text mining can help an organization gain valuable insights from text-based content such as word documents, email, and postings on social media sites like Facebook, Twitter and LinkedIn (Rossi, Malliaros & Vazirgiannis, 2015). Mining unstructured data with natural language processing (NLP), statistical modeling and machine learning techniques can be challenging because natural language text is usually inconsistent. It contains ambiguities caused by inconsistent syntax and semantics. Text analytics software can help by transposing words and phrases in unstructured data into numerical values which can then be linked with structured data in a database and analyzed with traditional data mining techniques. By using text analytics, an organization can successfully gain insight into content specific values such as emotion, sentiment, intensity and relevance. Text mining techniques include methods for corpus handling, data import, metadata management, preprocessing, and the creation of term-document matrices. The main structure for managing documents in is a corpus, representing a collection of text documents.

Behavioral Health

Behavioral health can be classified into several different categories, depending on the type and severity of the mental health disorder. Mental health care practitioners rely on specific evaluation criteria, such as that contained in the Diagnostic and Statistical Manual of Mental Disorders (DSM), as well as data gathered from one-on-one sessions with the patient in order to reach a diagnosis for these disorders. Currently, over 61.5 million Americans experiences a mental illness in any given year. One in 17, about 13.6 million, have a serious mental illness such as major depression, schizophrenia, or bipolar disorder (Matthews, Abdullah, Gay, & Choudhury, 2014). About 60 percent of adults and almost one half of youth ages 8 to 15 with a mental illness did not receive mental health services in 2013 (Keating, Campbell & Radoll, 2013).

Suicide

Many individuals at risk of suicide do not seek help prior to an attempt, and they do not remain connected to any mental health services following the attempt (Abboute et al., 2014). E- 12 health interventions are now being defined as a means to identify individuals who are at risk, offer self-help, or deliver interventions in response to user posts on the internet. Patterns found in users' social media usage can be especially indicative of suicide ideation. Research shows that there is some evidence to suggest that social media platforms can be used to identify individuals or geographical areas at particular risk for suicide. Specific language used in tweets can give practitioners and other Twitter users information about an individual's mental health status. Recent studies found specific tweets by users who both tweeted about suicidal ideations. One quote stated \people say \stop cutting! be happy with who you are." It's so much easier to say than do? i hate myself so much.." (Burton , Giraud-Carrier & Hanson, 2014) Another tweeter posted, \I'm so sick of being bullied. Everyone care about their problems and don't even bother to check on me. I'm going to kill myself!!" (Burton , Giraud-Carrier & Hanson, 2014) It is evident from these tweets that intervention is possible. The few studies done in this area have shown that it is possible to use computerized sentiment analysis and data mining to identify users at risk for suicide.

Depression

Many have begun turning towards online communities for help in understanding and dealing with symptoms. Nimrod (2012) examined the content online forum discussion of depression in order to explore the potential benefits they could offer people with depression. Quantitative content analysis of one year of data from 25 top online communities was per-formed, using the Forum Monitoring System. Content analysis revealed nine main subjects discussed in the communities, including (in descending order) \symptoms", \relationships", \coping", \life", \formal care", \medications", \ causes", \suicide", and \work". The results indicated that online depression communities serve as a place for sharing experiences and receiving techniques for coping (Nimrod, 2012). Searching for online health information and searching within social media sites are both on-going difficulties users face (White & Horvitz , 2009). There are many reasons that these social media platforms are a valuable source of health information. For example, social media provides an important tool for people with health concerns to talk to one another. Also, these sites are well known as a source of tacit information, that

is less common online. Wilson et al. (2014) focused their study on a prevalent mental health issue, depression. Depression has increased substantially in developed and developing countries (BBC, 2013), and it is estimated to affect over 350 million people (WHO, 2015). Depression affects more than 27 million Americans and is believed to be responsible for more than 30,000 suicides every year (CDC, 2015; Luoma, Martin & Pearson, 2002). Although discussing issues related to depression with others is seen to be an important facet of coping, personal factors discourage people from doing so in real life (Back et al., 2010). Therefore, social media sites provide an outlet for people to communicate with potentially millions of people, while reducing the consequences of real life disclosure (Efron & Winget, 2010; Pak & Paroubek, 2010). More users are choosing to share their thoughts and emotions that encompass their daily lives. The language and emotion used in social media posts may include feelings of worthlessness, helplessness, guilt, and self-hatred, which are all characteristic of depression. The characterization of social media activity can provide a measurement of depression symptoms in a manner that could help detect depression in populations. Choudhury et al. (2013) examined the use of social media as a behavioral assessment tool. In contrast to behavioral health surveys, social media measurement of behavior captures social activity and language expression in a naturalistic setting (Choudhury, Gamon, Ho, & Roseway, 2013).

Preliminary analysis and results

We have conducted some preliminary analysis on online behavioral health forum that educates the public about responsible drug use by promoting free discussion. The identity of the forum is not divulged for privacy consideration. The purpose of our research was to develop an automated technique to understand the content of such discussion forums automatically. The forum data was modeled as a graph which was partitioned into homogeneous groups/themes where each theme contained the discussion threads with strong correlations with each other measured by co-occurrence of common terms. The partitioning method followed our previous work described in (Yesha, Gangopadhyay & Siegel, 2015). The partitioned graph, shown in Figure 1, consisted of around 1000 nodes with around 120,000 edges, which indicates a much larger value for the average degree per node as compared to other real world networks (Leskovec, Kleinberg & Faloutsos, 2007). The nodes of each partition in the graph are shown in a different color and each partition represented
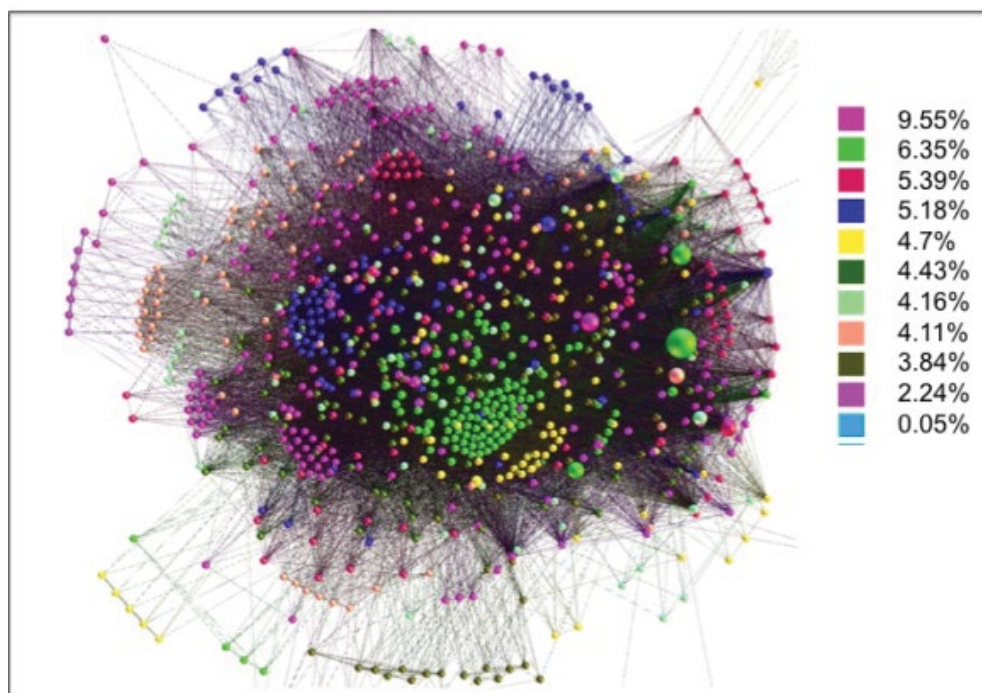

Figure 1. Partitioned Graph

a different theme. We describe the themes corresponding to the top five partitions of the graph shown in Figure 1. For example, the first partition contains personal experience and recommendations such as the Linden method for dealing with anxiety, panic attacks. The second partition was focused on other drugs such as Xanax and Benzodiazepine. The third partition contained clinical issues such as the discovery of cannabinoid receptors (link is external) in the amygdala. The other two partitions discussed semi-religious issues and a social interaction on a BBC documentary.

CONCLUSIONS

The prevalence of online social networks has enabled users to communicate, connect, and share content. This presents an unprecedented opportunity to extract the patterns that are hidden in the increasingly voluminous amounts of text in social media. Such patterns can be useful to users, clinicians, and researchers alike to determine the underlying factors that affect individuals, identifying the proper forum and searching for specific discussion threads. As an application of data analytics this presents challenges in dealing with big data, data visualization, pattern recognition, document clustering, and information retrieval.

REFERENCES

Abboute, A., Boudjeriou, Y., Entringer, G., Aze, J., Bringay, S., & Poncelet, P. (2014). Mining twitter for suicide prevention. In *Natural Language Processing and Information Systems - 19th International Conference on Applications of Natural Language to Information Systems, NLDB 2014, Montpellier, France, Proceedings*, pp. 250-253.

Back, M., Stopfer, J., Vazire, S., Gaddis, S., Schmukle, S., Egloff, B., et al. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*(3), 372-374.

BBC (2013). Available online at http://www.bbc.com/news/health-23192252.

CDC (2015). Available online at http://nccd.cdc.gov/s broker/WEATSQL.exe/weat/freq/ year.hsql.

Charu, C., Aggarwal, & Cheng, X.Z. (2012). Editors. Mining Text Data. *Springer*.

Chen, Z., & Liu, B. (2014). *Topic modeling using topics from many domains, lifelong learning and big data*. In Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, pp. 703-711.

Choudhury, M.D., Gamon, M., Hoff, A., & Roseway, A. (2013). "*Moon phrases": A social media faciliated tool for emotional reection and wellness*. In 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2013, Venice, Italy, pp. 41-44.

David, M.B. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84

Davidov, D., Tsur, O., & Rappoport, A. (2010). *Enhanced sentiment learning using twitter hashtags and smileys*. In COLING 2010, 23rd International Conference on Computational Linguistics, Posters Volume, 23-27 August 2010, Beijing, China, pp. 241-249.

Efron, M., & Winget, M. (2010). *Questions are content: A taxonomy of questions in a microblogging environment*. In Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47, ASIS&T '10, Silver Springs, MD, USA, 2010, *27*, 1-10.

Jashinsky, J., Burton, S.H, Hanson, C., West, J., Giraud-Carrier, C., Barnes, M.D., et al., (2014). Tracking suicide risk factors through twitter in the us. *Crisis, 35*(1) 51-59.

Johnsen, K.J., Rosenvinge, H.J., & Gammon, D. (2002). Online group interaction and mental health: An analysis of three online discussion forums. *Scandinavian Journal of Psychology, 43*(5), 445-449.

Keating, W.B., Campbell, A.J., & Radoll, P. (2013). *Evaluating a new pattern development process for interface design: Application to mental health services*. In Proceedings of the International Conference on Information Systems, ICIS 2013, Milano, Italy, December 15-18.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). *Graph evolution: Densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data, 1*(1).

Li, H., Mukherjee, A., Liu, B., Kornfield, R., & Emery, S. (2014). *Detecting campaign promoters on twitter using markov random fields*. In 2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, pp. 290-299.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

Luoma, J., Martin, C., & Pearson, J. (2002). Contact with mental health and primary care providers before suicide: a review of the evidence. *American Journal of Psychiatry, 159*(6), 909-916.

Matthews, M., Abdullah, S., Gay, G., & Choudhury, T. (2014). Tracking mental well-being: Balancing rich sensing and patient needs. *IEEE Computer, 47*(4), 36-43.

Newman, M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review*.

Nimrod, G. (2012). From knowledge to hope: online depression communities. *International Journal o Disability and Human Development, 11*(1), 23-30.

Pak, A., & Paroubek, P. (2010). *Twitter as a corpus for sentiment analysis and opinion mining*. In Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.

Paltoglou, G., & Thelwall, M. (2012). Twitter, myspace, digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology, 3*(4), *66*, 1-19.

Rossi, G.M., Malliaros, D.F., & Vazirgiannis, M. (2015). *Spread it good, spread it fast: Identification of influential nodes in social networks*. In Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, pp. 101-102.

Si, J., Mukherjee, A., Liu, B., Pan, J.S., Li, Q., & Li, H. (2014). *Exploiting social relations and sentiment for stock prediction*. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, pp. 1139-1145.

Tumasjan, A., Sprenger, O.T., Sandner, G.P., & Welpe, M.I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *International Conference on Weblogs and Social Media, 10*, 178-185.

White, W.R., & Horvitz, E. (2009). Cyberchondria: Studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems, 27*(4).

WHO (2015). Available online at http://www.who.int/mental health/prevention/ suicide/wspd/en/.

Wilson, R., Capuano, A., Boyle, P., Hoganson, G., Hizel, L., Shah, R., et al. (2014). Clinical-pathologic study of depressive symptoms and cognitive decline in old age. *Neurology, 83*(8), 702-709.