

An Overview of Big Data Applications in Mental Health Treatment

Neil Eliason
Indiana University Online
Anderson, Indiana 46012
nreliaso@iu.edu

ABSTRACT

Mental health treatment presents with complex informational challenges, which could be effectively tackled with big data techniques. However, as researchers and treatment providers explore these applications, they find a lack of infrastructure and ethical concerns hamper their progress. A unified approach of developing an ethically informed data infrastructure is necessary to proceed.

KEYWORDS

i523, Mental Health Treatment, Big Data, Data Analytics, Data Infrastructure, Data Ethics

1 INTRODUCTION

1.1 Big Data

There is no immutable or standardized definition of big data. However, most conceptualizations include data with high volume (amount of data stored), velocity (frequency of data input or update), and/or variety (number of data sources or types), known as the “three v’s”. As these factors increase, they reach the so called “three v tipping point”, where traditional methods of analysis do not meet operational needs. Here, big data analytic techniques are utilized to make these unruly collections of data useful. For example, text mining, audio analytics, video analytics, and social media analytics are specific techniques used to make low value data more organized, condensed, and useful. Then predictive analytics take this processed data, and create data models which can predict future outcomes. These can be divided into regression techniques, which identify ways groups rely on each other, and machine learning techniques, which look for patterns in validated test data and then apply them to an unvalidated sample [3].

1.2 Mental Health Treatment

Mental health difficulties are a common problem across the United States, and worldwide. Mental illness of some kind was prevalent among 17.9 % of Americans in 2015, and of that number 4% experienced serious functional impairment as a result [14]. A 2014 meta-analysis study estimated that the worldwide prevalence of mental illness was 17.6% and that 29.2% of the world population would experience mental illness at some point during their life [18]. The effects of these disorders on individuals and societies is costly. The US Center for Disease Control and Prevention estimated that 36,035 people died during a suicide attempt in 2008, and that 666,000 sought emergency room care for self harming behavior [2]. In 2013, the Social Security Administration reported that 1,947,775 persons received social security/disability benefits for either a mood or psychotic disorder, which is around 19% of all recipients [16]. It

is estimated that mental health issues had a \$100 billion cost on the US economy in 2002 [14], and in 2015 there were over 12,000 mental health treatment facilities in the US [20].

Mental health treatment attempts to address these pervasive and complex problems at an individual level. While this by nature results in a system that is heterogeneous and complex, treatment still follows a fairly consistent pattern. First the mental health issue is identified [1], then treatment interventions are assigned [21], and finally treatment progress is monitored [4].

The identification process involves mental health screening and assessment. Screening attempts to identify a person’s primary mental health risks and needs for the purpose of directing them to appropriate sources. They tend to be narrow in focus and brief, which allows them to be easily disseminated to help filter people to the right level of care. Similar to screening, assessment aims to identify a person’s mental health dysfunction, but does so in more clinically robust categories, typically resulting in a diagnosis [1]. Once a person’s mental health issues have been clinically identified, then interventions are assigned. Those traditionally take the form of talk-therapy to develop effective change strategies, medication to reduce symptoms of mental illness, and supportive services such as case management to help coordinate efforts towards the person’s goals [21]. Treatment monitoring is essential to the treatment life-cycle, as this is where clinicians receive feedback regarding the effectiveness of the chosen interventions. While it is natural for clinicians to do this informally, more intentional methods are often overlooked [4]. This process requires an extensive data gathering effort, which traditionally is labor intensive and requires a large team of clinicians.

1.3 Thesis

There are large numbers of people struggling with mental illness, and their treatment requires large amounts of frequent data from various sources. This process as traditionally done is inefficient and labor intensive. Big data analytic techniques are designed to target this kind of data, and could greatly increase treatment effectiveness and scope.

2 BIG DATA APPLICATIONS IN MENTAL HEALTH TREATMENT

2.1 Screening and Diagnosis

Mental health screening is the first chance to direct people in the appropriate direction to meet their mental health needs. Methods that can screen larger amounts of people effectively are critical, as many people with mental illness are not connected with treatment. Several studies explored using social media to identify mental illness in the general population, and demonstrated potential to identify

issues on a large scale. Many attempted to identify depression by analyzing the content of social media posts, and to create a predictive model which would predict variables of interest from dependent variables. By using public data from Twitter or mental health forums large sample sizes were possible, but also resulted in less reliable data. It is estimated that the ability to detect depression by machine driven predictive models running on big social media data was above that of unaided primary care clinicians, but below that of self-report surveys. [5].

Clinical assessment and diagnostic assignment follows screening. There is considerable interest in developing more effective diagnostic assessment using big data analytics. Models were created using techniques such as data mining, machine learning, and natural language processing to group people into diagnostic categories based on data from a variety of sources. [12]. In bipolar research, machine learning algorithms looked for patterns in neuroimaging, genetic analysis, neuropsychological tests, and protein biomarkers. They were able to create predictive models, but their performance was not greater than current diagnostic systems. While this task could not be completely automated via big data analytics any time soon, it may inform clinical diagnosis in the short-term [8].

Predictive models using machine learning techniques are also being constructed from a variety of data sources to estimate patient outcomes, which could be helpful in selection of interventions at the onset of treatment. [12] Predictive risk profiles for patient's with bipolar disorder were created by taking data from Electronic Medical Records and identifying patient characteristics connected to negative outcomes, such as relapse and hospital admission. Studies also explored models which predict patient mood states, based on past monitoring data and how patients will respond to specific interventions. While these examples were fairly accurate (68% to 99%), they were based on relatively small sample sizes [8]. Predictive models show promise of being an effective big data application in mental health treatment, but require further advances in machine learning techniques and validated on larger samples before they can be widely administered [12].

2.2 Interventions

Once a person's mental health issues have been clinically identified, then interventions are assigned. Traditional interventions are clinician driven, and are often limited in scope by clinician availability. Web-based interventions, which provide treatment activities via web-browser, have the potential to provide more flexible treatment options for patients. Initial attempts have seen some success, particularly if paired with a human coach. Few estimates of effectiveness exist, as these techniques have not been applied to large groups [10]. While big data approaches are not widely utilized, there is interest in using machine learning to predict content that a particular user would find helpful [11], which is a technique called a recommender system [17]. Also, as interactive interfaces are developed and used by large numbers of online users [10], big data analytics would be beneficial.

2.3 Treatment Monitoring

As a person receives treatment, tracking progress towards their goals is critical. Traditionally this is done by patient report via a

tracking log or by clinician inquiry during a session, and is often hindered by a lack of patient engagement. One solution to this is active monitoring utilizing mobile devices. Utilizing text message or application notifications, treatment goal reminders, symptom assessment questions, or encouraging messages are sent to the treatment participant [11]. Feedback from the patient can come in various forms from filling out a survey to voice response, and may be collected multiple times a day. The frequent collection of different types of data make active monitoring an application which could benefit from a big data approach. However, trouble with integrating data into the electronic medical record and a lack of widespread utilization have prevented such approaches from being extensively applied or reliably tested [12].

Another possibility is passive monitoring, which would access information from a mobile device, and connect those to patient behaviors, without any intentional action on the patient's part. This has been done using clinically informed algorithms or machine learning paired with self-report [11]. Devices used were not just smartphones, but including wearables and a sensor which is swallowed to detect medication adherence. Active monitoring has generated considerable research interest, but implementation at a big data level is challenged by lack of client engagement, clinician's ability to use, and difficulties integrating the large quantities and varieties of data [12].

3 DISCUSSION

3.1 Barriers

Overall, there is considerable interest in developing big data applications at every stage of the mental health process. However, this development has been slow and halting due to a number of issues inherent though not necessarily unique to human services.

For example, the issue of privacy is relevant with many big data applications, but in mental health the sensitive nature of an individual's mental health treatment data creates new difficulties. Typically privacy is preserved through de-identification of the data, but this is not always effective with large-scale data [12]. A specific privacy risk is big data analysis of social media, which captures large amounts of information, which can be used to infer mental health status [5]. When mental health privacy is breached, discrimination regarding employment, insurance, housing, etc. are possible [12]. On the other side of the privacy question, mental health professionals are mandated to report if someone is an imminent risk to themselves or others. Currently, there are no clear guidelines to follow, if this is discovered through public data [5].

Another challenge to capitalizing on big data is the variety of data sources, formats, and storage locations. The vast majority of mobile devices are not run on open source software, as they are sold as commercial products. This hinders collaboration and integration of the data with sources from other companies' products [13]. It is also unclear who owns the data in these situations, causing more disruption [12]. This is not just the case with private data. Large databases and research institutions often struggle to share data, and the decision to do so is often up to the individual researchers. This prevents the collaboration and coordination required to make good use of the available big data opportunities [6].

3.2 Future Directions

Considerable attention is being given to big data applications in mental health treatment, and some major initiatives seek to address some of the technical issues mentioned previously. The National Institute of Health's Office of Behavioral and Social Sciences Research has a strong focus on big data in its 2017 to 2021 strategic plan. It specifically called for the development of "data infrastructure that promotes data sharing, harmonization, and integration", and also to develop research methods which are designed for "data-rich" science [15]. There is a related call for treatment to inform research questions, and research questions to inform the structure and collection of big data, as opposed to primarily opportunistic research, which studies data that is most convenient [19]. The integration of private commercial data for big data analytics is also a goal of some researchers [13]. Concerning specific technologies, there is generally great optimism that the big data analytics techniques will continue to be refined, and that wider implementation will result in greater strides in treatment effectiveness.

Most of the research reviewed ended with a short description of ethical concerns in big data use for mental health treatment, and a call for someone to look into this in more detail. The problem is that there is a wide variety of perspectives about this topic. Some operate from the assumption that if data is publicly accessible, that resolves any privacy issues. Others point out cases where individual's privacy was seriously compromised by comparing data from multiple public databases [9]. This is a point where public policy has fallen behind technological innovation. An inter-disciplinary effort from legal, data science, and mental health experts may be required to strike the balance between science and citizen security [7].

4 CONCLUSION

At every stage, mental health treatment is a data intensive task. As electronic medical records, social media, and mobile devices continue to increase in data collection and storage capabilities, data relevant to mental health continues to grow larger, faster, and more varied. Many researchers and practitioners are eager to use big data analytics to tap into the potential insights of these data sets.

The first steps of development have already started, and show promise of making a significant positive impact in the field. Predictive analytics are being tested to screen for people with mental illness via social media, and machine learning techniques are being applied to improve the resolution of diagnosis and to inform treatment assignments through outcomes prediction. Though these results need replication with larger samples, they already demonstrate predictive power, which could soon equate with improved treatment in practice.

Applications utilizing mobile devices for active and passive monitoring of treatment participants are generating considerable attention, but are only early in development. As this approach is expanded to larger samples, big data analytics will be critical to managing the velocity and variety of data coming from smartphones and wearables. Integrating big data analytics in web-based mental health interventions, is even earlier in development. The potential to create interactive interfaces, utilizing artificial intelligence and

recommender systems is present, but currently web-based treatments are being tested themselves for viability.

While progress to develop algorithms and programs to process mental health big data continues, it is hindered by the current limitations of data infrastructure and research culture. Though large data sources are available, they are not integrated with one another, and are often prevented from doing so due to preferences of individual researchers or from corporate interest. The National Institute of Health and many researchers are calling for an integrated and open data sharing framework to address this issue.

Also of concern is a variety of ethical questions involved in applying big data analytics to mental health. Ownership of data is not well defined, and often data is sold and studied without the knowledge of its subjects. During this process, an individual's privacy may be compromised, even with de-identified data. This can lead to discrimination and stigma for the individual whose mental health data has been unmasked. While this problem is readily recognized, no major policy or legislative change has adequately addressed it.

As big data analytics continues to mature, mental health treatment should seek to benefit from the unlocking of new knowledge and insights. However, this cannot be done without consideration of how to create an environment that simultaneously encourages practice innovation and patient protection. Treatment seeks to provide effective help to those with mental illness, and big data may help with that aim, but to do this at the expense of the patient rights undermines any help they hoped to gain.

ACKNOWLEDGMENTS

The researcher would like to thank Professor Gregor von Laszewski, along with Teaching Assistants Juliette Zerick, Saber Sheybani Moghadam, and Miao Jiang, and the anonymous reviewers who helped with the present work.

REFERENCES

- [1] APA Practice Organization. 2017. Distinguishing Between Screening and Assessment for Mental and Behavioral Health Problems. Webpage. (2017). www.apapracticentral.org/reimbursement/billing/assessment-screening.aspx
- [2] Alex E Crosby, Beth Han, LaVonne A G Ortega, Sharyn E Parks, and Joseph Gfroerer. 2011. Suicidal thoughts and behaviors among adults aged 18 years–United States, 2008–2009. *Morbidity And Mortality Weekly Report. Surveillance Summaries (Washington, D.C.: 2002)* 60, 13 (2011), 1 – 22. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=22012169&site=eds-live&scope=site>
- [3] Amir Gandomi and Murtaza Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35, 2 (2015), 137 – 144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [4] Jessica D. Goodman, James R. McKay, and Dominick DePhilippis. 2013. Progress monitoring in mental health and addiction treatment: A means of improving care. *Professional Psychology, Research and Practice* 44, 4 (2013), 231. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsggo&AN=edsgcl.354463723&site=eds-live&scope=site>
- [5] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18 (2017), 43 – 49. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost-com.proxyiub.uits.iu.edu/login.aspx?direct=true&db=edselp&AN=S2352154617300384&site=eds-live&scope=site>
- [6] Diego Hidalgo-Mazzei, Andrea Murru, Mara Reinares, Eduard Vieta, and Francesc Colom. 2016. Big Data in mental health: a challenging fragmented future. *World Psychiatry: Official Journal Of The World Psychiatric Association (WPA)* 15, 2 (2016), 186 – 187. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=27265716&site=eds-live&scope=site>

- [7] Sharona Hoffman. 2015. CITIZEN SCIENCE: THE LAW AND ETHICS OF PUBLIC ACCESS TO MEDICAL BIG DATA. *Berkeley Technology Law Journal* 30, 3 (2015), 1741 – 1806. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=115393178&site=eds-live&scope=site>
- [8] Diego Librenza-Garcia, Bruno Jaskulski Kotzian, Jessica Yang, Benson Mwangi, Bo Cao, Luiza Nunes Pereira Lima, Mariane Bagatin Bermudez, Manuela Vianna Boeira, Flvio Kapczynski, and Ives Cavalcante Passos. 2017. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neuroscience and Biobehavioral Reviews* 80 (2017), 538 – 554. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edselp&AN=S0149763417300337&site=eds-live&scope=site>
- [9] Jacob Metcalf and Kate Crawford. 2016. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society* 3, 1 (2016), 2053951716650211. <https://doi.org/10.1177/2053951716650211> arXiv:<https://doi.org/10.1177/2053951716650211>
- [10] Thomas D. Meyer, Rebecca Casarez, Satyajit S. Mohite, Nikki La Rosa, and M. Sriram Iyengar. 2018. Novel technology as platform for interventions for caregivers and individuals with severe mental health illnesses: A systematic review. *Journal of Affective Disorders* 226, Supplement C (2018), 169 – 177. <https://doi.org/10.1016/j.jad.2017.09.012>
- [11] David C. Mohr, Michelle Nicole Burns, Stephen M. Schueller, Gregory Clarke, and Michael Klinkman. 2013. Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health. *General Hospital Psychiatry* 35, 4 (2013), 332 – 338. <https://doi.org/10.1016/j.genhosppsych.2013.03.008>
- [12] Scott Monteith, Tasha Glenn, John Geddes, Peter C. Whybrow, and Michael Bauer. 2016. Big data for bipolar disorder. *International Journal of Bipolar Disorders* 4, 1 (11 Apr 2016), 10. <https://doi.org/10.1186/s40345-016-0051-7>
- [13] Andrea Murru, Eduard Vieta, and Francesc Colom. 2016. Big Data in mental health: a challenging fragmented future. *World Psychiatry* 15, 2 (2016), 186. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edssoaf&AN=edssoaf.d1b40726336430fced614a54ac6d4b079039a17&site=eds-live&scope=site>
- [14] National Institute of Mental Health. 2017. (2017). <https://www.nimh.nih.gov/health/statistics/index.shtml>
- [15] William T Riley. 2017. Behavioral and social sciences at the National Institutes of Health: Methods, measures, and data infrastructures as a scientific priority. *Health Psychology: Official Journal Of The Division Of Health Psychology, American Psychological Association* 36, 1 (2017), 5 – 7. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=28045300&site=eds-live&scope=site>
- [16] Social Security Administration. 2013. (2013). <https://www.ssa.gov/policy/docs/statcomps/di.asr/2013/di.asr13.pdf>
- [17] Stanford Info Lab. 2017. Recommendation Systems. webpage. (2017). <http://infolab.stanford.edu/~ullman/mmds/ch9.pdf>
- [18] Z. Steel, C. Marnane, C. Iranpour, Tien Chey, J. W. Jackson, Patel Vikram, and D. Silove. 2014. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *International Journal of Epidemiology* 43, 2 (2014), 476 – 493. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=llh&AN=20143278163&site=eds-live&scope=site>
- [19] Robert Stewart and Katrina Davis. 2016. 'Big data' in mental health research: current status and emerging possibilities. *Social Psychiatry and Psychiatric Epidemiology* 51, 8 (2016), 1055. <http://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsgeo&AN=edsgeo.460296829&site=eds-live&scope=site>
- [20] Substance Abuse and Mental Health Services Administration. 2015. (2015). <https://www.samhsa.gov/data/sites/default/files/2015.National.Mental.Health.Services.Survey.pdf>
- [21] Substance Abuse and Mental Health Services Administration. 2017. Behavioral Health Treatments and Services. (2017). <https://www.samhsa.gov/treatment>

5 BIBTEX ISSUES

6 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

6.1 Assignment Submission Issues

Do not make changes to your paper during grading, when your repository should be frozen.

6.2 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

6.3 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

Other formatting issues

6.4 Writing Errors

Errors in title, e.g. capitalization

Spelling errors

Are you using *a* and *the* properly?

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

Do not use the word *I* instead use *we* even if you are the sole author

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

If you want to say *and* do not use & but use the word *and*

Use a space after . , :

When using a section command, the section title is not written in all-caps as format does this for you

\section{Introduction} and NOT \section{INTRODUCTION}

6.5 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Claims made without citations provided

Need to paraphrase long quotations (whole sentences or longer)

Need to quote directly cited material

6.6 Latex Errors

Erroneous use of quotation marks, i.e. use “quotes”, instead of ” ”

To emphasize a word, use *emphasize* and not “quote”

When using the characters & # % _ put a backslash before them so that they show up correctly

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}
```

6.7 Structural Issues

Acknowledgement section missing

Incorrect README file

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

Do not artificially inflate your paper if you are below the page limit

6.8 Details about the Figures and Tables

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

Do use *label* and *ref* to automatically create figure numbers

Wrong placement of figure caption. They should be on the bottom of the figure

Wrong placement of table caption. They should be on the top of the table

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

Do not submit eps images. Instead, convert them to PDF

The image files must be in a single directory named "images"

In case there is a powerpoint in the submission, the image must be exported as PDF

Make the figures large enough so we can read the details. If needed make the figure over two columns

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

Do not use `textwidth` as a parameter for `includegraphics`

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.