

Dealing with medication non-adherence expressions in Twitter

Takeshi Onishi[†], Davy Weissenbacher[‡], Ari Klein[‡], Karen O'Connor[‡], Graciela Gonzalez[‡]

[†]Toyota Technological Institute at Chicago, 6045 South Kenwood, Chicago, IL

[‡]University of Pennsylvania, Philadelphia, PA

[†]tonishi@tttic.edu

[‡]{dweissen, ariklein, karoc, gragon}@pennmedicine.upenn.edu

Abstract

Through a semi-automatic analysis of tweets, we show that Twitter users not only express Medication Non-Adherence (MNA) in social media but also their reasons for not complying; further research is necessary to fully extract automatically and analyze this information, in order to facilitate the use of this data in epidemiological studies.

1 Introduction

Past studies (Claxton et al., 2001) have shown that 50% of medications are not taken as prescribed by patients. This Medication Non-Adherence (MNA) increases morbidity and mortality with an estimated cost of 100–239 billion per annum to the US healthcare system. The patients' reasons to not comply with treatments are of diverse nature, such as high price for a drug or its negative adverse effect, not trusting the medication, or because they feel better or forget. Healthcare providers have to understand such reasons in order to influence patients' behavior. A major challenge, however, is the difficulty in identifying such reasons. Traditional methods, such as mining clinical records or using pharmacy claims data, or interacting directly with patients through surveys and intervention trials have been found limited to identify MNA reasons (Xie et al., 2017).

With the large adoption during the last decades of Social Media (SM) and the proneness of the SM users to discuss medical habits and share health issues, SM is increasingly regarded as an important source that can provide unique insights into Medication Non-Adherence reasons. For our study, we have chosen Twitter due to the large volume of easily-accessible data.

In this work, through a semi-automatic analysis of 4 million tweets, we show that not only Twitter users clearly express MNA in the social media

but also their reasons for not complying with their treatments, which calls for further research to fully automatize our process.

2 Methods

Tweets mention medications in various contexts such as advertising/selling drugs or personal drug experiences. Typically, accounts owned by a company or organization advertise and sell drugs, and individual persons post personal drug experiences such as their prescriptions, their reaction to the drugs, and sometimes their non-adherence.

To determine if Twitter users are mentioning their non adherence to their treatment and their reasons, we manually analyzed an existing corpus of four millions tweets, the Pregnancy corpus. This corpus is composed of ~112,500 timelines¹ of women posting during their pregnancy and collected for the needs of a previous epidemiologic study (Golder et al., 2018).

Two independent methods detected tweets mentioning a MNA. The methods rely on different features related to MNA and were applied in parallel.

Drug names matching: We compiled a list of 103 distinct names of drugs related to HIV and diabetes from Drug.com² and eMEDTV³. These lists include generic and brand names. We filtered out all tweets which did not contain any drug name from the drug list. We, then, removed all tweets containing a hyperlink, retweets, reply tweets and tweets not written in English. These heuristic rules were inspired by Adrover et al. (2015) and were based on the observation that a majority of tweets containing drug names and a hyperlink were posted by companies commenting web articles, whereas tweets posted by individuals

¹We call a *timeline* the exhaustive set of tweets posted by a user during a given period.

²<https://www.drugs.com>

³<http://cholesterol.emedtv.com>

	Drug	Pattern
Tweets matched	377	27
Tweets mentioning MNA	9	9
Reason in the tweet	6	8
Reason in the tweet vicinity	0	1

Table 1: Tweets mentioning a MNA and its reason manually discovered in tweets retrieved by Drug names and Patterns matching methods.

describing their experiences about drugs did not contain hyperlinks.

Patterns matching: We encoded our patterns in REs and searched for all tweets in the corpus. For this preliminary study, we searched for two patterns: all tweets which contain both phrases “stopped taking” and “made me”, regardless of the order. The previous heuristic rules, used to remove tweets posted by bots or companies, were not applied on the tweets retrieved by the patterns since, due to the semantic of the patterns, the tweets they retrieved were personal tweets.

Two annotators independently investigated the tweets obtained by both methods. Each annotator judged if the tweets were mentioning a MNA or not for precision. The recall was not estimated because MNA tweets are rare and estimating such frequency even from random samples is practically impossible. For the tweet mentioning an MNA, they looked for the reasons in the users’ timelines up to ten days before and after the MNA tweet. A third annotator resolved the disagreements.

3 Results

Table (1) details our results. Despite the limited number of drug names and the small size of our corpus, the drug names matching method retrieved 377 tweets including nine tweets mentioning an MNA. Six of the nine tweets were also describing the reason of the MNA in the tweet. The patterns matching method retrieved 27 tweets including nine MNA tweets. The 27 tweets are exclusive to the 377 tweets retrieved by the first method. The precision of the pattern matching is 9/27 which appears to be more precise compared to that of the drug name matching (9/377). Of these nine tweets retrieved by the pattern matching, one specifies a medication, two specifies a type of medication (*e.g.*, pain medication), and the other six use a generalization (*e.g.*, pills) or a pronoun to

refer to a medication mentioned elsewhere in the users’ timelines. Due to the patterns searched, all of the tweets also mention the reasons, except for one tweet that is truncated and mentions the reason in the subsequent post. The other 18 tweets retrieved by the patterns either did not refer to a type of medication (*e.g.*, birth control, prenatal vitamins) or used generalizations or pronouns to refer to medications for which we did not discover the referent.

4 Conclusion

Two semi-automatic processes successfully isolated 18 tweets in total from four millions of tweets where Twitter users explicitly report their MNA. Additionally, we found that users are also more likely to explain their failure to comply in the same MNA tweets or in the following tweets. These results showed the potential of Twitter for understanding patients’ behavior at a large scale and justify further research to extract and analyze automatically the MNA reasons. To increase the number of tweets retrieved, we will listen in real-time tweets from the stream of Twitter, searching for all drugs names using a Drug Name Recognizer and a manually expended set of patterns.

References

- Cosme Adrover, Todd Bodnar, Zhuojie Huang, Amalio Telenti, and Marcel Salathé. 2015. Identifying Adverse Effects of HIV Drug Treatment and Associated Sentiments Using Twitter. *JMIR Public Health and Surveillance*, 1(2).
- Ami J. Claxton, Joyce Cramer, and Courtney Pierce. 2001. A Systematic Review of the Associations Between Dose Regimens and Medication Compliance. *CLINICAL THERAPEUTICS*, 23(8):1296–1310.
- Su Golder, Stephanie Chiuve, Davy Weissenbacher, Ari Klein, Karen O’Connor, Martin Bland, Murray Malin, Mondira Bhattacharya, Linda Scarazzini, and Graciela Gonzalez-Hernandez. 2018. Pharmacoepidemiologic evaluation of birth defects from health-related postings in social media during pregnancy. *Drug Safety [Submitted]*.
- Jiaheng Xie, Xiao Liu, Daniel Dajun Zeng, and Xiao Fang. 2017. Understanding medication nonadherence from social media: A sentiment-enriched deep learning approach. *SSRN Electronic Journal*.