Digital Healthcare Empowering Europeans
R. Cornet et al. (Eds.)

doi:10.3233/978-1-61499-512-8-55

Exploring Brand-Name Drug Mentions on Twitter for Pharmacovigilance

Pablo CARBONELL[a,1], Miguel A. MAYER[a] and Àlex BRAVO[a]
[a] *Research Programme on Biomedical Informatics (GRIB), IMIM-Universitat Pompeu Fabra, Barcelona, Spain*

Abstract. Twitter has been proposed by several studies as a means to track public health trends such as influenza and Ebola outbreaks by analyzing user messages in order to measure different population features and interests. In this work we analyze the number and features of mentions on Twitter of drug brand names in order to explore the potential usefulness of the automated detection of drug side effects and drug-drug interactions on social media platforms such as Twitter. This information can be used for the development of predictive models for drug toxicity, drug-drug interactions or drug resistance. Taking into account the large number of drug brand mentions that we found on Twitter, it is promising as a tool for the detection, understanding and monitoring the way people manage prescribed drugs.

Keywords. Social Media, drug safety, Internet, prescription drugs

Introduction

Twitter is an Internet micro-blogging social media service that allows users to post short messages (140 characters) about facts, feelings and opinions but also as several studies show, users' health conditions [1]. In addition, Twitter has been proposed by several works as a means to track public health trends such as influenza outbreaks and more recently about Ebola by analyzing user messages in order to measure different population features and interests [2, 3, 4].

One promising application of social media platforms and in particular Twitter is to use them as tools for the detection, understanding and monitoring the way people manage prescribed drugs [5] and if it is the case, how Twitter users mention drug side effects or possible drug-drug interactions and therefore if this information can be considered as another channel of information of drug safety or pharmacovigilance [6, 7, 8].

In this work we proposed a method for the automated measurement of the number and features of drug brand names tweet mentions, using the Twitter API [9] and the Drugbank database, illustrating the potential usefulness of the analysis and detection of drug side-effects and drug-drug interactions on social media platforms such as Twitter.

[1] Pablo Carbonell. Email: pablo.carbonell@upf.edu

1. Methods

We compiled a list of drug brand names for FDA approved drugs from the Drugbank database [10]. In order to avoid false hits and overly similar repetitions, names for very similar sounding brand names for the same drug were removed, as measured by a Levenshtein distance of 3. In addition, common English names were filtered out by removing any word present in the Wordnet database [11]. After that filtering, 8,368 drug names corresponding to 1,242 substances were retained for screening. A list containing 27,246 drug-drug interactions and the classification of the drugs into 585 drug categories were obtained from the Drugbank database.

We used the twitter API in Python 2.7.8 to download, in periods of 1 week, mentions in that social media matching the list of drug names. When the rate of downloaded mentions for one drug name exceeded 1,000 tweets per hour, the word used was considered too common and discarded. Statistics were obtained by using R. Time series analysis was performed using the zoo package [12]. Each drug name was associated with a regularly sampled time series by counting the number of mentions within intervals of 30 minutes. We considered a time series as information-rich when it contained at least 100 tweets and no more than 10 empty time sample intervals. Correlation calculations between time series were restricted to only information-rich series.

The resulting messages on Twitter were mined using BeFree System [13], a text mining tool for information extraction. BeFree is composed of a module for Biomedical Named Entity Recognition (BioNER) [14] based on dictionaries using fuzzy and pattern matching methods to find and uniquely identify entity mentions in the biomedical literature, and a module for Relation Extraction (RE) based on Support Vector Machine (SVM). For this study, we only used the BioNER module for diseases including a lengthy diseases dictionary collected from the Unified Medical Language System (UMLS) [15]. Finally, message hits were crossed with drug categories to allow further analysis.

2. Results

For the present study, the period covered was a 3-week interval from October 6th, 2014 to October 27th, 2014. The total number of mentions that were downloaded was 1,456,961, corresponding to 946 drugs and 2,406 names in 53 languages, English being the most often used with approximately 30% of messages. The highest number of tweets (86,969) was posted on Monday, October 14th, while the lowest rates on the observed period occurred on Saturdays. After filtering out false hits as described in Methods, 99,485 tweets were kept for analysis. Drug names were associated with 390 categories. As shown in Table 1, anti-bacterial agents were the ones with the highest number of mentioned substances (85), followed by anti-inflammatory agents (80) and antineoplastic agents (75). The number of tweets associated with these three categories during the analyses period was of 1,140, 2,938 and 8,906, respectively.

We analyzed co-evolution of drug mentions on Twitter to test if those pairs of drugs that were often found simultaneously mentioned could correspond to cases of known drug-drug interactions. To that end, we associated each drug with a regularly sampled time series that followed drug mentions evolution as described in Methods. For each series, we determined its best correlation with the rest of drug series. In total,

we considered 151 drugs time series containing 114 pairs of drug-drug interactions within 32 drugs. In Figure 1, the distribution of the best correlations for each drug is compared with those that corresponding to drug-drug interactions, observing that, in general, high correlation between drug mentions could not be associated with the existence of a drug-drug interaction.

Table 1. Top drug categories (based on MeSH terminology) associated with mentioned drugs in function of the number of substances and number of counted tweets for each category

Drug category (MeSH term)	No. of substances	No. of tweets
Anti-Bacterial Agents	85	1140
Anti-inflammatory Agents	80	2938
Antineoplastic Agents	75	8906
Anti-anxiety Agents	73	8965
Hypnotics and Sedatives	73	9707
Antihypertensive Agents	70	1561
Anticonvulsants	64	9993
Antiemetics	60	16868
Vasodilator Agents	60	2040
Anti-Allergic Agents	58	9462

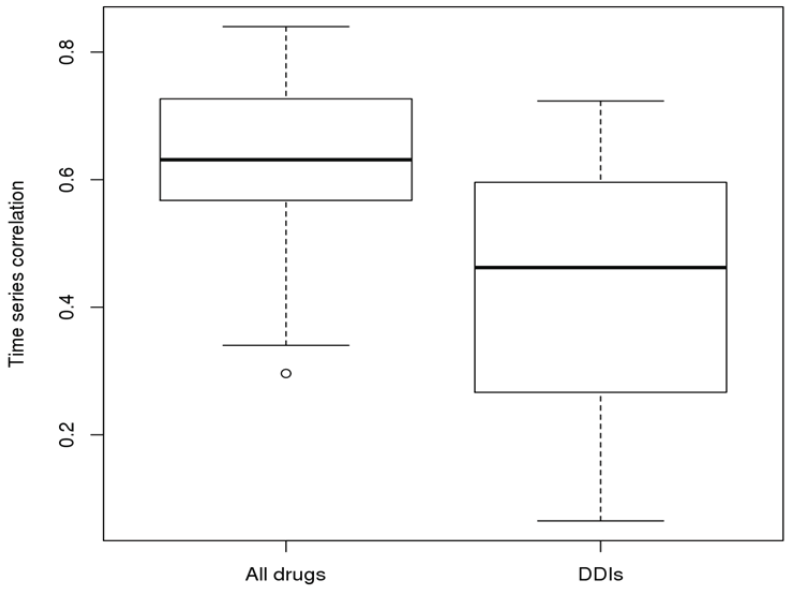

Figure 1. Distribution of maximum correlations between drug series and for drug-drug interaction pairs.

We also analyzed the co-occurrence of disease condition mentions with drug names in the messages, as well as the corresponding drug categories. A heat map relating drug categories with the highest frequency of hits for disease condition mentions is shown in

Figure 2. Interestingly, we observed that drug mentions were in a significantly high number of cases found in messages containing related disease conditions, including hypnotics and sedatives co-occurring with insomnia; antidepressives with depression, or antineoplastic agents with breast cancer.

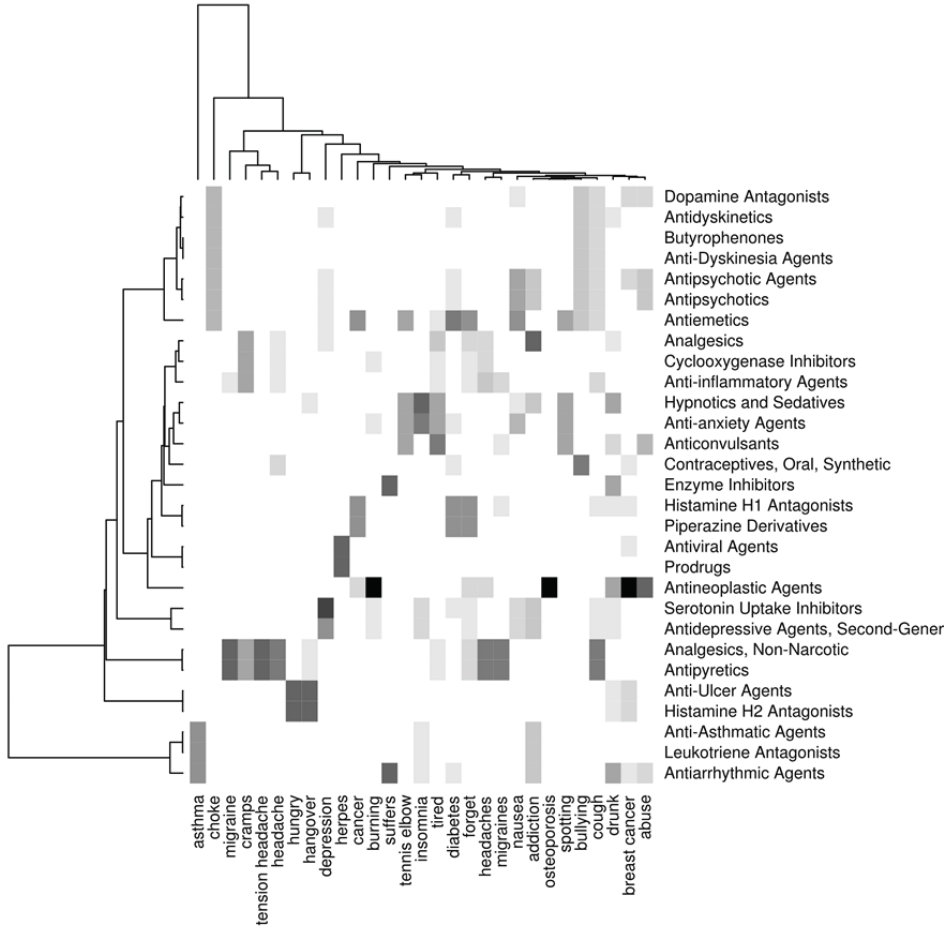

Figure 2. Heat map relating drug categories with frequency of hits for disease condition mentions. Yellow squares correspond to higher frequencies. Data shown is limited to 30 top categories and terms.

3. Discussion

Taking into account the large number of mentions that we found in a relatively short period of time, we believe that a promising application of social media platforms and in particular Twitter is to use them as tools for the detection, understanding and monitoring the way people manage prescribed drugs. Notably, people use Twitter to share and discuss feelings, expectations and opinions about their own health and about prescribed treatments they received more often than expected. This wealth of information could be used as a novel means for drug safety monitoring in parallel with

conventional pharmacovigilance reporting and surveillance systems. For instance, this information can be used for the development of predictive models for drug toxicity, drug-drug interactions or drug resistance [16]. Data showed that the most highly correlated pairs of drug mentions on Twitter could not be always attributed to drug-drug interactions. This unexplained high correlation would require more in depth analysis in order to identify additional factors and social context. The analysis based on drug categories may, in turn, provide useful insights into trends in drug associations. Results showed that in general there was a good correlation between drug categories and disease condition terms, suggesting that messages on Twitter often contain useful drug-related information. The limitations of this work include the fact that the mentions monitoring was performed within a short period of time that did not show long term trends, the difficulty in parsing content messages to avoid false positive and the inherent complexity of social media language.

References

[1] A. Leis, MA. Mayer. How Twitter is used in international health events: World Aids Day Case Study. *iProceedings Medicine 2.0. London (UK) Sept 23-24*, (2013), 220-221. Available at: http://www.medicine20congress.com/ocs/index.php/med/med2013/paper/view/1872.

[2] MJ. Paul, M. Dredze. You are what you tweet: analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, (2011), 265-272.

[3] C. Chew, G. Eysenbach. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One. **5** (2010), e14118.

[4] G. Chowell, H. Nishiura. Transmission dynamics and control of Ebola virus disease (EVD): a review. BMC Medicine, **12** (2014), 196.

[5] CL. Hanson, B. Cannon, S. Burton, C. Giraud-Carrier. An exploration of social circles and prescription drug abuse through Twitter. *Journal of Medical Internet Research*, **15** (2013), e189.

[6] A. Rodríguez-González, MA. Mayer, JT. Fernández-Breis. Biomedical information through the implementation of social media environments. *Journal of Biomedical Informatics*, **46** (6) (2013), 955-56.

[7] R. Ginn, P Pimpalkhute, A.Nikfarjam, A Patki, K. O'Connor, A Sarker, K. Smith, G. González. Mining Twitter for Adverse Drug Reaction mentions: a corpus and classification benchmark. *Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing. BioTxtM. Reykiavik* (2014). Available at: http://www.nactem.ac.uk/biotxtm2014/papers/Ginnetal.pdf.

[8] K. Jiang, Y. Zheng. Mining Twitter data for potential drug effects. *Advanced Data Mining and Applications,* **8346** (2013), 434-443.

[9] Twitter API. Available at: https://dev.twitter.com/rest/public/search

[10] V. Law et al DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. **42** (2014) D1091-1097.

[11] C. Fellbaum, Christiane, WordNet and wordnets. In: Brown, Keith et al. (eds.), *Encyclopedia of Language and Linguistics*, 2005.

[12] A. Zeileis, G. Grothendieck, zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, **14** (2005), 1-27.

[13] BeFree system. Available at: http://ibi.imim.es/befree.

[14] À. Bravo, M. Cases, N. Queralt-Rosinach, F. Sanz, and L. I. Furlong. A Knowledge-Driven Approach to Extract Disease-Related Biomarkers from the Literature. *BioMed Research International*, 11 (2014), 253128.

[15] 2013AA UMLS Full Release Files Jan. 2013 version. Available at: http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html.

[16] P. Carbonell, JY Trosset. Overcoming drug resistance through in silico prediction. Drug Discov Today: Tech, **11** (2014), 101-107.