Discovering Potential Effects of Dietary Supplements from Twitter Data

Keyuan Jiang
Purdue University Northwest
Hammond, Indiana, USA
kjiang@pnw.edu

Yongbing Tang
Cerner Corporation
Kansas City, Missouri, USA
yongbing.e.tang@gmail.com

G. Elliott Cook
Provider Resources Inc.
Erie, Pennsylvania, USA
ecook@provider-resources.com

Michael M. Madden
The Wright Center
Scranton, Pennsylvania, USA
maddenm@thewrightcenter.org

ABSTRACT

The U.S. Food and Drug Administration uses the Center for Food Safety and Applied Nutrition (CFSAN) Adverse Event Reporting System (CAERS) as the primary tool for identifying new and emerging dietary supplement adverse events. Despite mandatory and voluntary reporting of dietary supplement adverse events to CAERS, many continue to go unreported. Availability of social media has enabled dietary supplement consumers to freely share their concerns and experiences online. Such consumer generated information can be a useful source to further monitor the safety of dietary supplements. To study the usefulness of social media (Twitter in particular) for safety surveillance of dietary supplements, we developed a computational processing pipeline: 1) machine learning based identification of potential Twitter posts (tweets) of personal experiences related to the use of dietary supplements, 2) detection of potential supplement events from these tweets using the medpie open source tool, and 3) mapping detected events to effects through the taxonomy provided in SNOMED CT. Using our pipeline, we identified, from a group of 1,244,661 tweets collected, a total of 17,346 personal experience tweets pertaining to 4 dietary supplements. A total of 191 effects were mapped to SNOMED CT and we discovered that 48 of the 191 effects are not listed in either of the two online sources we referenced. However, the effects discovered from the social media data will need to be verified and confirmed with other sources and/or clinical evidences.

DH '17, July 02-05, 2017, London, United Kingdom

ACM ISBN 978-1-4503-5249-9/17/07...$15.00
https://doi.org/http://dx.doi.org/10.1145/3079452.3079467

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences** → **Consumer health**; • **Applied computing** → **Life and medical sciences** → **Health informatics**;

KEYWORDS

Health surveillance, dietary supplements, machine learning, social media mining, Twitter

ACM Reference format:
Keyuan Jiang, Yongbing Tang, G. Elliott Cook, and Michael M. Madden. 2017. Discovering Potential Effects of Dietary Supplements from Twitter Data. In *Proceedings of DH '17, London, United Kingdom, July 02-05, 2017,* 8 pages.
https://doi.org/http://dx.doi.org/10.1145/3079452.3079467

1 INTRODUCTION

Dietary supplements are among the most commonly used complementary or alternative medical therapies in the United States, representing a multibillion dollar industry. Approximately one-half of the U.S. population uses dietary supplements with the majority using them to improve overall health, with the assumption that they are both safe and effective[7, 8]. Dietary supplements cannot be marketed to diagnose, treat, cure or prevent any disease [FDA, nd], yet they are known to be associated with adverse events from mild events such as dizziness (melatonin) [6], to more severe events, such as, hypersensitivity reactions (Echinacea)[15], hepatotoxicity (valerian root)[11], and even contribute to severe drug interactions (St. Johns Wort)[23]. The U.S. Food and Drug Administration (FDA) has a significant history of monitoring such adverse effects with a defined systemic approach. In 2003, the FDA took a major step in more aggressively monitoring adverse events from dietary supplements through the creation of the CFSAN Adverse Event Reporting System (CAERS), a post-marketing voluntary reporting database for dietary supplements[4]. Subsequently further regulation titled the Dietary Supplement and Nonprescription Drug Consumer Protection Act was enacted. This act mandated that dietary supplement manufacturers, packers, and distributors report serious adverse events to the[3]. The FDA relies heavily on this form of mandatory reporting to

complete post-market surveillance and identify potential issues such as new and emerging adverse events. In March of 2013, the GAO found that only 6,307 adverse event reports were submitted to the agency between 2008 and 2011, with 71% constituting mandatory reports as prescribed in the Dietary Supplement and Nonprescription Drug Consumer Protection Act[20]. The remainder was voluntary reports provided by consumers or healthcare professionals. Overall, the GAO found that FDA could do more to capture additional adverse events from dietary supplements as many adverse events go unreported[20]. The GAO report specifically cites that more can be done and the FDA has taken some initiative to address concerns in the GAO report among others.

The FDAs Regulatory Science Priority Area 5: Harness Diverse Data through Information Sciences to Improve Health Outcomes has led to more interest in assessing new data sources and using innovative analytical methods and approaches[2]. Attempting to embody the call to action by the GAO as well as attempting to find innovative ways to improving health outcomes, we chose to determine if we could identify potential adverse events from 4 different dietary supplements (melatonin, Echinacea, valerian, and St. Johns Wort). The approach we utilized herein is an improved modification of our previous work with the Twitter platform [18]. We posit that Twitter could be utilized as another data source for FDA to identify new and emerging adverse events associated with dietary supplements providing a surveillance tool that is capable of generating potential safety signals in real time.

2 RELATED WORK

Authors performed searches in the U.S. National Library of Medicines PubMed abstract database, and found that there had been no reported effort on detecting potential effects associated with consuming dietary supplements using social media data. At this writing, authors conducted searches at pubmed.gov with "dietary supplement"[All Fields] AND Twitter[All Fields] and "dietary supplement"[All Fields] AND "social media"[All Fields] and yielded the No items found message. However, there have been a growing number of reported studies of detecting medicine effects, especially the adverse effects, in social media data, and our review of others work will primarily focus on detecting potential medicine effects from the social media data, Twitter in particular.

In studying adverse drug reactions from Twitter data, Gonzalezs group[19, 21] collected 187,450 tweets related to 74 carefully selected drug names, and retained 71,571 of these tweets after removing the retweets and discarding those containing URLs, which were considered as advertisements. Out of 71,571 tweets, 10,822 were randomly chosen with a cap of 500-800 per drug to achieve the balance. These 10,822 tweets were manually annotated later, and contained a significant number of noisy irrelevant tweets. After annotation by two domain experts, only about 1,200 tweets were found to contain text related adverse drug reactions (ARDs). Authors tested two classifiers (Nave Bayes and Support Vector Machine) with the annotated data, but it was not clear whether the authors applied the trained classifiers to classify tweets with or without adverse drug reactions.

To investigate whether Twitter data can be used to monitor the safety of medicinal products, Freifeld et al. [13] retrieved 6.9 million English tweets associated with 23 medicinal products from November 1, 2012 through May 31, 2013. A convenience sample of 61,402 tweets were chosen from the 6.9 million Twitter posts collected, and manually annotated. The annotation outcome showed that only 4,401 (7.2%) tweets were relevant to the study medicines. Both products and symptoms were identified by a tree-based dictionary-matching algorithm.

In search for adverse drug reactions related to 5 clinical trial medications, Bian et al. [10] identified 239 possible users after pre-processing 2 billion tweets randomly collected the method of how the 239 users were identified was not described in their work. The tweets of each of the 239 users were combined to form a text document for downstream analysis. The manual annotation of 239 sets resulted in only 27 positive cases of adverse event reporting.

In studying effects of 5 medicines, Jiang et al. [18] developed a machine learning-based computational approach to collect, process and analyze Twitter data. The approach first classifies personal experience tweets, and recognizes and extracts word phrases related to medicine effects with the help of U.S. National Library of Medicines MetaMap software. An analysis of 6,829 relevant tweets resulted in 102 discovered effects with matching rates ranging from 74% to 88% with the known effects.

3 METHOD

We designed a pipeline to collect, process, and analyze the Twitter data for discovering potential effects caused by the use of dietary supplements. Unlike quantitative and structured data commonly found in sources such as electronic medical records or drug claims, Twitter data are textual and unstructured, and this requires using natural language processing (NLP) techniques to process the data. To process a large amount of data efficiently, we used machine learning based classifiers to predict relevant tweets in order for us to extract clinical events related to the study dietary supplements. Figure 1 shows the data processing and analysis pipeline.

First, we collected data from Twitter. After preprocessing the raw data, we extracted a set of features from the tweets with help of natural language processing (NLP), and classified the data with an ensemble of machine learning based classifiers. Tweets classified as personal experience tweets (PETs) were further annotated. We then used the open source tool medpie [9] to identify potential dietary supplement effects which were mapped to clinical concepts with SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) [12]. Finally, the clinical concepts were compared with the effects of study dietary supplements listed in two online

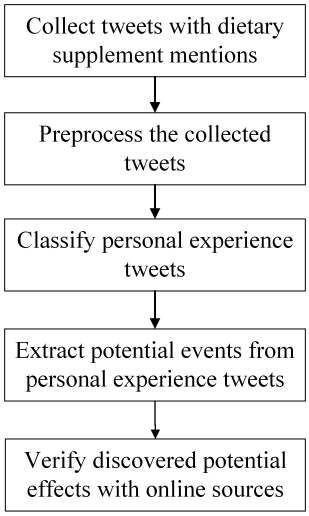

Figure 1: Data processing/analysis pipeline

sources: Natural Medicine Comprehensive Database (NMCD) and drugs.com.

3.1 Personal Experience Tweets (PETs)

The uniqueness of our method is that we first identify personal experience tweets from the data collected and then process and analyze these tweets. As defined in [17], a personal experience tweet (PET) is a Twitter post describing a persons encounter, observation, and important event related to his or her life. Specifically to health related activities, such experience pertains to the change of a persons health. Hence, PETs contain health related information generated by the patients or consumers who had experiences with health related events, products or services. This information is valuable in studying health related issues.

Below are the examples of personal experience tweets related to the effects of dietary supplements, in which the boldfaced word or phrase represents a dietary supplement, and the underscored word or phrase indicates the effects.

Valerian *triggers my depression, but I might have a look into the other stuff.*

St. John's Wort *always makes my eye twitch but it helps my anxiety. So its either a constant twitch or anxiety. #thestruggle*

The sense of the PET definition is broad in that a tweet without mentioning any clinical effect could still be considered a PET. For instance, in a PET, the user could merely state that a dietary supplement helped him or her. This type of tweets can be quite commonly seen, partially due to the 140 character limitation of the tweet text.

3.2 Data Collection

Although Twitter provides APIs to retrieve its data, there are limitations. The REST APIs only allow to query a fixed number of tweets within a short time window of the last several days, and the Streaming APIs can only collect ongoing, newly created Twitter data. We developed a method that crawls the twitter.com website to collect historical tweets for any given keywords. Our method mimics the scroll-down behavior of the Web browser to iteratively request more and older data from twitter.com. In each iteration, the gathered data (in HTML format) are parsed for each individual tweet. With the method, we were able to retrieve tweets since the onset of twitter.com.

3.3 Preprocessing

The collected tweets are dirty and require cleanup. In preprocessing, we removed emojis, emoticons, user mentions, URLs, and hashtags from the tweets. And we extracted a set of 16 features for classifiers. This set consists of textual features described in [17] without including features derived from the tweet metadata.

3.4 Classifying Personal Experience Tweets

Twitter data are known for their noisiness. Even in the tweets collected by keyword search, there can be a significant amount of irrelevant noisy tweets which are associated with sales promotions, news items, and even spam. For example, in studying 5 clinical trial medicines, [10] processed and analyzed 2 billion tweets only to discover 27 positive cases of adverse event reporting. If we were to manually annotate and examine all the collected tweets, it would require a huge amount of effort which is prohibitive in practice, because 1) annotation is a laborious time consuming process and 2) most of the collected tweets are irrelevant.

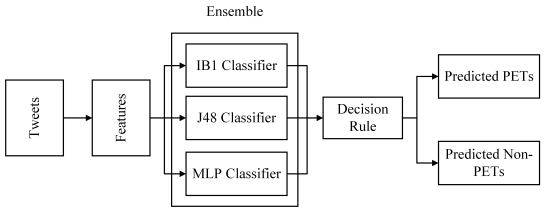

Figure 2: Classification of personal experience tweets

To identify personal experience tweets efficiently and effectively, we used an ensemble of machine learning based classifiers to filter out noisy irrelevant tweets, and only annotated the tweets which were classified as PETs. The ensemble consists of 3 machine learning based classifiers: K-nearest neighbors (IB1), decision tree (J48), and neural network (MLP multilayer perceptron), and they were chosen because of 1) their relatively better performance than other algorithms in our experiments and 2) their differences in performance which made us to consider using all three classifiers to improve and maintain the consistent performance. Weka software [14] which contains the implementation of the three classifiers in the ensemble was used to classify the tweets.

To classify the PETs, we first trained the ensemble of 3 classifiers using a personal experience tweet corpus constructed by [17]. The corpus is made up of 8,770 annotated tweets, of which 2,067 are PETs and 6,703 non-PETs, indicating a relatively balanced data set.

After training, the classifiers were fed with the features extracted from the cleaned tweets, and yielded two classes of tweets: predicted PETs and predicted non-PETs. Of the predicted PETs, only those that were considered positive by at least two classifiers were included in the potential PETs. However, to make sure the tweets to be analyzed are actual personal experience tweets, we manually annotated all the potential PETs and built a set of actual PETs after verification by the annotator.

3.5 Extracting Dietary Supplement Effects

This study was interested in knowing what effects related to dietary supplements were mentioned in PETs. We relied on one of the modules of an open source software, medpie [9], to extract the dietary supplement effects and the pairs of co-occurring dietary supplement and effect. Included in medpie is a controlled vocabulary of drug, dietary supplement, and event terms. The list of dietary supplements contains 507 terms and was compiled by an expert in complementary alternative medicine. There are about 27,000 terms in the event vocabulary that could either refer to a side effect or an indication of a medicine or a dietary supplement. The event terms were gathered from symptom terms from http://www.medicinenet.com/ and adverse effect terms from FDAs Adverse Event Reporting System (AERS) database. This list was augmented with laymans synonyms from the Consumer Health Vocabulary (CHV) [22] to improve the recognition of effect terms commonly used by non-healthcare professionals. In addition, medpie outputs the statistics of term pairs which we used to guide us in determining if the mentioned dietary supplements and potential effects are related.

3.6 Verifying Discovered Effects

In the extract dietary supplement effects, many of them were similar or identical but they were expressed in different ways. To unify different terms of the same clinical concept, we used Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) [12] to map effect terms to clinical concepts in the SNOMED CT taxonomy. SNOMED CT is a standardized, multilingual vocabulary of clinical terminology that is used by physicians and other health care providers for the electronic exchange of clinical health information.

To validate the effectiveness of our approach, we compared the discovered effects with those listed in the two online sources: Natural Medicines Comprehensive Database[1] (NMCD) and drug.com. NMCD is made up of multiple databases and its core database contains about 1,100 detailed, evidence-based monographs on individual natural ingredients (for example, Valerian and Echinacea, etc.). Drugs.com provides information on medicines and dietary supplements such as their uses, side effects and potential to interact with other medicines. According to drugs.com, its data sources include Micromedex®, Cerner Multum^{TM}, Wolters Kluwer^{TM} and others.

[1]http://naturaldatabaseconsumer.therapeuticresearch.com/home.aspx?cs=&s=NDC

4 RESULTS

In this study, four commonly used dietary supplements: St. Johns Wort, Echinacea, Valerian, and Melatonin, were identified for investigation, and their generic names were used as keywords in retrieving Twitter data. These dietary supplements were chosen from the top 100 dietary supplements sold in U.S. from 2006-2012 [16] and on the basis of their likely utilization by the average Twitter user and their likelihood of producing adverse events based on previously known adverse event profiles [1, 5].

4.1 Data Collection

Using the names of the study dietary supplements as keywords, we collected a total of 1,244,661 tweets which were posted to Twitter from March of 2006 through January of 2016. Figure 3 shows the statistics of collected tweets by dietary supplement and year.

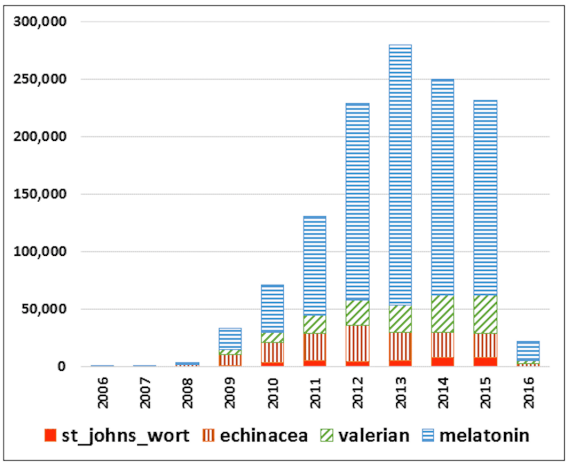

Figure 3: Counts of collected tweets related to 4 study dietary supplements. The 2016 counts only include the tweets collected in January when our data collection ended.

The breakdown of the collected dietary supplement tweets is shown in Table 1. As can be seen, most collected tweets are related to Melatonin.

4.2 Classifier Performance on Training Data

We trained three classifiers used in this study with a corpus of 8,770 annotated tweets [17]. The feature set fed to the

Table 1: Breakdown of the collected tweets

Dietary Supplement	#Tweets
St. Johns Wort	36,494
Echinacea	153,681
Valerian	143,288
Melatonin	919,863
Total	919,863

classifier is similar to what was used in constructing the corpus, and the following textual features were top-ranked in Weka:

1) Occurrences of automatically categorized frequent terms in *tweet text* of PET class
2) Occurrences of automatically categorized frequent terms in tweet text of non-PET class
3) Occurrences of automatically categorized frequent terms in *username* of PET class
4) Occurrences of automatically categorized frequent terms in *username* of non-PET class
5) Occurrences of frequent terms in tweet text of PET class
6) Count of first person pronouns
7) Count of personal pronouns
8) Count of proper nouns

The first five of these top-ranked features are mainly the terms used in the tweet texts or the Twitter usernames, indicating that the many terms in either of the PET class or non-PET class can differentiate its class from the other. The personal pronoun features were considered to be important in describing a persons own experience, and proper nouns may be related to the names of dietary supplements.

As shown in Table 2, all 3 classifiers performed well on the corpus data. The performance data were generated using 10 fold cross-validation. Even though that MLP had the best performance, we hypothesized that using this ensemble of classifiers would help achieve better performance. The performance on the corpus of training data provided us some confidence that they could do sufficiently well on the unlabeled unseen data.

Table 2: Overall performance of classifiers

Classifier	Precision	Recall	F1	ROC Area
IB1	0.905	0.905	0.905	0.868
J48	0.930	0.931	0.931	0.932
MLP	0.935	0.935	0.935	0.982

IB1 is a k-nearest neighbors classifier, J48 is a decision tree classifier, and MLP (multilayer perceptron) is a neural network classifier. The overall performance is the average performance of classifying both PET and non-PET tweets. Given that in the corpus non-PETs over-numbered PETs, all 3 classifiers demonstrated lower performance of classifying PETs than that of non-PETs (data not shown).

Table 3: Breakdown of the predicted PETs

Dietary Supplement	# Tweets	# Predicted PETs
St. Johns Wort	36,494	299
Echinacea	153,681	1,592
Valerian	143,288	1,807
Melatonin	919,863	33,208
Total	1,244,661	36,906

4.3 Personal Experience Tweets

Using the ensemble of three classifiers, we predicted a total of 36,906 potential PETs, and 1,207,755 non-PETs. The distribution of the number of predicted PETs by dietary supplement is shown in Table 3.

Table 4: Breakdown of actual PETs after annotation

Dietary Supplement	# Predicted PETs	# Actual PETs	Precision
St. Johns Wort	299	146	48.8%
Echinacea	1,592	746	46.9%
Valerian	1,807	837	46.3%
Melatonin	33,208	15,607	47.0%
Total	36,906	17,346	47.0%

All 36,906 predicted PETs were manually annotated to make certain that they are actual PETs. Table 4 shows the resultant actual PETs by each individual dietary supplement. In the table, Precision is equal to # Actual PETs / # Predicted PETs.

Table 5: Discovered dietary supplement effects and their mapping with those in SNOMED CT

Dietary Supplement	# Effect Mentions	# Potential Effects
St. John's Wort	130	20
Echinacea	170	17
Valerian	767	41
Melatonin	8,335	113
Total	9,402	191

4.4 Potential Effects

Through the use of medpie, we extracted mentions of effects (adverse and beneficial) associated with the study dietary supplements. These effect mentions were mapped to clinical concepts using SNOMED CT (Table 5), and the clinical concepts (non-duplicate normalized effects) were compared with dietary supplement effects extracted from NMCD and drugs.com (Table 6). Table 5 shows the statistics of potential effects associated with each dietary supplement.

The potential effects refer to those mapped with the taxonomy in SNOMED CT. As can be seen in the table, among all the PETs related to St. Johns Wort, there are 130 mentions

Table 6: Discovered dietary supplement effects and their matches with those listed in NMCD and at drug.com

Dietary Supplement	# Discovered Effects	# and % Matched with NMCD	# and % Matched with Drugs.com	# Overlapped Effects (NMCD and Drugs.com)	# of Effects in Neither Online Source
St. John's Wort	20	16	16(80%)	16	4
Echinacea	17	14(82.4%)	16(94.1%)	14	1
Valerian	41	34(82.9%)	35(85.4%)	33	5
Melatonin	113	75(66.4%)	66(58.4%)	66	38
Total	191	139(72.8%)	133(69.6%)	129	48

of effects which were mapped to 20 clinical concepts (effects) in SNOMED CT.

Table 6 summarizes the matching rates of discovered effects with those in the online sources. Finding matching rates was to indicate that our method is valid by its being able to discover the known effects in Twitter data. It was not our goal to investigate if potential effects discovered from Twitter data match all known effects this is an unrealistic goal because not all the dietary supplement effects would necessarily be posted to Twitter. Instead, a realistic goal is to see how discovered effects can match any known effects, which can assert partially that our approach is effective in finding some known effects.

In Table 6, overlapped effects are the effects that are discovered and listed in NMCD and at drugs.com.

5 DISCUSSIONS

Only about 3% (36,906/1,244,661) of the collected tweets were classified as PETs by the ensemble of classifiers. This small number of tweets can be attributed to two factors: 1) a significant number of noisy irrelevant tweets exist even in the Twitter data collected by simple keyword search, and 2) the imperfect performance of machine learning based classifiers may have missed a number of positive PETs.

In the section of Related Work above, we reported that in several studies only a small number of relevant tweets were discovered, and our finding in this project appears to be consistent with that of other studies.

5.1 Performance on Unlabeled Tweets

Good performance of classifiers on the annotated training data does not necessarily guarantee that the classifiers will perform the same on the unlabeled data. In most existing literature, authors only reported on the results of classifying the annotated data, which certainly helps consistently evaluate the performance of different classification algorithms on the same data set, but it does not truly reflect how well classifiers performed on the unlabeled data which contain unseen instances.

In this paper, we reported classifiers performance on unlabeled Twitter data (see Table 4). Although the results are not comparable with others, it shows the performance on the Twitter data which may be unseen by the classifiers. Unlike evaluating the performance on the labeled data, we could only evaluate the precision of classification because we only annotated predicted PETs (TP and FP) by ignoring the predicted non-PETs (TN and FN) to reduce the effort needed for annotation.

Precision is defined as

$$precision = \frac{TP}{TP + FP}$$

Although it is not as good as that on the labeled training data (Table 2), the performance of our classifiers on the unlabeled tweets (Table 4) was 47%, and we believe that it is a reasonably good performance on over 1.2 million of unlabeled tweets. In other words, near half the predicted PETs are actual PETs.

5.2 Potential Dietary Supplement Effects

We found a number of potential effects associated with using 4 study dietary supplements (see Tables 5 and 6), and these effects match well many known dietary supplement effects listed in two online sources. The matching rates ranged from 66.4% 82.9% with the effects listed in NMCD and varied from 58.4% to 94.1% with those provided at drugs.com (Table 6). This demonstrates the effectiveness of our method. Although Melatonin has the most potential effects, it has the lowest matching rates with the two online sources. In addition, we noticed that of 191 discovered effects, 48 had no match with effects listed in either of the online sources. These unmatched effects may be those that have not been reported or rare effects, and confirmation of these effects warrants rigorous clinical investigations.

5.3 Limitations and Future Directions

There are limitations of this work. First, to achieve the efficiency and reduce the amount of annotation work, we focused solely on the predicted PETs without considering any non-PETS where false negatives may exist. In other words, actual PETs which were classified as non-PETs could have been missed because of the imperfection of the classifiers. Given that only predicted PETs were annotated, it is impossible to know how many actual PETs were in the predicted non-PETs this made it impossible to measure the accuracy and other performance indicators of the classifiers. However, the improvement of the accuracy may be measurable by studying

the change of the numbers of TP and FP. One of our future direction is to improve both the accuracy and precision of the classifiers by using new and more prominent features and/or different machine learning algorithms. In addition, a new tweet corpus can be generated by combining [17] corpus with annotated tweets out of this study. Having a larger corpus will make more data samples available for training and may have better representation of the characteristics of Tweet data, yielding better performance of classification on unlabeled data.

Second, our project only concentrated on the generic name of the dietary supplements, but all 4 study dietary supplements have various brand names. It is possible that Twitter users shared their experiences of these dietary supplements using other brand names. To broaden this research, we plan to collect, process and analyze tweets containing the brand names of the study dietary supplements.

Furthermore, the open source tool medpie we used in the project does not help in identifying the relation between a dietary supplement and a clinical event, and manually verification was required. This leads to another future direction of this project: to develop an automated yet reliable method to identify the relations.

6 CONCLUSION

In this paper, we presented our work of detecting potential effects (adverse and beneficial) associated with the consumption of dietary supplements from social media data. Unlike structured quantitative data commonly found in sources such as electronic medical records and spontaneous report systems, Twitter data are unstructured, textual, and limited by the 140 character constraint. The results of our work demonstrate the efficiency and effectiveness of our approach: it was able to efficiently process a large number of tweets (> 1.2 millions) computationally and to be able to identify 191 effects associated with 4 study dietary supplements.

We believe that our approach can complement existing methods in better reporting adverse effects caused by the use of dietary supplements especially from the information reported by the consumers, and can be used for studying other dietary products. However, what is discovered from the social media will need to be verified and confirmed with other sources and/or clinical investigations.

7 HUMAN SUBJET RESEARCH

The protocol of this project was reviewed and approved for compliance with the human subject research regulation by the Institutional Review Board of Purdue University which the first author (KJ) is affiliated with.

8 COMPETING INTERESTS

The authors have declared that no competing interests exist.

9 AUTHOR CONTRIBUTIONS

KJ, GEC, and MMM initiated this project. KJ and YT designed data processing/analysis pipeline. YT collected, annotated and analyzed data under KJs supervision. KJ drafted this manuscript. GEC and MMM proofread the manuscript.

REFERENCES

[1] *An Exhaustive Study of Twitter users across the world.* Available at: http://www.beevolve.com/Twitter-statistics/.

[2] *FDA. Advancing Regulatory Science at the FDA: FDA; 2011.*

[3] *FDA. In: Administration USFaD, ed. Vol 21 CFR 101.93(c). United States Code of Federal Regulations.*

[4] *FDA. The CFSAN's Adverse Event Reporting System (CAERS). 2016;.* http://www.fda.gov/AboutFDA/WhatWeDo/History/FOrgsHistory/CFSAN/ucm083863.htm.Accessed10/21/2016.

[5] *National Institutes of Health. Office of Dietary Supplements. Dietary Supplement Fact Sheets.* Available at:http://ods.od.nih.gov/factsheets/list-all/.

[6] Lars Peter Holst Andersen, Ismail Ggenur, Jacob Rosenberg, and Russel J. Reiter. 2016. The Safety of Melatonin in Humans. *Clinical Drug Investigation* 36, 3 (2016), 1–7.

[7] R. L. Bailey, J. J. Gahche, C. V. Lentino, J. T. Dwyer, J. S. Engel, P. R. Thomas, J. M. Betz, C. T. Sempos, and M. F. Picciano. 2011. Dietary supplement use in the United States, 2003-2006. *Journal of Nutrition* 141, 2 (2011), 261–6.

[8] Regan L. Bailey, Jaime J. Gahche, Paige E. Miller, Paul R. Thomas, and Johanna T. Dwyer. 2013. Why US Adults Use Dietary Supplements. *Jama Internal Medicine* 173, 5 (2013), 355.

[9] A. Benton, J. H. Holmes, S. Hill, A. Chung, and L. Ungar. 2012. medpie: An information extraction package for medical message board posts. *Bioinformatics* 28, 5 (2012), 743–744.

[10] Jiang Bian, Umit Topaloglu, and Fan Yu. 2012. Towards large-scale twitter mining for drug-related adverse events. In *International Workshop on Smart Health and Wellbeing.* 25–32.

[11] D. L. Cohen and Toro Y Del. 2008. A case of valerian-associated hepatotoxicity. *Journal of Clinical Gastroenterology* 42, 8 (2008), 961.

[12] K Donnelly. 2006. SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in Health Technology & Informatics* 121, 121 (2006), 279–290.

[13] Clark C. Freifeld, John S. Brownstein, Christopher M. Menone, and Wenjie Bao. 2014. Digital Drug Safety Surveillance: Monitoring Pharmaceutical Products in Twitter. *Drug Safety* 37, 5 (2014), 343.

[14] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *Acm Sigkdd Explorations Newsletter* 11, 1 (2009), 10–18.

[15] A. L. Huntley, Coon J Thompson, and E Ernst. 2005. The safety of herbal medicinal products derived from Echinacea species: a systematic review. *Drug Safety* 28, 5 (2005), 387–400.

[16] Penton Media Inc. 2013. NBJs Supplement Business Report Analysis of markets, trends, competition, and strategy in the U.S. dietary supplement industry. *Nutrition Business Journal* (2013).

[17] Keyuan Jiang, Ricardo Calix, and Matrika Gupta. 2016. Construction of a Personal Experience Tweet Corpus for Health Surveillance. In *The Workshop on Biomedical Natural Language Processing.* 128–135.

[18] Keyuan Jiang and Yujing Zheng. 2013. *Mining Twitter Data for Potential Drug Effects.* 434–443 pages.

[19] Karen OConnor, Pranoti Pimpalkhute, Azadeh Nikfarjam, Rachel Ginn, Karen L Smith, and Graciela Gonzalez. 2014. Pharmacovigilance on Twitter? Mining Tweets for Adverse Drug Reactions. *AMIA Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* 2014 (2014), 924–933.

[20] United States. Government Accountability Office. 2013. Dietary Supplements: FDA May Have Opportunities to Expand Its Use of Reported Health Problems to Oversee Products. 7, 14 (2013), 1427.

[21] Abeed Sarker, Rachel Ginn, Azadeh Nikfarjam, Pranoti Pimpalkhute, Karen O'Connor, and Graciela Gonzalez. 2014. Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark. In *Biotxtm.*

[22] Q. T. Zeng, T Tse, G Divita, A Keselman, J Crowell, and A. C. Browne. 2005. Exploring lexical forms: first-generation consumer health vocabularies. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* 2006 (2005), 1155.

[23] S. Zhou, E. Chan, S. Q. Pan, M. Huang, and E. J. Lee. 2004. Pharmacokinetic interactions of drugs with St John's wort. *Journal of Psychopharmacology* 18, 2 (2004), 262.