Social Media Mining for Drug Safety Signal Detection

Christopher C. Yang
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104
Chris.Yang@drexel.edu

Haodong Yang
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104

Ling Jiang
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104

Mi Zhang
College of Information Science and Technology
Drexel University
Philadelphia, PA 19104

ABSTRACT
Adverse Drug Reactions (ADRs) represent a serious problem all over the world. They may complicate a patient's medical conditions and increase the morbidity, even mortality. Drug safety currently depends heavily on post-marketing surveillance, because pre-marketing review process cannot identify all possible adverse drug reactions in that it is limited by scale and time span. However, current post-marketing surveillance is conducted through centralized volunteering reporting systems, and the reporting rate is low. Consequently, it is difficult to detect the adverse drug reactions signals in a timely manner. To solve this problem, many researchers have explored methods to detect ADRs in electronic health records. Nevertheless, we only have access to electronic health records form particular health units. Aggregating and integrating electronic health records from multiple sources is rather challenging. With the advance of Web 2.0 technologies and the popularity of social media, many health consumers are discussing and exchanging health-related information with their peers. Many of this online discussion involve adverse drug reactions. In this work, we propose to use association mining and Proportional Reporting Ratios to mine the associations between drugs and adverse reactions from the user contributed content in social media. We have conducted an experiment using ten drugs and five adverse drug reactions. The FDA alerts are used as the gold standard to test the performance of the proposed techniques. The result shows that the metrics *leverage*, *lift*, and *PRR* are all promising to detect the adverse drug reactions reported by FDA. However, *PRR* outperformed the other two metrics.

Categories and Subject Descriptors
H.2.8 [Database Management]: Database applications – Data mining; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – Linguistic processing; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval; H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia

SHB'12, October 29, 2012, Maui, Hawaii, USA.

Keywords
Adverse Drug Reaction, Online Health Community, Association Mining, Proportional Reporting Ratios.

1. INTRODUCTION
Adverse Drug Reactions (ADRs) is defined as an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which predicts hazard from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product [1]. ADRs may complicate a patient's medical conditions and increase the morbidity, even mortality. It is found that ADRs contribute to 5% of all hospital admissions and represent the fifth most common cause of death in hospital. In year 2000, there were about 100,000 deaths in the U.S. due to medical errors, of which about 7,000 were attributed to drug reactions [2].

Even though during the pharmaceutical product development, pre-marketing review process is required to identify risks associated with drugs, pre-marketing review process is known to have limitations. Pre-marketing clinical trials are often conducted in selective patient populations, with relatively small numbers of patients, and a short duration of follow-up. Hence, the pre-marketing review process is too constrained in scale and time span to possibly identify all potential adverse effects. Therefore, drug safety currently depends heavily on post-marketing surveillance – the systematic detection and evaluation of medicines once they have been marketed – to detect latent ADRs.

In United States, FDA (Food and Drug Administration) is responsible for most of the post-marketing surveillance. Healthcare professionals, drug manufactures, consumers etc. spontaneously report suspected ADRs to FDA's Adverse Event Reporting System (AERS). However, AERS is a passive system in that it depends on voluntary, spontaneous reports. It was estimated that the reporting rate of AERS is lower than 10% [24]. The system relies on human recognition of potential links between drugs and adverse reactions, as well as on volunteers' wills to report the ADRs. Even if potential ADRs were reported to AERS, it would still take a long time for FDA to review the cases. Laser et al. [3] pointed out that many serious ADRs are discovered only after a drug has been on the market for years. Only half of newly discovered serious ADRs are detected within 7 years after drug approval. Therefore, current approaches employed in post-marketing surveillance are not efficient enough to detect potential ADRs timely to avoid unnecessary healthcare cost and even mortality.

With the rapid development of Internet, there are many online health communities booming and many patients go to these websites to seek or offer healthcare information. Previous study revealed that previously unreported ADRs can be identified from patients' reports through centralized reporting systems and that their quality is similar to those that health professional reports [4]. Therefore, these online health communities provide great platforms for the patients to discuss about the drugs they are taking, which provide enormous valuable information for detecting potential ADRs. If we can make good use of this information, we may detect ADRs much more timely and efficiently than existing reporting system. As a result, an effective adverse drug reaction signal detection system is desired to crawl, analyze, and identify signals from the health social media sites such as PatientsLikeMe and MedHelp or popular social media sites such as Facebook and Twitter, in supporting post-marketing surveillance.

In this work, we focus on harnessing social media for *signal detection* of *adverse drug reactions*. In order to explore the potential of detecting ADRs using online healthcare communities, we proposed to employ association mining and Proportional Reporting Ratios (*PRR*) to extract interesting associations of drugs and adverse reactions. When social media users contribute content regarding the ADRs of a specific drug, the co-occurrence of the drug and its ADR in the posts or comments of an online healthcare social media site could be regarded as an association, and its interestingness and impressiveness can be measured by investigating such metrics as *support*, *confidence*, *leverage* and *lift*. Association rule mining was first utilized in the field of data mining. Also, in the area of ADRs detection, this method was employed by several researchers to identify potential causal relationships between drugs and adverse reactions from electronic health data [5, 6, 7]. Developed by Evan et al. [8], *PRR* is a statistical indicator, which compares the proportion of all reactions to the drug of interest to the same proportion of all other drugs. In this work, we attempt to employ association mining with *PRR* to extract accurate adverse reactions associated with certain drugs from online healthcare communities.

These detected signals are not meant to be proven adverse effects but need to be further validated by signal analysis. The signal analysis includes determination of causality, evaluation of frequency, evaluation of biological gradient, and determination of health consequences through appropriate medical and epidemiological evaluation that exclude biases and confounding variables. Such signal analysis requires sophisticated clinical and laboratory evaluations, which are not intended to be part of this work. This work only focuses on signal detection. When a signal is confirmed by further analysis, subsequent appropriate actions must be taken to inform the pre-scribers and consumers.

2. RELATED WORK

As a practical problem, detection of ADRs has received a great deal of attentions from researchers and many studies have been done in this area. Edwards and Aronson defined ADRs as an appreciably harmful or unpleasant reaction, resulting from an intervention related to the use of a medicinal product, which is predictive of future adverse events from future administration and warrants prevention or specific treatment, or alteration of the dosage regimen, or withdrawal of the product [1]. Edwards and Aronson also defined *signal* as possible causal relation between an adverse event and a drug [1]. In this work, we focus on detecting signals of adverse drug reactions in general health consumer contributed content in social media sites.

Some previous studies employed observational methods to conduct ADRs assessment and detection, such as medical record review, solicited surveillance, patient survey, administrative data and laboratory and clinical values [9]. However, by reviewing these methods, Hakobyan et al. [9] showed that they were inadequate for identifying all possible ADRs and would not provide sufficient information about ADRs to clinicians and patients. Although Wu and Makuch incorporated external data such as established databases or pre-NDA (New Drug Application) data into observational cohort study and provided direct evidence for a reduction in sample size with these data [10], it didn't change the fact that observational studies are time-consuming and costly.

Recently, instead of concentrating on observational studies, many researchers and health professionals use database-related quantitative methods to detect and predict ADRs. In the United States, current postmarketing methods primarily rely on FDA's spontaneous reporting system MedWatch [1]. There are also spontaneous reporting centers in other countries such as England and Japan [8, 11], and based on these reporting data various data mining methods have been practically implemented. For example, the FDA currently adopts an algorithm called Multi-item Gamma Poissson Shrinker (MGPS) for detecting potential signals from its MedWatch data [12]; UK Medicines Control Agency employs *PRR* to recognize adverse reactions, events related to the underlying disease and signals requiring further evaluation by comparing the proportion of all reactions to a drug of interest to the same proportion of all other drugs in UK Yellow Card database [8]; the Uppsala Monitoring Center uses Bayesian Confidence Propagation Neural Network as its signal detection strategy with World Health Organization database [13]; the Netherlands Pharmacovigilance Foundation Lareb utilizes the method using the 95% confidence interval for the Reporting Odds Ratio [14]. The performance of these methods were compared by Kubota et al. using a Japanese spontaneous reporting database and the results showed that the ability of detecting a signal varies among these methods [11]. A number of other data mining methods such as empirical Bayes model [15, 16, 17] and pharmacovigilance map method [18] have also been used with spontaneous reporting dataset.

Although the approaches mentioned above performed more efficiently than observational methods, their performance is likely to be highly situation dependent because of the weakness and potential biases such as latency and inconsistency inherent in spontaneous reporting systems [16]. In addition, early generation of a new signal can be very difficult because a large number of interesting cases cannot be timely collected due to the underreporting nature of the current reporting system [7]. In order to solve this problem, electronic medical record which is more accessible in various healthcare organization has been used by many researchers. Ji et al. developed a fuzzy logic-based computational recognition-primed decision (PRD) model to calculate the extent of causality between a drug and some of its adverse effects [19]. Based on this model, Ji et al. proposed a novel intelligent agent software system approach for proactively monitoring and detecting potential ADRs of interest using electronic patient records [20]. On the basis of PRD model, another important signal-detection strategy is known as causal association mining algorithm in which a new interestingness measure, causal-leverage, is used to predict potential ADRs from electronic health databases [6, 7]. In addition to PRD model-based approaches, Pouliot et al. generated logistic regression models that correlate postmarketing ADRs with screening data from

[1] http://www.fda.gov/Safety/MedWatch/default.htm

the PubChem BioAssay database [21]. Jin et al. brought up a new interestingness measure, residual-leverage to mine unexpected temporal associations for generating ADRs signals from real-world healthcare administrative databases [5]. Nehemiah and Kannan proposed a diagnostic decision support system for adverse drug reaction using temporal reasoning. In the study, the analysis is carried out based on Modified Association Classification algorithm, which is a modified version of Apriori algorithm and uses Interestingness and Local Support measures to calculate the risk ratio and the odds ratio [22].

The studies mentioned above showed that data mining techniques based on electronic health data could generate earlier ADR signals than spontaneous reporting data. However, this kind of data is not available for every researcher but those who are cooperating with hospitals, clinics or any other health organizations and communities. Most researchers may only have one dataset of electronic health records depending on the affiliating or collaborating health unit. The integration of electronic health records from multiple resources is still a technical and policy challenge. A single dataset of electronic health records may have limitations on the patient records that it may cover. Therefore, the availability of large scale electronic patient data from multiple sources is a limitation for its application on ADRs research in spite of its usefulness. In addition, the electronic health records are submitted by health professionals. That means the data is collected only when the health consumers visit the health professional and the adverse drug effect is recorded by the health professional.

Nowadays, with the development of Internet and Web 2.0, more and more online healthcare community are emerging and flourishing such as MedHelp[2] and PatientsLikeMe[3]. Everyday tens of hundreds of users post topics or comment on other users' posts talking about their health conditions, treatment experience, drugs taken as well as ADRs of the drugs through these social media platforms. This cyber-based technique empowers patients and healthy individuals to play a substantial role in their own health and treatment and these social media data is available and accessible to public. If these data could be used effectively and efficiently, ADRs can be detected more accurately and earlier than using either spontaneous reporting data or electronic health data. In our knowledge, very few studies have employed social media to predict ADRs. For example, Chee et al. used machine learning method to classify drugs into FDA's watchlist and non-watchlist based on messages extracted from an online health forum - Health & Wellness Yahoo! Groups but it required a training dataset to train the ensemble classifiers [23]. Practically, it takes a tremendous amount of human effort to prepare training data for detecting the ADR signals and it may not be feasible if we have a large number of drugs and adverse drug reactions. Leaman et al. [24] used the DailyStrength health-related social network as the source of user comments. They extracted the adverse reactions by matching the terms in user comments with a lexicon that combined concepts and terms from four resources and compared the extracted adverse reactions with the annotated results generated by two annotators.

3. METHODOLOGY

3.1 Data Processing

To detect adverse drug reactions from online health communities, it is important to effectively identify drugs and side effects from discussions of social media. In common online communities, people discuss with each other in the form of "threads". Each thread includes an original post and a series of following comments focusing on the same discussion topic.

3.1.1 ADRs Lexicon

It is a challenge to extract adverse reactions from threads of drugs, because people use free and creative expressions in online healthcare communities, which is an open and casual platform for discussions. Moreover, in healthcare area, laypersons (consumers) and professionals use quite different vocabularies and expressions to describe the same health-related concepts, like symptoms and side effects. So, standard medical lexicon for professionals like UMLS cannot be applied directly to identify ADR terms in this study.

To better understand and match user expressions with related concepts of ADR in online healthcare communities, we apply Consumer Health Vocabulary (CHV) to generate ADR lexicon, which is a computerized collection of health expressions derived from actual consumer utterances (authored by consumers), linked to professional concepts and, reviewed and validated by professionals and consumers [25]. In this study, CHV is used to expand the lexicon of ADR terms based on FDA reports. Concretely, for each FDA-reported ADR to investigate in this study, we search for its expressions used by consumers from CHV wiki[4], and add them to the lexicon for this ADR, which is used to match thread texts. For example, for one ADR term – diarrhea which is reported by FDA, 11 consumer expressions are found from CHV, including loose bowel motion, diarrhea, diarrhea running, loose bowel movement, diarrheas, watery stool, bowels loose movement, diarrhea, water stools, bowel loose movements and diarrhea nos. The 11 expressions are added to the ADR lexicon to identify and indicate the adverse reaction of diarrhea in healthcare communities.

3.1.2 ADRs Matching

Before matching ADR lexicon with threads from the healthcare communities, punctuations and stopwords[5] are first removed from threads. Then thread contents are tokenized by splitting at whitespace. Stemming is not implemented to maintain the original meaning of users. After the data preprocessing, ADR terms are detected and identified by matching a sliding window of tokens from threads with each item in ADR lexicon. The sliding window is a multi-gram term generator, for which the size of window represents the number of grams of the term. Given the size of sliding window j, each thread would be represented as a list of j-gram terms combing tokens in that thread, and every term would be compared with ADR lexicon to look for the matching ADR.

3.2 Data Analysis Methodology

In our study, we use association mining and proportional reporting ratio to analyze the detected ADR for different drugs.

3.2.1 Association Mining

Counting all threads in the dataset to analyze, we regard all drugs and adverse reactions as items, and threads as transactions in association mining. Our goal is to mine rules in the form of $D \Rightarrow R$, where D is a 1-itemset containing a drug such as {Lansoprazole}, and R is another 1-itemset containing an adverse reaction like {diarrhea}.

For every possible combination of itemsets of D and R in the whole dataset, we calculate indicators of *leverage* and *lift* for the rule of

[2] http://www.medhelp.org

[3] http://patientslikeme.com

[4] http://consumerhealthvocab.chpc.utah.edu/CHVwiki

[5] http://norm.al/2009/04/14/list-of-english-stop-words

D ⇒ R. *Leverage* and *lift* are both based on the probabilities of D, R and D ∪ R appearing in threads of the dataset. Let $P(D) = \frac{\text{count(D)}}{\text{total count}}$, $P(R) = \frac{\text{count(R)}}{\text{total count}}$ and $P(D \cup R) = \frac{\text{count(D}\cup\text{R)}}{\text{total count}}$, where count(D) and count(R) denote numbers of threads associated with D and R respectively, count(D ∪ R) is the number of threads in which D and R coexist, and total count is the total number of threads in the dataset. *Leverage* and *lift* of D ⇒ R can be calculated as

$$\text{leverage}(D \Rightarrow R) = P(D \cup R) - P(D) \times P(R) \quad (1)$$

$$\text{lift}(D \Rightarrow R) = \frac{P(D \cup R)}{P(D) \times P(R)} \quad (2)$$

P(D ∪ R) denotes the actual probability of occurrence of drug D and adverse reaction R in threads of the dataset. P(D) × P(R) is the probability of their occurrence if D and R are absolutely independent. So, *leverage* reflects the difference between the actual occurrence probability and the theoretical occurrence probability if the drug and reaction is independent. Comparatively, *lift* reflects the division of the actual probability and theoretical probability. Both of them indicate the associations of the drug and adverse reaction for a pair of D and R.

3.2.2 Proportional Reporting Ratio

Proposed by Evans et al. [7][8] in 2001, *PRR* has been widely used to identify ADR from clinical or medical records. *PRR* is a statistical indicator, which compares the reaction proportion of a drug with that of other drugs. Compared to other statistical indicators for ADR detection in medical records, *PRR* has the advantage that it could be derived solely from spontaneous ADR data and is simple to calculated and interpreted. Traditionally, *PRR* is calculated based on a set of records about drug reactions. So for online health communities in this study, drug threads which are not related to adverse reactions are not used to calculate the *PRR* indicator. For a specific drug D and adverse reaction R, the *PRR* could be calculated by

$$PRR(D,R) = \frac{\frac{count(D \cup R)}{count(D)}}{\frac{count(!D \cup R)}{count(!D)}} \quad (3)$$

Where $count(D \cup R)$ is the number of threads containing both drug D and adverse reaction R, $count(!D \cup R)$ is the number of threads containing other drugs except D and adverse reaction R, $count(D)$ is the number of threads associated with drug D and $count(!D)$ is the number of threads associated with other drugs except D. Note that all the treads taken in to account in formula (3) are associated with at least one drug and one adverse reactions. Drug threads which are not associated with adverse reactions should be removed at the beginning.

3.3 Algorithms

We developed algorithms to identify ADRs from threads of drugs, and implemented association mining to calculate *leverage* and *lift* for each possible pair of drugs and adverse reactions in the dataset. At the same time, *PRR* is also calculated. To identify and match terms in ADR lexicon from threads of a drug, the maximum size of sliding window is given as n at the beginning. Then, the window size is set from 1 to n separately. For each window size j, each thread i of the drug is represented as a list of j-gram terms which would be compared with ADR lexicon. If a matching is detected, one count would be added to the co-occurrence of this drug and ADR. With the number of threads containing each pair of drug-adverse reaction k, the *leverage*, *lift* and *PRR* could be easily determined. Below is the pseudo code for the whole process.

Association Mining Algorithm
INPUT: Sliding window size N; all the threads for each drug; ADRs lexicon
OUTPUT: *drug-adverse reaction* association and its *support*, *confidence* and *leverage* values

```
1: for each thread i
2:   for each sliding window j
3:     generate a list of j-gram;
4:     compare each j-gram with each ADR;
5:     if matching then
6:       identify if this ADR has been detected before
7:         if this ADR is new then
8:           number of threads for this drug-adverse reaction
9:           association + 1;
10:        else continue;
11:    else continue;
12:  end for
13: end for
14: for each drug-adverse reaction association k
15:  p(k) = number of threads containing k / total number of threads for all drugs
16:  leverage (k) = p (k) – p (drug in k) × p (ADR in k)
17:  lift (k) = p(k) / (p(drug in k)×p(ADR in k))
18:  PRR(k) = (number of threads containing k / number of threads containing drug in k) / (number of threads containing ADR in k and drugs not in k / number of threads without drug in k)
```

4. EXPERIMENT

4.1 Dataset

To implement proposed techniques and evaluate the performance, dataset used in this study was collected from MedHelp. As the pioneer in online health community, MedHelp empowers over 12 million people each month to share medical information and find answers to their medical questions since it was founded in February 1994. Every day, members come to MedHelp to receive the support they need from other patients like them, to research information on drugs and health topics, to document their medical history, and to share their knowledge with others in need. The Drugs section is one of the sub-forums in MedHelp, and there are tens of thousands kinds of drugs included[6]. For each drug in this section, there is a brief introduction, and MedHelp users can start a thread of this drug with a post, on which all users can comment. There are up to thousands of threads under each drug. To perform our techniques on the dataset, the drug should have active discussion in MedHelp. Therefore we selected ten drugs with more than five hundred threads of discussion, and collected all the original posts and comments of these drugs.

To gather all the posts and comments of the ten drugs from MedHelp efficiently, we implemented an automatic web crawler. All data was obtained from the raw HTML using PHP codes since there is no open API provided by MedHelp. For each thread, we extracted the subject, username, timestamp and content that included both original post and all the following comments. However, in this study, we only use the content to mine the associations, while the other information can be analyzed in our future studies.

[6] http://www.medhelp.org/health_topics/drugs_list

4.2 Gold Standard

Currently in United States, FDA is responsible for the administration of post-marketing drug safety. Any information regarding ADRs is spontaneously reported to the FDA's Adverse Event Reporting System (AERS). According to FDA's website[7], some drugs have active safety alerts while some don't. Among the ten drugs selected for this study, five of them have active ADR alerts while the other five drugs' alerts are no longer active. Therefore, we used five types of active ADRs released by FDA for the five drugs as the ground truth to evaluate our experiment results.

Table 1 presents the ten selected drugs, the selected active-alerted adverse reactions of the drugs, and the number of threads available in MedHelp. There are in total seven *drug* ⇒ *adverse reaction* associations we are supposed to detect such as *Biaxin* ⇒ *Heart Disease*, *Lansoprazole* ⇒ *Diarrhea*, *Tacrolimus* ⇒ *Cancer* and so on.

Table 1 Adverse Reactions Reported by FDA

Drug Name	Active Adverse Reactions	Number of Threads
Biaxin	Heart Disease	686
Lansoprazole	Diarrhea	592
Luvox	Heart Condition; Suicidal	570
Prozac	Suicidal; Depression	718
Tacrolimus	Cancer	583
Adenosine	None	567
Cialis	None	745
Elidel	None	619
Lantus	None	601
Vyvanse	None	563

4.3 Experiment Results

In this study, we used our dataset to match the ADRs lexicon in order to find the adverse reactions of the ten drugs. We set the value of maximum size of sliding window as three because the longest term of ADR we obtained in CHV consisted of three words after pre-processing, and then we computed the *leverage*, *lift* and *PRR* of each pair of *drug* ⇒ *adverse reaction*.

Table 2 shows the *leverage*, *lift* and *PRR* of each pair of *drug* ⇒ *adverse reaction*. The pairs that have active alerts in FDA are highlighted and it can be seen that most of the drugs which are reported to cause the ADR have higher values than any other drugs.

For diarrhea, it is obvious that the pair *Lansoprazole* ⇒ *Diarrhea* has the highest *leverage*, *lift* and *PRR*, which has been alerted by FDA. For heart disease, the pairs *Biaxin* ⇒ *Heart Disease* and *Luvox* ⇒ *Heart Disease* have respectively the fifth and second highest *leverage* and *lift* as well as respectively the fourth and sixth highest value in *PRR*. These two drugs have been reported by FDA to be related to adverse reaction of heart diseases. As to depression, the *Prozac* ⇒ *Depression* association is ranked first in *leverage* and *PRR*, while it has the second highest *lift* comparing to all the other *Drug* ⇒ *Depression* associations.

For suicidal, the drugs Luvox and Prozac, which have been reported to cause suicidal thoughts or suicidal actions, have the first two highest value in *leverage* as well as the first and third highest value in *PRR* and *lift*. As to the last association which we are supposed to identify, however, *Tacrolimus* ⇒ *Cancer* appeared to be unimpressive. Both *leverage* and *lift* only ranked this association as the sixth; however, *PRR* is able to detect this association as the first. This result may due to the characteristics of cancer itself that a number of cancers cannot be diagnosed in time. Therefore, many cancer patients would not be aware of their situation at early stages. Thus it is less possible for them to discuss their cancers symptoms caused by the drugs they are taking in online health communities.

On the other hand, health consumers can assess their symptoms in diarrhea, heart disease, depression, and suicidal thoughts easily without consulting health professionals. Therefore, discussions on these ADRs can be easily identified. The reason that *PRR* has a better performance than other metrics could be ascribed to the fact that the calculation of *PRR* is only based on the threads which mentioned the ADRs.

Generally speaking, our algorithm can effectively detect the drugs which have active safety alert from FDA. All the *drug* ⇒ *adverse reaction* associations which we are interested in ranked among the first three places in terms of *leverage*, *lift* and *PRR* except *Biaxin* ⇒ *Heart Disease*, *Luvox*⇒ *Heart Disease* and *Tacrolimus* ⇒*Cancer* that only ranked in the middle. Although *Tacrolimus* ⇒*Cancer* cannot be effectively detected using *leverage* and *lift*, it can be identified with *PRR*.

It is found that we cannot simply apply one threshold on *leverage*, *lift*, or *PRR* for all drugs and ADRs to detect the drug-adverse reaction association. The *leverage*, *lift*, and *PRR* values vary substantially across different drugs and across different ADRs. By applying a simple threshold, we can easily miss the true drug-adverse reaction associations or identify many false drug-adverse reaction associations. This can be reflected by the diverse discussions on drugs and the variation of vocabulary usage in describing ADRs.

We can see from the results that for each adverse reaction in this experiment, most of the *drug* ⇒ *adverse reaction* pairs of interest, which are alerted by FDA, ranked highly among all the *drug* ⇒ *adverse reaction* pairs in terms of those three metrics especially *PRR* in our dataset. This indicates that our algorithm is effective in terms of detecting FDA alerted *drug* ⇒ *adverse reaction* pairs.

Moreover, in this experiment, we also found several other high-value associations such as *Biaxin* ⇒ *Diarrhea*, *Luvox* ⇒ *Depression*, *Adenosine* ⇒ *Heart Disease* etc. Actually, these adverse reactions have already been labeled as one of the sides effects of corresponding drugs, and that's why users also talk about these ADRs a lot in online health communities. In addition, we identified several other impressive associations that have not been either reported by FDA or labeled, such as *Biaxin* ⇒ *Cancer* and *Lansoprazole* ⇒ *Cancer*. These *drug* ⇒ *adverse reaction* pairs all have high value in *leverage*, *lift* and *PRR*, which means there are many MedHelp users are discussing about them. These may be potential ADR signals deserving our attention and further investigation. But current study is not addressing this problem, and it can be a part of our future study.

[7] http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111085.htm

Table 2 Leverage, Lift and PRR

Drug Name	Diarrhea	Heart Disease	Depression	Suicidal	Cancer
Leverage					
Biaxin	4.52E-03(2)	-1.89E-04(5)	-1.18E-02(9)	-1.84E-03(9)	8.31E-03(1)
Lansoprazole	8.47E-03(1)	-8.21E-07(3)	-1.01E-02(6)	-1.52E-03(7)	5.95E-03(2)
Luvox	1.40E-03(3)	2.78E-04(2)	3.16E-02(2)	7.20E-03(1)	1.13E-03(3)
Prozac	-1.29E-03(4)	-2.60E-03(9)	3.63E-02(1)	2.05E-03(2)	-5.91E-03(10)
Tacrolimus	-1.73E-03(6)	-2.51E-03(8)	-1.61E-02(10)	-1.49E-03(6)	-8.12E-04(6)
Adenosine	-2.59E-03(9)	1.05E-02(1)	-8.14E-03(5)	-1.76E-03(8)	-7.48E-04(5)
Cialis	-3.22E-03(10)	-1.88E-04(4)	-6.70E-03(4)	-1.24E-03(5)	-5.23E-04(4)
Elidel	-1.95E-03(7)	-2.87E-03(10)	-1.16E-02(7)	-1.94E-03(10)	-1.80E-03(8)
Lantus	-2.00E-03(8)	-2.10E-04(6)	-1.17E-02(8)	-1.23E-03(4)	-1.07E-03(7)
Vyvanse	-1.60E-03(5)	-2.25E-03(7)	8.19E-03(3)	1.78E-03(3)	-4.54E-03(9)
Lift					
Biaxin	2.05E+00(2)	9.49E-01(5)	5.56E-01(7)	2.07E-01(8)	1.86E+00(1)
Lansoprazole	3.29E+00(1)	1.00E+00(3)	5.60E-01(6)	2.40E-01(7)	1.72E+00(2)
Luvox	1.39E+00(3)	1.09E+00(2)	2.44E+00(1)	4.73E+00(1)	1.14E+00(3)
Prozac	7.13E-01(4)	3.30E-01(7)	2.31E+00(2)	1.84E+00(3)	4.13E-01(10)
Tacrolimus	5.27E-01(6)	2.03E-01(9)	2.84E-01(10)	2.43E-01(6)	9.01E-01(6)
Adenosine	2.71E-01(10)	4.44E+00(1)	6.28E-01(5)	8.34E-02(9)	9.06E-01(5)
Cialis	3.09E-01(9)	9.53E-01(4)	7.67E-01(4)	5.08E-01(4)	9.50E-01(4)
Elidel	4.96E-01(7)	1.43E-01(10)	5.15E-01(8)	7.64E-02(10)	7.93E-01(8)
Lantus	4.68E-01(8)	9.36E-01(6)	4.96E-01(9)	3.94E-01(5)	8.74E-01(7)
Vyvanse	5.45E-01(5)	2.63E-01(8)	1.38E+00(3)	1.93E+00(2)	4.26E-01(9)
PRR					
Biaxin	2.48E+00(2)	9.90E-01(4)	5.53E-01(9)	1.98E-01(8)	2.19E+00(2)
Lansoprazole	4.01E+00(1)	9.28E-01(5)	4.97E-01(10)	2.06E-01(7)	1.72E+00(3)
Luvox	7.03E-01(6)	5.34E-01(6)	1.38E+00(2)	3.67E+00(1)	5.62E-01(8)
Prozac	3.54E-01(9)	1.56E-01(10)	1.43E+00(1)	1.07E+00(3)	1.98E-01(10)
Tacrolimus	1.35E+00(3)	5.04E-01(7)	7.11E-01(7)	6.06E-01(6)	2.39E+00(1)
Adenosine	2.69E-01(10)	7.19E+00(1)	6.45E-01(8)	8.14E-02(10)	9.56E-01(7)
Cialis	3.78E-01(8)	1.27E+00(3)	9.96E-01(4)	6.38E-01(5)	1.26E+00(6)
Elidel	9.50E-01(4)	2.65E-01(8)	9.89E-01(5)	1.40E-01(9)	1.57E+00(4)
Lantus	7.76E-01(5)	1.63E+00(2)	8.25E-01(6)	6.47E-01(4)	1.51E+00(5)
Vyvanse	4.81E-01(7)	2.26E-01(9)	1.32E+00(3)	1.96E+00(2)	3.71E-01(9)

Note: ( ) denotes the ranking of the scores

4.4 Measurement

In order to evaluate our approaches and better compare the performance of *leverage, lift* and PRR, we used top-k recall as a measurement to see among the seven *drug ⇒ adverse reaction* associations we are supposed to identify, how many of them are ranked within top k (k = 2, 3, 4) places.

Table 3 Top-k Recall

	Top2	Top3	Top4	Avg
Leverage	0.714	0.714	0.714	0.714
Lift	0.571	0.714	0.714	0.666
PRR	0.571	0.714	0.857	0.714

According to Table which shows the top-k recall of our experiment results, we can see that each metric has diverse performance compared with other metrics in each top-k recall ratio. However, in terms of average top-k recall value, *leverage* and *PRR* are better than *lift*. Therefore, we can conclude that *leverage* and *PRR* is relatively superior to *lift* in identifying *drug ⇒ adverse reaction* associations that we are supposed to find out according to FDA's active safety alerts.

5. DISCUSSION

In our previous research [26], we compared the performance of metrics *support*, *confidence* and *leverage*. The results showed that *leverage* outperformed other two measures because it eliminated the portion of independent relationship between a drug and an adverse reaction. In this study, we find that *leverage* and *PRR* had similar performance. For ADR such as cancer, *PRR* is performing far better than *leverage* although both leverage and *PRR* are promising in detecting signals of adverse drug reactions.

According to formulas (1) and (2), *leverage* and *lift* both compare the probability of the co-occurrence of drug *D* and adverse reaction *R* with what to expect if they are independent. *Leverage* calculates their difference, and *lift* calculates their quotient. The values of *leverage* and *lift* are inconsistent, especially for items with low counts in the dataset. For rare item sets with low counts in the dataset, the *leverage* value would be low, but the *lift* value might be enormous. In this study, the thread numbers of different drugs and adverse reactions vary substantially, so the results indicated by leverage and lift are different for some cases.

PRR is calculated only based on threads related to ADR and it could be presented as

$$PRR(D,R) = \frac{\frac{count(D \cup R)}{count(D)}}{\frac{count(!D \cup R)}{count(!D)}} = \frac{\frac{P(D \cup R)}{P(D) \times P(R)}}{\frac{P(!D \cup R)}{P(!D) \times P(R)}} = \frac{lift(D \Rightarrow R)}{lift(!D \Rightarrow R)} \qquad (4)$$

Formula (4) indicates that *PRR* is closely related to the *lift* indicator. However, different from *lift*, *PPR* also considers the universality of the adverse reaction associated with other drugs ($lift(!D \Rightarrow R)$). So, *PRR* is more likely to detect specific adverse reactions rather than common reactions that most drugs are associated with.

Natural language processing (NLP) approaches can be used to further analyze drug-adverse reaction associations. For example, using NLP techniques such as part-of-speech tagging to explore each sentence which is talking about the adverse reactions, we might be able to understand what aspects of ADRs the online health consumers are discussing or to detect consumers' attitudes toward an ADR (positive or negative) by identifying the existence of negations. Many researchers have applied NLP techniques to the field of health informatics. For example, Genevieve et al. [27] used NLP to effectively detect adverse events defined in the New York Patient Occurrence Reporting and Tracking System (NYPORTS) using discharge summaries. Murff et al. [28] found out that among patients undergoing inpatient surgical procedures at Veterans Health Administration medical centers, natural language processing analysis of electronic medical records to identify postoperative complications had higher sensitivity and lower specificity compared with patient safety indicators based on discharge coding. However, all these studies used NLP techniques to process and analyze electronic health records which were reported by health professionals, meaning that these reports were formal texts and consisted of well-structured sentences. Therefore, NLP analysis can be easily and efficiently applied to these materials. Very few works are focused on the context of online health communities in which the contents are mostly contributed by health consumers. Qiu et al. [29] used machine learning and text mining techniques to perform sentiment analysis on an online forum dataset – American Cancer Society Cancer Survivors Network (CSN). This approach automatically estimated the sentiment of forum posts, discovered sentiment change patterns in CSN members, and allowed

investigation of factors that affect the sentiment change. Their sentiment analysis was based on Bag-of-Word to identify positive and negative words as well as the strength of those words. However, one the biggest disadvantages of Bag-of-Word approach is that it is difficult to identify what the sentiment is about and multiple sentiments could be expressed in different sentences of the thread. Therefore, sentence-level sentiment analysis is sometimes preferable. NLP techniques such as POS-tagging can be used to analyze the structure of sentences and then extract words related to sentiment as well as what the sentiment is about [30]. However, one big challenge of using NLP in health consumer-contributed contents is that we must deal with overwhelming spelling errors, abbreviations and especially incomplete sentences. For example, the following sentence was extracted from our dataset, which was talking about side effects of a drug.

> Side Effect: dry mouth, headache, nausea, sweating, somnolence, insomnia, diarrhea, constipation, decreased appetite, sexual dysfunction, fatigue, pyrexia, sinusitis, rash, pruritus, bradycardia, tachycardia, hypertension, hypotension, spontaneous abortion.

By using Stanford's POS tagger [31], the tagging results are shown below.

> Side/NN Effect/NN :/: dry/JJ mouth/NN ,/, headache/NN ,/, nausea/NN ,/, sweating/NN ,/, somnolence/NN ,/, insomnia/NN ,/, diarrhea/NN ,/, constipation/NN ,/, decreased/VBD appetite/NN ,/, sexual/JJ dysfunction/NN ,/, fatigue/NN ,/, pyrexia/NN ,/, sinusitis/NN ,/, rash/NN ,/, pruritus/NN ,/, bradycardia/NN ,/, tachycardia/NN ,/, hypertension/NN ,/, hypotension/NN ,/, spontaneous/JJ abortion/NN ./.

As we can see, there is no formal structure for this sentence which only consists of adjectives (including "decreased" in this case) and nouns, and it is very difficult to perform semantic and sentiment analysis on this kind of texts. However, sentiment analysis is a very interesting topic for our future research on the base of health consumer-contributed contents.

Another direction for future work is to expand current ADRs lexicon. By building an extensive ADRs lexicon, we could improve the accuracy of ADRs signals detection. After closely observing the dataset collected in this study, we found that consumers used diverse expressions in online drug forum when talking about the ADR diarrhea, such as "gassy", "loose stool", "early dumping", "Number 2", etc.. Following is an example extracted from the data. The terms with double underline are the terms that have already been in the lexicon, and the terms with single underline are related terms.

> I have had pain above my navel which radiates left and right, terrible nausea with retching and sometimes vomiting, and very loose and watery stools. I have had a headache just to the left of my forehead, and during the nausea/retching/diarrhea episodes, my skin becomes very clammy and I feel hot and like I'm going to pass out.

These expressions are all related to the symptoms of diarrhea. This means that people would describe their symptoms in details when discussing about the ADR they are suffering. Therefore, expressions of symptoms of certain type of ADR should be also included in the lexicon. In addition, people always mention nausea and vomiting along with diarrhea. It may be because that these three types of adverse reactions are all related to digestive system problems. These kinds of terms are also useful for expanding the lexicon. However, not every related term has the same significance for detecting an ADR. We could differentiate important terms from those with minor significance by assigning weights to the terms in the lexicon. By including more related terms and assigning weights to the terms, we could improve the lexicon and further enhance the performance of detection.

6. CONCLUSION

There is a huge potential to identify ADRs from social media, because more and more people participate in online health communities to share health experiences, and the data is open and public which could be accessed in real time. In this study, we collected posts and comments of 10 drugs from MedHelp, extracted 5 FDA-alerted adverse reactions of these drugs as ground truth, and employed association mining and *PRR* method to detect the ADR signals for each pair of drug and adverse reaction. *Leverage* and *lift* were used as indicators for association mining, and they were compared to results based on *PRR*. Although all of the three indicators worked effectively to detect ADR according to our experiment, *PRR* and *leverage* generated better results than *lift*.

In the future work, we plan to extend our dataset to include more drugs and adverse reactions for further investigation. To analyze the context of ADR signals to make better decision, content analysis and sentiment analysis would be employed. Moreover, it is still a challenge to identify and match ADR terms from expressions in social media. We will extend ADR lexicon based on other lexicons to capture ADR terms as much as possible.

7. REFERENCES

[1] Edwards, I. R. and Aronson, J. K. 2000. Adverse drug reactions: definitions, diagnosis, and management. *Lancet*, 356, 9237 (2000), 1255-1259.

[2] Kohn, L. T., Corrigan, J. M., and Donaldson, M. S.. To Error is Human: Building a Safer Health System. Washington, DC: *National Academy Press*, 1999.

[3] Lasser, K. E., Allen, P. D., Woolhandler, S. J., Himmelstein, D. U., Wolfe, S. M., and Bor, D. H. "Timing of new black box warnings and withdrawals for prescription medications," *J. Amer. Med. Assoc.*, vol. 287, pp. 2215–2220, 2002.

[4] Blenkinsopp A, Wilkie P, Wang M, Routledge PA. Patient reporting of suspected adverse drug reactions: a review of published literature and international experience. *Br J Clin Pharmacol*. 2007;63(2):148-156.

[5] Huidong, Jin., Jie, C., Hongxing, H., Williams, G. J., Kelman, C. and O'Keefe, C. M. 2008. Mining Unexpected Temporal Associations: Applications in Detecting Adverse Drug Reactions. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 12, 4 (2008), 488-500.

[6] Yanqing, J., Hao, Y., Dews, P., Farber, M. S., Mansour, A., Tran, J., Miller, R. E. and Massanari, R. M. A fuzzy recognition-primed decision model-based causal association mining algorithm for detecting adverse drug reactions in postmarketing surveillance. IEEE, City, 2010.

[7] Yanqing, J., Hao, Y., Dews, P., Mansour, A., Tran, J., Miller, R. E. and Massanari, R. M. 2011. A Potential Causal Association Mining Algorithm for Screening Adverse Drug Reactions in Postmarketing Surveillance. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 15, 3 (2011), 428-437

[8] Evans, S. J. W., Waller, P. C. and Davis, S. 2001. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and Drug Safety*, 10, 6 (2001), 483-486.

[9] Hakobyan, L., Haaijer-Ruskamp, F. M., de Zeeuw, D., Dobre, D. and Denig, P. 2011. A review of methods used in assessing

non-serious adverse drug events in observational studies among type 2 diabetes mellitus patients. *Health and quality of life outcomes*, 9, 1 (2011), 83-83.

[10] Wu, Y. T. and Makuch, R. W. 2006. Detecting rare adverse events in postmarketing studies: Sample size considerations. *DRUG INFORMATION JOURNAL*, 40, 1 (2006), 89-98.

[11] Kubota, K., Koide, D. and Hirai, T. 2004. Comparison of data mining methodologies using Japanese spontaneous reports. *Pharmacoepidemiology and Drug Safety*, 13, 6 (2004), 387-394.

[12] Szarfman, A., Tonning, J. M. and Doraiswamy, P. M. 2004. Pharmacovigilancein the 21st century: New systematic tools for an old problem. *Pharmacotherapy*, 24(2004), 1099-1104.

[13] Lindquist, M., Edwards, I. R., Bate, A., Fucik, H., Nunes, A. M. and Stahl, M. 1999. From association to alert - A revised approach to international signalanalysis. *Pharmacoepidemiol. Drug Safety*, 1(1999), 15-25.

[14] van Puijenbroek, E. P., Egberts, A. C. G., Meyboom, R. H. B. and Leufkens, H. G. M. 1999. Signalling possible drug–drug interactions in a spontaneous reporting system: delay of withdrawal bleeding during concomitant use of oral contraceptives and itraconazole. *Br J Clin Pharmacol*, 47(1999), 689-693.

[15] DuMouchel, W. 1999. Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *The American Statistician*, 53, 3 (1999), 177-190.

[16] Hauben, M. 2004. Early postmarketing drug safety surveillance: data mining points to consider. *The Annals of pharmacotherapy*, 38, 10 (2004), 1625-1630.

[17] Lanctot, K. L. and Naranjo, C. A. 1994. COMPUTER-ASSISTED EVALUATION OF ADVERSE EVENTS USING A BAYESIAN-APPROACH. *JOURNAL OF CLINICAL PHARMACOLOGY*, 34, 2 (1994), 142-147.

[18] Alan, M. H., Stephanie, J. R., Ronald, K. P., Donald, J. O. H. and Kevin, H. 2007. Using Data Mining to Predict Safety Actions From FDA Adverse Event Reporting System Data. *Drug Information Journal*, 41, 5 (2007), 633.

[19] Ji, Y. Q., Massanari, R. M., Ager, J., Yen, J., Miller, R. E. and Ying, H. 2007. A fuzzy logic-based computational recognition-primed decision model. *Inf. Sci.*, 177, 20 (2007), 4338-4353.

[20] Ji, Y., Ying, H., Farber, M. S., Yen, J., Dews, P., Miller, R. E. and Massanari, R. M. 2010. A distributed, collaborative intelligent agent system approach for proactive postmarketing drug safety surveillance. *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE*, 14, 3 (2010), 826-837.

[21] Pouliot, Y., Chiang, A. P. and Butte, A. J. 2011. Predicting Adverse Drug Reactions Using Publicly Available PubChem BioAssay Data. *Clinical Pharmacology & Therapeutics*, 90, 1 (2011), 90-99.

[22] Khanna, N. H. and Kannan, A. 2006. A DIAGNOSTIC DECISION SUPPORT SYSTEM FOR ADVERSE DRUG REACTION USING TEMPORAL REASONING. *AIML Journal*, 6, 2 (2006), 79-86.

[23] Chee, B. W., Berlin, R. and Schatz, B. 2011. Predicting adverse drug events from personal health messages. *Annual Symposium proceedings*, (2011), 217-226.

[24] Leaman, R.,Wojtulewicz,L., Sullivan, R., Skariah,A., Yang,J., and Gonzalez,G. 2010. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks. *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, ACL 2010, 117–125.

[25] Zeng, Q. T., & Tse, T. (2006). Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association : JAMIA, 13*(1), 24-29.

[26] C. C. Yang, L. Jiang, H. Yang, and X. Tang, "Detecting Signals of Adverse Drug Reactions from Health Consumer Contributed Content in Social Media," *Proceedings of ACM SIGKDD Workshop on Health Informatics*, Beijing, August 12, 2012

[27] Melton, G. B. and Hripcsak, G. 2005. Automated Detection of Adverse Events Using Natural Language Processing of Discharge Summaries. *Journal of the American Medical Informatics Association*, 12, 4 (2005), 448-457.

[28] Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., Dittus, R. S., Rosen, A. K., Elkin, P. L., Brown, S. H. and Speroff, T. 2011. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA : the journal of the American Medical Association*, 306, 8 (2011), 848.

[29] Qiu, B. J., Zhao, K., Mitra, P., Wu, D. H., Caragea, C. and Yen, J. Get Online Support，Feel Better--Sentiment Analysis and Dynamics in an Online Cancer Survivor Community. *The Proceeding of IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing,* 2011, pp. 274-281.

[30] Liu, B., Hu, M. Q. and Cheng, J. S. Opinion Observer: Analyzing and Comparing Opinions on the Web. *Proceedings of the 14th international World Wide Web conference (WWW-2005)*, May 10-14, 2005, in Chiba, Japan.

[31] Toutanova, K., Klein, D., Manning, C. and Singer, Y. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*, pp. 252-259.