

Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-Related Social Networks

Robert Leaman¹, Laura Wojtulewicz², Ryan Sullivan²
Annie Skariah², Jian Yang¹, Graciela Gonzalez²

¹School of Computing, Informatics and Decision Systems Engineering

²Department of Biomedical Informatics

Arizona State University, Tempe, Arizona, USA

{robert.leaman, whitz, rpsulli, annie.skariah,
jian.yang, graciela.gonzalez}@asu.edu

Abstract

Adverse reactions to drugs are among the most common causes of death in industrialized nations. Expensive clinical trials are not sufficient to uncover all of the adverse reactions a drug may cause, necessitating systems for post-marketing surveillance, or pharmacovigilance. These systems have typically relied on voluntary reporting by health care professionals. However, self-reported patient data has become an increasingly important resource, with efforts such as MedWatch from the FDA allowing reports directly from the consumer. In this paper, we propose mining the relationships between drugs and adverse reactions as reported by the patients themselves in user comments to health-related websites. We evaluate our system on a manually annotated set of user comments, with promising performance. We also report encouraging correlations between the frequency of adverse drug reactions found by our system in unlabeled data and the frequency of documented adverse drug reactions. We conclude that user comments pose a significant natural language processing challenge, but do contain useful extractable information which merits further exploration.

1 Introduction

It is estimated that approximately 2 million patients in the United States are affected each year by severe adverse drug reactions, resulting in roughly 100,000 fatalities. This makes adverse drug reactions the fourth leading cause of death in the

U.S, following cancer and heart diseases (Giacomini et al., 2007). It is estimated that \$136 billion is spent annually on treating adverse drug reactions in the U.S., and other nations face similar difficulties (van Der Hooft et al., 2006; Leone et al., 2008). Unfortunately, the frequency of adverse drug reactions is often under-estimated due to a reliance on voluntary reporting (Bates et al., 2003; van Der Hooft et al., 2006).

While severe adverse reactions have received significant attention, less attention has been directed to the indirect costs of more common adverse reactions such as nausea and dizziness, which may still be severe enough to motivate the patient to stop taking the drug. The literature shows, however, that non-compliance is a major cause of the apparent failure of drug treatments, and the resulting economic costs are estimated to be quite significant (Urquhart, 1999; Hughes et al., 2001). Thus, detecting and characterizing adverse drug reactions of all levels of severity is critically important, particularly in an era where the demand for personalized health care is high.

1.1 Definitions

An adverse drug reaction is generally defined as an unintended, harmful reaction suspected to be caused by a drug taken under normal conditions (World Health Organization, 1966; Lee, 2006). This definition is sufficiently broad to include such conditions as allergic reactions, drug tolerance, addiction or aggravation of the original condition. A reaction is considered *severe* if it “results in death, requires hospital admission or prolongation..., results in persistent or significant disability/incapacity, or is life-threatening,” or if it causes a congenital abnormality (Lee, 2006).

1.2 Pharmacovigilance

The main sources of adverse drug reaction information are clinical trials and post-marketing surveillance instruments made available by the Food and Drug Administration (FDA), Centers for Disease Control and Prevention (CDC) in the United States, and similar governmental agencies worldwide. The purpose of a clinical trial, however, is only to determine whether a product is effective and to detect common serious adverse events. Clinical trials, by their nature and purpose, are focused on a limited number of participants selected by inclusion/exclusion criteria reflecting specific subject characteristics (demographic, medical condition and diagnosis, age). Thus, major uncertainties about the safety of the drug remain when the drug is made available to a wider population over longer periods of time, in patients with co-morbidities and in conjunction with other medications or when taken for off-label uses not previously evaluated.

Recently, the regulatory bodies of both the U.S. and the U.K. have begun programs for patient reporting of adverse drug reactions. Studies have shown that patient reporting is of similar quality to that of health professionals, and there is some evidence that patients are more likely to self-report adverse drug reactions when they believe the health professionals caring for them have not paid sufficient attention to an adverse reaction (Blenkinsopp et al., 2007). In general, however, the FDA advocates reporting only serious events through MedWatch.

Self-reported patient information captures a valuable perspective that might not be captured in a doctor's office, clinical trial, or even in the most sophisticated surveillance software. For this reason, the International Society of Drug Bulletins asserted in 2005 that "patient reporting systems should periodically sample the scattered drug experiences patients reported on the internet."

1.3 Social Networks

Social networks focusing on health related topics have seen rapid growth in recent years. Users in an online community often share a wide variety of personal medical experiences. These interactions can take many forms, including blogs, microblogs and question/answer discussion forums. For many reasons, patients often share health experiences with each other rather than in a clinical

research study or with their physician (Davison et al., 2000). Such social networks bridge the geographical gap between people, allowing them to connect with patients who share similar conditions—something that might not be possible in the real world.

In this paper we propose and evaluate automatically extracting relationships between drugs and adverse reactions in user posts to health-related social network websites. We anticipate this technique will provide valuable additional confirmation of suspected associations between drugs and adverse reactions. Moreover, it is possible this technique may eventually provide the ability to detect novel associations earlier than with current methods.

2 Related Work

In the work closest in purpose to this study, two reviewers manually analyzed 1,374 emails to the BBC and 862 messages on a discussion forum regarding a link between the drug paroxetine and several adverse reactions including withdrawal symptoms and suicide (Medawara et al., 2002). The authors concluded that the user reports contained clear evidence of linkages that the voluntary reporting system then in place had not detected.

Not much work has been done to automatically extract adverse reactions from text, other than the SIDER side effect resource, which was created by mining drug insert literature (Kuhn et al., 2010). There is, however, significant literature support for mining more general concepts, such as diseases. MetaMap is a primarily lexical system for mapping concepts in biomedical text to concepts in the UMLS Metathesaurus (Aronson, 2001). The ConText system categorizes findings in clinical records as being negated, hypothetical, or historical (Harkema et al., 2009).

Most of the work on finding diseases concerns either biomedical text or clinical records. A notable exception is the BioCaster system, which detects infectious disease outbreaks by mining news reports posted to the web (Collier et al., 2008).

Health social networks have become a popular way for patients to share their health related experiences. A considerable amount of research has been devoted to this area (Moturu et al., 2008), but most of this work has focused on the study of social interactions and quality evaluation instead of text mining. Automated information extrac-

tion from health social network websites remains largely unexplored.

3 Data Preparation

We used the DailyStrength¹ health-related social network as the source of user comments in this study. DailyStrength allows users to create profiles, maintain friends and join various disease-related support groups. It serves as a resource for patients to connect with others who have similar conditions, many of whom are friends solely online. As of 2007, DailyStrength had an average of 14,000 daily visitors, each spending 82 minutes on the site and viewing approximately 145 pages (comScore Media Metrix Canada, 2007).

3.1 Data Acquisition

To efficiently gather user comments about specific drugs from the DailyStrength site, we implemented a highly parallelized automatic web crawler. All data was scraped from the raw HTML using regular expressions since the site has no open API. Users indicate a specific treatment when posting comments to DailyStrength, however we filter treatments which are not drugs. For each user comment we extracted the user ID, disease name, drug name, and comment text. While more information about each user is available at the site (gender, age, self-declared location, and length of membership at the site), we limited our data usage to just the comment data. The DailyStrength Privacy Policy states that comments made by users will be publicly available. All data was gathered in accordance with the DailyStrength Terms of Service, and to respect fair use the data will not be made publicly available without permission from the site.

3.2 Preparing the Lexicon

To enable finding adverse reactions in the user comments, we created a lexicon by combining terms and concepts from four resources.

The UMLS Metathesaurus is a resource containing many individual biomedical vocabularies (National Library of Medicine, 2008). We utilized a subset limited to the COSTART vocabulary created by the U.S. Food and Drug Administration for post-marketing surveillance of adverse drug reactions, which contains 3,787 concepts.

¹<http://www.dailystrength.org>

The SIDER side effect resource contains 888 drugs linked with 1,450 adverse reaction terms extracted from pharmaceutical insert literature (Kuhn et al., 2010). We used the raw term found in the literature and the associated UMLS concept identifier (CUI).

The Canada Drug Adverse Reaction Database, or MedEffect², contains associations between 10,192 drugs and 3,279 adverse reactions, which we used to create a list of adverse reaction terms. We found many adverse reaction terms with very similar meanings, for example “appetite exaggerated,” and “appetite increased,” which we grouped together manually.

We also included a small set of colloquial phrases we located manually in a subset of the DailyStrength comments and mapped to UMLS CUIs. This list is available³, and includes the terms “throw up,” meaning *vomit*, “gain pounds,” meaning *weight gain*, and “zoned out,” meaning *somnolence*.

We considered all terms which are associated with the same UMLS concept identifier (CUI) as synonymous and grouped them into a single concept. We also merged all concepts containing a term in common into a single unified concept. Our lexicon contains 4,201 unified concepts, each containing between one and about 200 terms.

4 Annotation

We annotated comments relating to the following 4 drugs: carbamazepine, olanzapine, trazodone, and ziprasidone. These drugs were chosen because they are known to cause adverse reactions and we could verify our results with close collaborators. We retained but did not annotate comments for the drugs aspirin and ciprofloxacin; these comments are used during evaluation. Our data contains a total of 6,890 comment records. User comments were selected for annotation randomly and were independently annotated by two annotators.

Annotator 1 has a BS in biology, 10 years nursing experience in the behavioral unit of a long term care facility, and has dispensed all of the drugs annotated. Annotator 2 has a BS and an MS in neuroscience, and has work experience in data management for pharmaceutical-related clinical research and post-marketing drug surveillance.

²<http://www.hc-sc.gc.ca/dhp-mps/medeff/index-eng.php>

³<http://diego.asu.edu/downloads/adrs>

Concept	Definition
<i>Adverse effect</i>	A reaction to the drug experienced by the patient, which the user considered negative
<i>Beneficial effect</i>	A reaction to the drug experienced by the patient, which the user considered positive
<i>Indication</i>	The condition for which the patient is taking the drug
<i>Other</i>	A disease or reaction related term not characterizable as one of the above

Table 1: The concepts annotated in this study and their definitions.

4.1 Concepts Annotated

Each comment was annotated for mentions of *adverse effects*, *beneficial effects*, *indications* and *other* terms, as defined in table 1. Each annotation included the span of the mention and the name of the concept found, using entries from the lexicon described in section 3.2. Each annotation also indicates whether it refers to an *adverse effect*, a *beneficial effect*, an *indication* or an *other* term, which we shall call its characterization.

4.2 Annotation Practices

There are four aspects which require careful consideration when characterizing mentions. First, the stated concept may or may not be actually experienced by the patient; mentions of concepts not experienced by the patient were categorized as *other*. Second, the user may state that the concept is the reason for taking the drug. If so, the mention was categorized as an *indication*. Third, the concept may be an effect caused by the drug. In this case, the mention is categorized as either an *adverse effect* or a *beneficial effect* based on whether the user considers the effect a positive one. This requires some judgment regarding what people normally view as positive – while sleepiness is normally an *adverse effect*, someone suffering from insomnia would consider it a *beneficial effect*, regardless of whether insomnia is the primary reason for taking the drug. Mentions of concepts which were experienced by the patient but neither an effect of the drug nor the reason for taking it were also categorized as *other*. Concepts were characterized as an *adverse effect* unless the context indicated otherwise.

Comments not containing a mention or that only indicated the presence of an adverse effect (“Gave

me weird side effects”) were discarded. If more than one mention occurred in a comment, then each mention was annotated separately.

Some comments clearly mentioned an adverse reaction, but the reaction itself was ambiguous. For example, in the comment “It did the job when I was really low. However, I BALLOONED on it,” the annotator could infer “BALLOONED” to mean either *weight gain* or *edema*. A frequent example is colloquial terms such as “zombie,” which could be interpreted as a physiological effect (e.g. *fatigue*) or a cognitive effect (e.g. *mental dullness*). In such cases, each mention was annotated by using both the context of the mention and annotator’s knowledge of the effects of the drug.

Spans were annotated by choosing the minimum span of characters from the comment that would maintain the meaning of the term. Locating the mention boundaries was straightforward in many cases, even when descriptive words were in the middle of the term (“It works better than the other meds ive taken but I am gaining some weight”). However some comments were not as simple (“it works but the pounds are packing on”).

4.3 Corpus Description

A total of 3,600 comments were annotated, a sample of which can be seen in table 2. We reserved 450 comments for system development. The annotators found 1,260 *adverse effects*, 391 *indications*, 157 *beneficial effects* and 78 *other*, for a total of 1,886 annotations.

We measured the agreement between annotators by calculating both kappa (κ) (Cohen, 1960) and inter-annotator agreement (IAA). For κ , we considered agreement to mean that the concept terms were in the same unified concept from the lexicon and the characterization of the mentions matched, since there is no standard method for calculating κ which includes the span. For IAA, we added the constraint that the annotation spans must overlap, since discussions of IAA typically include the span. Using these definitions, κ was calculated to be 85.6% and IAA to be 85.3%⁴.

5 Text Mining

Since the drug name is specified by the user when the comment is submitted to DailyStrength, no ex-

⁴ $\kappa > \text{IAA}$ here due to the different definitions of agreement.

Sample Comments	Annotations
hallucinations and weight gain	“hallucinations” - <i>hallucinations: adverse effect</i> ; “weight gain” - <i>weight gain: adverse effect</i>
This has helped take the edge off of my constant sorrow. It has also perked up my appetite. I had lost a lot of weight and my doctor was concerned.	“constant sorrow” - <i>depression: indication</i> ; “perked up my appetite” - <i>appetite increased: beneficial effect</i> ; “lost a lot of weight” - <i>weight loss: other</i>
It worked well, but doctor didn’t asked for the treatment to continue once my husband was doing well again.	<i>none</i>
ARGH! Got me nicely hypomaniac for two weeks, then pooped out on me and just made me gain a half pound a day so I had to stop.	“hypomaniac” - <i>hypomania: beneficial effect</i> ; “pooped out” - <i>tolerance: adverse effect</i> ; “gain a half a pound a day” - <i>weight gain: adverse effect</i>
Works to calm mania or depression but zonks me and scares me about the diabetes issues reported.	“mania” - <i>mania: indication</i> ; “depression” - <i>depression: indication</i> ; “zonks me” - <i>somnolence: adverse effect</i> ; “diabetes” - <i>diabetes: other</i>
Works for my trigeminal neuralgia. Increasing to see if it helps stabilize mood. Fatigue!	“trigeminal neuralgia” - <i>trigeminal neuralgia: indication</i> ; “stabilize mood” - <i>emotional instability: indication</i> ; “Fatigue” - <i>fatigue: adverse effect</i>
Take for seizures and bipolar works well	“seizures” - <i>seizures: indication</i> ; “bipolar” - <i>bipolar disorder: indication</i>
fatty patti!	“fatty” - <i>weight gain: adverse effect</i>

Table 2: An illustrative selection of uncorrected comments submitted to the DailyStrength health-related social networking website, and their associated annotations.

traction was necessary for drug names. To extract the adverse drug reactions from the user comments, we implemented a primarily lexical method, utilizing the lexicon discussed in section 3.2.

5.1 Methods Used

Each user comment was split into sentences using the Java sentence breaker, tokenized by splitting at whitespace and punctuation, and tagged for part-of-speech using the Hepple tagger (Hepple, 2000). Stop-words were removed from both user comments and lexical terms⁵. Tokens were stemmed using the Snowball implementation of the Porter2 stemmer⁶.

Terms from the lexicon were found in the user comments by comparing a sliding window of tokens from the comment to each token in the lexical term. The size of the window is configurable and set to 5 for this study since that is the number of tokens in the longest term found by the annotators. Using a sliding window allows the tokens to be in different orders and for there to be irrelevant tokens between the relevant ones, as in *weight gain* and “gained a lot of weight.”

Since user comments contain many spelling errors, we used the Jaro-Winkler measurement of string similarity to compare the individual tokens

(Winkler, 1999). We scored the similarity between the window of tokens in the user comment and the tokens in the lexical term by pairing them as an assignment problem (Burkard et al., 2009). We then summed the similarities of the individual tokens and normalized the result by the number of tokens in the lexical term. This score is calculated for both the original tokens and the stemmed tokens in the window, and the final score is taken to be the higher of the two scores. The lexical term is considered to be present in a user comment if the final score is greater than a configurable threshold.

We noted that most mentions could be categorized by using the closest verb to the left of the mention, as in “taking for seizures.” As this study focuses on *adverse effects*, we implemented a filtering method to remove *indications*, *beneficial effects*, and *other* mentions on a short list of verbs we found to indicate them. Verbs on this list include “helps,” “works,” and “prescribe” all of which generally denote *indications*. The complete list is available⁷.

5.2 Text Mining Results

We first evaluated the system against the 3,150 annotated comments not reserved for system development. Because our purpose is to find adverse drug reactions, we limited our evaluation to *ad-*

⁵http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words

⁶<http://snowball.tartarus.org>

⁷<http://diego.asu.edu/downloads/adrs>

verse effects. We used a strict definition of true positive, requiring the system to label the mention with a term from the same unified concept as the annotators. The results of this study are 78.3% precision and 69.9% recall, for an f-measure of 73.9%.

Since the purpose of this study is to determine if mining user comments is a valid way to find adverse reactions, we ran our system on all available comments and compared the frequencies of adverse reactions found against their documented incidence. We calculated the frequency that each *adverse effect* was found in the user comments for each of the drugs studied in this experiment. We then determined the most commonly found adverse reactions for each drug and compared them against the most common documented adverse reactions for the drug. Since the four drugs we chose for annotation all act primarily on the central nervous system, we added aspirin and ciprofloxacin for this study. The results of this evaluation contain encouraging correlations that are summarized in table 3.

6 Discussion

6.1 Error Analysis

We performed an analysis to determine the primary sources of error for our extraction system. We randomly selected 100 comments and determined the reason for the 24 false positives (FPs) and 29 false negatives (FNs) found.

The largest source of error (17% of FPs and 55% of FN) was the use of novel adverse reaction phrases (“liver problem”) and descriptions (“burn like a lobster”). This problem is due in part to idiomatic expressions, which may be handled by creating and using a specialist lexicon. This problem might also be partially relieved by the appropriate use of semantic analysis. However, this source of error is also caused by the users deliberately employing a high degree of linguistic creativity (“TURNED ME INTO THE SPAWN OF SATAN!!!”) which may require deep background knowledge to correctly recognize.

The next largest source of error was poor approximate string matching (46% of FPs and 17% of FN). While users frequently misspelled words, making lexical analysis difficult, the approximate string matching technique used also introduced many FPs. We note that spelling unfamiliar medical terminology is particularly difficult for users.

Correcting this important source of error will require improved modeling of the spelling errors made by users.

Ambiguous terms accounted for 8% of the FPs and 7% of the FN. While this is frequently a problem with colloquial phrases (“brain fog” could refer to *mental dullness* or *somnolence*), there are some terms which are ambiguous on their own (“numb” may refer to *loss of sensation* or *emotional indifference*). These errors can be corrected by improving the analysis of the context surrounding each mention.

Surprisingly, miscategorizations only accounted for 4% of the FPs. This small percentage seems to indicate that the simple filtering technique employed is reasonably effective. However this source of error can be seen more prominently in the frequency analysis, as seen in table 3. For example, one of the most frequent effects found in comments about trazodone was *insomnia*, which is one of its most common off-label uses. Other examples included *depression* with olanzapine, *mania* with ziprasidone, and *stroke* with aspirin. We note that since conditions not being experienced by the patient are always categorized as *other*, our techniques should profit somewhat from an extension to handle negation.

6.2 Analysis of Documented vs. Found Adverse Reactions

The experiment comparing the documented incidence of adverse reactions to the frequency they are found contained some interesting correlations and differences. We begin by noting that the adverse reaction found most frequently for all 6 of the drugs corresponded to a documented adverse reaction. There were also similarities in the less common reactions, such as *diabetes* with olanzapine and *bleeding* with aspirin. In addition, many of the adverse reactions found corresponded to documented, but less common, reactions to the drug. Examples of this included *edema* with olanzapine, *nightmares* with trazodone, *weight gain* with ziprasidone, *tinnitus* with aspirin, and *yeast infection* with ciprofloxacin.

One interesting difference is the relative frequency of “hangover” in the comments for ziprasidone. Since the users were not likely referring to a literal *hangover*, they were probably referring to the *fatigue*, *headache*, *dry mouth* and *nausea* that accompany a *hangover*, all of which are doc-

Drug name (Brand name)	Primary Indications	Documented Adverse Effects (Frequency)	Adverse Effects Found in User Comments (Frequency)
carbamazepine (Tegretol)	epilepsy, trigeminal neuralgia	dizziness, somnolence or fatigue , unsteadiness, nausea , vomiting	somnolence or fatigue (12.3%), allergy (5.2%), weight gain (4.1%), rash (3.5%), depression (3.2%), dizziness (2.4%), tremor/spasm (1.7%), headache (1.7%), appetite increased (1.5%), nausea (1.5%)
olanzapine (Zyprexa)	schizophrenia, bipolar disorder	weight gain (65%), alteration in lipids (40%), somnolence or fatigue (26%), increased cholesterol (22%), diabetes (2%)	weight gain (30.0%), somnolence or fatigue (15.9%), appetite increased (4.9%), depression (3.1%), tremor (2.7%), diabetes (2.6%), mania (2.3%), anxiety (1.4%), hallucination (0.7%), edema (0.6%)
trazodone (Oleptro)	depression	somnolence or fatigue (46%), headache (33%), dry mouth (25%), dizziness (25%), nausea (21%)	somnolence or fatigue (48.2%), nightmares (4.6%), insomnia (2.7%), addiction (1.7%), headache (1.6%), depression (1.3%), hangover (1.2%), anxiety attack (1.2%), panic reaction (1.1%), dizziness (0.9%)
ziprasidone (Geodon)	schizophrenia	somnolence or fatigue (14%), dyskinesia (14%), nausea (10%), constipation (9%), dizziness (8%)	somnolence or fatigue (20.3%), dyskinesia (6.0%), mania (3.7%), anxiety attack (3.5%), weight gain (3.2%), depression (2.4%), allergic reaction (1.9%), dizziness (1.2%), panic reaction (1.2%)
aspirin	pain, fever, reduce blood clotting	nausea, vomiting, ulcers , bleeding , stomach pain or upset	ulcers (4.5%), sensitivity (3.8%), stroke (3.1%), bleeding time increased (2.8%), somnolence or fatigue (2.7%), malaise (2.1%), weakness (1.4%), numbness (1.4%), bleeding (1.0%), tinnitus (0.7%)
ciprofloxacin (Cipro)	bacterial infection	diarrhea (2.3%), vomiting (2.0%), abdominal pain (1.7%), headache (1.2%), restlessness (1.1%)	abdominal pain (8.8%), malaise (4.4%), nausea (3.8%), allergy (3.1%), somnolence or fatigue (2.5%), dizziness (1.9%), weakness (1.6%), tolerance (1.5%), rash (1.3%), yeast infection (1.1%)

Table 3: List of drugs included in the subset for analysis, with their indications and 5 most common adverse effects together with their frequency of incidence in adults taking the drug over the course of one year, as listed in the FDA online drug library, <http://www.accessdata.fda.gov/scripts/cder/drugsatfda> (some frequency data is not available). Also the 10 most frequent adverse effects found in the the DailyStrength data using our automated system. Correlations are highlighted in bold.

umented adverse reactions to the drug.

Users frequently commented on *weight gain* and *fatigue* while ignoring other reactions such as *increased cholesterol*. While this may be because users are more conscious of issues they can directly observe, this hypothesis would not explain why other directly observable reactions such as *nausea* and *constipation* are not always reported. Determining the general trends in the differences between clinical and user reports is an important area for future work.

6.3 Limitations

The present study has some limitations. We did not analyze the demographics of the users whose comments we mined, though it is likely that they are predominantly from North America and English-speaking. In future work we intend to expand the range of users and compare their demographics against clinical studies of adverse reactions. Also, the drugs we annotated oper-

ate primarily on the central nervous system and therefore have different adverse reaction profiles than would other drugs with substantially different mechanisms. While the inclusion of aspirin and ciprofloxacin does provide some evidence these techniques are more generally applicable, we also intend to expand the range of drugs studied in future work.

6.4 Opportunities for Further Study

In addition to our current classification for adverse reactions, there are additional dimensions along which each user comment could be studied. For example, many comments describe the degree of the adverse reaction, which can be straightforward (“extremely”) or more creative (“like a pig”). Also, many users explicitly state whether they are still taking the drug, typically indicating whether their physician took them off or whether they took themselves off (non-compliance), and whether adverse reactions were the reason. User

comments can also be categorized as medically non-descriptive (“I took one tablet and could not get out of bed for days and felt like I got hit by a truck”), somewhat medically descriptive (“My kidneys were not functioning properly”), or medically sound (“I ended up with severe leg swelling”). Comments also typically indicate whether the user is the patient or a caretaker by being phrased in either the first person or third person narrative. Finally, users also frequently describe whether they thought the benefits of the drug outweighed the adverse effects. We believe these additional dimensions represent a fertile area for further research.

7 Conclusion

In summary, we have shown that user comments to health related social networks do contain extractable information relevant to pharmacovigilance. We believe this approach should be evaluated for the ability to detect novel relationships between drugs and adverse reactions.

In addition to the improvements discussed in section 6, we plan in future work to increase the scale of the study (additional drugs, additional data sources, more user comments), improve the characterization of reactions using rule-based patterns, and evaluate the improved system with respect to all characterizations.

Acknowledgments

The authors would like to thank Dr. Diana Pettiti for her early support and suggestions, Tasnia Tahsin for reviewing an earlier version, Skatje Myers for locating mergeable reaction concepts, and the anonymous reviewers for many useful suggestions. The authors are grateful for support from Science Foundation Arizona grant CAA 0277-08, the Arizona Alzheimers Disease Data Management Core under NIH Grant NIA P30 AG-19610, and the Arizona Alzheimers Consortium pilot grant.

References

Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In *Proceedings of the AMIA Symposium*, page 17. American Medical Informatics Association.

D.W. Bates, R.S. Evans, H. Murff, P.D. Stetson, L. Pizziferri, and G. Hripsak. 2003. Detecting ad-

verse events using information technology. *Journal of the American Medical Informatics Association*, 10(2):115–128.

A. Blenkinsopp, M. Wang, P. Wilkie, and P. A. Routledge. 2007. Patient reporting of suspected adverse drug reactions: a review of published literature and international experience. *British Journal of Clinical Pharmacology*, 63(2):148–156.

Rainer Burkard, Mauro Dell’Amico, and Silvano Martello. 2009. *Assignment Problems*. Society for Industrial and Applied Mathematics.

Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

Nigel Collier, Son Doan, Ai Kawazoe, Reiko Matsuda Goodwin, Mike Conway, Yoshio Tateno, Quoc-Hung Ngo, Dinh Dien, Asanee Kawtrakul, Koichi Takeuchi, Mika Shigematsu, and Kiyosu Taniguchi. 2008. BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics*, 24(24):2940–2941.

comScore Media Metrix Canada. 2007. Key Measures Report - Health.

K. P. Davison, J. W. Pennebaker, and S. S. Dickerson. 2000. Who talks? The social psychology of illness support groups. *The American Psychologist*, 55(2):205–217.

K.M. Giacomini, R.M. Krauss, D.M. Roden, M. Eichelbaum, M.R. Hayden, and Y. Nakamura. 2007. When good drugs go bad. *Nature*, 446(7139):975–977.

Henk Harkema, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. 2009. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42(5):839851.

Mark Hepple. 2000. Independence and commitment: Assumptions for rapid training and execution of rule-based POS taggers. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 277–278.

Dyfrig A. Hughes, Adrian Bagust, Alan Haycox, and Tom Walley. 2001. The impact of non-compliance on the cost-effectiveness of pharmaceuticals: a review of the literature. *Health Economics*, 10(7):601–615.

International Society Of Drug Bulletins. 2005. Berlin Declaration on Pharmacovigilance.

Michael Kuhn, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. 2010. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6:343–348.

Anne Lee, editor. 2006. *Adverse Drug Reactions*. Pharmaceutical Press, second edition.

- Roberto Leone, Laura Sottosanti, Maria Luisa Iorio, Carmela Santuccio, Anita Conforti, Vilma Sabatini, Ugo Moretti, and Mauro Venegoni. 2008. Drug-Related Deaths: An Analysis of the Italian Spontaneous Reporting Database. *Drug Safety*, 31(8):703–713.
- Charles Medawara, Andrew Herxheimer, Andrew Bell, and Shelley Jofre. 2002. Paroxetine, Panorama and user reporting of ADRs: Consumer intelligence matters in clinical practice and post-marketing drug surveillance. *The International Journal of Risk and Safety in Medicine*, 15(3):161169.
- S. T. Moturu, H. Liu, and W. G. Johnson. 2008. Trust evaluation in health information on the World Wide Web. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1525–1528.
- National Library of Medicine. 2008. UMLS Knowledge Sources.
- John Urquhart. 1999. Pharmacoeconomic consequences of variable patient compliance with prescribed drug regimens. *PharmacoEconomics*, 15(3):217–228.
- Cornelis S. van Der Hooft, Miriam C. J. M. Sturkenboom, Kees van Grootheest, Herre J. Kingma, and Bruno H. Ch. Stricker. 2006. Adverse drug reaction-related hospitalisations: a nationwide study in The Netherlands. *Drug Safety*, 29(2):161–168.
- William E. Winkler. 1999. The state of record linkage and current research problems.
- World Health Organization. 1966. International Drug Monitoring: The Role of the Hospital.