**By: Gupta Amit**

**Voruganti Avinash**

**Sethi Tanvi**

**Abstract**

Extraction of unreported adverse drug reactions using user tweets

**Extraction of unreported ADRs**

**Abstract**-

With increase in usage of social networking there is an exponential rise in online data. Twitter is one of such social networking site that provides a platform for everyone to share emotions, feelings, feedbacks and a lot more. Since the data present on twitter is directly from customers and available on very large scale, this information can be analyzed to understand market trends & emotions in different fields. Health is one such industry which has shown tremendous growth in utilizing digital technology & information. A drug requires to pass multiple clinical tests before it is approved for market circulation. So, it is very important to get proper feedbacks for every drug. A drug may also involve some unexpected reactions along with health benefits. Although all the drugs are released with some prior reaction warnings but there can be some other undiscovered reactions. On similar line of idea, this report contains the work where tweets were analyzed to extract unreported *Adverse Drug Reaction.*

**Introduction-**

An **adverse drug reaction** (**ADR**) is an injury caused by taking a [medication](https://en.wikipedia.org/wiki/Medication). ADRs may occur following a single dose or prolonged administration of a [drug](https://en.wikipedia.org/wiki/Drug) or result from the combination of two or more drugs. The meaning of this expression differs from the meaning of "[side effect](https://en.wikipedia.org/wiki/Side_effect)", as this last expression might also imply that the effects can be beneficial. During multiple clinical trials of a drug all the reactions may not be recorded because of limited number of trials or limited number of health settings**.** DirectFeedbacks based on tweets are used to evaluate potential of a drug further. A detailed analysis on the tweets may reveal important unreported outcomes of drug which can be utilized by pharmaceuticals to improve the drug. This report focuses on extracting those unreported ADRs of a drug which are not reported/mentioned by concerned pharmaceuticals. This report focuses on the working of the algorithm rather than outcome. This report offers easy and generic algorithm to extract ADRs of a drug.

**Architecture-**

The overall process of the ADR extraction involved multiple steps mentioned below-

1. Data(Tweets) Extraction for specific Drugs
2. Data Cleaning/Preprocessing
3. Tweets sentiment analysis
4. ADRs extraction for each drug
5. Classified not mentioned ADRs
6. **Data(Tweets) Extraction for specific Drugs**

Firstly, top 15 drugs were selected based on the maximum sales specifically for Cancer, High BP and Diabetes. Drugs were decided based on the [lowestmed](#Reference_1) & [top-50-prescription-drugs-filled websites](#Reference_2).

The drugs are mentioned in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Disease** | **Drug** | | | | |
| 1 | Cancer | Levothyroxine | Revlimid | Rituxan | Herceptin | Avastin |
| 2 | High BP | Losartan | Lisinopril | Amlodipine | Lasix | Carvedilol |
| 3 | Diabetes | Humalog | Levemir | Lantus | Canagliflozin | Atorvastatin |

**Drugs selected for ADRs extraction**

To extract tweets a twitter bot was created. A Twitterbot is a program that integrates with the Twitter platform, automatically posting, retweeting, liking, or following other users. To create twitterbot [digitalocean](#Reference_3) page was followed. To create a twitterbot one should have a twitter account linked to a mobile number along with following four credential keys. consumer\_key, consumer\_secret, access\_token, access\_token\_secret. After creating the bot, data was extracted using R packages; **rtweet** & **twitteR** based on the credentials created and the drug names as the search word**.** Thisanalysis is done on limited number of tweets since twitter doesn’t allow to extract tweets older than a week. Therefore, tweets were extracted multiple time covering 20 days of span. Total 44 columns were extracted for each tweet mentioning screen\_name, tweet\_id, Number\_of\_retweets etc. These columns helped to retrace and confirm the tweets on skeptical situations and confirming authenticity of tweets. The respective search term and the disease it cured were appended in two separate columns. The data was saved in a csv format and analyzed using Python. The columns in the extracted tweet sheet were as follows:

status\_id,created\_at,user\_id,screen\_name,text,source,reply\_to\_status\_id,reply\_to\_user\_id,reply\_to\_screen\_name,is\_quote,is\_retweet,favorite\_count,retweet\_count,hashtags,symbols,urls\_url,urls\_t.co,urls\_expanded\_url,media\_url,media\_t.co,media\_expanded\_url,media\_type,ext\_media\_url,ext\_media\_t.co,ext\_media\_expanded\_url,ext\_media\_type,mentions\_user\_id,mentions\_screen\_name,lang,quoted\_status\_id,quoted\_text,retweet\_status\_id,retweet\_text,place\_url,place\_name,place\_full\_name,place\_type,country,country\_code,geo\_coords,coords\_coords,bbox\_coords,searchTerm,disease

Out of the above columns, the following columns were chosen for further steps:

screen\_name, text, source, retweet\_count, hashtags, mentions\_screen\_name, searchTerm, disease

1. **Data Cleaning/Preprocessing**

This is one of the most important step to achieve unbiased and accurate results. Raw Data i.e. extracted tweets were processed properly to report genuine ADRs. The TAGS, URLs, special characters, extra spaces and ‘RT’ were removed from the tweet text. All the tweets were converted to lower case followed by removing the duplicate tweets. The rows corresponding to multiple marketing tweets or paid feedback tweets present in the data were also removed. Steps mentioned above were applied and then finally raw data was transformed for further sentiment analysis.

1. **Tweets sentiment classification**

Next step in the process was to train a classifier based on the sentiments of tweets. Most of the pretrained sentiment classifier fails with classification of health-related text. This is because other pretrained classifier are trained on different reviews which are governed by economical or entertainment aspects of the text which is completely different as compared to health domain. For instance, high sales of a drug convey neutral sentiment for this specific task but when a polarity based classification or other pretrained models were used to classify this statement it was classified as positive sentiment. For this reason, tweets were classified to positive, negative and neutral sentiment manually.

First 500 tweets were classified manually based on the health benefit aspect of the tweet. Based on those 500 tweets naïve Bayes classifier was trained to classify 50 new tweets. After classification of 50 new tweets, those tweets were reviewed and then the process was repeated to train 50 more new tweets. This time Naïve Bayes classifier was trained on 550 tweets and then next 50 new tweets were classified and then reviewed. This process was repeated till 1000 tweets were classified. The reason to classify tweets in smaller chunks was to achieve higher accuracy. After classification of 1000 tweets 100 new tweets were classified based on the Naïve Bayes classifier trained on 1000 classified tweets. Then after careful manual review 200 new tweets were classified. This process was repeated to achieve 1500 accurately classified tweets.

Naïve Bayes classifier was used based on the reference of the paper “[Efficient Adverse Drug Event Extraction Using Twitter Sentiment Analysis](#Reference_5)”*.* As per the analysis Naïve Bayes classifier performance better than SVM or another classifier. Following sentiment classification were achieved for all the drugs-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Drug** | **Negative** | **Neutral** | **Positive** | **Grand Total** |
| **Amlodipine** | 5 | 69 | 2 | 76 |
| **Atorvastatin** | 12 | 41 | 14 | 67 |
| **Avastin** | 1 | 100 | 2 | 103 |
| **Canagliflozin** | 7 | 25 | 22 | 54 |
| **Carvedilol** | 3 | 45 | 23 | 71 |
| **Herceptin** | 4 | 154 | 23 | 181 |
| **Humalog** | 16 | 61 | 28 | 105 |
| **Lantus** | 13 | 88 | 25 | 126 |
| **Lasix** | 29 | 170 | 71 | 270 |
| **Levemir** | 4 | 13 | 9 | 26 |
| **Levothyroxine** | 27 | 96 | 26 | 149 |
| **Lisinopril** | 4 | 58 | 12 | 74 |
| **Losartan** | 6 | 40 | 4 | 50 |
| **revlimid** | 3 | 69 | 7 | 79 |
| **Rituxan** | 12 | 107 | 16 | 135 |
| **Grand Total** | 146 | 1136 | 284 | 1566 |

**Drug wise Sentiment Classification Statistics of Tweets**

It was observed that the classifier classified the neutral tweets with 100% accuracy. However, the negative sentiments were not correctly classified attributing to the fact that the training data had a very less percentage of the negative tweets. Hence manual intervention was required to reclassify the incorrectly classified tweets with negative sentiments.

1. **Treatment of Research Tweets**

During the manual sentiment classification of the tweets, it was observed that there were many negative tweets which were related to clinical trials or research study. Since the project is about the adverse drug reactions reported by a drug user, the tweets with clinical trials or a study were not considered for further ADR extraction. Although study related tweets were included to check other possible unreported reactions but no significant difference was found since only 9 tweets had negative sentiments.

1. **Negative Bag of words-**

After classification of all the tweets, negative bag of words was prepared to cover all the possible negative/reaction words. A list of adverse reactions was prepared. Firstly, words were extracted from “*Bidirectional LSTM for Labeling Adverse Drug Reactions in Twitter Posts”* published on [github](#Reference_4). After all the possible ADRs were extracted, all possible variations of the ADRs were required. For example, for word ache, aching, aches, ached etc. were required to extract all possible reactions from the classified tweets. It was difficult to create all possible variations of reaction words using automated process/coding. Therefore, instead of creating reactions to all possible variations, possible variations in tweets were transformed to roots words using lemmatization & stemming. Although all the possible variations for those words were also created manually to capture all the possibilities.

1. **ADRs extraction from tweets**

Based on the bag of words, ADRs were extracted from negative sentiment tweets. All the tweets were already preprocessed, all the possible ADRs or negative words were extracted from negative sentiment tweets. Two approaches were followed to do so.

In the first approach lemmatization was applied on the tweets and bag of ADRs. Lemmatization is very similar to stemming. The major difference between these is, stemming can often create non-existent words, whereas lemmas are actual words. Lemmatization (or lemmatisation), is the process of grouping together the different inflected forms of a word so they can be analyzed as a single item. So, lemmatization was applied on both tweets to get all the possible reactions from the tweets. Lemmatization was also applied on bag of ADRs since reactions from tweets were transformed to lemma, it was necessary to apply lemmatization on bag of ADRs to extract transformed reactions.

In the second approach Porter stemmer was also applied on the negative sentiment tweets. Along with lemmatization. The idea of stemming is a sort of normalization. Many variations of words carry the same meaning, other than when tense is involved. The reason why we stem is to shorten the lookup, and normalize sentences. Consider:

I was taking a ride with my friend.  
I was riding with my friend.

This sentence means the same thing. The phrase ‘with my friend’ is the same. ‘I was’ is the same. ‘the ing’ denotes a clear past-tense in both cases, so is it truly necessary to differentiate between ride and riding, in the case of just trying to figure out the meaning of what this past-tense activity was? Having individual dictionary entries per version would be highly redundant and inefficient, especially since, once we convert to numbers, the "value" is going to be identical. This was the reason porter stemmer was used. Again, both lemmatization and porter stemmer were applied to bag of ADRs for similar reasons mentioned above.

Better results were achieved when both lemmatization and porter stemmer were applied. More number of ADRs were extracted.

Since classifier was trained manually on the complete sentence aspect and not just on the meaning of the word, negation aspect of the tweets was handled. To cover phrases on the tweets, part of sentence around the ADRs were captured. Three words on each side of the ADRs were extracted to capture the sense of possible phrases present. At this level manual intervention was required to go through phrases and finalize on authenticity of possible ADRs. In the final file, all the phrases containing the ADR extracted is mentioned against each tweet.

1. **Extracting unreported ADRs for each drug**

After extracting all the ADRs from tweets along with the phrases, those ADRs were identified which were not mentioned under the drug warnings by the pharmaceutical companies. To identify unreported ADRs, reported reactions were compared against the extracted ADRs. It was easier to compare single word, mentioned ADRs against single word extracted ADRs. But it was challenging to compare extracted phrases against multi-word warnings. Therefore, to extract unreported ADRs without risking any potential loss of unreported reactions, only those phrases were removed which matched exactly with mentioned warnings. Rest of the ADRs were identified even if one of the word from phrase was not matched. All those phrases were reported and the tweets related to those ADRs were traced back and reported for un-reported ADRs.

1. **Input Output Files**

|  |  |  |
| --- | --- | --- |
| **S.No** | **Files Info** | **File Name** |
|  |  |  |
| 1 | Raw Data | twitter\_data\_All\_Final |
| 2 | R Code (Tweet Extraction) | Tweet\_extraction |
| 3 | Python Code(ADR Analysis) | ADR\_Twitter.ipynb |
| 4 | Bag of ADRs | General ADR |
| 5 | Reported ADRs | FDA\_Drugs\_effects |
| 6 | Sentiment Classified Data | DatasetsWithSentiments |
| 7 | ADRs Extraction | IDS 594\_ADR\_extraction |
| 8 | Test data for final sentiment Classification | sentiment\_test50 |
| 9 | Final Output | TweetsWithUnidentifiedSideEffectsStem |
| 10 | Final ADR Extraction Report | IDS 594\_ADR\_Report |

**Key Findings**

* Out of 3941 original tweets, sentiment classification was done on 1457 unique non-marketing tweets.
* 146 negative sentiment tweets were classified.
* On treating the preprocessed text with the two approaches, it was observed that lemmatizing and stemming the preprocess text gave better extraction of ADR as compared to just lemmatization.
* 90 tweets were captured by the lemmatization technique as compared to 97 tweets from lemmatization and stemming techniques.
* Out of the 15 drugs, 9 drugs had unspecified adverse drug effects based on the 2 weeks data collected.
* With the current technique, human intervention is required for analyzing whether the ADR extracted is unidentified or not. It may be a variation of a word already warned by a pharmaceutical company.
* The adverse drug effects on the present data for the drugs are as follows:

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Drugs** | **Unspecified Reactions** |
| 1 | Levothyroxine | Bad heart palpitations, growth hormone deficiency, increase the risk of death if taking for treating borderline thyroid, addictive, mood changing |
| 2 | Rituxan | infusion reaction (throat closing), occasional bumps, low BP |
| 3 | Lasix | diuretic, dehydration electrolyte imbalance |
| 4 | Humalog | acts slow |
| 5 | Levemir | Blood sugar goes up |
| 6 | Lantus | burns bad |
| 7 | Canagliflozin | increase ketoacidosis risks, foot, toe and lower limb amputations |
| 8 | Atorvastatin | risk of ALS, short term memory, slurring |
| 9 | Amlodipine | addiction, swelling in legs, risk of oedema |

**Future Scope-**

The current process involves manual intervention because of restrictive data and references. The next step is to minimize the manual intervention and work on larger data sets and automate the complete process. This will help to report unspecified ADRs more efficiently and help the health industry.

**References-**

1. <https://www.lowestmed.com/top-50-prescription-drugs-filled/>
2. http://clincalc.com/DrugStats/
3. https://www.digitalocean.com/community/tutorials/how-to-create-a-twitterbot-with-python-3-and-the-tweepy-library

1. <https://github.com/chop-dbhi/twitter-adr-blstm>

1. <https://ieeexplore.ieee.org/document/7752365/>