

Pharmacovigilance using social media..

Saeed Ahmad
Department of Computer Science
FAST-NUCES Islamabad, Pakistan.
saeed.ahmad.sd7@gmail.com

ABSTRACT

Twitter is a great platform contain a wealth of information regarding user preferences and experiences over approximately every product and domains. We can use this raw information to obtain valuable insights using data mining approaches such as sentiment analysis and context analysis. So in this project we process twitter reviews in context of pharmacovigilance field to find Adverse drug reaction (ADR). These reviews contains information related to multiple aspects such as effectiveness, dosage, usage duration, conditions, results and reactions of drugs. This automatic analysis seems very interesting but also challenging. However, analyzing various aspects of drug reviews and sentiments definitely gave us valuable insights,

Helps for decision making and real time and cost effective monitoring public health by revealing collective experience.

In this project we will do multiple tasks over drug reviews with data collect by twitter api for some drugs(will use the bio-labs data). We first perform sentiment analysis to predict the sentiment concerning overall satisfaction, side effects observed by the user and effectiveness on specific drug. In this project we will analyse our results with knows drugs side effects to that we could says that our results are trusted if our system generated side effects have higher frequency with known side effects of that drug.

KEYWORDS

Text classification; Sentiment Analysis; Adverse drug reaction (ADR); Pharmacovigilance; Twitter data processing;

1. INTRODUCTION

Pharmaceutical product safety presently depends on clinical trials and these are short studies conducted in a very few hundred patients in standardised conditions and short time span before promoting a drug. Therefore, additional investigation on ADRs should be pursued within the post-marketing part.

As a consequence, the discrepancies in patient choice and treatment conditions will have important impact on the effectiveness and potential risks of adverse drug reactions (ADRs).

Therefore, post-marketing drug police investigation, i.e. pharmacovigilance plays a significant role regarding drug safety once a drug has been free. moreover, treatment selections are expected to play an progressively vital role in health care.

Approaches like medical care recommender systems, that aim at serving to to seek out optimum customised medical care choice for a given patient and time, benefit of feedback on

medical care. These systems, however, generally accept structured knowledge, i.e. knowledge classified into variety of categories on predefined scales. the quantity of such knowledge typically is proscribed as a result of it needs intense preparation that isn't customary in clinical routine. Here, alternative sources, like user reviews, supply nice potential.

However, one major demand for automatic process and analysis of tweets, knowledge contained in massive amounts of unstructured information. One typical approach of doing this is sentiment analysis, its extensively studied domain in process free-text in net media analyses. several approaches to sentiment analysis. These approaches acknowledge sentiment terms and patterns of sentiment expressions in language texts by matching matter units with opinion words in google word to vec corpus annotated for sentiment polarity. However, studies showed that sentiment analysis is usually domain-dependent since the polarity of single terms will disagree looking on the context they're utilised in. moreover, the language in on-line forums is extremely informal and user-expressed medical ideas are typically nontechnical, descriptive, and difficult to extract. that is why typical nlp processes use for drug review analyses. another approach treats the task as classification downside. approaches have the extra advantage to be capable of playacting medical sentiment analysis over multiple aspects, i.e. sentiments may be learned on specific aspects like effectiveness and facet effects.

Sentiment analysis of patient knowledge generally and on drug expertise above all may be a difficult analysis downside that's presently receiving sizeable attention. one in all the most problems, however, is that the lack of annotated knowledge, that is crucial for correct sentiment classification. Especially, tagged knowledge handling distinct aspects is rare. Moreover, the provision of tagged knowledge is extremely domain dependent.

2. RELATED WORK

Literature on drug reviews and pharmacovigilance will primarily be divided into studies on identification of aspects like automatic detection of ADRs or aspect effects and such works coping with overall or aspect-based sentiment analysis.

Most approaches attempt ADR or aspect result identification area unit lexicon-based and have faith in mapping relevant terms and phrases from user information to specific vocabulary from numerous individual or combined lexicons. However, lexicon-based approaches suffer from phonetic and trade misspellings. Therefore, recent works have additionally targeted on machine learning techniques to over- come back such limitations. Nikfarjam et al. applied association rule mining to search out pattern, i.e. combos in terms or conditional random fields (CRFs), to extract mentions of ADRs. supported the underlying assumption that patients' posts regarding ADRs generally specific negative sentiments, Korkontzelos et al. studied the result of enriching a lexicon-based ADR identification technique with sentiment analysis options. Cavalcanti et al. demonstrate the extraction and classification of multiple aspects in drug reviews, e.g. adverse reactions, effectuality of a drug, symptoms and conditions, employing a technique supported grammar dependency ways. an in depth review on pharmacovigilance and ADR extraction techniques will be found in. Works on drug review sentiment analysis will primarily be divided into approaches applying lexicons with sentiment scores or such approaches learning sentiments using supervised classification. In one in all the earliest works on drug review sentiment analysis. developed a subject classifier from patient information to eventually apply many polarity classifiers, one per topic demonstrate a clause-level sentiment analysis algorithmic program considering multiple review aspects as overall satisfaction, effectiveness, aspect effects and condition. Here, a rule-based approach is utilized that takes grammatical relations and

linguistics annotation into account and computes sentiment orientation of individual clauses supported a lexicon aspect based sentiment analysis

3. DATASET

We used data from twitter scripted data provided by bio lab. This dataset has 1.5 lakh tweet records of

REFERENCE

[1] Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. **Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning**. DH '18 Proceedings of the 2018 International Conference on Digital Health. 121-125
<http://kdd.cs.ksu.edu/Publications/Student/kallumadi2018aspect.pdf>

[2] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared. 2017. Approaches to CrossDomain Sentiment Analysis: A Systematic Literature Review. IEEE Access 5 (2017), 16173–16192.
<https://doi.org/10.1109/ACCESS.2017.2690342>

english and data is misspelled and not 100% accurate. We have only one col of tweets text and we don't and it is not annotated.

We have cyder known drug reviews so, we can compare our extracted drug ADRs from the data and the frequency and compare with known ADRs.

[3] Diana Cavalcanti and Ricardo Prudêncio. 2017. Aspect-Based Opinion Mining in Drug Reviews. In Progress in Artificial Intelligence, Eugénio Oliveira, João Gama, Zita Vale, and Henrique Lopes Cardoso (Eds.). Springer International Publishing, Cham, 815–827.

[4] World Health Organization(WHO)
<https://www.who.int/>

[5] Adverse drug reaction (ADR)
https://en.wikipedia.org/wiki/Adverse_drug_reaction

[6] <https://code.google.com/archive/p/word2vec>