Deep Neural Networks Ensemble for Detecting Tweets Mentioning Medications

Davy Weissenbacher, Ph.D.1, Abeed Sarker, Ph.D.2, Ari Klein, Ph.D.1, Karen O’Connor, MS1, Arjun Magge, Ph.D.3, Graciela Gonzalez-Hernandez, Ph.D.1

1 University of Pennsylvania, Philadelphia, PA, USA; 2 Emory University, Atlanta, GA, USA; 3 Arizona State University, Tempe, AZ, USA

Introduction

Twitter posts are now recognized as an important source of patient-generated data, providing unique insights into population health. A fundamental step towards incorporating Twitter data in pharmacoepidemiological research is to automatically recognize medication mentions in tweets. A common approach is to search for tweets containing lexical matches of drug names occurring in a manually compiled dictionary. However, this approach has several limitations. Many tweets contain drugs that are misspelled and, even when a match is found, oftentimes, the referent is not actually a drug; for example, tweets that mention *Lyrica* are predominantly about the singer, Lyrica Anderson, and not about the antiepileptic drug. In this study, when using the lexical match approach on a corpus where names of drugs are rare, we retrieved only 71% of the tweets that we manually identified as mentioning a drug, and more than 45% of the tweets retrieved were false positive. Enhancing the utility of social media for public health research requires methods that are capable of improving the detection of posts that mention drugs.

Methods

We present *Kusuri* (薬, “medication” in Japanese), an Ensemble Learning classifier able to identify tweets mentioning drug products and dietary supplements. *Kusuri* is composed of two modules: Module 1, four different filters (lexicon-based, spelling-variant-based, pattern-based, and a weakly-trained neural network) are applied in parallel to discover tweets potentially containing medication names; Module 2, an ensemble of deep neural networks (DNNs) encoding morphological, semantic, and long-range dependencies of important words in the tweets makes the final decision. The architecture of *Kusuri* is presented in Figure 1.
Each DNN of the ensemble in Module 2 starts by independently encoding each sequence of characters composing the tokens of a tweet through 3 layers sequentially connected: a recurrent layer (GRU), an attention layer, and a densely connected layer. All resulting vectors are then concatenated with their respective pretrained word embedding vectors. The concatenated vectors are passed to a bidirectional-gated recurrent unit (GRU) layer, followed by an attention layer. A final dense layer computes the probability for the tweet to contain a mention of a drug.

To evaluate Kusuri, we collected all publicly available tweets posted by 112,500 Twitter users (their timelines). To decide which users to include, we used a set of manually defined keywords and a simple classifier to detect tweets announcing a pregnancy. Once a tweet announcing a pregnancy was identified, we collected the timeline of the author of the tweet. Following this process, we collected a total of 421,5 million tweets.

Building a corpus of tweets containing drug names to train and evaluate the Module 2 of Kusuri was a challenging task. Tweets mentioning drug names are extremely rare. We found that they only represent 0.26% of the tweets in the UPennHLP Twitter Pregnancy Corpus (see below), and are often ambiguous with common and proper nouns. Therefore, to build a gold-standard corpus we had to rely on a more sophisticated method than simply lexicon matching. We obtained positive examples by selecting tweets retrieved by at least two filters of Module 1, as they were most likely to mention drug names. We obtained negative examples by selecting tweets detected by only one filter of Module 1, given that if these tweets did not contain a drug name, they were non-obvious negative examples. Following this process, we created and manually annotated the UPenn HLP Twitter Drug Corpus, a corpus of 15,005 tweets containing ~50% of positive and ~50% of negative examples. Two annotators annotated the corpus in its entirety, with a high interannotator agreement (IAA) measured as Cohen's kappa of .892. In November 2018, we organized the 3rd Social Media Mining for Health Applications shared task (SMM4H) with Task 1 of our challenge dedicated to the problem of the automatic recognition of drug names in Twitter. Eleven teams tested multiple approaches on the UPenn HLP Twitter Drug Corpus.

Due to the mechanism of its construction, the UPennHLP Twitter Drug Corpus does not represent the natural distribution of tweets mentioning drugs on Twitter. Consequently, an evaluation made on the corpus will not be indicative of the performance to expect of the classifier in the presence of a large proportion of tweets not containing a drug mention, as is expected in studies based on tweets collected over time. In order to further assess whether Kusuri could reliably be used in such a study, we manually identified all drugs mentioned in the timelines of 113 users (98,959 tweets) from our collection and evaluate Kusuri on this corpus, the UPennHLP Twitter Pregnancy Corpus. An IAA of 0.88 Cohen’s kappa was computed over 12 dual annotated timelines of the UPennHLP Twitter Pregnancy Corpus.

Results

On the UPennHLP Twitter Drug Corpus, Module 2 of Kusuri demonstrated performances close to human annotators with 93.7% F1-score (95.1% Precision, 92.5% Recall) and outperformed the best system of the Task 1 of the SMM4H 2018 competition. We analyzed randomly selected labeling errors made by Module 2 and found that most errors were made on tweets discussing medical topic without actually mentioning a drug or were caused by the ambiguity, not only of common English words (e.g. airborne), but also of dietary supplements and food products sometimes consumed for their medicinal properties (e.g. clove, arnica or aloe).

On the UPennHLP Twitter Pregnancy Corpus, Kusuri (i.e. both modules applied sequentially) obtained 78.8% F1-score (94.6% Precision, 67.5% Recall), a score comparable to the score obtained on the most frequent types of NEs by the best systems competing in well-established challenges, despite our corpus having only 0.26% positive instances in it.

Discussion

Kusuri identifies tweets mentioning drug names with performance high enough to ensure its usefulness, and is ready to be integrated in pharmacovigilance, toxicovigilance, or more generally, public health pipelines that depend on medication name mentions. Kusuri is freely available at https://healthlanguageprocessing.org/kusuri/

References