Predicting Effectiveness of Drug from Patient's Review

Abhishek Das1, Koushik Deb2, Shuvendu Das2

Abstract

In this paper, we have made multiple reviews and finally made a comparative study of machine learning and deep learning to predict the effectiveness of drugs. We have used both approaches and found that deep learning approaches are providing better results than machine learning approaches. We have made two types of classifications (good and bad) and got an efficiency rate with the help of a confusion matrix of more than 80% in the case of deep learning and less than 80% in the case of machine learning approaches.

Keywords

machine learning, deep learning, effectiveness, comparative study, drug utilisation

1. Introduction

The objective behind the classification of drugs is to ensure that people can safely intake them with utmost benefit.

Although the effects of the drugs are meant for treatment therapy, it can also cause side effects that may be harmful. Moreover, having multiple drugs may sometimes cause more side effects than benefits[6]. By noting the classification of a drug, a person and his doctor can understand what to expect when the person intakes a particular drug, what are the probable risks, and to which drug the person can switch to if required[2].

2. Background and Related Work

The increase in prescription rate during some critical human conditions, that is, say during the gestation period of females, is caused by an increase in prescription rate of medicinal drugs for pregnancy-related symptoms[1]. However, drugs cannot be

always avoided during the gestation period. For women with severe physical illness and medical conditions like asthma, diabetes, etc., the different drugs prescribed by the doctor are unpreventable[5]. The indirect side effects of many drugs sometimes cause more damage than benefits. Intaking severe high dosage of drugs often results in fatal problems in case of child birth[4]. So, we have made a comparative study on machine learning and deep learning on the effectiveness of different medicinal drugs that will help the doctors to prescribe medical therapy for their patients with ease.

3. Experiment

We have taken this data from Team NDL from Penn State's Nittany Data Labs. In the first phase, bad symbols were replaced with null. After removal of the bad symbols, feature extraction was done. The feature uses textual data for predictive modeling. The text must also be parsed to remove certain words which means we have extracted the words such that only 5000 words were taken as featured. After that, we concatenated the review.

Next, a vectorizer was created. Vectorizer is used to transform a collection of text documents to a vector of term/token counts. It also authorizes the pre-processing of text data prior to generating the vector representation[3].Below are the approaches we have made.

3.1 Classification with Keras

Keras is a Python library for deep learning and machine learning that wraps the efficient numerical libraries TensorFlow and Theano. Keras allows you to quickly and simply design and train neural network machine learning and deep learning models.

Long Short Term Memory (or LSTM) neural network classification is used in this case. Other specifications, activation function, the embedding dimension is kept the same for all the code. The embedding dimension is taken as 100, and the maximum sequence as 256. In this method, we are taking the tokenizer and we are creating a token from the full data set. The maximum number of words that are used is 50,000. Now, from the created token, we found out the index of the data set

used in this code. Finally, we have used the Confusion Matrix to find out the efficiency of the whole code, based on Keras tokenizer.
We obtained an f1 score as 0.79, precision score as 0.83, and recall score as 0.77.

3.2 Google News Word Embedding

Word embedding is a collective term used for models that learn to map a set of words or phrases in a vocabulary to vectors of numerical values. Neural Networks are designed to learn from numerical data. Here also, the same architecture has been used with a little bit of difference. Here, the google news trained vector API has been downloaded. In this API, word vectors are stored. So here we have tokenized based on this word vector and then we have trained our model.
Using the Confusion matrix, we obtained an efficiency that is somewhat similar to that of the above-mentioned classification with Keras tokenizer where the value of the f1 score is 0.80, 0.82 for precision score, and the recall score as 0.79.

3.3 Continuous Bag of Words

The way Continuous Bag of Words or CBOW work is that it tends to predict the probability of a word given a context. A context may be a single word or a group of words.
In our code, we have used skip-gram as zero, which means continuous bag of words has been used in this training.
We have created our model using the embedding dimension, which has been taken as 300. So, using the Confusion matrix, we found the ultimate value is lesser than the previous approach because Google has more stock of words and the library is trained in a better way than us.
Using the Confusion matrix, we obtained f1 score as 0.79, precision score as 0.78 while recall score as 0.80.

3.4 Skip gram embedding

In terms of the architecture, Skip-gram is a simple neural network model with only one hidden layer. The input to the network is a one-hot encoded vector representation of a target-word — all of its dimensions are set to zero, apart from the dimension corresponding to the target-word.

Here also we have used a model from our data set with 300 dimensions. In this approach, with the help of the skip-gram embedding algorithm, we have made the word embedding and then we have tokenized it and then it was classified. Here we got f1 score as 0.79, precision score as 0.83, and recall score as 0.77 from the confusion matrix which is again almost the same as CBOW (a little better than CBOW).

3.5 Simple Regression

We have used machine learning methods to check if we can find any improvement. Random forest is an ensemble learning where we can create many decision tree and predict based on the highest voting. Here, we have taken 100 estimators or trees and 1000 depths to make predictions. Tf Ifd vectorizer has been used for feature extraction. We have received 0.83 accuracy.f1 score reached 0.75 and precision, recall stopped at 0.89 and 0.72 . It has been noticed that all scores are better than naïve bayes technique.

3.6 Naive Bayes Method

Multinomial Naive Bayes has been used to check review prediction. This should work well in machine learning as word count of text follows multinomial distribution. It is a binomial classification where greater than 5 is good and lesser equals to 5 is bad review. 0.75 accuracy has been found.0.58 f1 score, 0.82 precision and 0.59 recall we received after running a multinomial naïve bayes algorithm.

Conclusion

We have worked with various methods, namely, classification with Keras tokenizer, Google news word embedding, continuous bag of words, skip-gram embedding, etc. and made a detailed comparison on the efficiencies obtained from the confusion matrix.

From experiment, it can be seen that the skip-gram embedding is having the least efficiency while the Google news word embedding is having the highest average efficiency. In the future, perhaps trying to do some feature exploration in a different way would help develop insights and meaningful conclusions. Exploring

different neural network architecture could have been very beneficial, as recurrent nets are known to work very well.

References

[1] MK Bakker, J Jentink, F Vroom. PB Van Den Berg, HEK De Walle, LTW De Jong-Van Den Berg. Drug prescription patterns before, during and after pregnancy for chronic, occasional, and pregnancy-related drugs in the Netherlands.BJOG 2006 May;113(5):559-68. DOI: 10.1111/j.1471-0528.2006.00927.x.

[2] Gauld N, Kelly F, Emmerton L, Bryant L, Buetow S. Innovations from 'down-under': a focus on prescription to non-prescription medicine reclassification in New Zealand and Australia. SelfCare 2012;3(5):88-107

[3] Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In Proceedings of the 2018 International Conference on Digital Health (DH '18). ACM, New York, NY, USA, 121-125.

[4] A. Gudivada and N. Tabrizi, "A Literature Review on Machine Learning-Based Medical Information Retrieval Systems," *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, 2018, pp. 250-257, doi: 10.1109/SSCI.2018.8628846.

[5] C. Jimenez, M. Molina and C. Montenegro, "Deep Learning-Based Models for Drug-Drug Interactions Extraction in the Current Biomedical Literature," *2019 International Conference on Information Systems and Software Technologies (ICI2ST)*, Quito, Ecuador, 2019, pp. 174-181, DOI: 10.1109/ICI2ST.2019.00032.

[6] Sidey-Gibbons, J., Sidey-Gibbons, C. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* **19,** 64 (2019). https://doi.org/10.1186/s12874-019-0681-4