Neural DrugNet

Nishant Nikhil
IIT Kharagpur
Kharagpur India
nishantnikhil@iitkgp.ac.in

Shivansh Mundra
IIT Kharagpur
Kharagpur India
coolshivansh8@iitkgp.ac.in

Abstract

In this paper, we describe the system submitted for the shared task on Social Media Mining for Health Applications by the team Light. Previous works demonstrate that LSTMs have achieved remarkable performance in natural language processing tasks. We deploy an ensemble of two LSTM models. The first one is a pretrained language model appended with a classifier and takes words as input, while the second one is a LSTM model with an attention unit over it which takes character tri-gram as input. We call the ensemble of these two models: Neural-DrugNet. Our system ranks 2nd in the second shared task: Automatic classification of posts describing medication intake.

1 Introduction

In recent years, there has been a rapid growth in the usage of social media. People post their day-to-day happenings on regular basis. Weissenbacher et al. (2018) propose four tasks for detecting drug names, classifying medication intake, classifying adverse drug reaction and detecting vaccination behavior from tweets. We participated in the Task2 and Task4.

The major contribution of the work can be summarized as a neural network based on ensemble of two LSTM models which we call Neural DrugNet. We discuss our model in section 2. Section 3 contains the details about the experiments and pre-processing. In Section 4, we discuss the results and propose future works.

2 Model

Detection of drug-intake depends highly on:

- Whether the sentence conveys an intake.

- Whether a drug is mentioned in the sentence.

Long Short-Term Memory networks (Hochreiter and Schmidhuber, 1997) have been found efficient in tasks which need to learn structure of a sequential data. To learn a model which can value the first condition, we use LSTM based neural networks. Our first model is inspired from (Howard and Ruder, 2018), an LSTM model whose encoder is taken from a language model pre-trained on Wikipedia texts and fine-tuned on the tweets. After which a dense layer is used to classify into the different categories. And to also take into account the mentioning of a drug, which has not been there in the training data, we exploit the word structure of drug-names. Most of the drug-names have the same suffix. Example: melatonin, oxytocin and metformin have the suffix '-in'. We use a LSTM based model trained on the trigrams to learn that. Then, we take an ensemble of these two models. We give equal importance to both the models. That is, the prediction probability from Neural-DrugNet is the mean of the prediction probabilities from the two LSTM models. The predicted class is the one having the maximum prediction probability.

The training for the pre-trained language based LSTM model follows the guidelines given in the original paper (Howard and Ruder, 2018). They use discriminative fine-tuning, slanted triangular learning rates and gradual unfreezing of layers. For the character n-gram based LSTM model, as no fine-tuning is required, we train the model end-to-end.

3 Experiments

The data collection methods used to compile the dataset for the shared tasks are described in Weissenbacher et al. (2018).

3.1 Preprocessing

Before feeding the tweets to Neural-DrugNet, we use the same preprocessing scheme discussed in Nikhil et al. (2018). Then for the character trigram based model, we add an special character '$' as a delimiter for the word. That is, the character trigrams of 'ram' would be: '$ra', 'ram' 'am$'.

3.2 Results

We experimented with different type of architectures for both the tasks. Although any classifier like random forest, decision trees or gradient boosting classifier can be used. But due to lack of time, we used only support vector machine with linear kernel as baseline (denoted as LinearSVC). During development phase, we mistakenly used only accuracy as the metric for task2. The given results are based on a train-validation split of 4:1.

System	Accuracy
LinearSVC	0.675
LSTM with attention (words as input)	0.703
1D-CNN	0.651
Bi-LSTM with attention (words as input)	0.714
Bi-LSTM with attention (3-grams as input)	0.709
LSTM with encoder from Language Model	0.754
Neural DrugNet	**0.771**

Table 1: Results on validation data for Task2

System	F1-score
LinearSVC	0.751
Neural DrugNet	0.805
Neural DrugNet with LM fine-tuned on data from task3 also	**0.812**

Table 2: Results on validation data for Task4

The final results on best performing variant on test data for both the tasks are:

Precision	Recall	F1-score
0.520	0.491	0.505

Table 3: Results on test data for Task2

Accuracy	Precision	Recall	F1-score
0.857	0.824	0.897	0.859

Table 4: Results on test data for Task4

4 Conclusion and Future Work

In this paper, we present Neural-DrugNet for drug intake classification and detecting vaccination behavior. It is an ensemble of two LSTM models. The first one is a pretrained language model appended with a classifier and takes words as input, while the second one is a LSTM model with an attention unit over it which takes character tri-gram as input. It constantly outperforms the vanilla LSTM models and other baselines, which supports our claim that drug-intake classification and vaccination behavior detection rely on both the sentence structure and the character tri-gram based features. The performance reported in this paper could be further boosted by using a language model pretrained on tweets rather than the wikipedia texts. Furthermore, the ensemble module can be learned end-to-end by using a dense layer.

References

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780.

Jeremy Howard and Sebastian Ruder. 2018. Fine-tuned language models for text classification. *CoRR*, abs/1801.06146.

N. Nikhil, R. Pahwa, M. K. Nirala, and R. Khilnani. 2018. LSTMs with Attention for Aggression Detection. *ArXiv e-prints*.

Davy Weissenbacher, Abeed Sarker, Michael Paul, and Graciela Gonzalez-Hernandez. 2018. Overview of the third social media mining for health (smm4h) shared tasks at emnlp 2018. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.