

# Deep Learning for social media health text classification

T.Y.S.S.Santosh Vaibhav Gambhir Animesh Mukherjee

IIT Kharagpur  
West Bengal – 721302  
India

{santoshtyss, v.gambhir}@gmail.com  
{animeshm}@cse.iitkgp.ac.in

## Abstract

This paper describes the systems developed for 1st and 2nd tasks of the 3rd Social Media Mining for Health Applications Shared Task at EMNLP 2018. The first task focuses on automatic detection of posts mentioning a drug name or dietary supplement, a binary classification. The second task is about distinguishing the tweets that present personal medication intake, possible medication intake and non-intake. We performed extensive experiments with various classifiers like Logistic Regression, Random Forest, SVMs, Gradient Boosted Decision Trees (GBDT) and deep learning architectures such as Long Short-Term Memory Networks (LSTM), jointed Convolutional Neural Networks (CNN) and LSTM architecture, and attention based LSTM architecture both at word and character level. We have also explored using various pre-trained embeddings like Global Vectors for Word Representation (GloVe), Word2Vec and task-specific embeddings learned using CNN-LSTM and LSTMs.

## 1 Introduction

The tasks (Davy Weissenbacher, 2018) involve NLP challenges on social media mining for health monitoring and surveillance and in particular pharmaco-vigilance. This requires processing noisy, real-world, and substantially creative language expressions from social media. The proposed systems should be able to deal with many linguistic variations and semantic complexities in various ways people express medication-related concepts and outcomes. The tasks present several interesting challenges including the noisy nature of the data, the informal language of the user posts, misspellings, and data imbalance.

Deep learning has the potential to improve analysis of social media text because of its ability to learn patterns from unlabelled data (Arel et al., 2010). This property has enabled deep learn-

ing to produce breakthroughs in the domain of image, text and speech recognition. Moreover, deep learning has the ability to generalize learnt patterns beyond data similar to the training data, which can be advantageous while dealing with social media text. Despite the breakthroughs brought by deep learning, improvements are still to be made to further optimise it and improve its performance (LeCun et al., 2015). This paper proposes to explore how the emerging advantages of deep learning can be expanded upon to address the pertinent challenges for social media text analysis.

For Task 1, tweets are required to be distinguished those that mention any drug names or dietary supplement. For Task 2, the data-set contains tweets mentioning a drug and the objective is to classify the tweet into three classes. The class descriptions are as follows: personal medication intake tweets in which the user clearly expresses a personal medication intake/consumption; possible medication intake tweets that are ambiguous but suggest that the user may have taken the medication; non-intake tweets that mention medication names but do not indicate personal intake.

## 2 Method

This section describes the deep learning architectures we used for the tasks, described as follows: 1) CNN-LSTM 2) LSTM with attention mechanism. The subsections give a brief description of both of the approaches.

### 2.1 CNN-LSTM

With the development of deep learning, typical deep learning models such as CNNs and recurrent neural networks (RNNs) have achieved remarkable results in computer vision and speech recognition. Word embeddings, CNNs (Kim, 2014) and RNNs (Graves, 2012) have been applied to text classification and got good results. CNN and RNN are two mainstream architectures for such model-

ing tasks, which adopt totally different ways of understanding natural languages. In this system, we combine the strengths of both architectures and use a novel and unified model called CNN-LSTM (Zhou et al., 2015) for sentence classification. CNN-LSTM utilizes CNN to extract a sequence of higher-level phrase representations, and are fed into an LSTM to obtain the sentence representation. We take the word embeddings as the input of our CNN model in which windows of different length and various weight matrices are applied to generate a number of feature maps. After convolution and pooling operations, the encoded feature maps are taken as the input to the LSTM model. The long-term dependencies learned by LSTM can be viewed as the sentence-level representation. The sentence-level representation is fed to the fully connected network and the softmax output reveals the classification result. The deep learning algorithm we put forward to use for these tasks differs from the existing methods in that our model takes advantage of the encoded local features extracted from the CNN model and the long-term dependencies captured by the LSTM model.

## 2.2 LSTM with attention mechanism

A limitation of the usual LSTM architecture is that it encodes the input sequence to a fixed length internal representation. This imposes limits on the length of input sequences that can be reasonably learned. A recently proposed method for easier modeling of long-term dependencies is attention. Attention mechanisms allow for a more direct dependence between the state of the model at different points in time. Attention-based RNNs have proven effective in a variety of sequence transduction tasks, including machine translation (Bahdanau et al., 2014), image captioning (Xu et al., 2015), and speech recognition (Chan et al., 2016). This is achieved by keeping the intermediate outputs from the LSTM from each step of the input sequence and training the model to learn to pay selective attention to these inputs and relate them to items in the output sequence.

## 3 Experiment

This section details how the proposed approach is applied to Task 1 and Task 2 data sets. Task 1 is a binary classification problem and task 2 is a multi-class classification problem. The dataset statistics are given in Table 1 and Table 2. The dataset for

each task includes training data and test data.

As baselines, we experimented with several classifiers like Logistic Regression, Random Forest, SVMs, Gradient Boosted Decision Trees (GBDT). We have used TF-IDF to extract the feature values. We then used the CNN-LSTM and attention based LSTM networks and are trained (fine-tuned) using labeled data with back-propagation. We have also experimented with CNN-LSTM and attention based LSTM networks by using pre-trained embeddings such as GloVe and Word2vec for word level and we have also experimented them at character level. These networks also learn task-specific word embeddings. Therefore, for each of the networks, we also experimented by using these embeddings as features and trained various classifiers like Logistic Regression, Random Forest, SVMs, GBDT.

## 4 Results

We have submitted the top 3 systems for each task on validation data. Table 3 and 4 describes the precision, recall and F1-score on the validation data and test data for Task 1 respectively. We have selected top 3 based on cumulative score of recall, precision and F1-score. On test data character level LSTM-CNN gave the good precision and F1-score whereas word level LSTM with attention embeddings trained on Naive bayes classifier gave the good recall. Table 5 and 6 describes the precision, recall and F1-score on the validation data and test data for the Task 2. On test data character level LSTM-CNN gave highest micro-averaged precision, recall and F1-score.

## 5 Conclusion

In this paper we described briefly our two systems CNN-LSTM and LSTM with attention. We have experimented both at character level and at word level. We have also explored using different pre-trained embeddings like Word2Vec, GloVe and also with embeddings learned from deep neural network models combined with several classifiers.

## References

Itamar Arel, Derek C Rose, Thomas P Karnowski, et al. 2010. Deep machine learning-a new frontier in artificial intelligence research. *IEEE computational intelligence magazine*, 5(4):13–18.

<b>Data</b>	<b>Presence of drug</b>	<b>Absence of drug</b>
Train	3834	3572
Validation	959	893

Table 1: Task 1 Data Statistics

<b>Data</b>	<b>personal medication intake</b>	<b>possible medication intake</b>	<b>non-intake</b>
Train	2460	3932	5426
Validation	615	984	1357

Table 2: Task 2 Data Statistics

<b>Method</b>	<b>Precision</b>	<b>Recall</b>	<b>F Score</b>
LSTM-CNN with GloVe (word level)	0.8537	0.8537	0.8537
LSTMattention with GloVe and Naive Bayes classifier (word level)	0.8718	0.8718	0.8718
LSTM-CNN (character level)	0.8864	0.8864	0.8864

Table 3: Validation Data Results for Task 1

<b>Method</b>	<b>Precision</b>	<b>Recall</b>	<b>F Score</b>
LSTM-CNN with GloVe (word level)	0.8963	0.82433	0.85881
LSTMattention with GloVe and Naive bayes classifier (word level)	0.86264	0.87202	0.86731
LSTM-CNN (character level)	0.91833	0.83976	0.87229

Table 4: Test Data Results for Task 1

<b>Method</b>	<b>Micro-averaged Precision</b>	<b>Micro-averaged Recall</b>	<b>Micro-averaged F Score</b>
LSTM-CNN with GloVe and GBDT (word level)	0.683	0.683	0.683
LSTM attention with Word2Vec (word level)	0.706	0.694	0.694
LSTM-CNN (character level)	0.715	0.715	0.715

Table 5: Validation Data Results for Task 2

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

William Chan, Navdeep Jaitly, Quoc Le, and Oriol

Vinyals. 2016. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 4960–4964. IEEE.

Method	Micro-averaged Precision	Micro-averaged Recall	Micro-averaged F Score
LSTM-CNN with GloVe and GBDT (word level)	0.350	0.365	0.358
LSTM attention with Word2Vec (word level)	0.409	0.363	0.385
LSTM-CNN (character level)	0.408	0.407	0.408

Table 6: Test Data Results for Task 2

Michael Paul Graciela Gonzalez-Hernandez, Davy Weissenbacher, Abeer Sarker. 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)*, 2018.

Alex Graves. 2012. Supervised sequence labelling. In *Supervised sequence labelling with recurrent neural networks*, pages 5–13. Springer.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521(7553):436.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057.

Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A c-lstm neural network for text classification. *arXiv preprint arXiv:1511.08630*.