A Multi-channel Convolutional Neural Network for Multilabel Sentiment Classification Using Abilify Oral User Reviews

Tina Esther Trueman[1], Ashok Kumar Jayaraman[*,2], Jasmine S[2], Gayathri Ananthakrishnan[3] and Narayanasamy P[4]
[1]Department of Computer Science, University of the People, Pasadena, United States
[2]Department of Information Science and Technology, Anna University, Chennai, India
[3]Department of Information Technology, Vellore Institute of Technology, Vellore, India
[4]Department of Electrical and Electronics Engineering, PSG College of Technology, Coimbatore, India
E-mail: tina.trueman@uopeople.edu, jashokkumar83@auist.net, jasminemtech7@gmail.com, gayathri.a@vit.ac.in, drpnsam@gmail.com
[*]Corresponding author

Nowadays, patients and caregivers have become very active in social media. They are sharing a lot of information about their medication and drugs in terms of posts or comments. Therefore, sentiment analysis plays an active role to compute those posts or comments. However, each post is associated with multilabel such as ease of use, effectiveness, and satisfaction. To solve this kind of problem, we propose a multichannel convolution neural network for multilabel sentiment classification using Abilify oral user comments. The multichannel represents the multiple versions of the standard model with different strides. Specifically, we use the pre-trained model to generate word vectors. The proposed model is evaluated with multilabel metrics. The results indicate that the proposed multichannel convolutional network model outperforms the traditional machine learning algorithms.

Povzetek: Razvita je konvolucijska mreža za preučevanje izmenjav mnenj o boleznih na socialnih omrežjih.

1 Introduction

Social media has become an active part of drugs and medication users. They share the advantage or disadvantages of their medication and drugs. This information may give some insightful information about the reaction of the drug. Therefore, sentiment analysis plays a wide role to compute the opinions of drug users and caregivers. The sentiment analysis can be performed at the document level, sentence level, or aspect level [1, 2]. The document and sentence level computes the overall opinion. But, the aspect level computes opinion at a specific target or an entity. In this paper, we aim to focus on aspect level sentiment. A comment may be associated with a single label or multilabel [3]. The single label problem has only one label. However, It has two classification methods namely, binary classification or multiclass classification [4]. The binary classification problem belongs to a binary set such as true and false or positive and negative. The multiclass classification problem belongs to a set of more than two elements such as positive, neutral, and negative. In these problems, algorithms assign only one label to comment or instance. Multilabel classification problem belongs to a set of multiple target labels where each label maybe belongs to a binary class or multiclass.

Traditionally, the multilabel classification problems are solved using problem transformation, adapted algorithms, and ensemble learnings [3]. The problem transformation problem is further solved using the binary relevance, classifier chain, and label powerset methods [5, 6]. However, these methods use the traditional bag of words (BoW) method to represent features. These features fail to represent semantic meaning between words. Therefore, deep learning models are proposed to capture the semantic meaning between words in the input sequence. It is also proven that they outperform in many tasks such as image classification, text classification, etc [7, 8, 9]. In this paper, we propose a multichannel convolution neural network for multilabel sentiment classification using Abilify oral user comments. The multichannel model represents the multiple versions of the standard model with different strides. Particularly, we use the GloVe pre-trained model [10] to generate word vectors. We then evaluate the proposed multilabel metrics.

This paper is organized as follows. Section 2 briefly describes the related works. The proposed multichannel convolutional neural network for multilabel sentiment classification is presented in Section 3. In Section 4, the results and their comparison is presented. Finally, Section 5 concludes the paper.

2   Related works

In recent years, researchers widely studied clinical text and user text using natural language processing (NLP). They used both machine learning and deep learning to solve their problems. In this paper, we present the existing works on biomedical texts. Baumel et al. [11] investigated four models such as SVM, CNN, CBOW, and hierarchical attention-based recurrent neural network models for the extreme multilabel task using the MIMIC datasets. The authors indicated that the hierarchical attention-based recurrent neural network model achieves a 55.86% F1 score. Wang et al. [12] developed a rule-based algorithm to generate labels that are weakly supervised. Then, the authors used the pre-trained word embeddings to represent deep features. They employed SVM, random forest, MLPNN, and CNN algorithms. Their study indicated that the CNN model achieves the best performance score. Singh et al. [13] developed an attentive neural tree decoding model for tagging structured bio-medical texts with multilabel. This method decodes an input sequence into a tree of labels. The authors suggested that the proposed model outperforms on SOTA (sate-of-the art) approaches with biomedical abstracts. Citrome [14] reviewed the treatment of Abilify oral users with bipolar I disorder and schizophrenia. The author indicated that the tolerability of Abilify with schizophrenia appears superior to haloperidol, risperidone, and perphenazine. Rios et al. [15] demonstrated the biomedical text classification task using CNN. They indicated that they achieved a 3% improvement over the SOTA results.

Moreover, Gargiulo et al. [16] presented a deep neural network (DNN) for extreme multilabel and multiclass text classification tasks. The authors used two models: the first one uses a word embedding with two dense layers, and the second uses the convolution, word embedding, and the dense layers. Kolesov et al. [17] performed multilabel classification on incompletely labeled biomedical texts using the SVM and RF. They used soft supervised learning and weighted k-nearest neighbor algorithms for modifying the training set. Their study indicated that both algorithms perform better. Parwez et al. [18] presented the CNN model for multilabel text classification. The authors used the domain-specific and generic based pre-trained model to predict class labels. In summary, the above authors used SVM, NB, RF, and CNN to perform multilabel classification tasks on various biomedical texts (Table 1). In this paper, we propose multichannel convolutional neural network for multilabel sentiment classification using Abilify oral user comments.

3   The proposed method

In this section, we present a multichannel convolutional neural network for a multilabel sentiment classification model using Abilify oral user comments. The system architecture is shown in Fig.1. It includes data pre-processing, word embedding, multichannel CNN, merge layer, fully connected layer, and an output layer. Each of these processes is explained as follows.

3.1   Abilify oral dataset

We obtained this Abilify oral dataset from the IEEE Dataport [19, 20]. It contains 1722 user comments with their age group, gender, treatment condition, patient type, treatment duration, and labeled sentiment on satisfaction, effectiveness, and ease of use.

3.2   Pre-processing

The dataset is converted from upper case to lower case, removed punctuation lists and stop words, and retained the numbers where it describes the drugs in grams. Then, each instance is split into separate words using the tokenization method.

3.3   Multichannel convolutional neural network

The multichannel convolutional neural network represents the multiple version of the standard convolutional neural network model with different sizes of kernels. This representation allows the instance or document to process in different n-grams such as 4-gram, 6-gram, and 8-grams at the same time [22]. In particular, we define the standard convolutional neural network model with a word embedding layer, one-dimensional convolutional layer, dropout layer, max-pooling, and flatten layer. This standard version is defined with three channels for different n-grams. Each component of the channel is explained as follows.

3.3.1   Word embedding

In NLP, word embedding represents a feature learning technique where it maps the vocabulary of words or phrases into a vector space. Specifically, we use the GloVe word embedding [10] technique to generate word vectors in a fixed dimension with the semantic relationship between words.

3.3.2   Convolutional layer

Convolutional neural networks perform well in image classification and computer vision-related tasks. The convolutional layer is an important part of the convolutional neural network. It slides over an input sequence with a fixed kernel size to generate feature maps [15, 16, 18, 22, 23]. In this work, we use one-dimensional convolutional layers to move the kernel in one direction. This layer is mostly used to perform NLP tasks. The input and output of the 1D convolutional layer are 2D. The convoluted feature maps output the maximum, minimum, or average values using pooling layers.

| Authors	Dataset	Models	Accuracy	Key Findings
Baumel et al. [9]	MIMIC Datasets	HA-GRU	55.86%	Classification of patient notes on ICD code assignment
Wang et al. [10]	Mayo Clinic smoking status	CNN	92.00%	A rule-based algorithm to generate labels that are weakly supervised
Singh et al. [11]	Articles describing randomized controlled trials	NTD-s	32.70%	An attentive neural tree decoding model for tagging structured bio-medical texts with multilabel
Rios et al. [13]	MED-LINE Citations	CNN-Vote2	64.69%	Biomedical text classification
Gargiulo et al. [14]	PubMed Dataset	CNN-Dense	20.15%	Extreme multilabel and multiclass text classification tasks
Kolesov et al. [15]	AgingPortfolio Dataset	SVM	30.59%	Multilabel classification on incompletely labeled biomedical texts
Parwez et al. [16]	Tweets dataset	CNN-PubMed	94.12%	Domain-specific and generic based pre-trained model to predict class labels

Table 1: Summary of the related works.
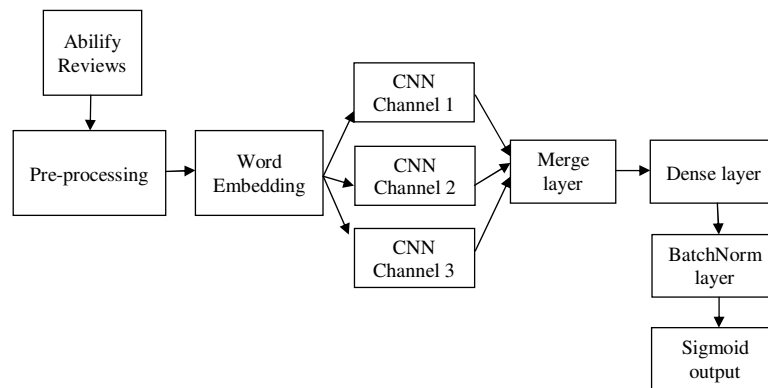
Figure 1: A multichannel convolutional neural network model.

3.3.3 Dropout layer

This layer is used to regularize the neural networks in terms of overfitting and underfitting. Specifically, it ignores some of the outputs in the neural network during the training process.

3.3.4 Max-pooling

The max-pooling layer is applied over each feature map to select the maximum value based on the filter size. It is smaller in size than the feature map. The output of this layer contains the most important feature values of the previous feature map [15, 16, 18, 22].

3.3.5 Flatten layer

The flatten layer converts the pooled feature map into a single column or one-dimensional array. This result is passed to a merged layer.

3.4 Merge layer

The merged or concatenate layer combines the output of each channel. These combined results passed to a fully connected or dense layer.

3.5 Fully connected layer

A fully connected or dense layer connects the input of the flatten layer to all units of the next layer. It works the same as the feed-forward neural network.

3.6 Batch normalization layer

The batch normalization layer allows all layers of a network to learn more independently. Specifically, it standardizes or normalizes the result of previous layers. Also, this layer acts as a regularization parameter to avoid overfitting.

3.7 Sigmoid output layer

The sigmoid output function predicts the probability-based output for each label as shown in equation 1. It is successfully applied in multilabel classification problems [24].

$$f(x) = \frac{1}{1 + exp^x} \tag{1}$$

4 Results and discussion

We implemented the proposed multilabel multichannel model on Abilify oral dataset. This dataset contains 1722 instances associated with a set of labels, namely, ease of use, satisfaction, and effectiveness. We split the dataset into training (1394), validation (155), and testing (173). The data instances are preprocessed with various tasks such as removing punctuations, stopwords, upper case to lower case, and tokenization. Then, word vectors are generated to the input sequences using the GloVe word embedding model. The proposed multichannel convolutional network model was applied to this dataset to perform the multilabel classification task. This model represents three multiple versions of the standard convolutional neural network with different kernel sizes. The standard CNN model consists of a word embedding layer, 1D-convolutional layer, dropout layer, max-pooling layer, and a flattened layer. The output

Data	Accuracy score	Hamming loss	F1 micro score	Accuracy per label		
				0	1	2
Validation	0.548	0.275	0.839	0.820	0.931	0.750
Testing	0.538	0.303	0.820	0.815	0.912	0.715

Table 2: Performance of the proposed multichannel CNN model.

Data	Accuracy score	Hamming loss	F1 micro score	Accuracy per label		
				0	1	2
BR_NB	55.2	0.379	71.5	60.5	68.9	57.0
BR_DT	59.6	0.343	73.9	61.9	79.7	55.3
BR_SVM	60.7	0.338	75.2	62.6	78.3	57.8
CC_NB	55.0	0.378	71.5	60.5	68.7	57.5
CC_DT	61.0	0.347	74.6	61.9	78.3	51.5
CC_SVM	60.5	0.346	75.0	62.6	77.5	55.7
LP_NB	54.1	0.385	69.4	566.	72.0	55.7
LP_DT	60.3	0.354	74.5	60.9	77.7	55.1
LP_SVM	**62.8**	0.334	76.5	63.4	79.8	56.4
Proposed	53.8	**0.303**	**82.0**	**81.5**	**91.2**	**71.5**

Table 3: Model comparison.

of each channel is combined through a merged layer and it is passed to a dense layer, batch normalization layer, and the sigmoid output layer. Specifically, we fixed the following hyperparameters using the random approach such as input length with 150 units, 100 embedding dimension, three kernel sizes (4, 6, and 8), ReLU activation, 0.8 dropouts, pooling size 2, 10 units in the fully connected layer, 20 epochs, and Adam optimizer with a binary cross-entropy loss function. The proposed multichannel CNN model for multilabel classification is evaluated using various multilabel metrics, namely, accuracy or exact match, hamming loss, F1-micro average score, and accuracy per label [3, 5, 20, 21]. Table 2 shows the performance of the proposed multichannel CNN model for multilabel classification. This result is compared with the problem transformation approaches, namely, binary relevance, label powerset, and classifier chains with NB, DT, and SVM [20] as shown in Table 3. The existing researchers in the Table 1 have addressed the multilabel classification using different biomedical texts. In this work, we used the patients and caregivers' opinion on drugs and medications dataset. In particular, we have compared the results of our proposed method with various baselines as shown in Table 3. The proposed multichannel CNN model achieves better results in terms of Hamming loss (30.3%), F1 micro score (82.0%), and accuracy per label (81.5%, 91.2%, 71.5%).

5   Conclusion

In this paper, we proposed a multichannel convolution neural network for multilabel sentiment classification using Abilify oral user comments. A pre-trained model was used to generate word vectors. Then, the proposed model was evaluated with the multilabel classification metrics. The results showed that the proposed multichannel CNN model achieves the better result in terms of Hamming loss, F1 micro score, and accuracy per label than the problem transformation approaches. In future work, we study the trend of drugs and medications in different age groups using patient and caregiver reviews.

Acknowledgement

References

[1] Ronen Feldman. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82-89, 2013. https://doi.org/10.1145/2436256.2436274

[2] Bing Liu. Sentiment analysis: Mining opinions, sentiments, and emotions. *Cambridge university press*, 2020. https://www.cs.uic.edu/~liub/FBS/sentiment-analysis-tutorial-2012.pdf

[3] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, 3(3):1-13, 2007. https://doi.org/10.4018/978-1-59904-951-9.ch006

[4] Sadri Alija, Edmond Beqiri, Alaa Sahl Gaafar, Alaa Khalaf Hamoud. Predicting Students Performance

Using Supervised Machine Learning Based on Imbalanced Dataset and Wrapper Feature Selection. *Informatica*, 47(1), 2022. `https://doi.org/inf.v47i1.4519`

[5] Read J, Pfahringer B, Holmes G, and Frank E. Classifier chains for multi-label classification. *In Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 254-269, 2009. Springer, Berlin, Heidelberg. `https://doi.org/10.1007/978-3-642-04174-7_17`

[6] Tsoumakas G, Katakis I, and Vlahavas I. Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering*, 23(7):1079-1089, 2010. `https://doi.org/10.1109/TKDE.2010.164`

[7] LeCun Y, Bengio Y, and Hinton G. Deep learning. *Nature*, 521(7553):436-444, 2015. `https://doi.org/10.1038/nature14539`

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. *Cambridge: MIT press*, 1(2), 2016. `http://www.deeplearningbook.org`

[9] Patel R, Tanwani S, and Patidar C. Relation Extraction Between Medical Entities Using Deep Learning Approach. *Informatica*, 45(3), 2021. `https://doi.org/10.31449/inf.v45i3.3056`

[10] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543, 2014. `https://doi.org/10.3115/v1/d14-1162`

[11] Baumel T, Nassour-Kassis J, Cohen R, Elhadad M, and Elhadad N. Multi-label classification of patient notes a case study on ICD code assignment. 2017. `https://arxiv.org/abs/1709.09587`

[12] Wang Y, Sohn S, and Liu S et al. A clinical text classification paradigm using weak supervision and deep representation. *BMC medical informatics and decision making*, 19(1):1, 2019. `https://doi.org/10.1186/s12911-018-0723-6`

[13] Singh G, Thomas J, Marshall IJ, Shawe-Taylor J, and Wallace BC. Structured multi-label biomedical text tagging via attentive neural tree decoding. 2018. `https://arxiv.org/abs/1810.01468`

[14] Leslie Citrome. A review of aripiprazole in the treatment of patients with schizophrenia or bipolar I disorder. *Neuropsychiatric Disease and Treatment*, 2(4):427, 2006. `https://doi.org/10.2147/nedt.2006.2.4.427`

[15] Rios A and Kavuluru R. Convolutional neural networks for biomedical text classification: application in indexing biomedical articles. *In Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, 258-267, 2015. `https://doi.org/10.1145/2808719.2808746`

[16] Gargiulo F, Silvestri S, and Ciampi M. Deep Convolution Neural Network for Extreme Multi-label Text Classification. *In HEALTHINF*, 641-650, 2018.

[17] Kolesov A, Kamyshenkov D, Litovchenko M, Smekalova E, Golovizin A, and Zhavoronkov A. On multilabel classification methods of incompletely labeled biomedical text data. *Computational and mathematical methods in medicine*, 2014. `https://doi.org/10.1155/2014/781807`

[18] Parwez MA and Abulaish M. Multi-label classification of microblogging texts using convolution neural network. *IEEE Access*, 7:68678-68691, 2019. `https://doi.org/10.1109/ACCESS.2019.2919494`

[19] Ashok Kumar J, Abirami S, and Tina Esther Trueman. Abilify Oral user reviews. *IEEE Dataport*, 2020. `https://dx.doi.org/10.21227/p1jp-2m84`

[20] Kumar JA, Abirami S, and Trueman TE. Multilabel Aspect-Based Sentiment Classification for Abilify Drug User Review. *In 2019 11th International Conference on Advanced Computing (ICoAC)*, IEEE, 376-380, 2019. `https://doi.org/10.1109/ICoAC48765.2019.246871`

[21] Baadel S, Thabtah F, Lu J, and Harguem S. OMCOKE: A Machine Learning Outlier-based Overlapping Clustering Technique for Multi-Label Data Analysis. *Informatica*, 46(4), 2022. `https://doi.org/10.31449/inf.v46i4.3476`

[22] Yoon Kim. Convolutional neural networks for sentence classification. 2014. `https://arxiv.org/abs/1510.03820`

[23] Oo SH, Hung ND, and Theeramunkong T. Justifying convolutional neural network with argumentation for explainability. *Informatica*, 46(9), 2023. `https://doi.org/10.31449/inf.v46i9.4359`

[24] Burkhardt S and Kramer S. Online multi-label dependency topic models for text classification. *Machine Learning*, 107(5):859-886, 2018. `https://doi.org/10.1007/s10994-017-5689-6`