

# *Team BJUT-BJFU at BioCreative VII LitCovid Track: A Deep Learning based Method for Multi-label Topic Classification in COVID-19 Literature*

Shuo Xu<sup>1</sup>, Yuefu Zhang<sup>1</sup>, Xin An<sup>2</sup>

<sup>1</sup> College of Economics and Management, Beijing University of Technology, Beijing 100124, P.R. China  
Email: xushuo@bjut.edu.cn, yaogeng\_z@163.com

<sup>2</sup> School of Economics and Management, Beijing Forestry University, Beijing 100083, P.R. China  
Email: anxin@bjfu.edu.cn, Corresponding author

**Abstract—Background:** The rapid growth of COVID-19-related articles poses a significant challenge for manual curation and interpretation. It is very crucial to be able to increase the accuracy of automated topic prediction in COVID-19-related literature. The LitCovid track in BioCreative VII was specially designed to optimize the multi-label annotation for COVID-19 literature.

**Materials and Methods:** The following fields are utilized to predict the multi-label classification of COVID-19 literature: title, abstract, keyword, journal name, and the resulting labels. To benefit from powerful deep learning models, four models are involved in our submissions: FastText, TextRCNN, TextCNN, and Transformer.

We combine the training and development sets into our training set, which is further grouped into ten disjoint subsets with nearly equal size and similar label distribution. Then, the resulting parameters in each model above are tuned with 10-fold cross-validation procedure. Finally, to ensure that the posterior probability of each instance's predicted label follows the same format of a provided sample, the posterior probability is adjusted properly.

**Results:** The best performance of our five runs is 85.56%, 78.47% and 87.01% in terms of label-based micro-average F1, label-based macro-average F1, and instance-based F1, respectively. This outperforms the baseline method ML-Net.

**Conclusions:** A deep learning based approach for multi-label classification in COVID-19 literature is proposed in this study. Our approach mainly consists of three components: preprocessing (dataset division, formatting), modeling (10-fold cross validation, prediction), and post-processing (threshold adaption and probability conversion). According to the results disclosed by the organizer, our method is superior to the baseline method. This indicates that our method is valuable in dealing with multi-label classification problem.

**Keywords—Multi-Label Learning; Deep Learning; Cross Validation; Label Relationship; Stratification Method**

## I. INTRODUCTION

The rapid growth of biomedical literature poses a significant challenge for manual curation and interpretation. This challenge has become more evident during the COVID-19 pandemic: the number of COVID-19-related articles in the literature is growing by about 10,000 articles per month (1).

The LitCovid corpus aims to help users to follow the latest SARS-CoV2 and COVID-19 studies (1).

The LitCovid track in BioCreative VII challenge (2) calls for a community effort to tackle automated topic annotation for COVID-19 literature. The LitCovid challenge in BioCreative VII was specially designed to optimize the multi-label classification for COVID-19 literature annotation. In this track, each scholarly article can be assigned one or more labels from a set of seven elements (mechanism, transmission, diagnosis, treatment, prevention, case report, or epidemic forecasting).

In the literature, many multi-learning methods have been put forward, which can be divided into three groups (7): problem transformation, algorithm adaptation, and ensemble. With the rise of deep learning technique in recent years, many deep learning based methods with promising performance for multi-learning learning are also raised, such as ML-Net (3), FastText (4), TextRCNN (6), TextCNN (5), and so on. In general, most of these methods use a similar framework, which often consists of 2 modules: a neural network and a label predictor. It is worth noting that these methods actually fall into the category of algorithm adaptation.

Therefore, we participate this track with a deep learning based method, in which several neural network models are embedded as our core module. Here, we present the method, the results and recognition system from our participation in the LitCovid challenge. In more details, we combine the training and development sets into our training set. Different from conventional splitting procedure for the single-label task, the stratification method in Sechidis et al. (8) is used here. Then, the resulting parameters in each model above are tuned with 10-fold cross-validation procedure. Furthermore, we model the label relationship (10). Finally, to ensure that the posterior probability of each instance's predicted label follows the same format of a provided sample, the posterior probability is adjusted properly.

The organization of the rest of the article is as follows. After the materials and methods are described, we present the results of our submissions on the LitCovid track. The last section concludes this contribution with next step for improvement.

## II. MATERIALS AND METHODS

As shown in Figure 1, our method comprises three major components: preprocessing (dataset division, formatting), modeling (10-fold cross validation, prediction), and post-processing (threshold adaption and probability conversion)

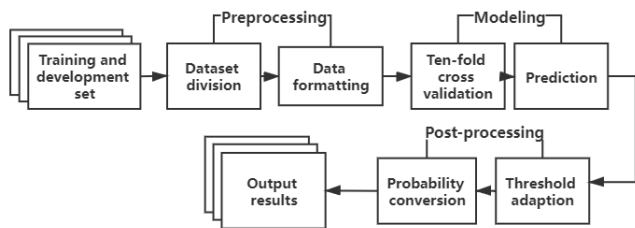


FIGURE 1 THE SYSTEM PROCESSING PIPELINE

In the provided dataset, there are seven unique labels: treatment, mechanism, case report, prevention, diagnosis, transmission, and epidemic forecasting. The resulting distribution of instances is 25.5%, 12.9%, 6.0%, 32.4%, 18.1%, 3.1%, and 2.0%, respectively. No hierarchical structure between labels is observed. Similar to Chen et al. (1), three representative topic co-occurrence patterns were observed: diagnosis-treatment, mechanism-treatment, and transmission-prevention. These observations follow two main characteristics in multi-label learning problems (10): imbalanced label distribution and label correlations.

For the sake of dealing with this problem, the following fields are utilized in our submissions: title, abstract, keyword, journal name, and the resulting labels. To facilitate further data processing, the dataset files are transformed into the JSON format. It should be noted that we have not excluded data, even though some instances miss keywords. In addition, we pool the training and development data sets into our training set, which is further grouped into ten disjoint subsets with nearly equal size and similar label distribution. Due to intrinsic characteristic of multi-label learning problem, it is not trivial to group a dataset into several disjoint subsets (11). For this purpose, the stratification method in Sechidis et al. (9) is used here.

The participating teams are allowed to generate up to five different predictions ("runs") for the test set and to submit the predictions to the organizer. Hence, our submissions utilize four models with different parameter settings: FastText (4), TextRCNN (6), TextCNN (5), and Transformer (9). The resulting parameters in each model are tuned with 10-fold cross-validation procedure (12).

To evaluate the performance of submitted results, the evaluation script was provided by the organizer. In this script, two groups of measures are involved: label-based indices and instance-based ones. In general, models that achieve higher performance in most metrics, especially the measures related to F-score, will rank higher. On closer examination, we find an implicit assumption seems be made in the evaluation script (default threshold: 0.5): it is better to transform this multi-label topic classification problem to multiple binary classification problems.

Since this study considers high-order correlations among labels (10), the posterior probabilities of seven labels of each instance are summed to 1. To ensure that the posterior probability of each instance's predicted label is compatible with the format of a provided sample for problem transformation approaches, the posterior probability is adjusted properly. In more details, based on the differences of two types of approaches, the threshold for the test dataset is adjusted to control the imbalanced label distribution. Finally, on the basis of the improved threshold, the probabilities are linearly transformed so that the probabilities under this threshold can be evaluated to reflect the real performance of our method.

## III. RESULTS

In this participation, we submit five groups of results in total, as illustrated Table 1. It is noteworthy that our submissions 1-2 correspond to FastText Model with different parameter settings, submission 3 to TextRCNN model, submission 4 to TextCNN model, and submission 5 to Transformer model. From Table 1, it is not difficult to see that our run 4 ranks the first in terms of all balanced F-measures. More specifically, for the LitCovid track, the best performance of our runs is 85.56%, 78.47% and 87.01% in terms of label-based micro-average F1, label-based macro-average F1, and instance-based F1, respectively. The baseline model ML-Net (3) achieved 84.37%, 76.55%, and 86.78% in terms of label-based micro-average F1, label-based macro-average F1 and instance-based F1. It is obvious that our method is superior to the baseline.

TABLE 1 THE PERFORMANCE OF OUR METHOD AND THE BASELINE

	Run1	Run2	Run3	Run4	Run5	Baseline
label-based micro avg. precision	0.8775	<b>0.8982</b>	0.8555	0.8704	0.8474	0.8756
label-based micro avg. recall	0.8003	0.8028	0.8153	<b>0.8413</b>	0.8213	0.8142
label-based micro avg. f1	0.8371	0.8478	0.8349	<b>0.8556</b>	0.8342	0.8437
label-based macro avg. precision	0.8492	<b>0.8750</b>	0.8162	0.8232	0.8230	0.8364
label-based macro avg. recall	0.7107	0.6913	0.7050	<b>0.7596</b>	0.6744	0.7309
label-based macro avg. f1	0.7626	0.7528	0.7492	<b>0.7847</b>	0.7241	0.7655
instance-based precision	0.8490	0.8670	0.8404	0.8731	0.8508	<b>0.8849</b>
instance-based recall	0.8308	0.836	0.8333	<b>0.8672</b>	0.8447	0.8514
instance-based f1	0.8398	0.8512	0.8368	<b>0.8701</b>	0.8477	0.8678

## IV. CONCLUSIONS

To deal with the multi-label topic classification for COVID-19 literature in BioCreative VII challenge, this study

develops a deep learning based approach. Our method consists of three major components: preprocessing, modeling, and post-processing. According to the results disclosed by the organizer, our method is superior to the baseline method. This indicates that our method is valuable in dealing with this problem. Though, there is still room for improvement. In the near future, we will leverage more information attached to scientific publications, such as MeSH terms, chemical entities (13) and so on. In addition, other advanced deep learning model with tuned hyper-parameters will be incorporated into our method for performance comparison.

## V. ACKNOWLEDGEMENTS

This work was supported partially by the Natural Science Foundation of China [Grant Number 72074014 and 72004012].

## VI. REFERENCES

- [1] Qingyu Chen, Allot Allot, Zhiyong Lu, (2021). LitCovid: An open database of COVID-19 literature. *Nucleic Acids Research*. Jan 8;49(D1):D1534-D1540.
- [2] Qingyu Chen, Alexis Allot, Robert Leaman., et al, (2021). Overview of the BioCreative VII LitCovid Track: multi-label topic classification for COVID-19 literature annotation. *Proceedings of the seventh BioCreative challenge evaluation workshop*.
- [3] Jingcheng Du, Qingyu Chen, Yifan Peng., et al, (2019). ML-Net: multi-label classification of biomedical texts with deep neural networks. *Journal of the American Medical Informatics Association : JAMIA*. Nov 1;26(11):1279-1285.
- [4] Armand Joulin, Edouard Grave, Piotr Bojanowski., et al, (2016). FastText.zip: compressing text classification models. arXiv preprint arXiv: 1621.03651.
- [5] Yoon Kim, (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746-1751.
- [6] Siwei Lai, Liheng Xu, Kang Liu., et al, (2015). Recurrent convolutional neural networks for text classification. *Proceedings of the 29th Conference on Artificial Intelligence*, pp. 2267-2273.
- [7] Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj., et al, (2012). An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9), pp. 3084-3104.
- [8] Konstantinos Sechidis, Grigorios Tsoumakas, Ioannis Vlahavas, (2011). On the stratification of multi-label data. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pp. 145-158.
- [9] Ashish Vaswani , Noam Shazeer , Niki Parmar., et al, (2017). Attention is all you need. *Proceedings of the 31st Conference on Neural Information Processing System*. arXiv:1706.03762
- [10] Shuo Xu , Xin An , (2019). ML<sup>2</sup>S-SVM: multi-label least-squares support vector machine classifiers. *The Electronic Library*, 37(6), pp. 1040-1058.
- [11] Shuo Xu , Ling Li , Liyuan Hao., et al, (2021). An author interest discovery model armed with authorship credit allocation scheme. *Proceedings of the 16th International Conference on Information (iConference)*, pp. 199-207.
- [12] Shuo Xu , Fujing Ma , Lan Tao , (2007). Learn from the Information contained in the false splice sites as well as in the true splice sites using svm. *Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering*, pp. 1360-1366.
- [13] Shuo Xu , Xin An , Lijun Zhu ., et al, (2015). A CRF-based system for recognizing chemical entity mentions (CEMs) in biomedical literature. *Journal of Cheminformatics*, 7(Suppl. 1), pp. S11.