Clinical Text Classification with Rule-based Features and Knowledge-guided Convolutional Neural Networks

Liang Yao, Chengsheng Mao and Yuan Luo*
Northwestern University
Chicago, Illinois 60611
{liang.yao, chengsheng.mao, yuan.luo}@northwestern.edu
*Corresponding Author

***Abstract*—Clinical text classification is an important problem in medical natural language processing. Existing studies have conventionally focused on rules or knowledge sources-based feature engineering, but only a few have exploited effective feature learning capability of deep learning methods. In this study, we propose a novel approach which combines rule-based features and knowledge-guided deep learning techniques for effective disease classification. We evaluated our method on the 2008 Integrating Informatics with Biology and the Bedside (i2b2) obesity challenge. The results show that our method outperforms the state of the art methods.**

***Index Terms*—clinical text classification, obesity challenge, convolutional neural networks, word embeddings, entity embeddings**

I. Introduction

Clinical records are an important type of electronic health record (EHR) data and often contain valuable and detailed patient information and clinical experiences of doctors. As a fundamental task of natural language processing, text classification plays an important role in clinical records organization and retrieval, and it can support cohort identification and clinical decision [1], [2].

In this work, we propose a new method which combines rule-based features and knowledge-guided deep learning techniques for disease classification. We evaluated our method on the 2008 i2b2 obesity challenge [3], a multilabel classification task focused on obesity and its 15 most common comorbidities (diseases). The results demonstrate that our method outperforms state of the art methods for the challenge.

II. Method

Our method consists of three steps: identifying trigger phrases, predicting classes with very few examples using trigger phrases and training a knowledge-guided convolutional neural network for more populated classes[1]. We base our method on Solt's system [4] to identify trigger phrases and predict classes with very few examples. Solt's system is a very strong rule-based system. It ranked the first place in the intuitive task and the second place in the textual task and overall first place in i2b2 Obesity challenge. Solt's system can discover very informative trigger phrases with positive, negative or uncertain contexts. We use a Perl implementation[2] of Solt's system provided by the authors.

A. Trigger Phrases Identification

We follow Solt's system [4] to identify trigger phrases. We first do the same preprocessing including abbreviation resolution and family history removing. We then use the same disease names (e.g., Gallstones), their directly associated terms (e.g., Cholelithiasis) and negative/uncertain words to identify trigger phrases. The trigger phrases are disease names and their alternative names with or without negative/uncertain words.

B. Predicting Classes with Very Few Examples using Trigger Phrases

As the classes in obesity challenge are very unbalanced, and some classes even don't have training examples. Therefore we could not predict these classes using machine learning methods and resort to rules defined in Solt's system [4]. We exclude classes with very few examples in training set of each disease. Specifically, we remove examples with Q or N label for textual task and remove examples with Q label in intuitive task. Then for examples in the test set, we use trigger phrases to predict their labels. Following Solt's system [4], we assume positive trigger phrases (disease names/alternatives without negative/uncertain words) are prior to negative trigger phrases, and negative trigger phrases are prior to uncertain trigger phrases. Therefore, if a clinical record has uncertain trigger phrases and dose not have positive/negative trigger phrases, we label it as Q. Similarly, if a clinical record has negative trigger phrases and dose not have positive trigger phrases, we label it as N.

C. Knowledge-guided Convolutional Neural Networks

After removing classes with very few examples, there are only two classes in the training set of each disease (Y and U for textual task, Y and N for intuitive task). We train a Convolutional Neural Network (CNN) on positive trigger

[1]We released the implementation at https://github.com/yao8839836/obesity.

[2]https://github.com/yao8839836/obesity/tree/master/perl_classifier

978-1-5386-6777-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICHI-W.2018.00024

phrases and Unified Medical Language System (UMLS) Concept Unique Identifiers (CUIs) of training records, and classify test examples using the learned CNN model.

For each disease, we use its positive trigger phrases with word2vec word embeddings as the input of CNN. We used the pre-trained 200 dimensional word embeddings learned from MIMIC-III clinical notes. We tried word embeddings with 100, 200, 300, 400, 500 and 600 dimensions, and found using 200 dimensional word embeddings performs the best.

We also use medical knowledge source to enrich the input of CNN model. We use MetaMap to link the full clinical text to CUIs in UMLS. After entity linking, a clinical record is represented as a bag of CUIs. We choose 13 types of CUIs which are closely related to diseases as the input entities of CNN. We found using the subset of CUIs leads to better results than using all CUIs. We use pre-trained CUIs embeddings made by De Vine et al [5] as the input entity vectors of CNN.

The input layer of our CNN model consists of word embeddings of positive trigger phrases and CUIs embeddings of selected CUIs in each clinical record. A one-dimensional convolution layer is built on the input word embeddings and entity embeddings. We use max pooling to select the most important feature with the highest value in the convolutional feature map. We then concatenate the max pooling results of word embeddings and CUIs embeddings. The concatenated hidden features are fed into a fully connected layer, then a dropout and ReLU activation layer. Finally, a fully connected layer is fed to a softmax output layer, whose output is the probability distribution over labels.

We implement our CNN model using TensorFlow. We set the following parameters for our model: the number of convolution filters: 256, the convolution kernel size: 5, the dimension of hidden layer in the fully connected layer: 128, dropout keep probability: 0.8, learning rate: 0.001, batch size: 64, the number of learning epochs: 30. We also tried other settings of these parameters but do not find much difference. We use softmax cross entropy loss as the loss function and Adam algorithm as the optimizer.

III. Results

Table I shows Macro F_1 scores and Micro F_1 scores of our method. From the table, we can observe that when using CNN with word embeddings and CUIs embeddings as inputs, F_1 scores for different diseases are improved, and the overall F_1 scores are higher than Solt's system in both textual task and intuitive task. This is due to the fact that the disambiguated CUIs are closely related to diseases, and word & CUIs embeddings contain more semantic information, which is helpful for disease classification. To the best of our knowledge, we have achieved the best results in intuitive task so far. Note that the results of Solt's system remain the same, while our method produces slightly different results in different runs. We run our model ten times and found the overall Macro F_1 scores and Micro F_1 scores are significantly higher (p value < 0.05 based on student t test) than Solt's system.

TABLE I: Macro F_1 scores and Micro F_1 scores of our method with word and entity embeddings. Scores in bold font means they are higher than the corresponding scores of Solt's system.

Disease	Textual		Intuitive	
	Macro F_1	Micro F_1	Macro F_1	Micro F_1
Asthma	0.9434	0.9921	0.9784	0.9894
CAD	0.8551	0.9235	**0.6233**	**0.9345**
CHF	0.7939	0.9355	0.6236	0.9315
Depression	0.9716	0.9842	**0.9602**	**0.9727**
DM	**0.9056**	**0.9801**	**0.9731**	**0.9770**
Gallstones	0.8141	0.9822	0.9689	0.9837
GERD	0.4880	0.9881	0.5768	0.9131
Gout	0.9733	0.9881	0.9771	0.9900
Hypercholesterolemia	0.7922	0.9721	**0.9113**	**0.9118**
Hypertension	0.8378	0.9621	**0.9240**	**0.9484**
Hypertriglyceridemia	0.9434	0.9961	0.7092	0.9630
OA	**0.9626**	**0.9781**	**0.6307**	**0.9610**
Obesity	**0.4885**	**0.9696**	**0.9747**	**0.9754**
OSA	0.8781	0.9920	0.8805	0.9939
PVD	0.9682	0.9862	0.6314	0.9742
Venous insufficiency	**0.8816**	**0.9882**	0.8083	0.9625
Overall	**0.8016**	**0.9763**	**0.6768**	**0.9624**

IV. Conclusion

In this study, we present a new clinical text classification method which combines rule-based features and knowledge-guided deep learning techniques. The evaluation results on the i2b2 obesity challenge show that our method outperforms the state of the art methods for the challenge. We plan to develop more principled methods and evaluate the methods on more clinical records datasets in our future work.

Acknowledgment

We would like to thank i2b2 National Center for Biomedical Computing funded by U54LM008748, for providing the clinical records originally prepared for the Shared Tasks for Challenges in NLP for Clinical Data organized by Dr. Ozlem Uzuner. We thank Dr. Uzuner for helpful discussions. We would like to also thank NVIDIA GPU Grant program for providing the GPU used in our computation. This work was supported in part by NIH Grant 1R21LM012618-01.

References

[1] D. Demner-Fushman, W. W. Chapman, and C. J. McDonald, "What can natural language processing do for clinical decision support?" *Journal of biomedical informatics*, vol. 42, no. 5, pp. 760–772, 2009.

[2] C.-C. Huang and Z. Lu, "Community challenges in biomedical text mining over 10 years: success, failure and the future," *Briefings in bioinformatics*, vol. 17, no. 1, pp. 132–144, 2015.

[3] Ö. Uzuner, I. Goldstein, Y. Luo, and I. Kohane, "Identifying patient smoking status from medical discharge records," *Journal of the American Medical Informatics Association*, vol. 15, no. 1, pp. 14–24, 2008.

[4] I. Solt, D. Tikk, V. Gál, and Z. T. Kardkovács, "Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier," *Journal of the American Medical Informatics Association*, vol. 16, no. 4, pp. 580–584, 2009.

[5] L. De Vine, G. Zuccon, B. Koopman, L. Sitbon, and P. Bruza, "Medical semantic similarity with a neural language model," in *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. ACM, 2014, pp. 1819–1822.