Interdiscip Sci Comput Life Sci
DOI 10.1007/s12539-016-0196-1

SHORT COMMUNICATION

Rectified-Linear-Unit-Based Deep Learning for Biomedical Multi-label Data

Pu Wang[1,2,3] · Ruiquan Ge[1,2] · Xuan Xiao[3] · Yunpeng Cai[1] · Guoqing Wang[4] · Fengfeng Zhou[5,6]

Received: 27 May 2016 / Revised: 3 October 2016 / Accepted: 25 October 2016

Abstract Disease diagnosis is one of the major data mining questions by the clinicians. The current diagnosis models usually have a strong assumption that one patient has only one disease, *i.e.* a single-label data mining problem. But the patients, especially when at the late stages, may have more than one disease and require a multi-label diagnosis. The multi-label data mining is much more difficult than a single-label one, and very few algorithms have been developed for this situation. Deep learning is a data mining algorithm with highly dense inner structure and has achieved many successful applications in the other areas. We propose a hypothesis that rectified-linear-unit-based deep learning algorithm may also be good at the clinical questions, by revising the last layer as a multi-label output. The proof-of-concept experimental data support the hypothesis, and the community may be interested in trying more applications.

Keywords Multi-label classification · Single-label classification · Deep learning · Rectified linear unit · Clinical diagnosis

Deep learning tries to model high-level data structure abstraction by multiple inner transformation layers and has been widely employed in various machine learning problems since Hinton proposed it in 2006 [1]. Generally speaking, deep learning can discover minor similarities within a high-dimensional dataset and abstract them into the significant combinations. Such high-level abstractions can be automatically inferred through convolutional neural networks (CNNs), deep belief networks (DBNs), deep stacking networks (DSNs), etc. [2].

Deep learning has been employed in many areas, including image processing, speech recognition, and natural language processing. In the recent years, deep learning algorithms also successfully solved quite a few biomedical

Pu Wang and Ruiquan Ge contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s12539-016-0196-1) contains supplementary material, which is available to authorized users.

✉ Yunpeng Cai
yp.cai@siat.ac.cn

✉ Guoqing Wang
qing@jlu.edu.cn

✉ Fengfeng Zhou
FengfengZhou@gmail.com; ffzhou@jlu.edu.cn;
http://www.healthinformaticslab.org/ffzhou/

[1] Shenzhen Institutes of Advanced Technology, and Key Lab for Health Informatics, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, People's Republic of China

[2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong 518055, People's Republic of China

[3] Computer Department, Jingdezhen Ceramic Institute, Jingdezhen, Jiangxi 333403, People's Republic of China

[4] Department of Pathogenobiology, Basic Medical College of Jilin University, Changchun, Jilin 130012, People's Republic of China

[5] College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, People's Republic of China

[6] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, People's Republic of China

Published online: 11 November 2016

classification problems, e.g. long non-coding RNAs prediction [3, 4], protein structure prediction [5, 6], chemoinformatics [7, 8], and RNA-binding protein prediction [9]. All these biomedical problems have a common feature that one sample has only one class label, and they are defined as single-label classification problems. But the real-world situation is much more complicated, and one sample may have multiple class labels, e.g. a patient may have more than one disease. Little attention has been paid to the application of deep learning algorithms in multi-label biomedical classification problems [10–12]. Lipton et al. [13] employed the Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) hidden units to model the multi-label diagnosis problem using the time-series data from the Intensive Care Unit (ICU) patients. This study demonstrated that deep learning may also contribute to the multi-label non-time-series biomedical data.

While the deep RNN algorithm performs well for the single-label biomedical time-series data, here we demonstrate that a deep learning algorithm featuring a new cost function and the rectified linear unit (ReLU) activation function performs efficiently and accurately on multi-label biomedical datasets (Fig. 1). The major advantage of the proposed model is ReLU's high computational efficiency, and independent experiments demonstrated that ReLU outperforms the traditional sigmoid or hyperbolic tangent activation function, even without the pre-training layer [2]. Generally speaking, the output classifier of a deep network is SoftMax, which cannot be used for multi-label learning, because the output class labels are not mutually exclusive to each other, and the sum of all the output values is no longer expected to be 1. So we employ the multiple linear regression model (MLR) in the output layer, and the cost function is formulated as:

$$J(W,b) = \frac{1}{2m}\|O - Y\|_F^2 + \frac{\lambda}{2}\sum_{l=1}^{L-1}\|W_l\|_F^2$$

where O is the actual outputs for all the m training samples, Y is the target labels for them, W_l ($l = 1,\ldots,L-1$) are the connection weights between layers, L is depth of a network, λ is the weight decay parameter, and F indicates Frobenius Norm.

Given a training sample (x, y), where x and y are the feature vectors and 0–1 label vectors of the sample, the parameters in the network can be updated by error back-propagation.

1. Calculate the activation values of all layers by forward propagation.
2. For the output layer, the error is

$$\delta_L = \boldsymbol{o} - \boldsymbol{y}$$

where $\boldsymbol{o}$ is the real output vector and $\boldsymbol{y}$ is the expected vector.

3. For the hidden layer when $l = L-1,L-2,\ldots,2$, the error is

$$\delta_l = (W_l^T \delta_{l+1}) \cdot f'(z_l)$$

where z_l is the input of layer l, f is the activation function, and $\cdot$ is the element-wise product.

4. Calculate the partial derivatives as below

$$\nabla_{W_l} J(W,b;\boldsymbol{x},\boldsymbol{y}) = \delta_{l+1}(a_l)^{\mathrm{T}},$$
$$\nabla_{b_l} J(W,b;\boldsymbol{x},\boldsymbol{y}) = \delta^{(l+1)},$$

where a_l is the activation of layer l.

Since $J(W,b) = \frac{1}{m}\sum_{i=1}^{m} J(W,b;\boldsymbol{x}^i,\boldsymbol{y}^i) + \frac{\lambda}{2}\sum_{l=1}^{L-1}\|W_l\|_F^2$, it is easy to get the batch gradients $\nabla_{W_l} J(W,b)$ and $\nabla_{b_l} J(W,b)$ ($l = 1,\ldots,L-1$), and then the model parameters will be optimized by the gradient descent algorithm. The proposed deep learning framework is denoted as bioReLU.

As shown in Fig. 2, the deep-learned features (outputs of the last hidden layer) may be further fed to the other multi-label classifiers. In our experiments, the DWkNN algorithm [14] was employed. Given a query sample, it

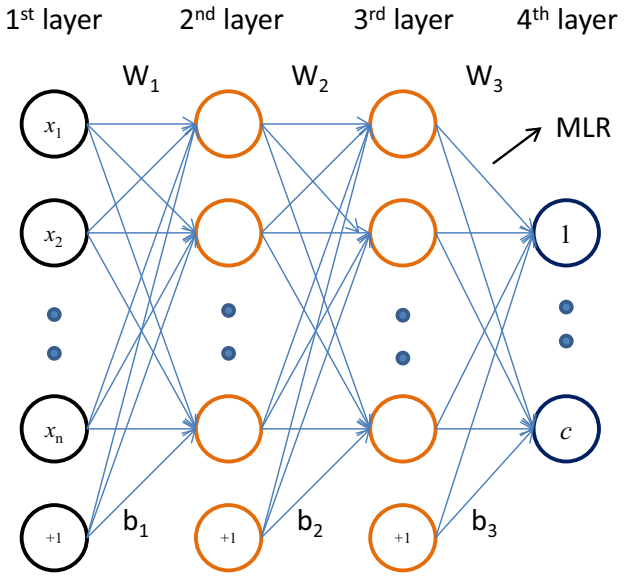

Fig. 1 Diagram of the proposed multi-label deep learning method

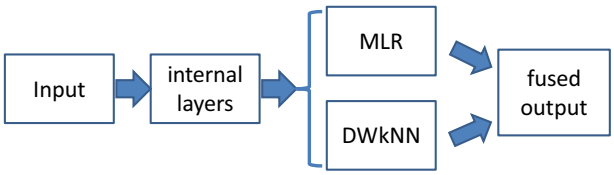

Fig. 2 Combination of two multi-label classifiers based on the deep-learned features

firstly computes the weights of k nearest neighbours based on the distances; then, the output of each label is defined as the ratio of sum of weights that is relevant with that label to the sum of all weights. Finally, the two classifiers are combined by averaging their outputs, and the relevant and irrelevant labels are separated by threshold value 0.5. This combination model is denoted as bioReLUc.

The performance of the compared methods was evaluated by the Hold-Out validation on the two publicly available multi-label datasets Medical and Yeast, which were pre-divided into training (about 2/3) and testing parts at the Mulan website: http://mulan.sourceforge.net/datasets-mlc.html [15–17]. Five widely used multi-label algorithms are used for comparing, *i.e.* MLKNN [18], RAkEL [15–17], Ensemble Classifier Chain (ECC) [19], HOMER [20], and RF-PCT [21]. The multi-label classifiers were evaluated using eight commonly used performance measurements, *i.e.* Hamming Loss, Subset Accuracy, Accuracy, micro-F1, macro-F1, Ranking Loss, One Error, Coverage, and Average Precision, as defined in [15–17]. In our methods, two ReLU hidden layers with size 500 were used, the maximum iterations was 1000, and the weight decay parameter λ, learning rate in gradient descent, and k value in DWkNN are determined by validation on the training set. The experiment results for the proposed deep learning algorithm and the combination model are listed in Tables 1 and 2. The results of the other methods come from [22], because all the parameters have been optimized in the experiments. From the tables, we can found that the combination model based on deep-learned features outperforms the unmixed one nearly in all metrics, and both the proposed methods are more competitive than the state-of-the-art algorithms.

The proposed algorithms are very different from the pre-existing multi-label neural network [23] or deep learning methods [24], including cost function, activation function of hidden layers, and the combination way. The same Hold-Out validation strategy and three publicly available biomedical datasets were utilized, and the classification

Table 3 Accuracy of different multi-label deep learning methods

Dataset	bioReLU	bioReLUc	DBN_{bp}	DBN_{ECC}
Yeast	0.519	0.530	0.529	**0.531**
Genbase	0.987	**0.993**	0.984	0.985
Medical	0.769	**0.789**	0.746	0.742

The best results for each dataset are in bold

Table 1 Evaluation results by independent test on medical dataset

Metrics	bioReLU	bioReLUc	ML-kNN	RAkEL	ECC	HOMER	RF-PCT
Hamming loss↓	**0.010**	**0.010**	0.017	0.012	0.014	0.012	0.014
Subset accuracy↑	0.678	**0.715**	0.462	0.607	0.526	0.610	0.538
Accuracy↑	0.769	**0.789**	0.528	0.673	0.611	0.713	0.591
MicroF1↑	0.817	**0.820**	0.634	0.714	0.714	0.773	0.693
MacroF1↑	**0.338**	0.334	0.192	0.210	0.203	0.282	0.207
Ranking loss↓	**0.024**	**0.024**	0.045	0.159	0.152	0.090	**0.024**
One error↓	0.141	**0.126**	0.279	0.312	0.315	0.216	0.174
Coverage↓	1.793	1.910	2.844	8.520	7.994	5.324	**1.619**
Average precision↑	0.889	**0.897**	0.784	0.676	0.684	0.786	0.868

↓ indicates lower is better, ↑ means higher is better, and the best results for each metric are in bold

Table 2 Evaluation results by independent test on Yeast dataset

Metrics	bioReLU	bioReLUC	ML-kNN	RAkEL	ECC	HOMER	RF-PCT
Hamming loss	0.192	**0.190**	0.198	0.192	0.207	0.207	0.197
Subset accuracy	0.190	**0.225**	0.159	0.201	0.215	0.213	0.152
Accuracy	0.519	0.530	0.492	0.531	0.546	**0.559**	0.478
MicroF1	0.650	0.658	0.625	0.656	0.658	**0.673**	0.617
MacroF1	0.370	0.406	0.336	0.359	0.350	**0.447**	0.322
Ranking loss	0.171	**0.163**	0.172	0.259	0.224	0.205	0.167
One error	0.229	**0.223**	0.234	0.254	0.249	0.248	0.248
Coverage	6.449	6.249	6.414	7.983	7.153	7.285	**6.179**
Average precision	0.764	**0.773**	0.758	0.715	0.734	0.740	0.757

↓ indicates lower is better, ↑ means higher is better, and the best results for each metric are in bold

accuracies are shown in Table 3, similar to [24]. Read and Hollmen demonstrated that the classifier chain strategy performs similarly well on the multi-label learning problem, compared with the binary relevance (BR) and back-propagation (BP) strategies [25]. Based on the only measurement microF1 utilized by both Read and Hollmen [25] and this study, ReLU proposed in this study outperforms the multi-label models in [25]..

So the proposed ReLU-based algorithms perform similarly to or better than the others, and the combined version bioReLUc tends to perform better. It should be noted that the DBN-based methods are very time-consuming, because it has to be trained layer by layer. From the perspective of accuracy, DBN_{ECC}, in which the deep-learned features are fed to ECC, seems an inefficient combination way, because the DBN and ECC are both intricate, and there is little improvement after combination.

Our proof-of-concept work and a few other recent studies suggest that deep learning algorithms may be applied on the multi-label classification problems, and appropriate cost and activation functions may achieve satisfactory multi-label classification performances in the biomedical area. The majority of the literature focuses on the single-label classification, which is frequently challenged by the multi-label issues in the real-world situation. This is a particularly serious issue in the biomedical area, where a gene may have multiple functions or a patient may suffer from multiple diseases. Accompanied by the rapid accumulation of biomedical big data, we believe that more researchers will choose to conquer the multi-label classification problems directly, and the deep learning techniques will play an important role in solving these complicated challenges.

Acknowledgements This work was supported by the China 863 program (SS2015AA020109), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13040400), the startup grant from the Jilin University, the grants from the National Natural Science Foundation of China (No. 31560316, No. 31260273, No. 61462047 and No. 61261027), and the Department of Education of JiangXi Province (GJJ13641). Constructive comments from the anonymous reviewers are greatly appreciated.

References

1. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
2. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
3. Fan XN, Zhang SW (2015) IncRNA-MFDL: identification of human long non-coding RNAs by fusing multiple features and using deep learning. Mol BioSyst 11:892–897
4. Liu F et al (2015) De novo identification of replication-timing domains in the human genome by deep learning. Bioinformatics
5. Heffernan R et al (2015) Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep 5:11476
6. Spencer M, Eickholt J, Jianlin C (2015) A deep learning network approach to ab initio protein secondary structure prediction. IEEE/ACM Trans Comput Biol Bioinf 12:103–112
7. Lusci A, Pollastri G, Baldi P (2013) Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. J Chem Inf Model 53:1563–1575
8. Nguyen SP, Shang Y, Xu D (2014) DL-PRO: A novel deep learning method for protein model quality assessment. In: Proceeding of International Joint Conference on Neural Networks, pp. 2071–2078
9. Zhang S et al (2015) A deep learning framework for modeling structural features of RNA-binding protein targets. Nucleic Acids Res 44(4):e32
10. Chakraborty C et al (2016) Rising strengths Hong Kong SAR in bioinformatics. Interdiscip Sci Comput Life Sci 1–13
11. Liu GP et al (2014) Deep learning based syndrome diagnosis of chronic gastritis. Comput Math Methods Med 2014:938350
12. Wang P et al (2016) hMuLab: a biomedical hybrid MUlti-LABel classifier based on multiple linear regression. IEEE/ACM Trans Comput Biol Bioinf
13. Lipton ZC et al (2015) Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511.03677
14. Dudani SA (1976) The distance-weighted k-nearest-neighbor rule. IEEE Trans Syst Man Cybern SMC-6 325–327
15. Tsoumakas G, Katakis I, Vlahavas I (2011) Random k-labelsets for multilabel classification. IEEE Trans Knowl Data Eng 23(7):1079–1089
16. Tsoumakas G et al (2011) Mulan: a java library for multi-label learning. J Mach Learn Res 12:2411–2414
17. Tsoumakas G et al (2011) MULAN: a java library for multi-label learning. J Mach Learn Res 12:2411–2414
18. Zhang ML, Zhou ZH (2007) ML-KNN: a lazy learning approach to multi-label learning. Pattern Recogn 40:2038–2048
19. Read J et al (2011) Classifier chains for multi-label classification. Mach Learn 85:333–359
20. Tsoumakas G et al (2008) Effective and efficient multilabel classification in domains with large number of labels. In: Proceedings of ECML/PKDD 2008 workshop on mining multidimensional data. pp 1–15
21. Kocev D et al (2007) Ensembles of multi-objective decision trees. In Kok JN et al (eds) Proceedings of machine learning: ECML 2007: 18th European conference on machine learning, Warsaw, Poland, September 17–21, 2007. Springer, Berlin, Heidelberg, pp. 624–631
22. Madjarov G et al (2012) An extensive experimental comparison of methods for multi-label learning. Pattern Recogn 45: 3084–3104
23. Min-Ling Z, Zhi-Hua Z (2006) Multilabel neural networks with applications to functional genomics and text categorization. IEEE Trans Knowl Data Eng 18(10):1338–1351
24. Jesse Read FP-C (2014) Deep learning for multi-label classification. arXiv eprint arXiv:1502.05988
25. Read J, Hollmen J (2014) A deep interpretation of classifier chains. In: Blockeel H, van Leeuwen M, Vinciotti V (eds) Advances in Intelligent Data Analysis XIII. Springer, Switzerland, pp 251–262