Deep Neural Network on classification of Twitter Buzz Data

Ning LI

ningli30@uw.edu

Department of Statistics, University of Washington

UNIVERSITY of WASHINGTON

Introduction

In this study, we are going to classify binary case(Buzz/NonBuzz) from Online Social Media benchmark data[1] by neural network. This data set has serious imbalance property, Among these 10,000 instances in this dataset, only 856 Buzz data points (0.856%), which is much fewer than the number of non-Buzz Data.

Model and Methods

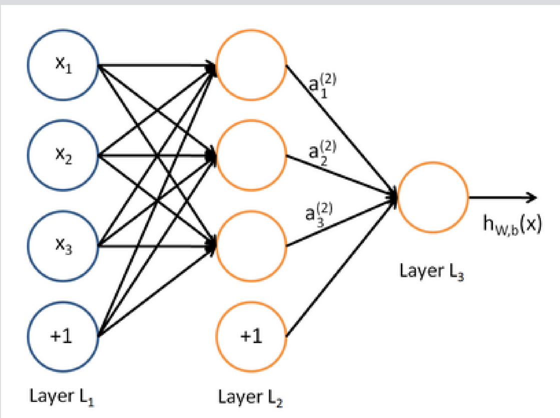

Figure 1: shallow neural network

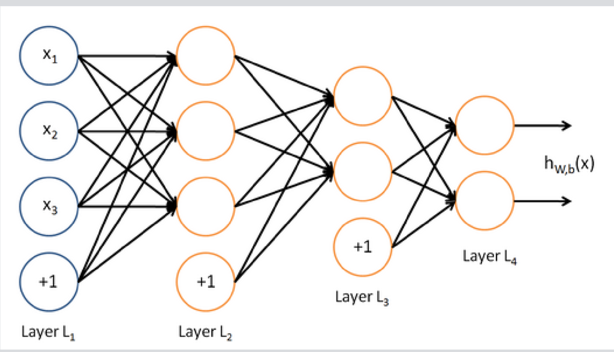

Figure 2: deep neural network

Because of imbalance property in our data, we use True Positive rate (TP), which is also called "recall", and False Positive rate(FP) to measure the performance of a classifier, instead of commonly used L_{01}

Conclusion

Based on experiments results, we find that deep neural network performs well than neural network. Dropout strategy and improvement in stochastic gradient descent solve this imbalanced data problem without resampling and pretraining.

In fact, deep neural network performs significantly well in high dimension data, such as image recognition field. Our study data contains only 77 dimensions, which can be considered as "low" dimension data. Therefore, the advantage of deep neural network cannot be seen clearly. But, based on the comparison of shallow neural network and deep neural network, we can still tell the deep neural network can give us more robust classification results.

References

[1] François Kawala, Ahlame Douzal-Chouakria, Eric Gaussier, Eustache Dimert, et al. Prédictions d'activité dans les réseaux sociaux en ligne. In *Actes de la Conférence sur les Modèles et lâĂš Analyse des Réseaux: Approches Mathématiques et Informatique (MARAMI)*, 2013.

[2] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[3] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Pascal Lamblin. Exploring strategies for training deep neural networks. *The Journal of Machine Learning Research*, 10:1–40, 2009.

Acknowledgements

I'd like to express my special thanks of gratitude to Prof. Marina Meila for nice instructions,and Haoran Cai who kindly offer computation resources and academic discussion.

Improvements

- **dropout**[2] is randomly temporarily dropping a unit out during training with a fixed probability p independent of other units.
- **Rectified linear function(ReL)**[3] $f(x) = max(0, x)$ replace traditional activation functions, like sigmoid functions or thanh functions
- **Minibatch and Momentum** improve Stochastic Gradient Descent(SGD) efficiency and avoid over-fitting

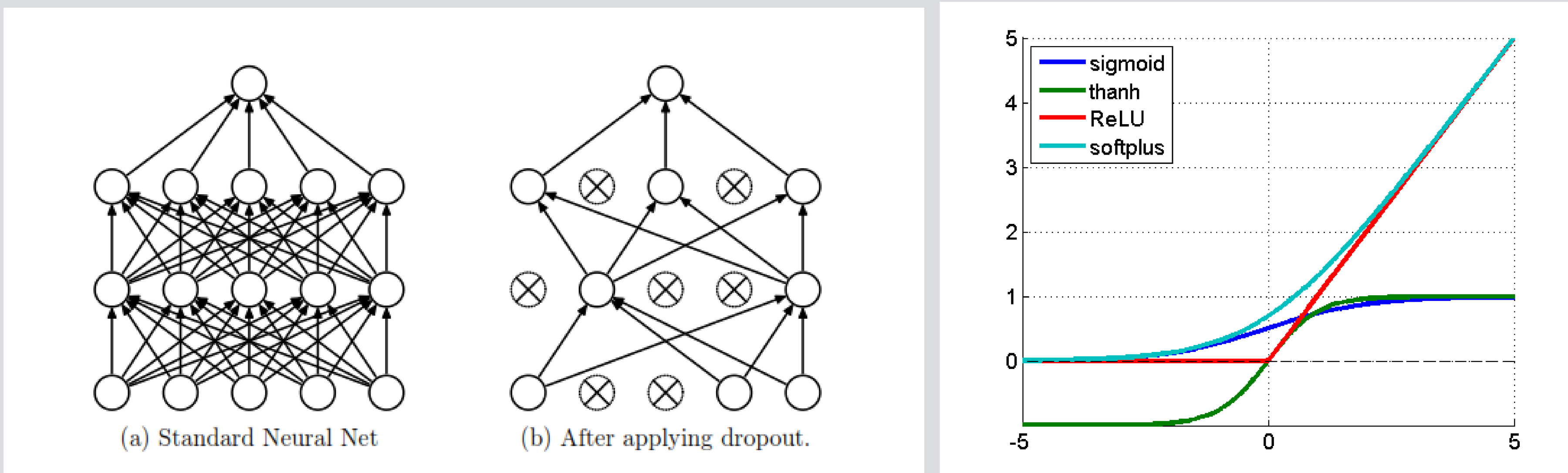

Figure 3: ROC curves for shallow and deep neural network (left); probability of being classified as 1 comparison (right)

Empirical Results

we fix the former 80% of twitter 10,0000 data as training data with 669 Buzz data(approximately 8.36%) , and the left 20 % of the data as validation data with 187 buzz records(approximately 9.35 %). Hence, the buzz data distributions for training and validation are almost as same as that for original data.

In this project, we conducted 11 experiments (9 deep models and 2 shallow models), classifying validation data by these models, and all of these results are listed in Table[1].

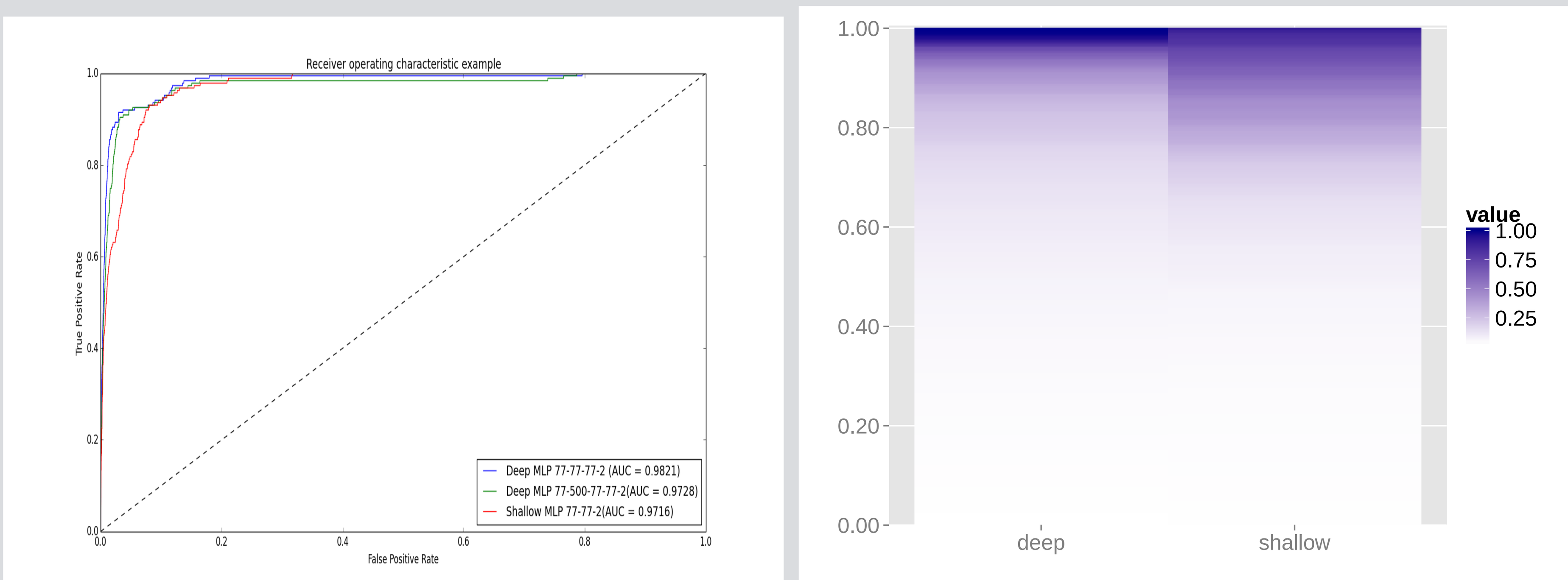

Figure 4: ROC curves for shallow and deep neural network (left); probability of being classified as 1 comparison (right)

Table[1] shows all the AUC values calculated from fitting validation data in fitted models. Among 9 deep neural network, the highest AUC value for deep neural network model is 0.9821. While among the two shallow neural network models,the largest AUC value on validation data, which is 0.9716.

Figure[2] shows us that the two deep neural network models (green line and blue line) increase faster than shallow neural network (red line) when the false positive rate is small, which means deep model tends to give us higher recall(which can be measured by true positive rate) when false positive rate is fixed. In other words, deep model can return extreme probabilities for classification. Figure[2] shows the last 500 largest points with probability of being classified as 1. The darker color is, the higher probability will be. As we can see, deep neural network tends to return extreme probabilities for each point, which convince us to get clear boundaries for Twitter Buzz data from deep neural model.

Layer structure (input-hidden-output)	Batch size	number of epoch	Learning rate	Drop-out	Momentum*	Activation function	AUC
77-77-77-2	100	100	0.1	0.5	0.5-0.99	rectifier	0.9821
77-77-77-2	10	100	0.1	0.5	0.5-0.99	rectifier	0.9735
77-77-77-2	100	100	0.1	0.5	0.9-0.99	rectifier	0.9766
77-77-77-2	100	100	0.1	0.5(drop out input)	0.5-0.99	rectifier	0.9799
77-77-77-2	100	100	1	0.5	0.5-0.99	rectifier	0.9789
77-500-77-2	100	100	0.1	0.5	0.5-0.99	rectifier	0.9774
77-77-11-2	100	100	0.1	0.5	0.5-0.99	rectifier	0.9806
77-500-77-77-2	100	100	0.1	0.5	0.5-0.99	rectifier	0.9728
77-77-77-77-2	100	100	0.1	0.5	0.5-0.99	rectifier	0.9816
77-11-2	100	100	0.1	-	-	sigmoid	0.9716
77-77-2	100	100	0.1	-	-	sigmoid	0.9652

Table 1: Experiment results for shallow and deep neural network