Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 5 Issue 10, October - 2018

A Study of the Effect of Dropout on Imbalanced Data Classification using Deep Neural Networks

Rukshima Dabare, Kok Wai Wong, Polychronis Koutsakis, Mohd Fairuz Shiratuddin
School of Engineering and Information Technology
Murdoch University
Western Australia 6150, Australia
{rukshima.dabare,k.wong, p.koutsakis f.shiratuddin}@murdoch.edu.au

***Abstract*— Learning from real-world data and classifying them according to their characteristics is an active area of research in machine learning. However, the imbalanced nature of the real-world data is still a challenging task whether it is a binary class or a multi-class problem. In recent years, the use of Deep Neural Network (DNN) for classification has been very popular among different fields. Although DNN is suitable to handle complex classification tasks, co-adaptation among the units of the DNN model leads to overfitting. Therefore, various regularization techniques have been introduced to reduce overfitting and improve generalization. In this paper, a study is conducted to investigate the impact of the Dropout rates of the DNN with imbalanced data. To our best knowledge, there is not much prior study on this. In this study, the imbalanced data is balanced using three popular techniques, and thereafter the Dropout regularization technique is applied. With the use of the evaluation metric ROC AUC (Receiver Operating Characteristic Area Under Curve), this study has identified the impact of the dropout rates for imbalanced data classification using DNNs.**

Keywords*— *Deep Neural Networks, Dropout, Imbalanced data

I. INTRODUCTION

Many real-world data by nature is imbalanced; meaning that the number of objects for each category (class) will not be the same. This imbalanced nature of real-world data needs to be considered when applying classifiers. Otherwise, the classifiers will be biased to the majority class. Therefore, it is a practice in the machine learning community to apply balancing techniques to handle it. Data level, algorithm-level, and hybrid methods are the three types of imbalanced data handling approach heavily used in this regard [1]. While the data level methods will be focused on modifying the data sets so that it will be suitable for the training using a standard classification algorithm, algorithm-level methods will be focused on modifying the algorithm so that the cost involved in misclassifying is reduced.

In this study, since we need to identify the impact on the original algorithm, we will be focusing on data-level balancing techniques in this paper such as oversampling, under sampling and hybrid techniques. Any classification task can either be a binary or a multi-class problem. In this paper, we will be focusing only on binary classification.

Binary classification is common in most real-life applications in different fields. In computer vision, binary classification task can be used to classify defected items from a manufacturing plant; in medicine, it can be used to identify a sick person from one's health records, and to identify a spam tweet in social network security [2] [3] [4]. For the classification tasks, the use of Deep Neural Network (DNN) is a popular supervised approach. However, only a limited number of studies have focused on imbalanced data classification using DNN. When DNN algorithms are used for classification, they are prone to overfit which could lead to poor generalization. Hence, regularization techniques like Dropout [5], Dropconnect [6], and Binary connect [7] can be used to reduce overfitting. In almost all these methods, different hyperparameters need to be adjusted according to the problem that needs to be solved.

The work in [5] introduces the term "Dropout" referring to dropping out units (hidden and visible) in a neural network, and they also interpret Dropout as a way of regularizing the model by the addition of noise to its hidden units. In their work, they state that the choice of which units to drop is random. Further, in their study, they suggested values for different hyperparameters involved with the use of Dropout. However, all the experiments and suggestions in their study are based on balanced datasets and have not focused on imbalanced datasets. The focus of this paper is to study the impact of the Dropout rates have in the DNN when handling imbalanced data.

The rest of the paper is organized as follows. Section two gives an overview of the imbalanced learning domain, an overview of the deep learning and regularization in general. In section three, we present the experimental setup, evaluation matrices used, datasets and the results of our comparison study with a detailed analysis of the results. The final section concludes the paper.

II. Learning from Imbalanced Data

A. *Imbalanced learning domain*

Learning and classifying from imbalanced datasets have been extensively studied in the literature because real-world data is on most occasions imbalanced in nature. Therefore traditional classification algorithms are biased towards the majority class [1]. According to [8] these traditional classification algorithms are built to minimize the overall error and to maximize the classification accuracy. Therefore, different imbalance-handling techniques have been introduced in these studies to classify from the learned model based on mainly three approaches, Data-level, Algorithm level and Hybrid methods [1]. Data level methods comprise methods used to alter the training datasets such as oversampling, under sampling and hybrid techniques. Some commonly used techniques are Random oversampling, Synthetic Minority Over-Sampling (SMOTE) [9] and its variations, Random under sampling, Tomek links [10] and its variations, Synthetic Minority Over-Sampling combined with Tomek links (SMOTETomek) and Complementary Fuzzy Support Vector Machine (CMTFSVM) combined with SMOTE [11]. However, some of these methods suffer from various drawbacks, e.g., adopting an oversampling-only method might introduce overfitting and adopting an under sampling-only method might introduce a loss of important information needed for classification [12]. Also, some of these methods have not been tried with DNN algorithms.

B. *A brief overview of classifiers and Deep Learning*

Classification tasks can be of binary or multi-class. Different machine learning techniques can be applied for classification tasks whether binary or multiclass, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), k-Nearest Neighbour (k-NN) and Naïve Bayes. According to [13], conventional machine-learning techniques suffer from various disadvantages, and therefore, Deep learning can be considered a major breakthrough technique. There are many deep learning models introduced like Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Deep Belief Networks (DBFs), Restricted Boltzmann Machine (RBM), Recurrent Neural Networks (RNNs) and many more.

In recent years, DNN models have been introduced for different classification tasks in many fields such as Medicine [3], Engineering [2,14], Vision [15], Speech recognition [16], Network security [17] and many more. According to the study of [12], although the application of DNN in classification is well studied and applied, only a few works in the literature focus on a systematic study on how to handle imbalanced data when using DNN for classification.

The work in [18] on training DNNs on imbalanced datasets discusses the need for the consideration of imbalanced nature in real data and the necessity to rethink about applying traditional DNN models to imbalanced data. The authors further introduce a novel loss function to handle imbalanced data when using DNNs. In [19], the proposed approach is similar to [18], as the authors introduce a new Algorithm-level change to CNN. In [20] a cost-sensitive Deep CNN which can be used in imbalanced data classification was proposed. The work in [21] proposes a novel data-level approach to handle imbalanced data using CNN. However, all these methods focus on introducing novel approaches to handle imbalance data whereas in our study we have concentrated on identifying the effect of the hyper-parameter Dropout rate on imbalanced data. This will help us understand more on how the dropout rate impact on generalization in imbalanced datasets, given that oversampling and undersampling may affect the generalization ability of the DNN.

III. Experimental Setup, Results and Result Evaluation

A. *Datasets and Experimental Setup*

In this study, we used imbalanced datasets with numerical attribute values where the classification task is binary. Both of the datasets used are popular in the imbalanced data literature and are used for performance evaluation in many recent studies such as the work described in [11], [22] and [23]. The Yeast3 dataset is from KEEL [24], and the ADULT datasets are from the UCI [25]. Table I gives the information on the datasets used for this study.

For the purpose of analysis, we used three balancing techniques for each dataset: oversampling, under sampling and oversampling followed by under sampling. Therefore, the balancing methods used are SMOTE, TOMEK links and SMOTE with Tomek links respectively. We used 10-fold cross-validation for each experiment, and the average of the 10 experiments is used for evaluation. In each experiment, 20% of the dataset was randomly dedicated for testing, and the rest was used for training.

TABLE I. Details of the imbalanced datasets

Characteristics	**Dataset**	
	Yeast3	***ADULT***
Imbalanced Ratio (IR)	8.1	3.33
No of Instances	1484	45222
No of Majority Instances	1321	34014
No of Minority Instances	163	11028
No of Attributes	8	15
No of Training Instances	1187	36178
No of Testing Instances	297	9044

The classifier used for this study is a DNN with two hidden layers following what has been reported at [26]. The number of nodes in the hidden layers was increased by 0.25% in the first layer and then decreased by 0.25% in the second layer for both datasets. For other than the final layer, all other layers used the Rectified Linear Unit (ReLU) as the activation function and the last layer used the Sigmoid activation function. The DNN model used the loss function binary cross entropy and used a batch size of 10 for each training with 150 epochs, as by increasing the number of epochs for this case did not improve the accuracies.

B. *Results and Discussions*

The results of the experiments are produced with the use of the evaluation metric ROC AUC (Receiver Operating Characteristic Area Under Curve) [27] which is one of the most acceptable and popular evaluation matrices for imbalanced data classification. If the classification accuracy in terms of precision/accuracy is used, the classification accuracy will be biased to the majority class, therefore, will result in high accuracy level, which is not what is expected from an imbalanced classification model.

The results of the experiments conducted using the DNN classifier with and without the application of the balancing techniques are presented in Table II. As mentioned in the previous section, the DNN used is a network with two hidden layers. The accuracy without the application of the Dropout and with the Dropout of 0.5 for each layer is given in Table II with the Imbalance ratio of each dataset and their balancing Technique (BT). The Dropout rate of 0.5 is chosen to start off with as it is the recommended conventional values used for most applications. [5]

As observed from the results in Table II, the use of Dropout regularization shows to provide better results for imbalanced datasets in most cases when the imbalanced ratio is high. Further, as we wanted to investigate the effect of the Dropout rate for imbalanced datasets, we carried out experiments for different combinations of Dropout rates for each layer of the DNN. Here we used the three balancing techniques mentioned in the earlier section, SMOTE, Tomek Links and SMOTE followed by Tomek Links. Table III and Table IV show the classification performance in terms of AUC of the trained DNN model with varying values of Dropout rates for each layer of the DNN for the datasets Yeast3 and ADULT respectively. From the results in Tables III and IV, the AUC is improved after applying lower values of Dropout rates such as 0.2 for input layer and for hidden layer. This is due to the reduction of coadoption among neurons of each layer with the use of a lower Dropout rate.

From the results of Tables III and IV, it can be observed that the identified balancing technique, SMOTE+Tomek performed well for the DNN. A possible explanation for this is that the combination of both the methods will decrease the possibility of overfitting and also reduce the chance of losing information. Therefore, further experiments were carried out to identify the effect of Dropout using this balancing technique. The relevant results are tabulated in Table V. It is clear from these results that a lower value of Dropout rate would be able to give better AUC in classification.

Model accuracy curve and model loss curve given in Fig.1 and Fig. 2, depicts the accuracy against epochs when Dropout rate is 0.2 and 0.3 for each hidden layer for the Yeast3 dataset after the use of SMOTE + Tomek balancing technique to handle imbalance. Model accuracy curve and model loss curve given in Fig. 3 and Fig.4, depicts the accuracy against epochs when Dropout rate is 0.2 and 0.2 for each hidden layer for the ADULT dataset after the use of SMOTE + Tomek balancing technique to handle imbalance.

TABLE II. EXPERIMENTAL RESULTS WITH THE USE OF DROPOUT AND WITHOUT DROPOUT

Dataset	**IR**	**BT**	**AUC**	
			Without Dropout	***Dropout (0.5)***
ADULT	8.1	No Balance	0.7609	0.7630
		SMOTTomek	0.8173	**0.8194**
		SMOTE	0.8214	0.8192
		Tomek links	0.7910	0.7917
Yeast3	3.33	No Balance	0.8547	0.8423
		SMOTTomek	0.9206	0.9282
		SMOTE	0.9186	**0.9297**
		Tomek links	0.8831	0.8473

TABLE III. EXPERIMENTAL RESULTS – YEAST3 DATASET

BT	**Dropout Rate**			
	0 & 0	***0.2 & 0.2***	***0.5 & 0.5***	***0.2 & 0.5***
SMOTE + Tomek	0.9206	**0.9360**	0.9282	0.9329
SMOTE	0.9186	**0.9347**	0.9297	0.9333
Tomek links	0.8831	**0.8936**	0.8473	0.8469

TABLE IV. EXPERIMENTAL RESULTS - ADULT DATASET

BT	**Dropout Rate**			
	0 & 0	***0.2 & 0.2***	***0.5 & 0.5***	***0.2 & 0.5***
SMOTE + Tomek	0.8173	**0.8222**	0.8194	0.8197
SMOTE	0.8201	**0.8204**	0.8191	0.8202
Tomek links	0.7910	0.7837	**0.7917**	0.7910

TABLE V. EXPERIMENTAL RESULTS – SMOTE+TOMEK

Dataset	**Dropout Rate for input and hidden layers**							
	0 & 0.2	**0.2 & 0**	**0 & 0.3**	**0.3 & 0**	**0.2 & 0.3**	**0.3 & 0.2**	**0.2 & 0.4**	**0.4 & 0.2**
Yeast3	0.9298	0.9361	0.9348	0.9343	**0.9374**	0.9301	0.9374	0.8905
ADULT	0.8198	0.8170	0.8176	0.8179	0.8197	0.8196	0.8181	0.8182

From the test accuracy curves given in below figures, it is clear that the learning model does not overfit the model but learns and therefore achieves good generalization. These plots are used only for the purpose to identify that the models do not overfit considering the imbalanced nature of the dataset.

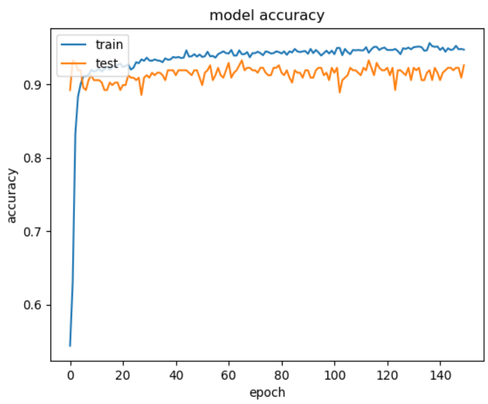

Fig. 1. Model Accuracy Model Loss when SMOTE + Tomek links are applied for balancing and dropout rates 0.2 & 0.3 applied for each hidden layer of the DNN for Yeast3 dataset

Fig. 2. Model Loss when SMOTE + Tomek links are applied for balancing and dropout rates 0.2 & 0.3 applied for each hidden layer of the DNN for Yeast3 dataset

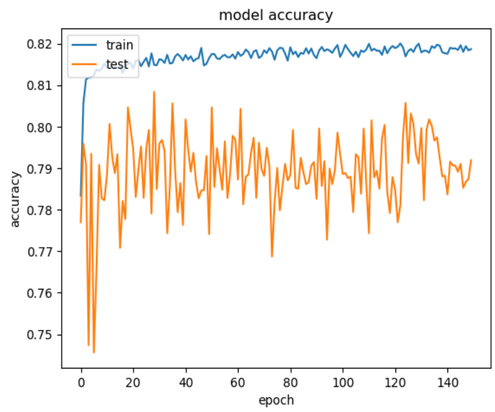

Fig. 3. Model Accuracy when SMOTE + Tomek links are applied for balancing and dropout rates 0.2 & 0.2 applied for each hidden layer of the DNN for the ADULT dataset

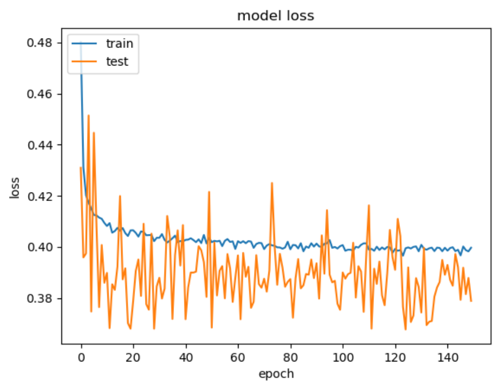

Fig. 4. Model Loss when SMOTE + Tomek links are applied for balancing and dropout rates 0.2 & 0.2 applied for each hidden layer of the DNN for the ADULT dataset

According to [5], the whole idea of Dropout is to minimize the co-adoption of all neuron units. As many studies show, including our own, Dropout is beneficial for Deep learning models however the Dropout rate for each layer is also important. So, for imbalanced datasets, the idea of dropping units will work only to a certain level. Otherwise, while dropping units, connections will also drop. Then, the learning model will miss important co-adaptations that are necessary for its learning.

IV. Conclusion

Although DNN and the use of Dropout for regularization is studied thoroughly in the machine learning community, the imbalanced nature of the data has not been a subject of most of the relevant literature. In our study, we provided the investigation of the impact of Dropout rate applied for each layer in a binary classification task for imbalanced datasets using a DNN model. The study presented in this paper has provided insight into the impact of the Dropout rates for the hidden layers when dealing with imbalanced datasets. This has also suggested that the use of Dropout rates has a role to play in dealing with the possible overfitting and the possible loss of important information when using oversampling and undersampling for imbalanced data as reported in [12].

References

[1] Krawczyk, B.: Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence 5(4), 221-232 (2016). doi:10.1007/s13748-016-0094-0

[2] Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D.: Deep learning for smart manufacturing: Methods and applications. Journal of Manufacturing Systems (2018).

[3] Greenspan, H., van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 35(5), 1153-1159 (2016).

[4] Wu, T., Liu, S., Zhang, J., Xiang, Y.: Twitter spam detection based on deep learning. In: Proceedings of the Australasian Computer Science Week Multiconference 2017, p. 3. ACM

[5] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.: Dropout: A simple way to prevent neural networks from ovrfitting. The Journal of Machine Learning Research 15, 1929-1958 (2014).

[6] Wan, L., Zeiler, M., Zhang, S., Le Cun, Y., Fergus, R.: Regularization of neural networks using dropconnect. In: International Conference on Machine Learning 2013, pp. 1058-1066

[7] Courbariaux, M., Bengio, Y., David, J.-P.: Binaryconnect: Training deep neural networks with binary weights during propagations. In: Advances in neural information processing systems 2015, pp. 3123-3131

[8] Ali, A., Shamsuddin, S.M., Ralescu, A.L.: Classification with class imbalance problem: a review. Int. J. Advance Soft Compu. Appl 7(3) (2015).

[9] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research 16, 321-357 (2002).

[10] Tomek, I.: Two modifications of CNN. IEEE Trans. Systems, Man and Cybernetics 6, 769-772 (1976).

[11] Pruengkarn, R., Wong, K.W., Fung, C.C.: Imbalanced data classification using complementary fuzzy support vector machine techniques and SMOTE. In: Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on 2017, pp. 978-983. IEEE

[12] Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. arXiv preprint arXiv:1710.05381 (2017).

[13] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436 (2015).

[14] Jia, F., Lei, Y., Lin, J., Zhou, X., Lu, N.: Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems and Signal Processing 72-73, 303-315 (2016). doi:10.1016/j.ymssp.2015.10.025

[15] Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S.: Deep learning for visual understanding: A review. Neurocomputing 187, 27-48 (2016).

[16] Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H.G., Ogata, T.: Audio-visual speech recognition using deep learning. Applied Intelligence 42(4), 722-737 (2015).

[17] Kwon, D., Kim, H., Kim, J., Suh, S.C., Kim, I., Kim, K.J.: A survey of deep learning-based network anomaly detection. Cluster Computing, 1-13 (2017).

[18] Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J.: Training deep neural networks on imbalanced data sets. In: Neural Networks (IJCNN), 2016 International Joint Conference on 2016, pp. 4368-4374. IEEE

[19] Huang, C., Li, Y., Change Loy, C., Tang, X.: Learning deep representation for imbalanced classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, pp. 5375-5384

[20] Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R.: Cost-Sensitive learning of deep feature representations from imbalanced data. IEEE transactions on neural networks and learning systems (2017).

[21] Guan, S., Chen, M., Ha, H.-Y., Chen, S.-C., Shyu, M.-L., Zhang, C.: Deep Learning with MCA-based Instance Selection and Bootstrapping for Imbalanced Data Classification. 288-295 (2015). doi:10.1109/cic.2015.40

[22] Yang, Y., Xiao, P., Cheng, Y., Liu, W., Huang, Z.: Ensemble Strategy for Hard Classifying Samples in Class-Imbalanced Data Set. In: Big Data and Smart Computing (BigComp), 2018 IEEE International Conference on 2018, pp. 170-175. IEEE

[23] Awasare, V.K., Gupta, S.: Classification of imbalanced datasets using partition method and support vector machine. In: Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on 2017, pp. 1-7. IEEE

[24] Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic & Soft Computing 17 (2011).

[25] Dua, D.a.K.T., E.: UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. In. (2017)

[26] Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practices for convolutional neural networks applied to visual document analysis. In: null 2003, p. 958. IEEE

[27] Fawcett, T.: An introduction to ROC analysis. Pattern recognition letters 27(8), 861-874 (2006).