

Analytic Approach: Jointly learns the noise statistics and better feature representations of a given sentence.

INTRODUCTION

Deep networks have made **huge gains in performance** over traditional methods on large datasets with very clean labels. However **large real-world datasets often contain label errors**. It is crucial to be able to train deep neural networks in the presence of label noise.

Our approach introduces a non-linear processing layer (noise model) that models the statistics of the label noise into a Deep neural network (DNN) architecture.

OBJECTIVES

- To train **deep neural networks** that are **robust to label noise**.
- To study the **effect of different initialization and regularization** on noise model and **different batch sizes** when training with noisy labels.
- To study the effect of **different types of label noise; Uniform label flipping (Uni), and Random label flipping (Rand)**. Where a clean label is swapped with another label from the given number of labels uniformly or randomly.

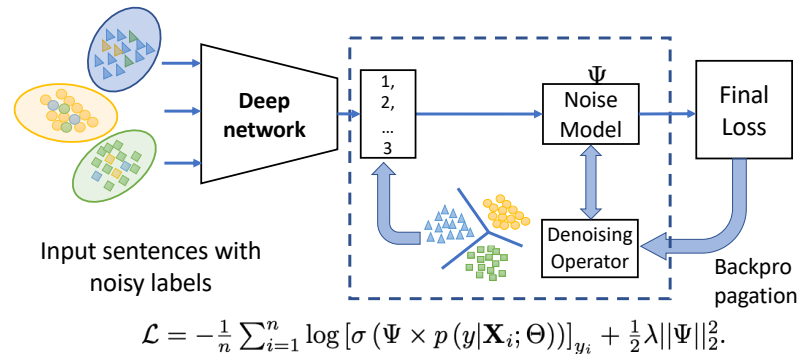
t-SNE Embedding for Trec dataset

DATASETS

Summary of text classification datasets used; K: denotes the number of classes, L: represents the average length of sentence, N: denotes the number of training samples, T: represents the number of test samples, Type: describes whether the dataset is balanced.

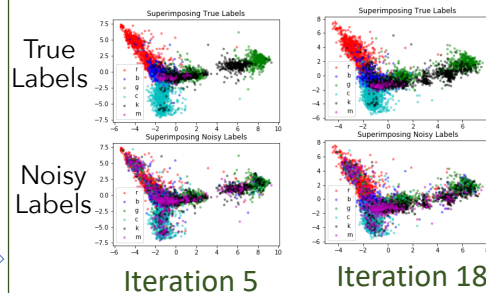
Text Data	Dataset	K	L	N	T	Type
	SST-2	2	19	76961	1821	Balanced
	Trec	6	10	5000	500	Not Balanced
	AG-News	4	110K	10K	Balanced	
	DbPedia	14	29	504K	70K	Balanced

LABEL NOISE ROBUST DEEP NETWORK

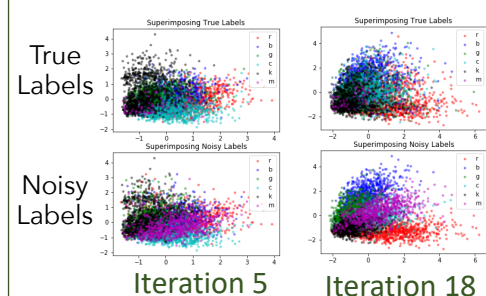


NUMERICAL RESULTS AND ANALYSIS

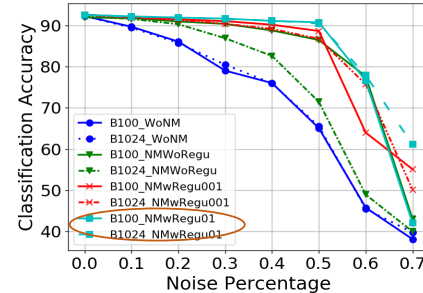
Proposed Method



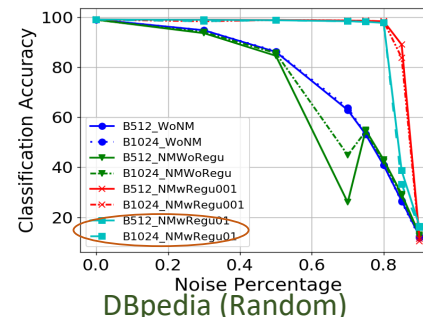
Base Model Alone



Effect of Batch Size



AG-News (Random)



Review Sentence

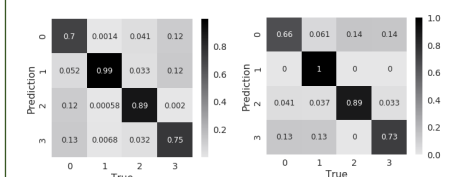
Label

Given	This laptop should be sold with SSD. HDD makes this laptop very slow. In contrast of that, this laptop's CPU is very powerful. You need to add or replace SSD to this laptop if you want to get real potential of this laptop.	Entertainment
Prediction	This laptop should be sold with SSD. HDD makes this laptop very slow. In contrast of that, this laptop's CPU is very powerful. You need to add or replace SSD to this laptop if you want to get real potential of this laptop.	Electronics

WoNM: Standard Deep Learning Architecture
NMwRegu: Noise model stacked on top of Std. DNN but no regularization applied.
NMwReguXX: Noise model stacked on top of Std. DNN and XX L2 regularization applied.

LEARNED NOISE MODEL

The learned noise model is able to capture the input label noise statistics and is highly correlated to the input noise distribution with Pearson Correlation Coefficient 0.988.



True label noise

Learned label noise

CONCLUSIONS

The proposed framework enables a Deep Neural Network to learn better sentence representations in the presence of label noise for text classification tasks. With proper initialization and regularization, the noise model is able to absorb most of the label noise and helps the base model to learn better sentence representations.

REFERENCES

- Ishan Jindal, Matthew Nokleby, and Xuewen Chen. 2016. Learning deep networks from noisy labels with dropout regularization. In Data Mining (ICDM), 2016 IEEE 16th International Conference on, pages 967-972. IEEE.
- Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. 2017. Making deep neural networks robust to label noise: A loss correction approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1944-1952.