Label Noise Detection System Against Label Flipping Attack

Ning Cheng

Zhengzhou University

Hongpo Zhang (✉ zhp@zzu.edu.cn)

Zhengzhou University <https://orcid.org/0000-0003-3485-8470>

Zhanbo Li

Zhengzhou University

Research Article

Keywords: lass noise detection, Machine learning, AdaBoost algorithm, Semi-supervised

Posted Date: April 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-176698/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The label flipping attack is a special poisoning attack in the adversarial environment. By adding noise to the data, it destroys the learning process of the model and affects the decision-making performance of the model. Recent literature work uses semi-supervised learning techniques to defend against label flipping attacks. However, these methods require a clean dataset to achieve their goals. This study proposes a novel label noise processing framework to correct the labels of contaminated samples in the data preprocessing stage. Based on five real UCI datasets, we evaluated the effectiveness of the semi-supervised defensive label noise correction algorithm based on the AdaBoost algorithm(AdaSSL). With a noise ratio of 0~20%, we compared the classification performance of six classic machine learning algorithms (NB, LR, SVM, DT, KNN and MLP) under the AdaSSL defense algorithm. The results show that compared with the most advanced semi-supervised defense algorithms in the literature, our algorithm does not need to use an additional dataset. At a noise ratio of 10%, the AdaSSL algorithm is significantly better than state of the art label noise defense technology.

Full Text

This preprint is available for [download as a PDF](#).

Figures

Figure 1

Label noise detection framework

Figure 2

Accuracy results on UCI datasets

Figure 3

F1-score results on UCI datasets

Figure 4

AUC results on UCI datasets