Active Learning from Unreliable Data
Zilong Zhao, Sophie Cerf, Robert Birke, Bogdan Robu, Sara Bouchenak,
Sonia Ben Mokhtar, Lydia Chen

To cite this version:
Zilong Zhao, Sophie Cerf, Robert Birke, Bogdan Robu, Sara Bouchenak, et al.. Active Learning from Unreliable Data. European Conference on Computer Systems (EuroSys) - 13th EuroSys Doctoral Workshop (EuroDW 2019), Mar 2019, Dresde, Germany. <hal-02045455>

HAL Id: hal-02045455
http://hal.univ-grenoble-alpes.fr/hal-02045455
Submitted on 22 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Active Learning from Unreliable Data

Zilong Zhao*, Sophie Cerf*, Robert Birke†, Bogdan Robu*, Sara Bouchenak‡, Sonia Ben Mokhtar‡, Lydia Y. Chen§

*Univ. Grenoble Alpes, France
†ABB Research, Switzerland
‡INSA Lyon, France
§TU Delft, Netherlands

{zilong.zhao, sophie.cerf, bogdan.robu}@gipsa-lab.fr
{robert.birke}@ch.abb.com
{sara.bouchenak, sonia.benmokhtar}@insa-lyon.fr
{y.chen-10}@tudelft.nl

Abstract—Classification algorithms have been widely adopted in big recommendation systems, e.g., products, images and advertisements, under the common assumption that the data source is clean, i.e., features and labels are correctly set. However, data collected from the field can be unreliable due to careless annotations or malicious data transformation. In our previous work, we proposed a two-layer learning framework for continuous learning in the presence of unreliable anomaly labels, it worked perfectly for two use cases, (i) detecting 10 classes of IoT attacks and (ii) predicting 4 classes of task failures of big data jobs. To continue this study, now we will challenge our framework with image dataset.

The first layer of quality model filters the suspicious data, where the second layer of classification model predicts data instance’s class. As we focus on the case of images, we will use widely studied datasets: MNIST, Cifar10, Cifar100 and ImageNet. Deep Neural Network (DNN) has demonstrated excellent performances in solving images classification problems, we will show that two collaborating DNN could construct a more robust and high accuracy model.

Index Terms—Unreliable Data; Images; Attacks; Machine Learning; Deep Neural Network

I. INTRODUCTION

A large amount of user-generated data powers up machine learning based applications in our daily life. Labeled data (especially for images) from search engine provides a fast and cheap way to build the large-scale dataset. But it also inevitably introduces some incorrect labels. For examples, if we want to construct an image dataset of "airplane", with keyword "airplane", Google Image can give us a very good first result. But we can also notice that there are not only airplane images, they have also images: airplane interior, airplane runway, airplane structure sketch and Cartoon airplane.

To solve this problem, a simple way to clean the data, it’s to find a domain expert to remove or relabel the suspect data in a preprocessing stage. However large-scale annotated datasets with high-quality label annotations are not always available for new domain, due to the significant time and efforts it takes, not to mention that for some online-tuning systems, the user-generated data can be infinite. There is no doubt that DNN is delivering superior results on image classification, but this success is highly tied to the availability of large-scale annotated datasets. So we can see a big contradiction here.

Standard machine learning algorithms typically assume clean labels and overlook the risk of noisy labels. But recent studies have shown that learning from high proportion noisy labels can significantly degrade the DNN’s classification accuracy [9].

Existing works of learning from noisy data can be roughly divided into two aspects: (1) filtering out noisy labels and learning only from the predicted clean data, (2) designing noise-aware classification algorithms. The first builds one or multiple filter models, e.g., SVM [8] to clean the data, and only the data instances that predicted label by one or more filters matches its original label, we believe this data instance is clean and can be used to train the classification model. The second type of approach can be summarized by several works, one proposed method is to correct noisy labels to their true labels via a clean label inference step. These methods assume the availability of a small clean dataset to be used [3], [10]. A different approach is to learn directly through noisy data, and in parallel, running a local intrinsic Dimensionality [2] measurement to monitor the stage of training process, to make sure at the end, the test accuracy stabilizes at its highest level [4].

While the recent state-of-the-art solution can have a very good result on noisy label issues, little focus has been given to the online setting with data instances having fluctuating noise levels. Our study presents the initial results on how to build a classifier by selectively and continuously learn from high quality data that leads to a strong classifier.

Figure 1 shows training and predicting process of our system. Training process is triggered by \( D_i \): \( i \)th training batch, the quality model will predict labels for each data instance, if the predicted label matches its original label, we believe this data instance is clean, and will pass it to train classification model, these "clean" data will also be used to train quality model itself, if the predicted label doesn’t match its original label, we will throw the data instance. From \( P_i \) to \( Y_i^{\mathcal{P}} \) represents predicting process, only classification model is involved in this process.
We study the impact of noisy data on anomaly classification detection. For standard classifiers, like KNN and nearest centroid, the detection accuracy decays faster than MLP that is more robust to the noisy labels. Such an observation holds for both use cases. In IoT attacks, MLP can even achieve a similar accuracy as the case of no label noises, when there is 50 percent of label classes are altered.

### References


