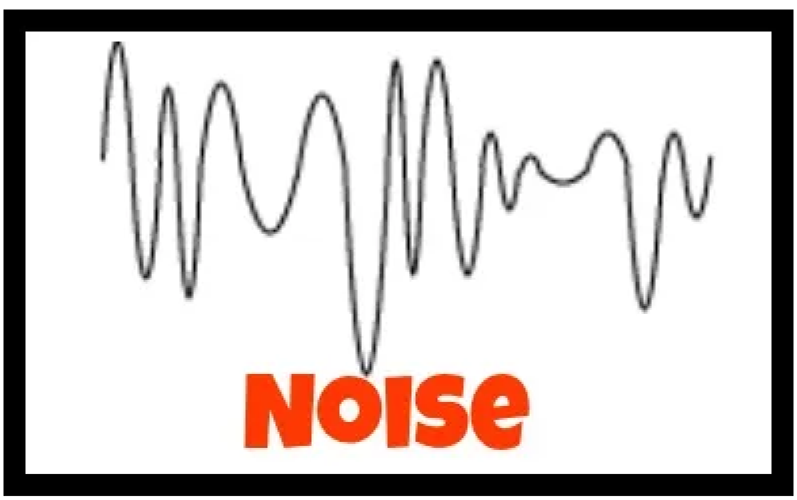
Noise Types & Their Effects on DL.

Mustafa Adel Ibrahim · Following
8 min read · Sep 23

A detailed analysis of the negative impacts of different types of label noise & the characteristics of each type.



Abstraction

It's the 1st work devoted to fully investigate the negative effects of all types of label noise on DNNs. **The contribution of this work can be summarized as follows:**

- This work proposes a generic framework to generate feature-dependent label noise to emphasize similarities among data instances by sparsely distributing them in the feature domain.
  This framework can be applied to any clean dataset to synthesize feature-dependent noisy labels.
- Label corruption algorithm that utilizes data similarities in feature domain by creating sparse representation of data (with synthetic feature-dependent label noise).

What is Label Noise ?

It refers to errors or inconsistencies in the assigned labels of data points in a dataset, which can occur due to human error, ambiguity, or subjective interpretation.
i.g., Imagine a group of students taking a multiple-choice test, and the teacher ***accidentally marks the wrong answers as correct or vice versa***.
This mislabeling of the answers represents label noise in the dataset, making it challenging for a ML algorithm to accurately learn from the data.

Noise can significantly impact the performance of a ML model, as it introduces incorrect information during the training process. The model may learn incorrect patterns or associations based on the mislabeled data, leading to reduced accuracy and reliability in making predictions.

Introduction.

The success of deep learning (DL) is mostly due to the availability of big datasets with clean annotations. ***But, gathering a cleanly annotated dataset is not always feasible.***
Numerous methods are proposed to train deep neural networks (DNNs) in

the presence of noisy labels to prevent performance degradation caused by noisy labels.

> - *Since each work generates a different kind of label noise, it's problematic to test and compare those algorithms fairly. As, there's no ground truth available.*

Therefore, commonly adopted methodology is to add synthetic label noise to the training set while keeping test set clean. Since there's no generic framework to corrupt labels of the given dataset in a systematic way to **mimic real-world noise**, each work adds their own characteristic label.

This results in subjective evaluation of algorithms and prevents fair comparison of the methods. Even though literature generally considers noisy labels phenomenon as one compact problem.

> ***By this approach, samples that are more likely to be mislabeled are detected from their softmax probabilities, and their labels are flipped to the corresponding class.***

DNNs have an impressive ability to generalize well; these powerful models have a great tendency to memorize even complete random noise. **Avoiding memorization** is an important challenge to be overcome in order to obtain representative NNs and it gets even more crucial in the presence of noise.

Types of Noise

> · ***Feature noise.***
> · ***Label noise.***

> Generally, label noise is considered to be more harmful than feature noise.

Label Noise.

> · *Uniform noise*
> · *Class-dependent noise*
> · *Feature-dependent noise*

> Each has its own characteristics and the effect on learning performance, therefore each of them should be evaluated under their own category.

However, *uniform label noise* **can easily be handled by NNs without an extra modification**. it can be misleading to draw conclusion about label noise by experimenting only with uniform label noise**.**

The main reason for this approach is the ability to generate corrupted datasets from toy datasets in purpose of quick application and testing of proposed algorithm. Yet, question of how to add synthetic noise in order to test noise robust algorithms stays to be an open question. For this purpose, most human-like noise type is feature-dependent noise, in which features of each instance effects probability of being mislabeled.

e.g., in **cars dataset**, some sport cars are more similar to classic cars than others. These specific instances have greater chance to be mislabeled by a human annotator. However, this relation among specific instances is not considered as **uniform or class-dependent noise**. This work proposes a feature-dependent label corruption algorithm,
that is inspired by *knowledge-distillation* technique.

Our methodology aims to learn representations which would result in sparse distribution of data instances in feature domain.
At the end, similarities among data samples are extracted and labels are flipped for uncertain samples. Additionally, we provide pre-generated noisy labels for commonly used datasets in purpose of other researchers to test their label noise robust algorithms.

Uniform label noise:

> *Flipping probability of label from its true class to any other class is equally distributed. meaning that each label has an equal probability of being mislabeled.*

e.g., say we've a dataset of images of cats VS dogs. So, each image is labeled either as "cat" or "dog". However, due to some error, a percentage of the labels are incorrect. In the case of uniform label noise, the mislabeled images are spread evenly across both classes, so there is an equal chance of a cat being labeled as a dog and vice versa.

To mitigate uniform label noise, various techniques can be employed, such as data cleaning or correction methods, using robust loss functions, or employing ensemble methods to combine multiple models and reduce the impact of mislabeled data.

Many works in literature use synthetic uniform label noise by just flipping labels randomly for a given percentage of data instances.

Class-dependent noise:

- It refers to the presence of noise in a dataset that is specific to certain classes. In other words, the noise in the data may have different characteristics or patterns for different classes. This is mostly represented by a “*confusion matrix”*

e.g., consider cats VS dogs dataset, class-dependent noise could be present if the images of cats have more variations or distortions compared to the images of dogs. This noise can make it more challenging to accurately analyze the data, as the noise may be more prominent in certain classes and may affect the performance of ML algorithms differently for each class.

Dealing with **class-dependent noise** requires careful consideration during data preprocessing and model training. Techniques such as data augmentation, noise filtering, or class-specific handling of noisy samples can be employed to mitigate the impact of class-dependent.

The easiest way is to attain inter-class transition probabilities just random; so that there’s still class dependence since transition probabilities are given according to classes but without any correlation to class similarities.
In a more structured way, noise transition matrix can be designed in a way that similar classes have a bigger probability to be flipped to each other. Some works use pairwise noise, in which transition from one class can only be defined to one another class.

• Feature-dependent noise:

The probability of mislabeling depends on features of instances. In order to generate feature-dependent noise, features of each instance should be extracted, and their similarities to other instances from different classes should be evaluated. Unlike uniform and class-dependent noise, there're much fewer implementations of synthetic feature-dependent label noise. One particular work in this field is [37], where data is clustered with the KNN algorithm, and labels are flipped randomly for clusters of data. This method provides concentrated noise in the feature space. But, this type of synthetic noise doesn't evaluate the instance similarities and therefore different from our proposed approach.
Alternatively, in case there is a surrounding text for each image in the dataset, some works create noisy labels from the interpretations of these texts [38]–[41], assuming surrounding texts are related to features of data. But this approach is restricted to datasets with surrounding user-defined texts, which is not the case for most of the time.
In this work, we focus on the closed set problem, in which all data instances are from the given class set. However, some works also investigated open set problem [42], where the dataset is polluted with instances that don't belong to any class from the class set.

It refers to the noise or variability in a dataset that is dependent on specific features or characteristics of the system being analyzed. In other words, the amount or nature of the noise present in the system or data may vary based on the specific features or attributes being considered.

i.e., in image processing, feature-dependent noise could refer to variations in noise levels or patterns that are specific to certain regions or structures within an image.

- For example, in a digital photograph, noise may be more pronounced in areas with low light or high contrast.
  Similarly, in speech recognition, feature-dependent noise could refer to variations in background noise that are dependent on the phonetic content or acoustic characteristics of the speech signal.

By considering the specific features that influence the noise, researchers and practitioners can develop more effective algorithms and techniques to mitigate or remove the noise and improve the accuracy or reliability of their analyses or models.

- In image processing, feature-dependent noise refers to the noise that is dependent on the specific features or characteristics of an image.
  - In audio processing, feature-dependent noise can occur when certain frequencies or tones are more susceptible to noise interference. For instance, a recording of a musical performance may have more noise in the high-frequency range due to the limitations of the recording equipment.
  - In natural language processing, feature-dependent noise can be observed in the presence of specific linguistic features. For instance, speech-to-text systems may struggle with accurately transcribing words or phrases with heavy accents or dialects, resulting in noise in the transcribed text.
  - In machine learning, feature-dependent noise can occur when certain features of a dataset are more prone to errors or inconsistencies. For example, in a dataset of handwritten digits, some digits may have more ambiguous or similar representations, leading to higher noise levels in the classification task.

- In image processing, feature-dependent noise refers to the noise that is present in different regions or features of an image. For example, in a photograph, the noise may be more prominent in the darker areas compared to the brighter areas.
- In speech recognition, feature-dependent noise can occur when certain phonetic features of speech are more prone to noise interference. For instance, fricative sounds like "s" and "sh" may be more affected by background noise than vowel sounds.
- In machine learning, feature-dependent noise can be observed when certain features of a dataset are more prone to measurement errors or inconsistencies. This can affect the accuracy of the machine learning model, as the noise in specific features may lead to incorrect predictions or classifications.

GENERATING SYNTHETIC NOISE.

The methodologies to produce different types of synthetic label noise. *Uniform noise* and *class dependent noise* can be represented with noise transition matrix *N* where *Nij* represents the probability of flipping label from class *i* to *j*. Since noise transition matrix consists of probabilities, P*j Nij* = 1.

On the other hand, in *feature-dependent noise*, each instance has its own transition probability depending on its features. Therefore, it cannot be generated using a noise transition matrix. The following sections will describe the process of generating these types of noises. Generated noisy labels are visualized with T-SNE plots in Figure III.

A. Uniform Noise For this type of noise, each entry in the noise transition matrix, besides diagonal ones are equally distributed. Noise transition matrix can be defined as follows:

CONCLUSION

It uses distillation technique to create a sparse distribution of data in the learned feature domain and therefore emphasizes similarities among data samples.

Interestingly, it's observed that feature dependent noise shows similar behavior in the training and validation phase while resulting in much lower test accuracy. Therefore, it's much harder to evaluate the progress of the network in case of feature-dependent noise by just checking outputs on noisy training and validation sets.

Moreover, it's seen that noise-robust algorithms behave differently for different types of noises. Therefore, it's important to consider the noise type while evaluating the proposed methodology.

Noise | Deep Learning | Labels | Label Errors

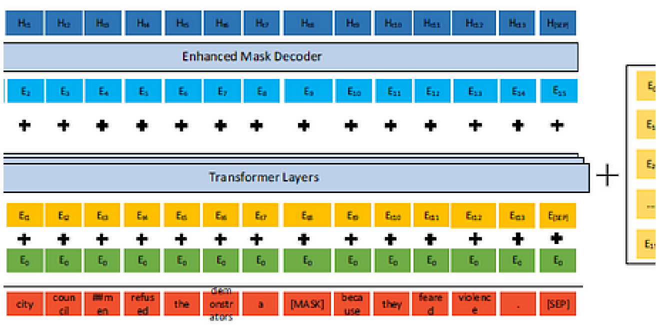
Written by Mustafa Adel Ibrahim

12 Followers

Following

Data Scientist | AI Engineer. https://www.linkedin.com/in/mustafa604/

More from Mustafa Adel Ibrahim

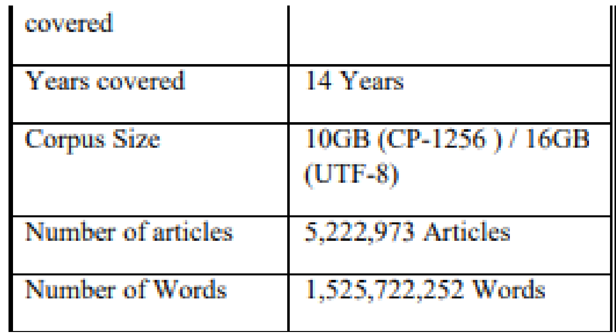


Mustafa Adel Ibrahim

DeBERTa: Decoding-enhanced BERT with Disentangled Attention

DeBERTa, 1st Single Model to Surpass Human Performance on SuperGLUE

6 min read · May 29



covered	
Years covered	14 Years
Corpus Size	10GB (CP-1256) / 16GB (UTF-8)
Number of articles	5,222,973 Articles
Number of Words	1,525,722,252 Words

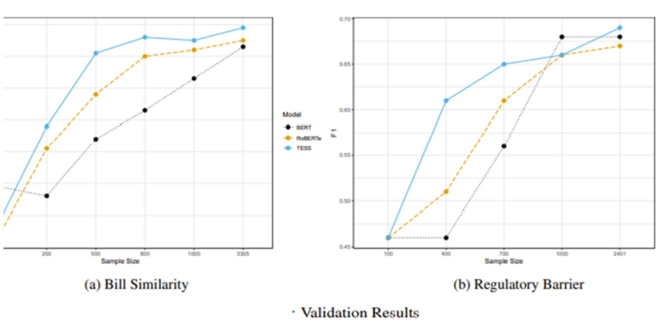
Mustafa Adel Ibrahim

1.5 billion words Arabic Corpus

Building a contemporary linguistic corpus for Arabic language from different countries…

5 min read · Sep 16

6

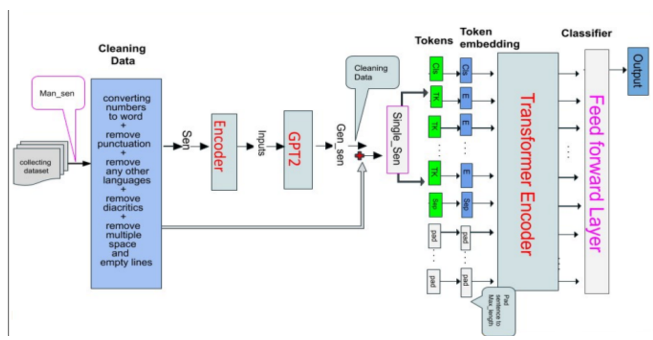


Mustafa Adel Ibrahim

Transformers for Social Science.

Transformer encoder designed for the social science research context, that outperformed…

6 min read · Oct 27, 2022



Mustafa Adel Ibrahim

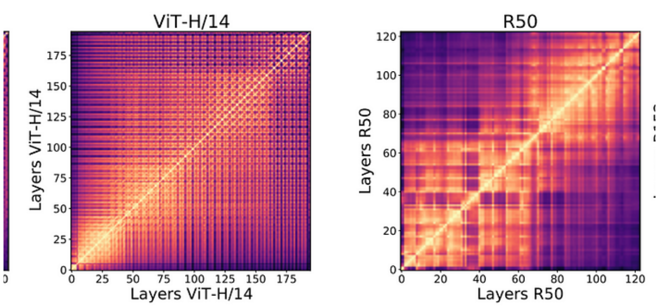
Detect Deepfake Text.

Automatically detect generated deepfake text using GPT2-Small-Arabic to distinguish…

4 min read · May 28

See all from Mustafa Adel Ibrahim

Recommended from Medium

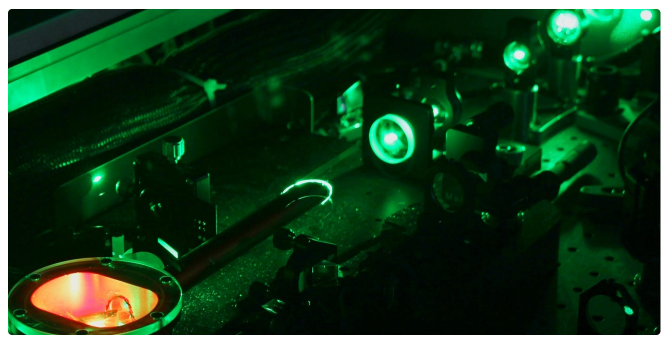


Ilias Papastratis

Comparison of Convolutional Neural Networks and Vision…

Introduction

19 min read · Sep 30



Ethan Siegel in Starts With A Bang!

Attosecond spectroscopy wins 2023's Nobel Prize in Physics

Our greatest tool for exploring the world inside atoms and molecules, and specifically…

· 11 min read · 2 days ago

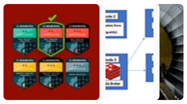
Lists

Natural Language Processing

698 stories · 309 saves

Practical Guides to Machine Learning

10 stories · 554 saves

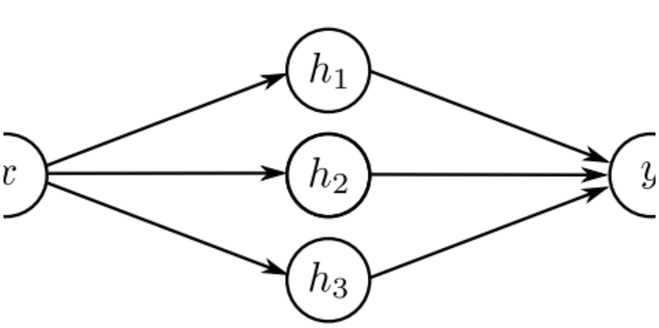


New_Reading_List

174 stories · 147 saves

Staff Picks

474 stories · 343 saves



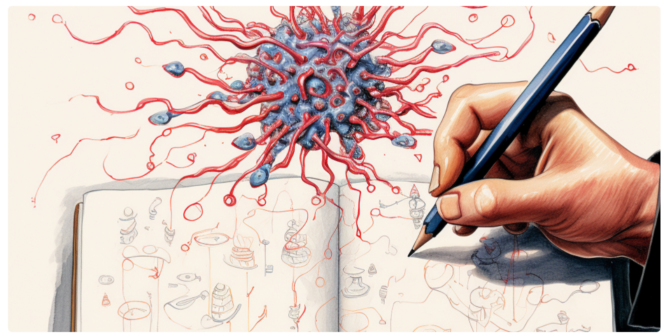
Raagulbharatwaj K in Towards AI

Mastering Deep Learning: The Art of Approximating Non-Linearities...

In the past year, we've witnessed an explosive surge in the popularity of Deep Learning....

6 min read · 5 days ago

21



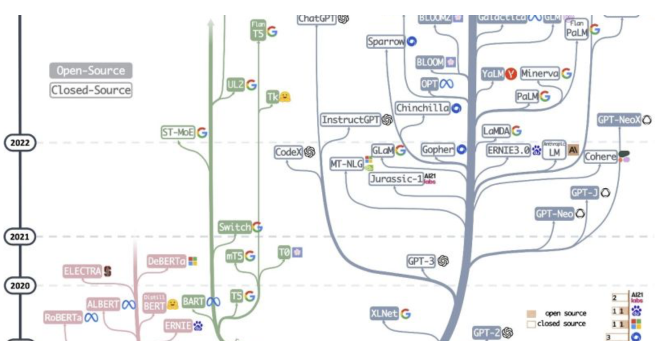
Thomas Smith in The Generator

ChatGPT Vision is Changing How I Interact With the Real World

I'm using the system every day—but not how I expected

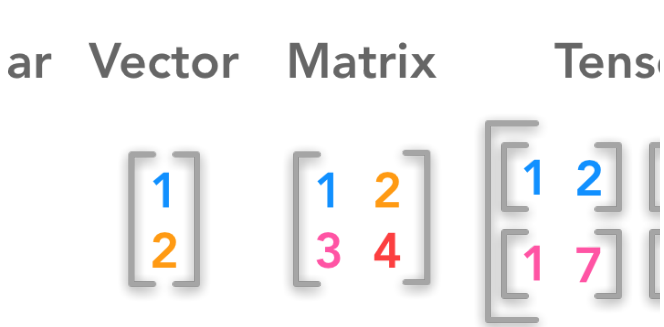
· 8 min read · Oct 7

2.9K 49



Haifeng Li

A Tutorial on LLM



Bibhuti Baibhav Borah

Matrices in machine learning

Generative artificial intelligence (GenAI), especially ChatGPT, captures everyone's...

15 min read · Sep 14

834

If I had to learn ML for the starting I would learn the Linear Algebra. Mathematics is the...

3 min read · Jun 9

77

See more recommendations