Semi-supervised learning

Lesson No. 11

Thiago Gomes Marçal Pereira *

July 8, 2020

1 Introduction

In Machine Learning we have two main type of models. Supervised and Unsupervised.
In Supervised Machine Learning, most of the time, we need labeled data. We make a prediction on the data, and calculate the difference between the prediction and the actual label, and try to minimize this difference, though improving our model.
On the other side, we have Unsupervised Machine Learning, in which we don't have labels, and let our model find similarities between the data and cluster them in several categories.
In the middle of them, there is another class of algorithms that help when only part of our data is labeled, and we want to use them to help in classifying the unlabeled data. These algorithms are called Semi-Supervised Machine Learning.
The main goal is to use the unlabeled data to make supervised learning better.

1.1 Why to use Semi-supervised Learning

Even though we may have access to label data, and also there are companies specialized in labeling data, this kind of data is still expensive, and labeling new data can be slow. On the other hand, there are lots of unlabeled data, that might be explored.
Once you collect some amount of labeled data, some state-of-the-art Semi-supervised Learning algorithms can behave better than Supervised Learning ones.

2 Core Concepts

There are a few core concepts needed to fully understand how Semi-supervised Learning works.

*RA: 189691 - t189691@g.unicamp.br

2.1 Confidence and Entropy

We want to make sure that when classifying the unlabeled data using labeled data, we have minimum entropy. Meaning that the classifier is confident on the classification of the unlabeled data.
With this, comes the possibility of Pseudo-learning, when based on the confidence of the classifier on the unlabeled data, we may start using this as the ground-truth label.

2.2 Label Consistency

When performing Data Augmentation on the unlabeled data, we must make sure that the new data has the similar predictions. This helps in regularizing the training data.
On labeled data, it's just the case of making sure that the augmented data receives the same label as the original data.

3 Methods for Label Consistency

These are some approaches to applying the Label Consistency on Unlabeled Data

3.1 Pi-Model

- Get the original image
- Generate two different entries with Stochastic Data Augmentation
- Pass through the model
- Get two different latent variables
- Calculate the Squared difference between them and Cross-Entropy related to the label
- Weight-sum to calculate the loss

It's possible to control which loss dominates each phase of learning, the supervised loss or the unsupervised one. Allowing for example for the supervised part have more weight in the beginning of the training.

3.2 Temporal Ensembling

Similar to Pi-model, but instead of using two different data augmented entries, we use only one, and compare it to some historical inverse of same samples in the past. Then, the losses are calculated similarly to Pi-Model.

3.3 Mean Teacher

Similar to Temporal Ensembling, but it averages the model weights, instead of the label predictions. This improves learning in large datasets, and also allows to learn with less labels.

3.4 Virtual Adversarial Training

The loss is defined as the robustness of the conditional label distribution around each input data point against local perturbation.
With simple improvement using entropy minimization principle, it achieves state-of-the-art performance for semi-supervised tasks.

4 Conclusion

When used correctly, Semi-supervised learning can improve efficiency in comparison with supervised learning.
Using unlabeled data can reduce costs of labeling data, but we must make sure that the unlabeled data comes from the same distribution as the labeled data, otherwise it won't be very useful.
And most of the models behave better with more labeled data.
Finally, it's a category of learning that is worth knowing, because it has been showing good results, and even though conclusions still come from small datasets, results are promising and evolution of state-of-the-art is coming fast.