

進捗報告

1 今週やったこと

- SSL の手法調査

2 SSL の手法調査

SSL におけるモチベーションとして、少量のラベル付きデータと大量ラベルなしデータがあるといった状況に利用されることはもちろんあるが、一方でラベルが付いたデータが絞られることで各ラベルに対する複雑化したデータ分布を狙ったものにしやすくなる可能性がある。

2.1 分類器による手法

ラベル付きデータによる分類器からラベルなしデータの疑似ラベルを生成する。そして確信度などの指標を用いて閾値を超えた疑似ラベルを実際のラベルとみなすことで、ラベル付きデータを水増しする手法。

- 自己訓練 (Self Training)
- 共訓練 (Co-Training)
- FixMatch

2.2 データ分布に基づく手法

ラベル付きデータによって得られる決定境界に対し、ラベルなしデータを用いて滑らかにする手法。

- Virtual Adversarial training(:VAT)
- Unsupervised Data Augmentation(:UDA)
- Ladder Network

2.3 感想

SSL において前者のようなラベルなしデータにラベル付けをして学習を進める手法より、後者のようなラベルなしデータをラベル付けを行わずに学習をするもののほうが cifar10 や MNIST などのタスクで

精度を上げていた。より複雑な問題ではどうか分からないがラベルなしデータにラベル付けをしたうえで学習する方法だけではあまりいいものが得られない気がする。

FixMatch や UDA においてノイズとして Rand Augment を用いた Data Augmentation が用いられており、一部の GA を組み込まれたものとなっている。

3 今後の方針

すでに考えられているだろうが、ラベルなしデータの一部に対し前者の方法でラベル付きデータを水増ししたうえで後者の方法で学習することはできないか。もし、有効であったとしてどの程度まで増やしていいのか。

GA によってどのラベルなしデータをラベル付きデータにするのかを探索する。

4 来週の課題

- SSL の実装及び実験