

目 次

内容

1	Section1_勾配消失問題	2
2	Section2_学習率最適化手法	3
3	Section3_過学習	4
4	Section4_畳み込みニューラルネットワーク	5
5	Section5_最新の CNN	6

1 Section1_勾配消失問題

誤差逆伝播法が階層（出力層から入力層）に向かって進んでいくにつれ勾配がどんどん緩やかになっていく。そのため勾配降下法による更新では、階層のパラメータはほとんど変わらず、訓練は最適値に収束しなくなる。

- 1) 活性化関数の選択
- 2) 重みの初期値設定
- 3) バッチ正規化
- 4) 学習率最適化手法
- 5) RNN における勾配消失問題

多層パーセプトロンや CNN においては、層が深くなると勾配消失によって入力層付近ほど学習ができなくなるといった課題。

時系列データを用いて RNN 学習を行う時は、過去の時系列を遡りながら誤差を計算する通時的誤差逆伝播を用いて勾配が計算されるが、この時計算される勾配に対して、時系列の古いデータほど勾配消失しやすい。

2 Section2_学習率最適化手法

1) RAdam

最適化アルゴリズムのデファクトスタンダードは Adam である。これに代わる RAdam。

①. 画像分類の ImageNet や CIFAR-10 から機械翻訳の IWSLT'14 や WMT'16 などの幅広いタスク及びデータセットで RAdam が Adam より優れてた性能

②. 学習の初期段階で Adam の適応学習率の分散が大きくなり過ぎるという問題があることを指摘。

③. ヒューリスティックな手法で、ハイパラ調整を必要とする Warmup は上記の問題を緩和することを発見

④. 適応学習率の分散を自動的に抑えられるような機構を Adam に組み込んだものが RAdam

- ・ステップ数が 4 以下：適応学習率を使わない
- ・ステップ数が 4 より大きい：Adam に補正項を掛けることで適応率の分散を抑える

2) 最適化手法

①. 確率的勾配降下法 (SDG)

②. MomentumSDG

③. ネストロフの加速法 (NAG)

④. Adagrad

⑤. RMSprop

⑥. AdaDelta

⑦. Adam

⑧. RMSpropGraves

⑨. SMORMS3

⑩. AdaMax

⑪. Nadam

⑫. Eve

⑬. Santa

⑭. GD by GD

⑮. AdaSecant

⑯. AMSGrad

⑰. AdaBound、ASMBound

3 Section3_過学習

1) 過学習とは予測がうまくできなかった状態

コンピュータが手元にあるデータから学習しすぎた結果、予測がうまくできなくなったという状態

2) 過学習に気づくために

中途半端なモデルを量産する悪循環にはまらないように、「モデルを作ってみる→検証する→改善する→同じ手法でよいモデルを作る」というサイクルをまわして過学習に気づき改善していきこと。

3) 過学習を解決する方法

過学習の解決は、基本的にモデルの自由度に制限をかけるもの。過学習とは全体の傾向が読み取れずに1つ1つのデータにフィットしてしまうこと。

4) 過学習の発見・解決手法

交差検証と正則化を用いて過学習を解決する

4 Section4_畳み込みニューラルネットワーク

1) データ拡張

モデルの汎化性能を高める最も単純な方法は、より多くのデータで学習を行うことである。しかし現実的に利用可能なデータには限界がある。そこで、現在手元にあるデータに何らかの変換を施し、それをデータ集合に追加することで疑似的にデータ量を増加させること。

2) 転移学習

人間が視覚によって判断する際、エッジや形状、質感などに着目することが多く、これらの特徴は異なるデータ集合においても共通のものであると考えられる。そのため大量のデータ集合で学習したモデルを用いて、サンプルが少量である、ほかのデータ集合の汎化性能を向上させる方法が数多く提案されている。

3) CNN の入力層の設計をする際に注意すべき、データの次元とチャンネル

- ・モノラル音声データ

1次元の波形データ1つで構成されている。

- ・ステレオデータ

1次元2チャンネル

- ・カラー動画データ

複数チャンネル3次元データ

5 Section5_最新の CNN

- 1) CNN は物体の移動の対応には強いが、回転の縮拡等の視点変換には向いていない。
- 2) 物体の不変性より、物体の対価性を探すべき
- 3) CNN は物体内のパーツ同士の関係が理解できない
- 4) 人間の物体認識法と根本的に違う
- 5) 座標を使わない。

結論

- 1) 座標なしの学習
- 2) Part->Whole の学習経路

1 ピクセルと隣のピクセルたちから画像を理解し始めている。これは人間が一般的な大範囲から小範囲 (Whole->Part) とは違う。