

深層学習 day3

Section1 : 再帰型ニューラルネットワークの概念

○RNNとは

- ・ 時系列データに対応可能なニューラルネットワーク

○時系列データとは

- ・ 時間的な繋がりがあるデータ
- ・ 音声データ、株価のデータ、テキストデータ、観光者の数など

○RNNの構造

- ・ 基本的な構造はニューラルネットと同じ。入力層 → 中間層 → 出力層
- ・ 中間層の出力を次の中間層の入力に加える
- ・ 重み：入力層 → 中間層の重み、中間層 → 出力層の重み、中間層 → 次の中間層への重み
- ・ 中間層 u を導出するときに、前の中間層の重みを加算している

$$u_t = W(in)x_t + Wz_{t-1} + b$$

- ・ 過去の状態を保持し、そこから次の時間での状態を再起的に計算する構造が必要
- ・ バイナリデータの足し算
繰り上がりを過去 → 未来の時系列データとして捉えられる

○BPTT(BackPropagation Through Time)

- ・ RNNにおける誤差逆伝播の手法
- ・ $W(in)$, $W(out)$, W それぞれを微分して逆伝播する

Section2 : LSTM

○RNNの課題

- ・ 時系列を遡れば遡るほど、勾配が消失していく。
バイナリデータの例だと8ステップ分
RNNの時系列のステップ分がNNの層の扱いと似ている
→ 構造自体を変えて解決する：LSTM
- ・ 連鎖律に伴い微分の回数が増えると、sigmoid関数のようなものだと誤差が微小になってしまう
- ・ 勾配爆発：勾配が層を逆伝播するごとに指数関数的に大きくなっていく
学習率の不適切な設定など

○CEC

- ・ CEC(Constant Error Carousel)
- ・ RNNの課題：中間層で過去情報を使う際に活性化関数を使うタイミングで勾配消失が起
こる
→ CECに過去の情報を記憶させる
考える機能と記憶する機能を分離させる

- ・勾配が1であれば、勾配消失や勾配爆発が起きない
- ・課題：CECにはNNの学習特性がない
→取り巻きで学習機能を配置し、何を入力から覚えるか、何を出力するかを学習させる

○入力ゲートと出力ゲート

- ・入力ゲート：CECへ入力値の覚えさせ方を学習する
今回の入力値(重み W_i)と前回の出力値(重み U_i)からCECにどう覚えさせるか(出力 V)
- ・出力ゲート：CECの出力をどう使うかを学習する
今回の入力値(重み W_o)と前回の出力値(重み U_o)からCECの出力をどう使うか

○忘却ゲート

- ・CECに過去の情報を記憶するが、不要になった過去の情報を削除できない
→忘却ゲート
- ・今回の入力(W_f)と前回の出力(U_f)から過去の情報をどう削除するのか学習する
$$c(t) = i(t) * a(t) + f(t) * c(t-1)$$

○覗き穴結合

- ・CECの現在の状態を判断材料に使う方法

Section3 : GRU

- ・LSTMではパラメータ数が多くなり、計算負荷が高くなる課題があった
→GRU
- ・GRUではパラメータ数を大幅に削減しつつ、精度を同等orそれ以上の構造にする。
- ・隠れ層でCECのように過去の情報を保持する
- ・入力ゲート、出力ゲート、忘却ゲートの代わりにリセットゲート、更新ゲートを使用

Section4 : 双方向RNN

- ・Bidirectional RNN(BiRNN)
- ・過去の情報だけでなく、未来の情報を加味することで、精度を向上させるモデル
ex.) 文章の推敲、機械翻訳など
- ・順方向、逆方向にRNNの処理を適用し、双方向の出力を結合する。

Section5 : Seq2Seq

○Seq2Seq全体像

- ・主に自然言語処理でよく使われる
- ・機械翻訳など
- ・単語の列をRNN系列モデルに入力→ベクトル表現(ex文の意味等)として出力結果を保持
ベクトル表現からRNN系列モデルを使い単語の列を出力
- ・Encoder-Decoderモデルの一種

○Encoder RNN

- ・文章を単語の列等に区切って渡していく
- ・単語を one-hot ベクトルに変換 (数万次元) → embedding 表現に変換 (数百次元)
- ・embedding 表現を RNN の入力値として使う

○Decoder RNN

- ・単語のトークンごとにアウトプットデータを生成する
各トークンの生成確率を出力していく

○HRED

- ・Seq2seq の課題：一問一答しかできない
→ HRED
- ・一文一文に対しても context RNN を適用することで文脈を反映する

○VHRED

- ・HRED の課題：確率的な多様性に乏しく、会話の流れが乏しい
→ VHRED
- ・VHRED：HRED に VAE を適用したもの

○オートエンコーダ

- ・教師なし学習の一つ。入力から同じ出力を出すことを学習させる。
- ・Encoder-Decoder モデル
- ・VAE：オートエンコーダにおける潜在変数 z を作る時に正則化を行い、確率分布の構造を導入する。
- ・encoding 時にノイズを加えることで、モデルの頑強性が生まれる。

Section6 : Word2vec

- ・RNN の課題：単語のような可変長の文字列を NN に与えることができない
→ 固定長形式で単語を表す必要がある (embedding)
- ・Word2Vec：embedding 表現を得る手法の一つ。
- ・CBOW(周囲の単語からある単語を予測), Skip-Gram(ある単語からその周囲の単語を予測) のニューラルネットワークモデルを使用

Section7 : Attention Mechanism

- ・seq2seq の課題：長い文章への対応が難しい。
文章の長さに関わらず、固定長の特徴ベクトルへ変換してしまう。
- ・Attention Mechanism：入力と出力のどの単語が関連しているのか、を学習させる
- ・Transformer の内部で Attention Mechanism が使われている
- ・現在の自然言語処理の代表的なモデルである BERT はその Transformer をベースとしたモデル