

目 次

内容

1 .	section1_再起型ニューラルネットワークの概念	2
2 .	section2_LSTM.....	4
3 .	section3_GRU.....	5
4 .	section4_双方向 RNN	6
5 .	section5_Seq2Seq	7
6 .	section6_Word2vec	8
7 .	section7_Attention Mechanism.....	9

1. section1_再起型ニューラルネットワークの概念

Recurrent Neural Network(RNN)

時間方向に状態を引き継ぎながら計算を進めることができるため、自然言語や音声などのように時間方向に順番に並んでいるデータを扱うタスクに向いている。

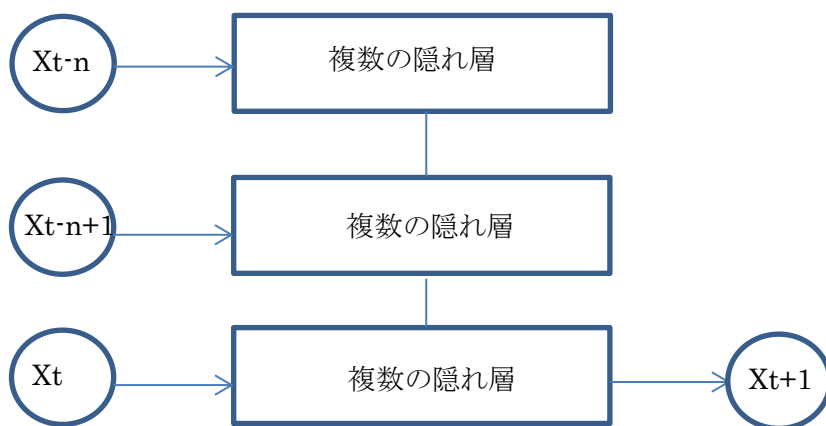
1) 自己回帰モデル（全結合型ニューラルネットワーク）



時刻 $t + 1$ における X の値を予測するために、時刻を n 個遡った X の過去の値 (X_{t-n}, \dots, X_t) を説明変数として用いている。

このような構成をとることで、全結合型ニューラルネットワークでも系列データを扱うことは可能であるが、入力系列数が固定される。自然言語処理では1単語を1つの X として扱うことが多いが、1度に扱う単語数が固定されることは制約となる。

2) 自己回帰モデル（再起型ニューラルネットワーク）



時刻ごとの X を順番に隠れ層に入力し、その隠れ層の出力が時間方向につながるという構成。このとき入力される X の個数は、データに合わせて変化させることができるため、入力の個数が固定されない。

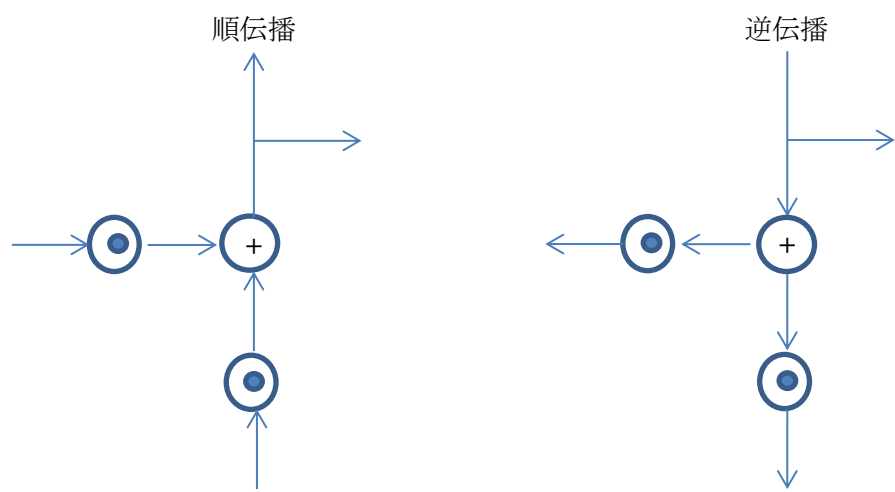
3) 勾配クリッピング

勾配爆発を回避するための方法

$$g_{\text{clipped}} = g \cdot (V / ||g||)$$

L2 ノルム

4) シンプルな再帰型ニューラルネットワークの計算グラフ



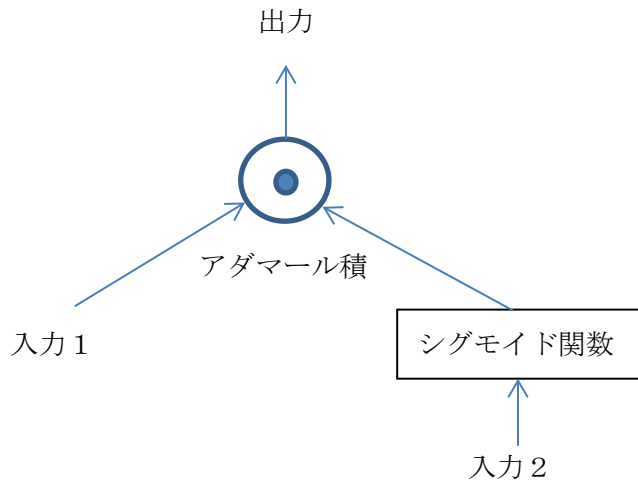
2. section2_LSTM

Long Short-Time Memory

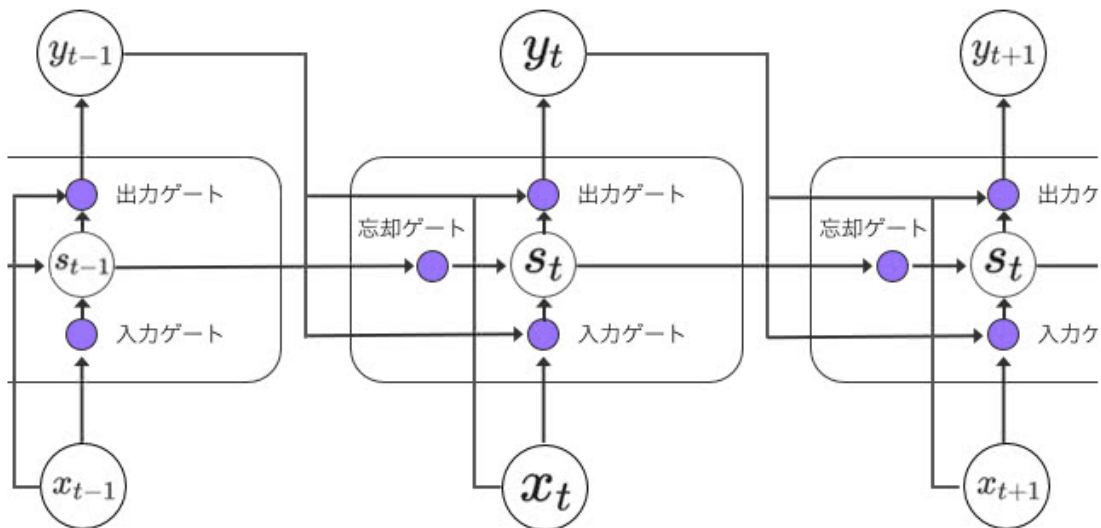
1) 3つのゲートと記憶セル

- ①. 入力ゲート
- ②. 出力ゲート
- ③. 忘却ゲート
- ④. 記憶セル

2) ゲートの基本構成



3) LSTM の構造



シンプル RNN でも理論上は、上手くいくが現実的にはかなり前の古い情報を考慮できなかった。

LSTM の画期的な特徴は、「ゲート」と呼ばれる情報の取捨選択機構を持ったことである。

各ゲートの取捨選択はシグモイド関数で行われる。

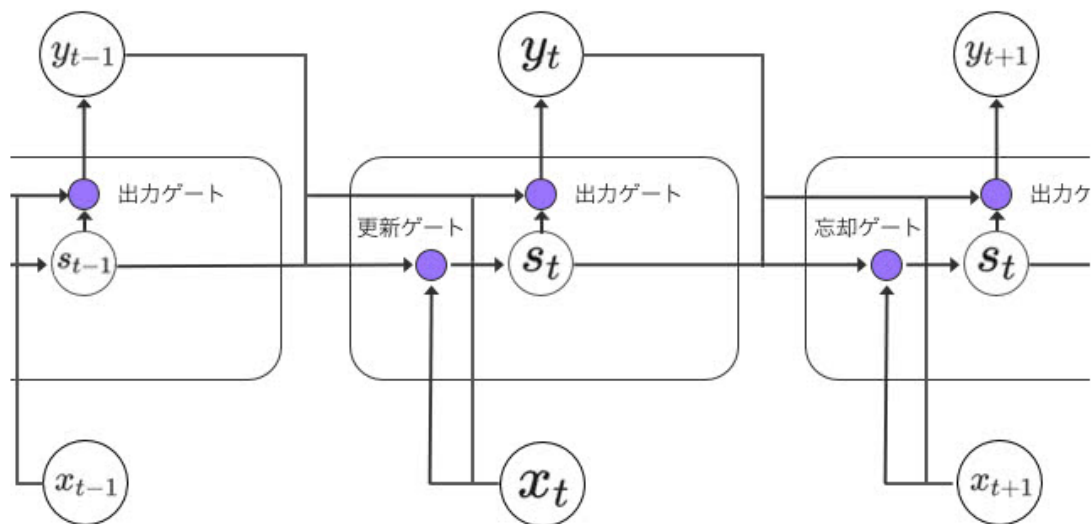
3. section3_GRU

Gated Recurrent Unit

1) 2つのゲート

- ①. リセットゲート
- ②. 更新ゲート

2) GRUはLSTMをもう少しシンプルにしたもの

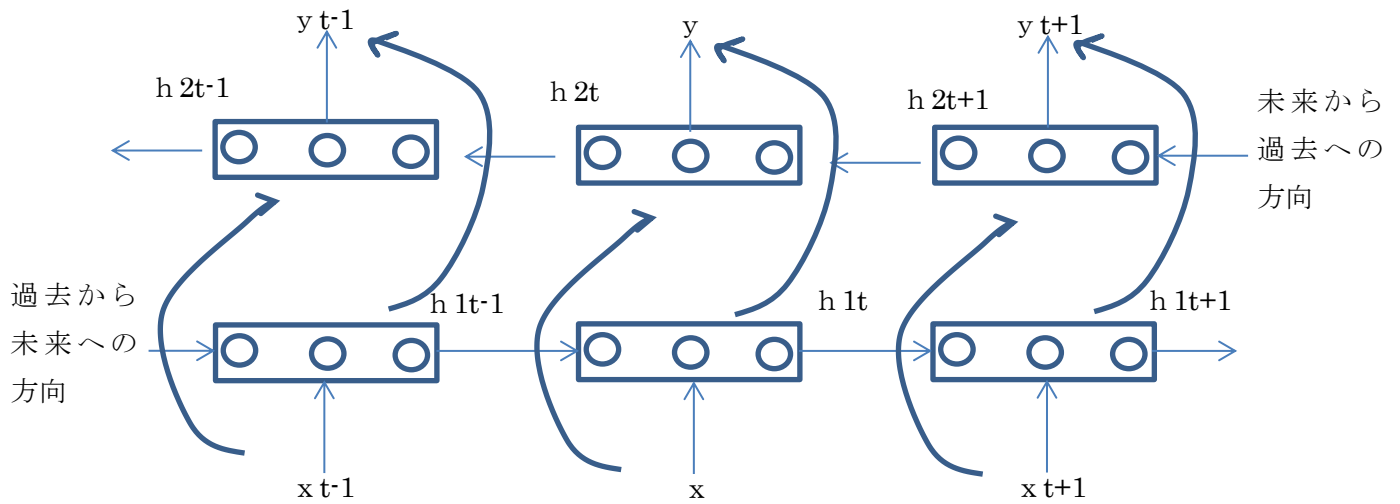


忘却・更新ゲートを導入することで、長いステップ前の出来事の特徴の記憶を維持しやすくなる。

4. section4_双方向 RNN

過去から未来だけでなく、未来から過去へ方向も考慮した再起型ニューラルネットワークである。

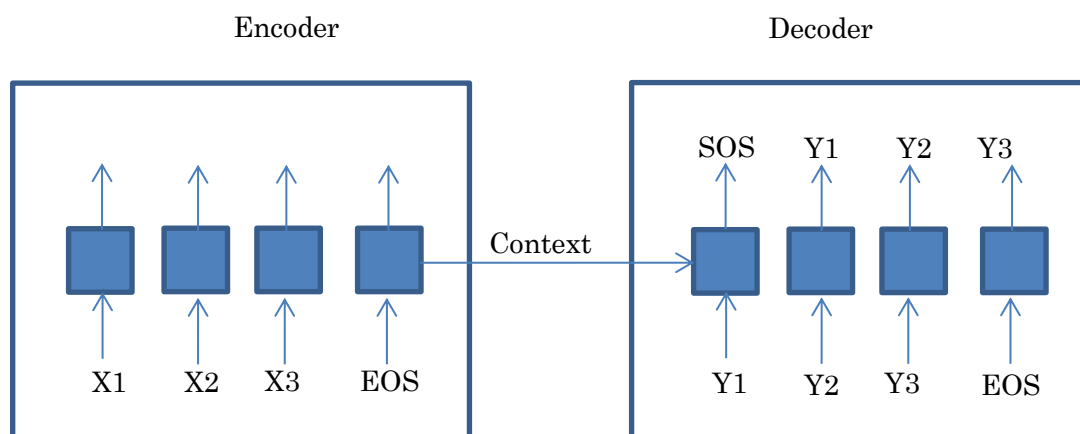
例) 文書内の文字の穴埋めタスクなど、過去から現在だけでなく、未来から現在までの系列情報を用いることが有効と考えられるタスク



5. section5_Seq2Seq

ブレークスルーな技術

- 1) 翻訳
- 2) 自動字幕技術
- 3) チャットボット (Question Answering)
- 4) 連続値などのシーケンシャルなデータの取り扱い



6. section6_Word2vec

ニューラルネットワークによる学習を通して、単語の分散表現を得るためのツールである。

Word2Vec には、 Continuous Bag-of-Words(CBW)と Skip-gram の 2 つのニューラルネットワークが実装されている。

CBW : 周辺単語から対象となる 1 単語を予測する。

Skip-gram : 1 単語から周辺単語を予測する。

いずれも入力層側の重み **Win** と出力層側の重み **Wout** をパラメータとして持つが、一般に **Win** を埋め込み行列として用いる。

Word2Vec で巨大なコーパスを扱う場合、通常の多クラス分類を行うと、計算速度が問題になる。

そのため、負例サンプリング (negative sampling) という手法を用いている。

通常の **CBOW** 又は **Skip-gram** の出力層では、全ての語集を対象としてソフトマックス関数と交差エントロピー誤差関数の計算を行う。

負例サンプリングを適用した場合の出力層では、1 個の正例と k 個の負例だけを対象として、シグモイド関数と交差エントロピー誤差関数の計算を行う。

通常の **CBOW** 又は **Skip-gram** では、多クラス分類問題を解いているが、それらに負例サンプリングを適用した場合は、 $k + 1$ 個の 2 クラス分類問題を解いていることになる。

k 個の負例は、コーパスの中から、単語の出現頻度に基づいてサンプリングする。

7. section7_Attention Mechanism

ニューラルネットワークにおいて、何らかの特徴があった時に、その特徴のどこを重視すればよいかを学習する機構

1) ソフト・アテンションは、ソフトマックス関数などで確率分布を求め、その確率分布を用いて重みづけ平均を得る方法である。

2) ハード・アテンションは、ソフトマックス関数などで確率分布を求め、その確率分布に従って抽出された1点だけを得る方法である。ただし確率分布を考えずに1点を決める場合もある。

3) 2017年に提案された **Transformer** という自然言語処理向けのモデルでは、ソースターゲット・アテンションやセルフ・アテンションなどの機構が用いられている。

$$Y = \sum \alpha(x, x_i) y_i$$

ここで

X : Query (検索文)

X_i : Key (辞書のインデックス)

Y_i : Value (辞書の本文)

辞書から **Query** と似ている **Key** を探して **Value** を取り出す。