# 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）

Xt-n

Xt+1

Xt-n+1

複数の隠れ層

Xt

　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（Xt-n,・・・Xt）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

あるが入力の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）

複数の隠れ層

Xt-n

Xt-n+1

Xt

複数の隠れ層

Xt+1

複数の隠れ層

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になる。

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

　ｇclipped=g・(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

　　　　　　　　　　　　順伝播　　　　　　　　　　　　　　　　逆伝播

+

+

# 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

出力

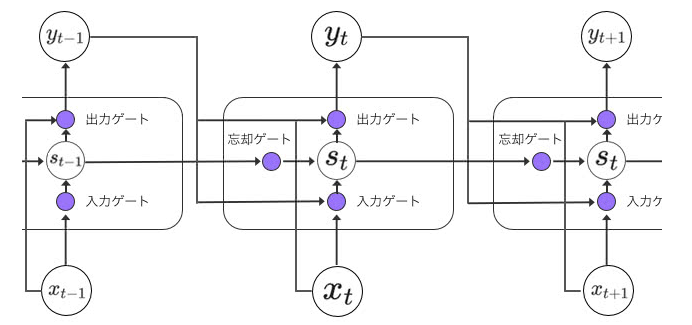
アダマール積

シグモイド関数

入力１

入力２

（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

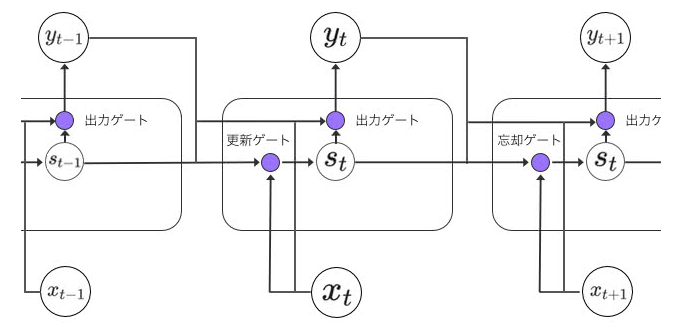
Gated Recurrent Unit

（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、未来から過去への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

ｙt-1

ｙt+1

ｙ

ｈ2t+1

ｈ2t

ｈ2t-1

未来から過去への方向

ｈ1t-1

ｈ1t+1

ｈ1t

過去から未来への方向

ｘt+1

ｘ

ｘt-1

# ５． section5\_Seq2Seq

　ブレークスルーな技術

　（１）翻訳

　（２）自動字幕技術

　（３）チャットポット（Queestion Answering）

　（４）連続値などのシーケンシャルなデータの取り扱い

Decoder

Encoder

SOS

Y3

Y2

Y1

Context

Y1

Y2

Y3

EOS

EOS

X3

X2

X1

# 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。

いづ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で巨大なコーバスを扱う場合、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、コーバスの中から、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=Σα（x,xi）yi

ここで

　　X：Query（検索文）

　　Xi：Key（辞書のインデックス）

　　Yi：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。