ラビットチャレンジ課題　作成者:内山 貴雄

深層学習後編day3の各章の要約及び確認テスト等の考察

●再帰型ニューラルネットワーク

　本章では、再帰型ニューラルネットワーク（Recurrent Neural Network：以下、RNN）を学んだ。これは、ニューラルネットワークを拡張して時系列データを扱えるようにしたものである。ここで言う時系列データとは、ある時間の経過とともに値が変化していくようなデータを指し、店舗の日次売上データやホームページのアクセス数履歴、工場設備のセンサデータなど、多種多様なデータが時系列データとして表現される。

〇確認テスト

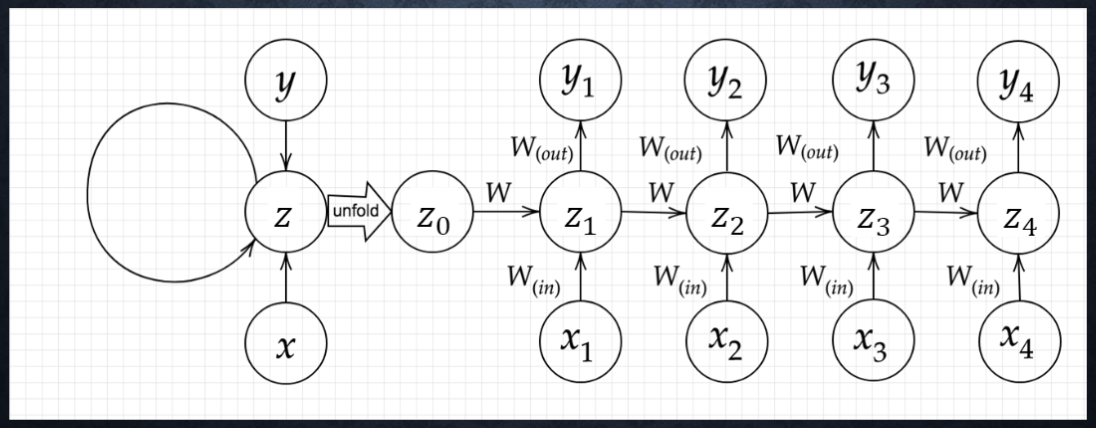
RNNのネットワークには大きくわけて3つの重みがある。1つは入力から現在の中間層を定義する際にかけられる重み、1つは中間層から出力を定義する際にかけられる重みである。残り1つの重みについて説明せよ。

・答案

過去の時間t-1の状態における中間層の出力にかけ、時間tの中間層に入力する際の重み。

○確認テスト

下図のy1をx・s0・s1・win・w・woutを用いて数式で表せ。※バイアスは任意の文字で定義せよ。※また中間層の出力にシグモイド関数g(x)を作用させよ。



・答案:

y1=g(Wout•Z1+c)

Z1=Win•x1+W•S0+b

●LSTM

RNNアーキテクチャーの一つである長・短期記憶（Long short-term memory、略称: LSTM）を学ぶ。

一般的なLSTMユニットは、セル、入力ゲート、出力ゲート、および忘却ゲートから構成される。セルは任意の時間間隔にわたって値を記憶し、3つの「ゲート」はセルを出入りする情報の流れを制御する。

LSTMネットワークは時系列データに基づく分類、処理、予測によく適している。これは、LSTMネットワークが時系列中の重要な事象間の未知の期間の時間差となることができるためである。LSTMは、従来のRNNを訓練する際に遭遇しうる勾配爆発および消失問題（英語版）に対処するために開発された。

○確認テスト

以下の文章をLSTMに入力し空欄に当てはまる単語を予測したいとする。文中の「とても」という言葉は空欄の予測においてなくなっても影響を及ぼさないと考えられる。このような場合、どのゲートが作用すると考えられるか。

「映画面白かったよね。ところで、とてもお腹が空いたら何か--」

・答案

忘却ゲート

●GRU

LSTMはRNNで長期的特徴の学習を可能だが、計算コストが大きいため、下記の様に状態ベクトル数とゲートを減らし、LSTMと同性能かつ計算量が少なく、高速での学習を可能としたものをGRUという。

・状態ベクトル数を減らす。

LSTMではCECとh(前の時間t-1の出力)の2つが状態を保持していたものをまとめた。

・ゲートコントローラ数を減らす。

LSTMでは忘却ゲートと入力ゲートの操作を1つのコントローラで操作するように変更。

状況変更に伴い、別途下記変更を行う。

CECとhをまとめたため、出力ゲートによる出力制限は不要となり削除したが、それにより重み衝突が発生するリスクがあるため、入力値を求めるニューロンに再帰される経路上にリセットゲートを設ける。

○確認テスト

LSTMとCECの課題をそれぞれ簡潔に述べよ。

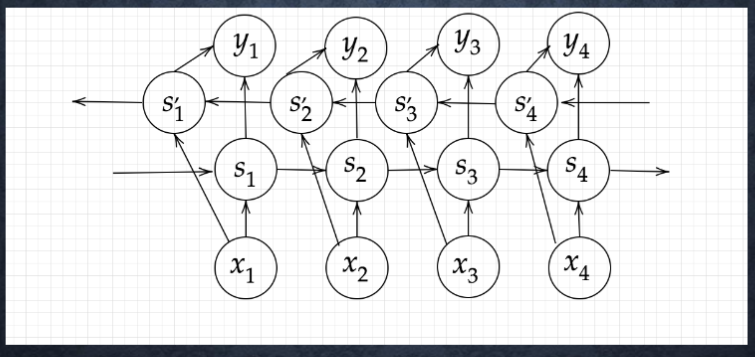
・答案

LSTMは各時間の中間層に、2つの隠れ層と入力ゲート、忘却ゲート、出力ゲートを持つなど構造が複雑であり、学習計算コストが大きいことが課題である。

CECに関しては、各時間における記憶能力のみを持ち学習能力がないことが課題である。そのためLSTMでは、前述の各ゲートにて学習を受け持つ構造としたため、その構造の複雑化に貢献していると言える。

●双方向RNN

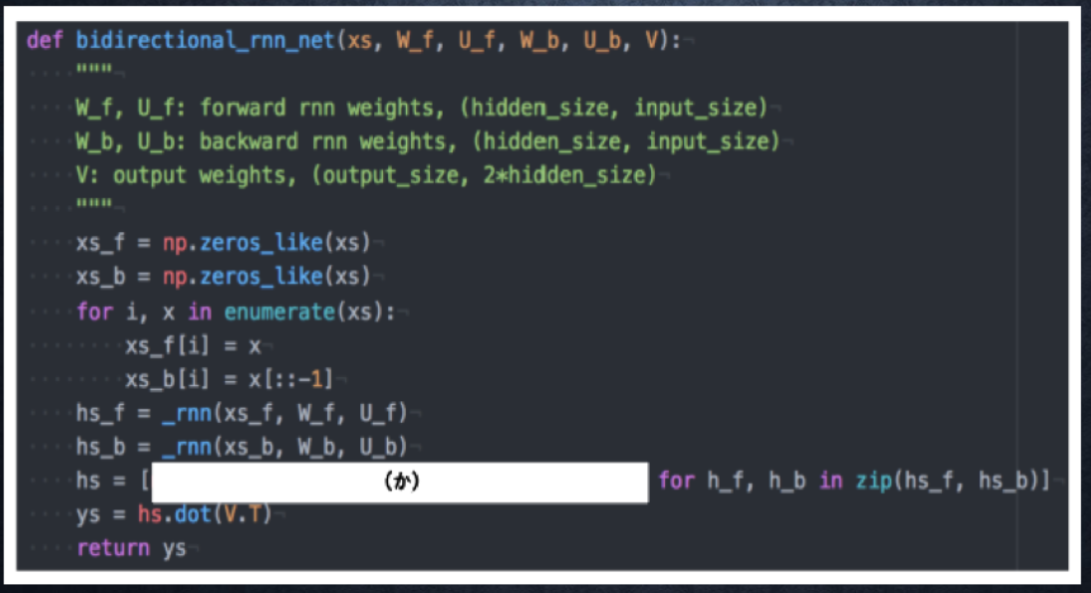
双方向RNN (Bidirectional RNN) は，系列の最初のステップt=0から繰り返して順方向に予想することに加えて，系列の最後のステップt=Tからの逆方向の予測も行う，RNNを双方向形に拡張したモデルである．以下の図に双方向RNNのダイアグラムを示す．



双方向RNNも，毎フレームtにおいて観測Xtを入力して潜在変数をもとに出力otを予測するところは通常の（単方向の）RNNと同じである一方で異なるのは，潜在変数が順方向htと逆方向のhtの２種類使用される点であり，これにより双方向のコンテキストを集約しての出力の予測が可能となる．もう少し正確に言うと，順方向側はステップ1からステップt-1までのコンテキストを知っており，逆方向側はステップ2からステップt+1までのコンテキストを「双方向」とも学習を通じて予測できるようになる。

○演習チャレンジ

以下は双方向RNNの順伝播を行うプログラムである。順方向については、入力から中間層への重みW\_f, 一ステップ前の中間層出力から中間層への重みをU\_f、逆方向に関しては同様にパラメータW\_b, U\_bを持ち、両者の中間層表現を合わせた特徴から出力層への重みはVである。\_rnn関数はRNNの順伝播を表し中間層の系列を返す関数であるとする。(か)に当てはまるものは何か？



・答案

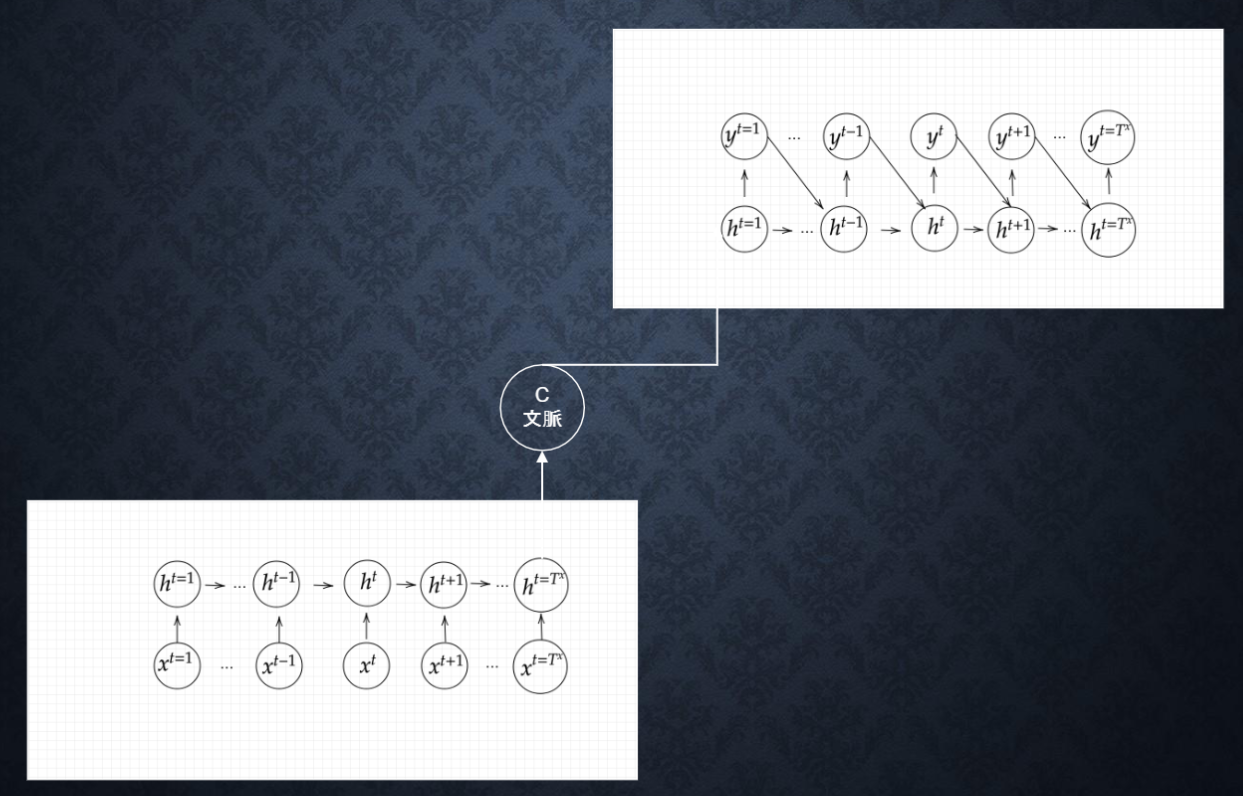
(4)

双方向RNNでは、順方向と逆方向に伝播したときの中間層表現をあわせたものが特徴量となるので、np.concatenate([h\_f, h\_b[::-1]], axis=1)である。

●Seq2Seq

　Seq2Seqとは、RNNを用いたEncoder-Decoderモデルの一種であり、機械翻訳などのモデルに使用される。

　本章では、前述Encoder-DecoderRNNでの処理内容の詳細とseq2seqモデルを組み合わせたHREDとその改良手法であるVHRED、VHREDにて用いられるVAEの詳細を学んだ。



○確認テスト

下記の選択肢から、seq2seqについて説明しているものを選べ。

（1）時刻に関して順方向と逆方向のRNNを構成し、それら2つの中間層表現を特徴量として利用するものである。

（2）RNNを用いたEncoder-Decoderモデルの一種であり、機械翻訳などのモデルに使われる。

（3）構文木などの木構造に対して、隣接単語から表現ベクトル（フレーズ）を作るという演算を再帰的に行い（重みは共通）、文全体の表現ベクトルを得るニューラルネットワークである。

（4）RNNの一種であり、単純なRNNにおいて問題となる勾配消失問題をCECとゲートの概念を導入し解決したものである。

答案:

(2)

○確認テスト

seq2seqとHRED、HREDとVHREDの違いを簡潔に述べよ。

答案：

seq2seqでは、ある一文や一単語の入力に対しての返答を生成するが、HREDは連続した複数の文や単語を入力し、それらの繋がりや文脈をRNNにて学習し返答を行うことが出来る。また、HREDにおいて、入力の各文・単語への返答をEncoder-Decoderモデルにて学習する際の潜在変数がone-hotベクトルの様な情報が少ないものとなり、返答がありきたりな短いものとなる課題がある。その課題を、Encoder-Decoderモデルにて学習する際の潜在変数の要素を確率分布z∼N(0,1)に従う値とすることで、それらに入力の情報量をより多く取り入れ解決したものがVHREDである。

○確認テスト

VAEに関する下記の説明文中の空欄に当てはまる言葉を答えよ。

自己符号化器の潜在変数に\_\_\_\_を導入したもの。

答案:

確率分布z∼N(0,1)

●Word2vec

　単語を固定長のベクトルで表現することを「単語の分散表現」と呼び、単語をベクトルで表現することができれば単語の意味を定量的に把握することができるため、様々な処理に応用することができる。Word2Vecも単語の分散表現の一つの手法であり、seq2seq内のRNNにおける計算負荷の軽量化に用いられるため本章にて学習した。

〇確認テスト

本章に確認テストが設定されていなかったため、割愛する。

●Attention Mecanism

　seq2seqの課題は長い文章への対応が難しい点です。具体的には入力の文章が3単語のとても短い文であっても、50単語あるとても長い文であっても、その意味をある固定次元ベクトルの中に押し込む必要がある点です。その課題を、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ことで、単語列の出力時に対応する入力の単語を引っ張ってくることで長い文書でも翻訳の精度を上げる手法がAttention Mecanismであり本章でそれを学んだ。

○確認テスト

RNNとword2vec、seq2seqとAttentionの違いを簡潔に述べよ。

　seq2seqなどのRNNでの学習を行う際に、単語や文章からなる大規模データの入力における計算、メモリ負荷が課題となる。その課題を、単語を固定長のベクトルで表現し、入力データの軽量化することで解決するとともに、単語の意味の定量的に把握することを可能とした手法をWord2Vecという。

　seq2seqの課題は長い文章への対応が難しい点です。その課題を、入力と出力のどの単語が関連しているのかを学習し、単語列の出力時に関連性の強い入力の単語を引用することで、長い文章の入力においても翻訳の精度を上げる手法がAttention Mecanismである。