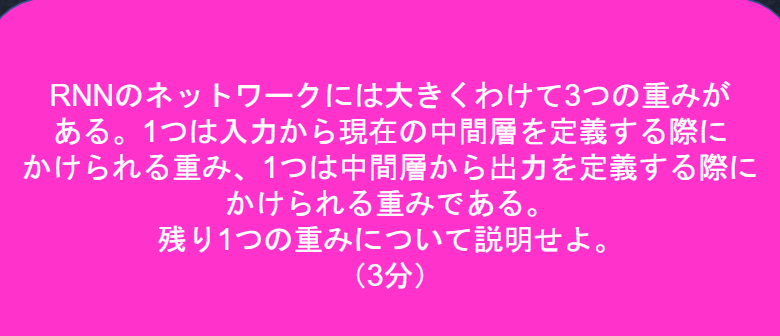
**まとめ要点１：**

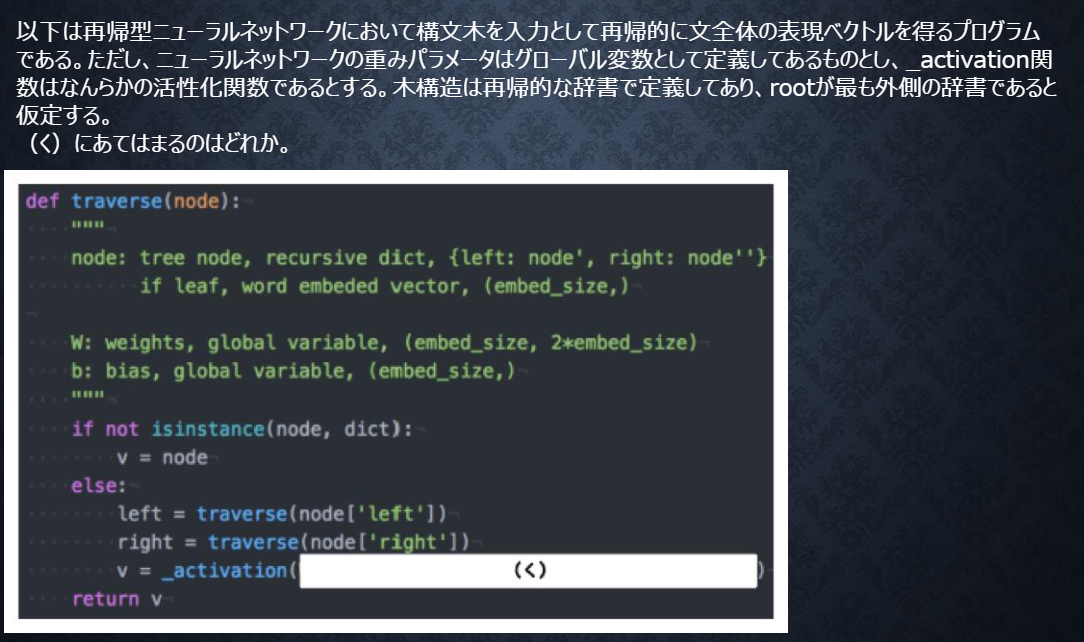
　再帰型ニューラルネットワークの概念

　RNNは時系列データに対応可能なNWのことである。時系列データは、時間順を追って一定間隔ごとに観察され、相互に統計的依存関係が認められるよなデータのシリーズ。典型的な時系列データは、音声データやテキストデータなど。BPTTは誤差逆転播の一種である。

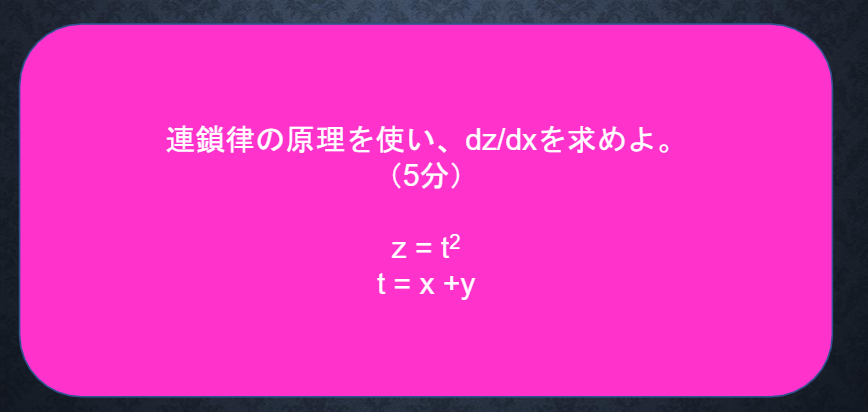
**練習問題：**

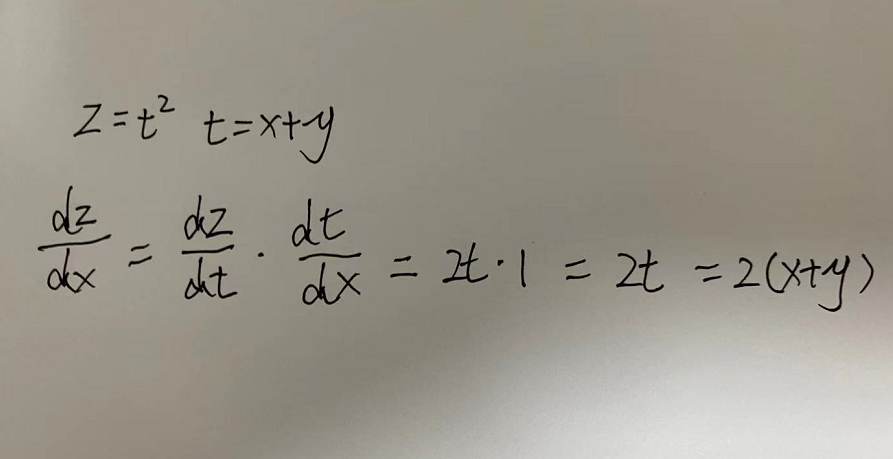


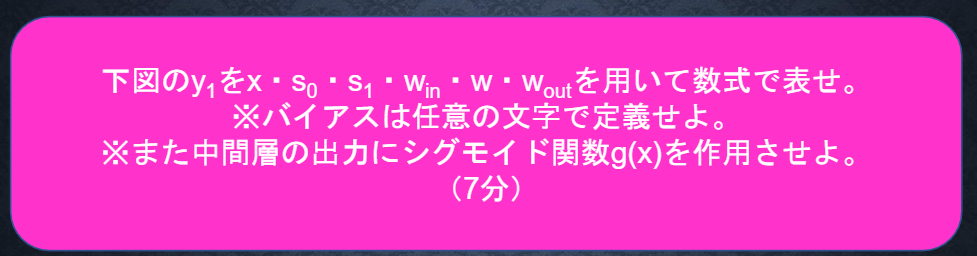
**答え：中間層の間にある重みである。**

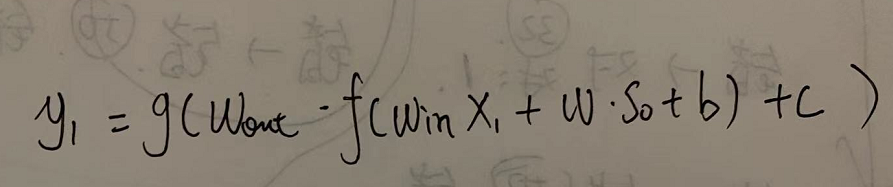


**答え：現ベクトルを作るという処理は、隣接している表現leftとrightを合わせたものを特徴量としてそこに重みを掛けることで実現出来るので、正解は２である。**

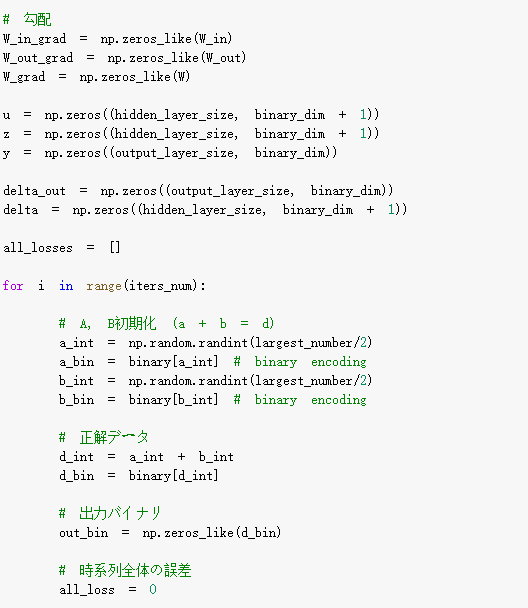


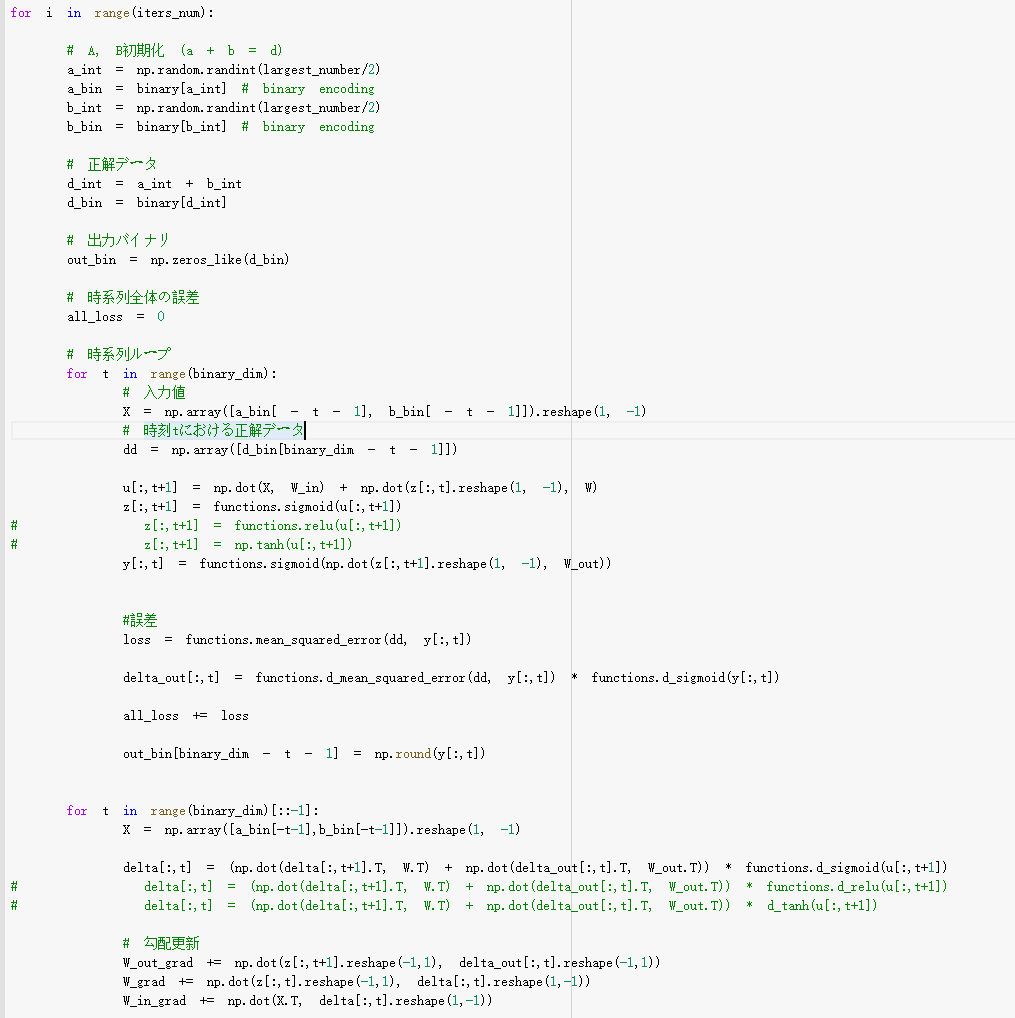
**答え：**

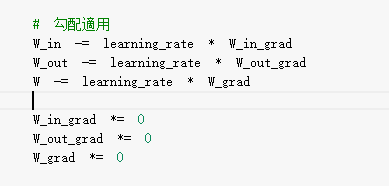


答え：

**実装演習の確認：**







**気になる点：**

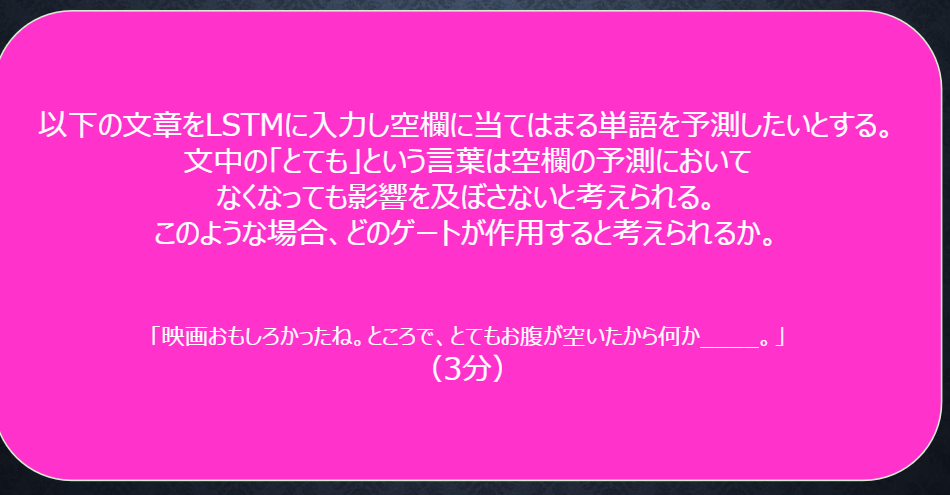
**RNN内の重みは2種類が存在していると気がする。種類１は通常通り、該当するデータが影響する重み(ｔ番目)、種類２はＮＷ全体の重みとして後ろの層に転送している重み、種類２の重みが存在するこそ、ＲＮＮは記憶みたいな能力を持つことにした。**

**まとめ要点２：**

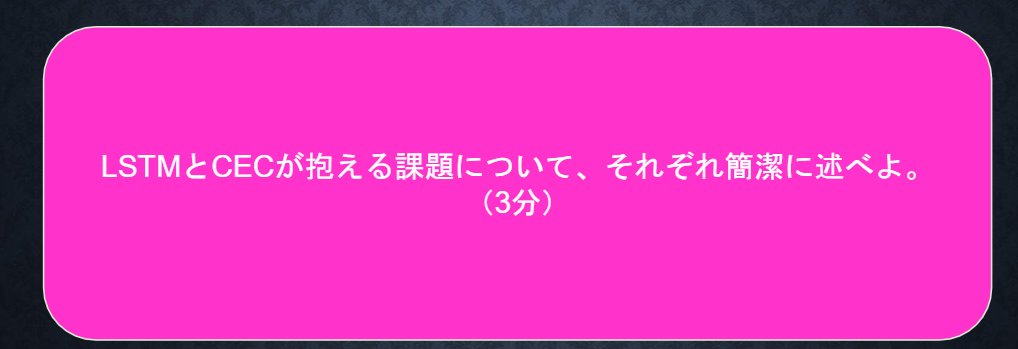
　LSTM

LSTMモデルは入力ゲートと出力ゲートと忘却ゲート、三種類のユニットが存在する。入力・出力ゲートを追加することで、それぞれのゲートへの入力値の重みを、重み行列W,Uで可変可能とした、過去の情報が要らなくなった場合、そのタイミングで情報を忘却する機能が必要。

**練習問題：**



**答え：忘却ゲート**

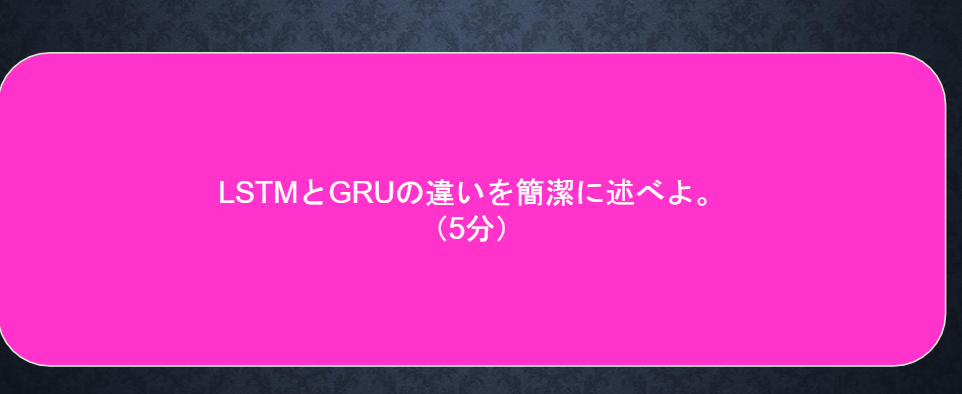


**答え：LSTMの課題は計算量が大きい、CECは記憶能力がない。**

**まとめ要点３：**

　GRU

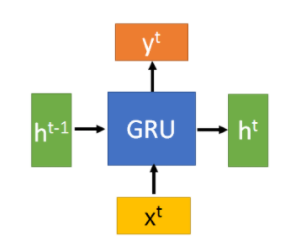
GRUはLSTMと比べてパラメータを大幅に削減し、精度は同等またはそれ以上が望める様になった構造である。LSTMの入力、出力、忘却３ゲットの機能のような、GRUはリッセト、更新ゲットを用いて、学習を行われていた。一般的、LSTMより計算量が少ないかつ精度は同等のGRUはよりいい選択である。



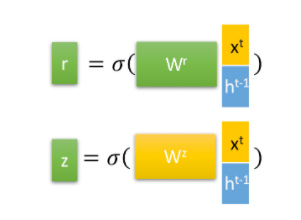
**答え：LSTMは三ゲット入力、出力、忘却、GRUはリッセト、更新、２ゲットしかない。GRUの方はLSTM計算量が大きいという問題を改善した。**

**自己考察：**

**GRUの構造とLSTMほぼ同じである。イメージ図は以下：**



前の層はからの状態ｈｔ－1と入力値ｘｔがリセットゲットと更新ゲットの状態をコントロールする：



**更新ゲットが求めらてたので、同時に忘却ゲットと選択記憶の機能を実現出来るので、LSTMは複数のゲットが必要、ここはGRUの構造はLSTMよりいい原因だと思います。**

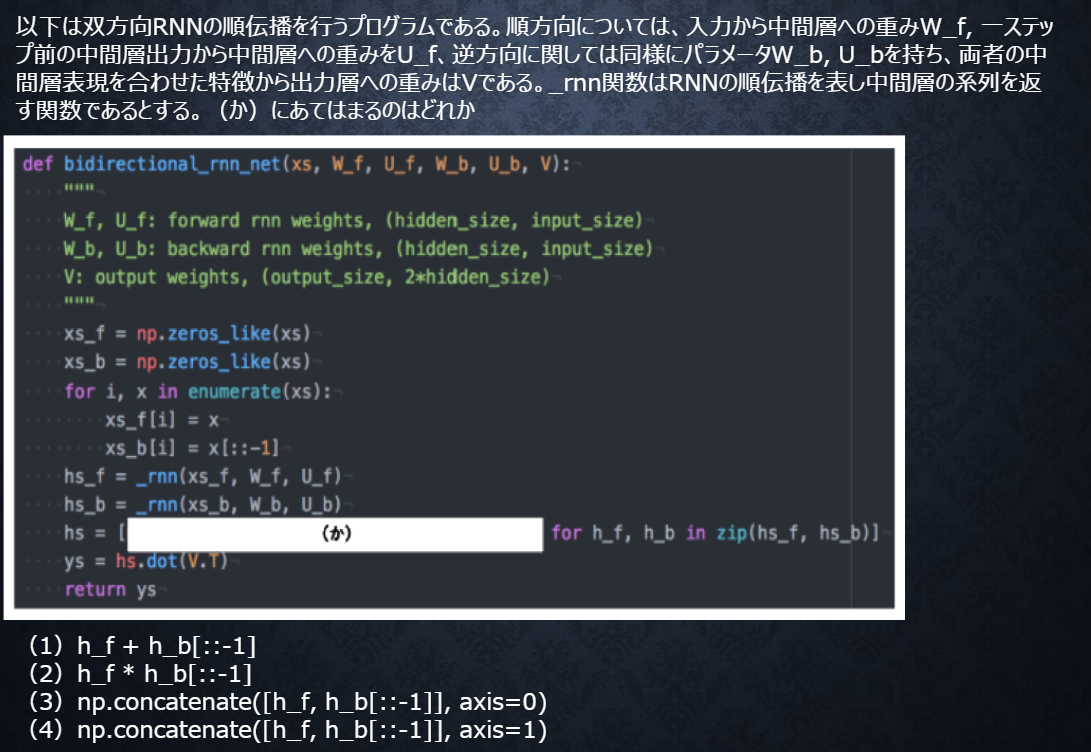
**更新式：**

**まとめ要点４：**

　双方向RNN

双方向RNNの特徴は過去の情報だけでなく、未来の情報を加味することで、精度を向上させるためのモデル、データの前後関係も考えて予測結果をもっと正確に出来るため工夫するモデルである。でも、前後関係も考えていた特化のモデルなので、実際の応用場面として、文章の推敲や、機械翻訳などの問題を解決する時、使われているですが、問題によって、使えない場合もある。

**演習チャレンジ：**

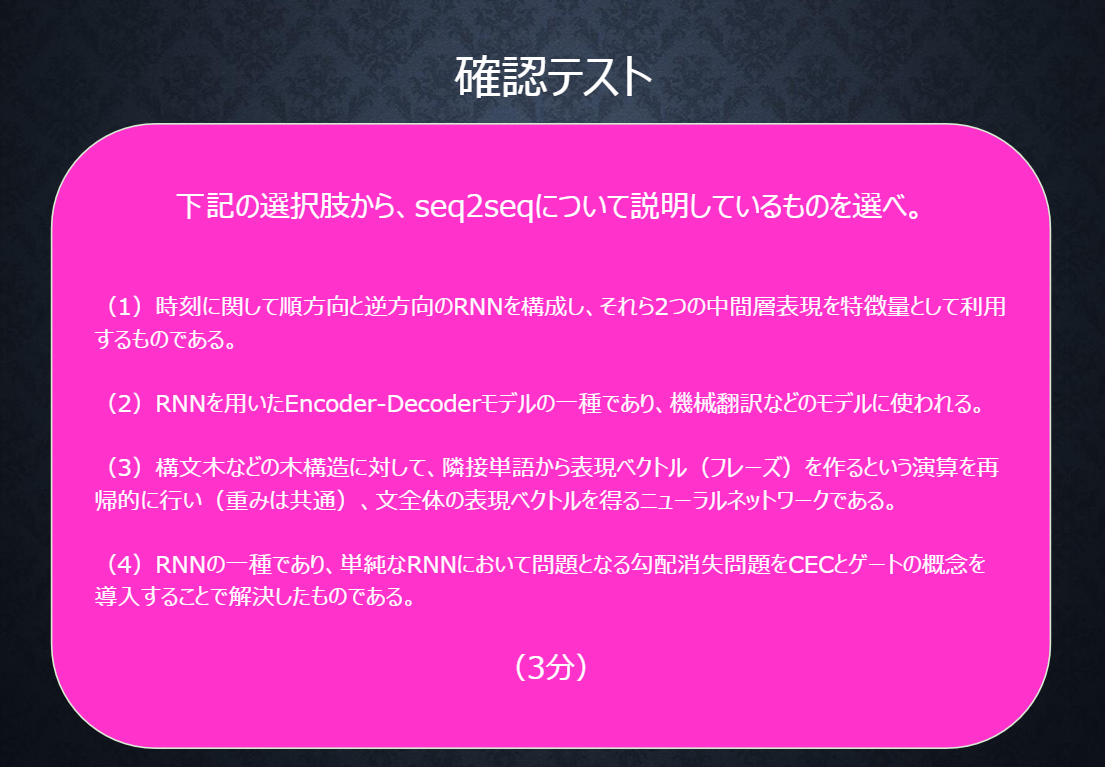


**答え：双方向RNNでは、順方向と逆方向に伝播したときの中間層表現をあわせたものが特徴量となるので、４を選択する。**

**まとめ要点５：**

　Seq2seq

Seq2seqとは機械対話や、機械翻訳などに使用されているEncoder-Decoderモデルの一種である。Encoderとは入力された言葉をRNNに入力して、パラメータを更新して、最後出力は入力した文の意味を表すベクトルとなる。Decoderは生成した意味を表すベクトルを文字に変換するものである。



**答え：まとめのように、２は正解である**



**答え：単語wはone-hotベクトルであり、それを単語埋め込みにより別の特徴量に変換する。これは埋め込み行列Eを用いて、E.dot(w)と書けるので、１は正解である。**

**まとめ要点６：**

　word2vec

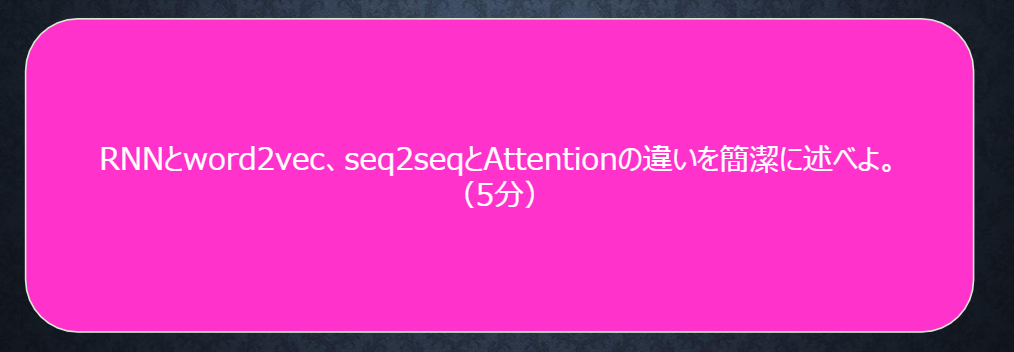
学習データからボキャブラリを作成して、one-hotベクトルに変換して大規模データの分散表現の学習が、現実的な計算速度とメモリ量で実現可能にしたモデルである。単語の言語コンテキストを再構築するように訓練された浅い2層ニューラルネットワークであり、大きなコーパスを受け取って一つのベクトル空間を生成出来る

**まとめ要点7：**

　Attention Mechanism

seq2seq は長い文章への対応が難しいという問題の解決策として、文章が長くなるほどそのシーケンスの内部表現の次元も大きくなっていく、仕組みが必要になります、それは入力と出力のどの単語が関連しているのかの関連度を学習する仕組み、それはAttention Mechanismである

確認テスト：



**答え：RNNは時系列データを処理するためのNWである。word2vec単語の分散表現をベクトルを得る手法である。seq2seqはある時系列データから別の時系列データを得る手法である。Attention Mechanismは時系列データの関連性をあらわすため、重みをつける手法である。**