

Seq2seq

Encoder-Decoderモデルの理解と実装

2019/3/11

BERTまでのロードマップ

BERTを理解するために必要な材料

Encoder-Decoder Model



Transformer
(Encoder-Decoder x Attention)



BERT

Seq2seqとは？

系列(Sequence)を入力として、系列を出力するもの

- Encoder-Decoderモデルとも呼ばれる
- 入力系列がEncode(内部状態に変換)され、内部状態からDecode(系列に変換)する
- 実応用上も、入力・出力共に系列情報なものは多い
 - 翻訳 (英語→日本語)
 - 音声認識 (波形→テキスト)
 - チャットボット (テキスト→テキスト)

Seq2Seqの理解に必要な材料

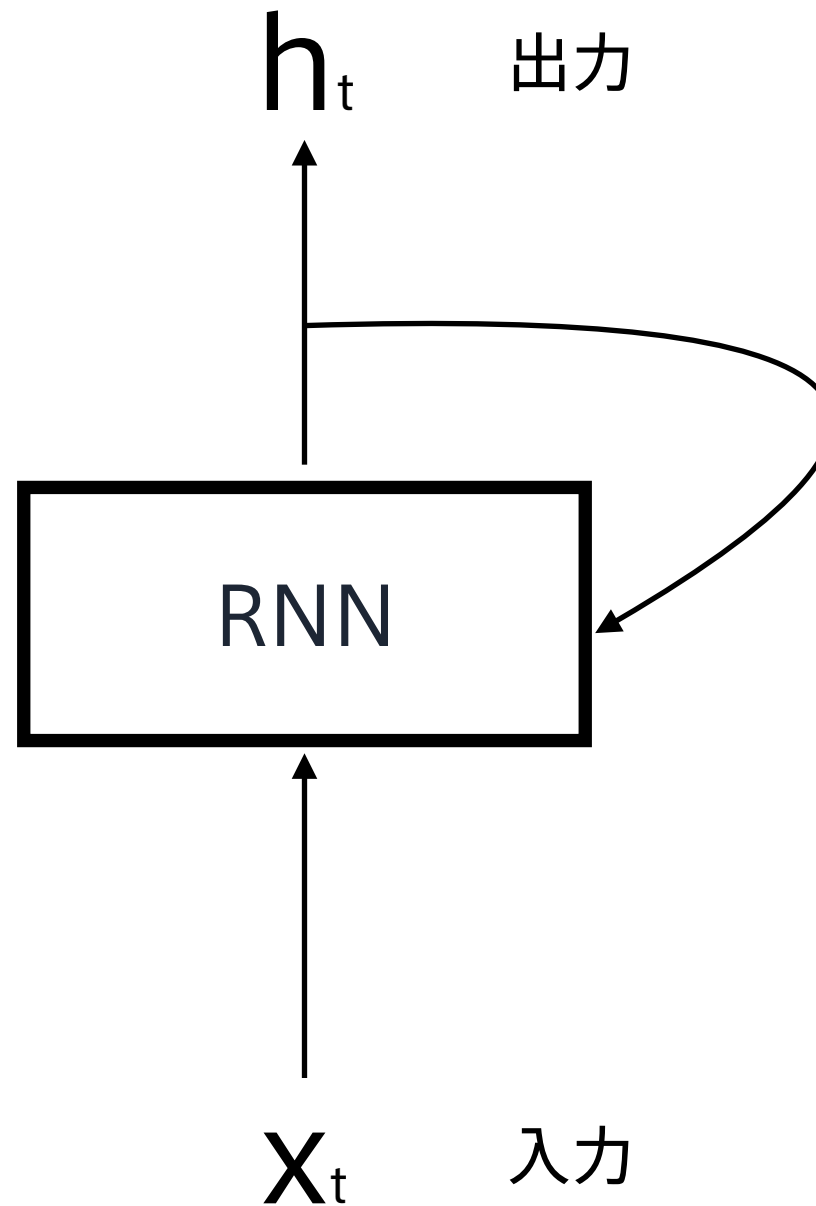
要は言語モデルをふたつ連結した形になっている

- RNNの理解
 - RNNの動作原理
 - LSTMなどの改良版RNNの理解
- 言語モデルの理解

RNNとは

系列データを読み込むために再帰的に動作するNN

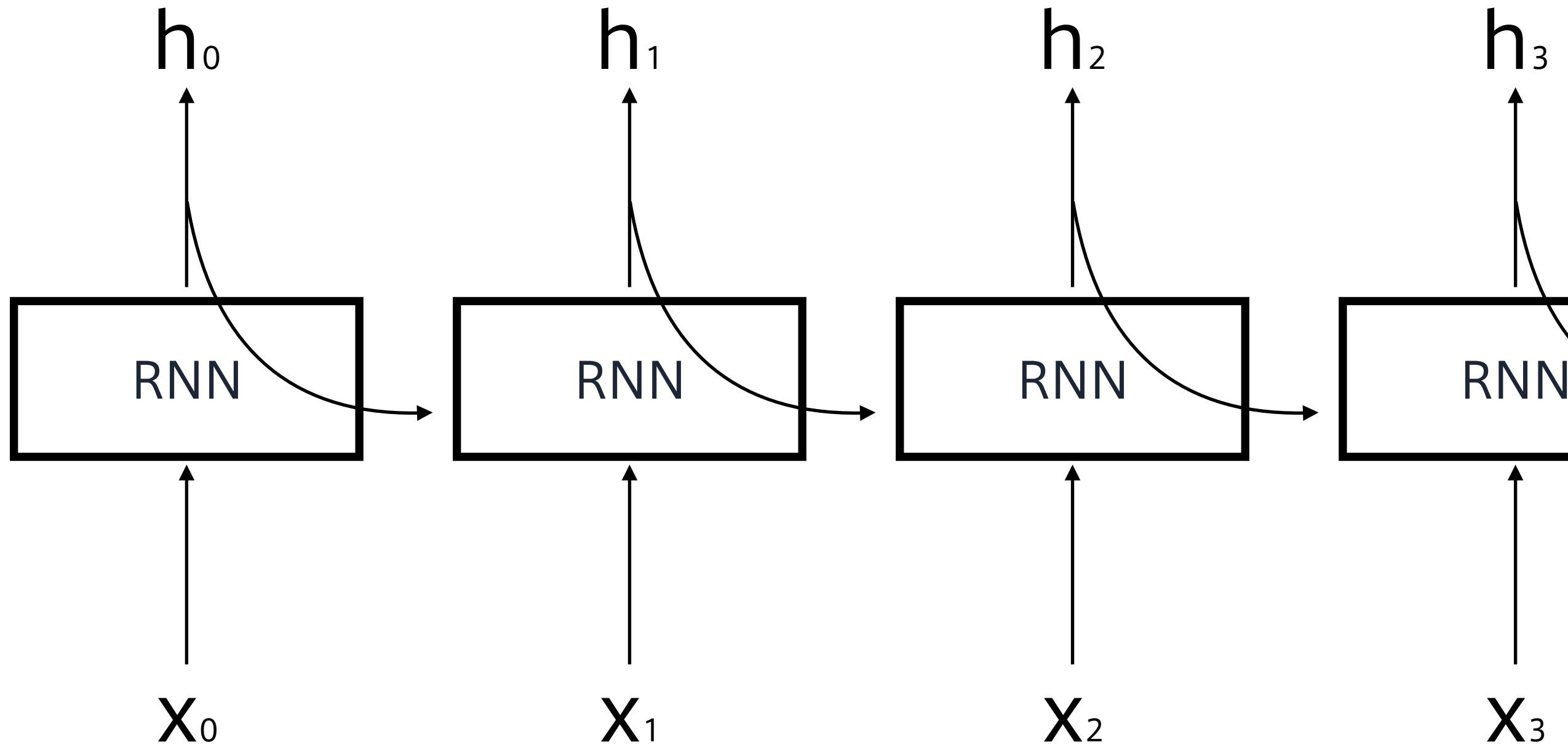
- 再帰的とは？



- 出力が再び入力になる

RNNとは

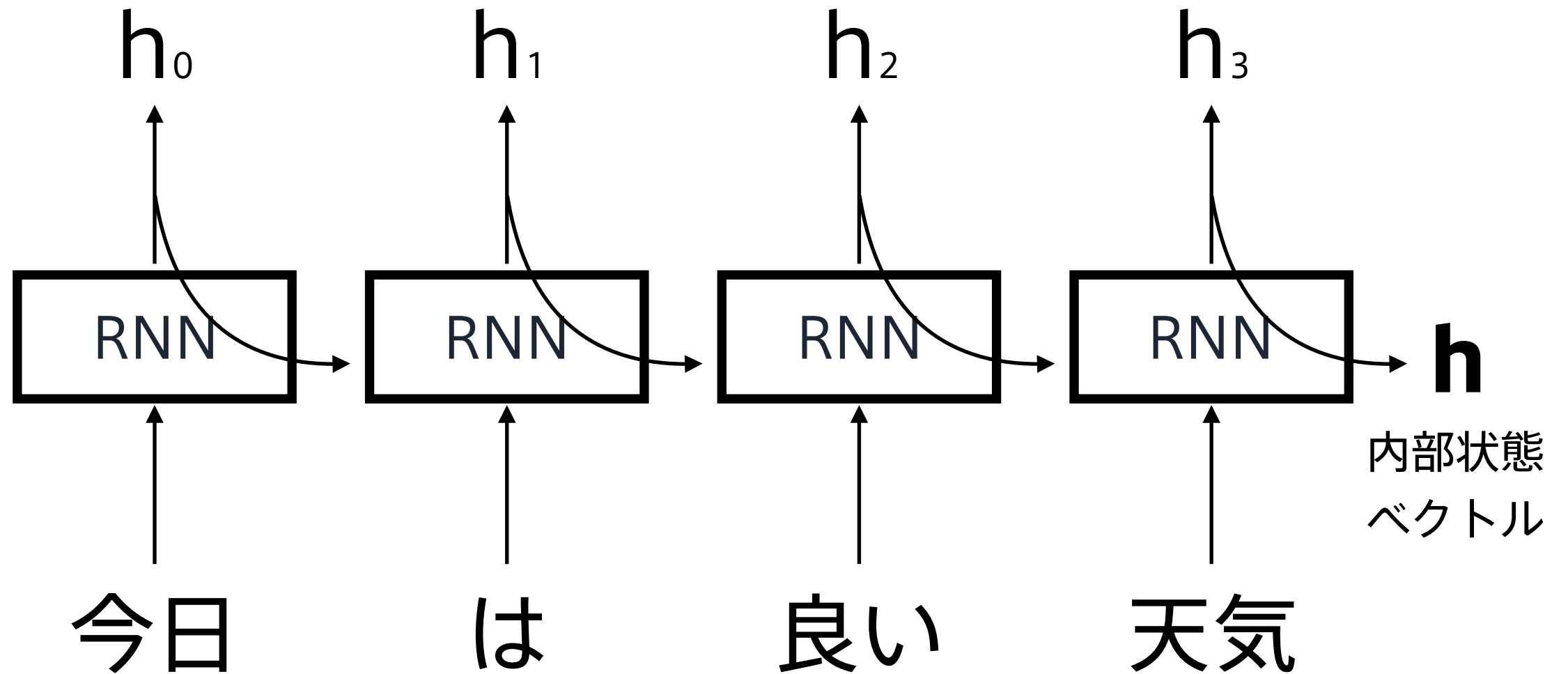
再帰処理は時間軸方向に展開できる



前の時刻の出力を現在の時刻の入力にする

RNNとは

系列情報を舐めて内部状態に変換できる



言語モデルとは

言語モデルは単語の並びに確率を与える

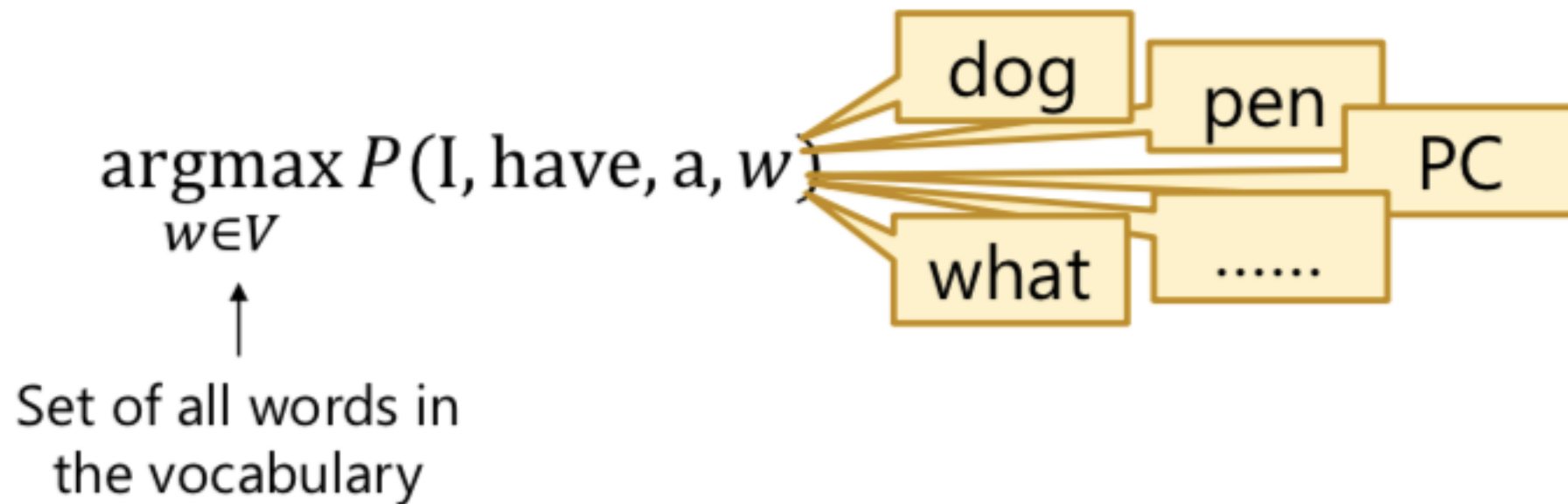
- 単語の並びに対して尤度(それがどれだけ起こり得るか)、すなわち、文章として自然かを確率で評価する
- 例)
 - You say goodbye → 0.092 (自然)
 - You say good die → 0.00000032 (不自然)
- 数式的には同時確率を事後確率に分解して表せる

$$P(w_1, \dots, w_m) = \prod_{i=1}^m P(w_i | w_1, \dots, w_{i-1})$$

言語モデルとは

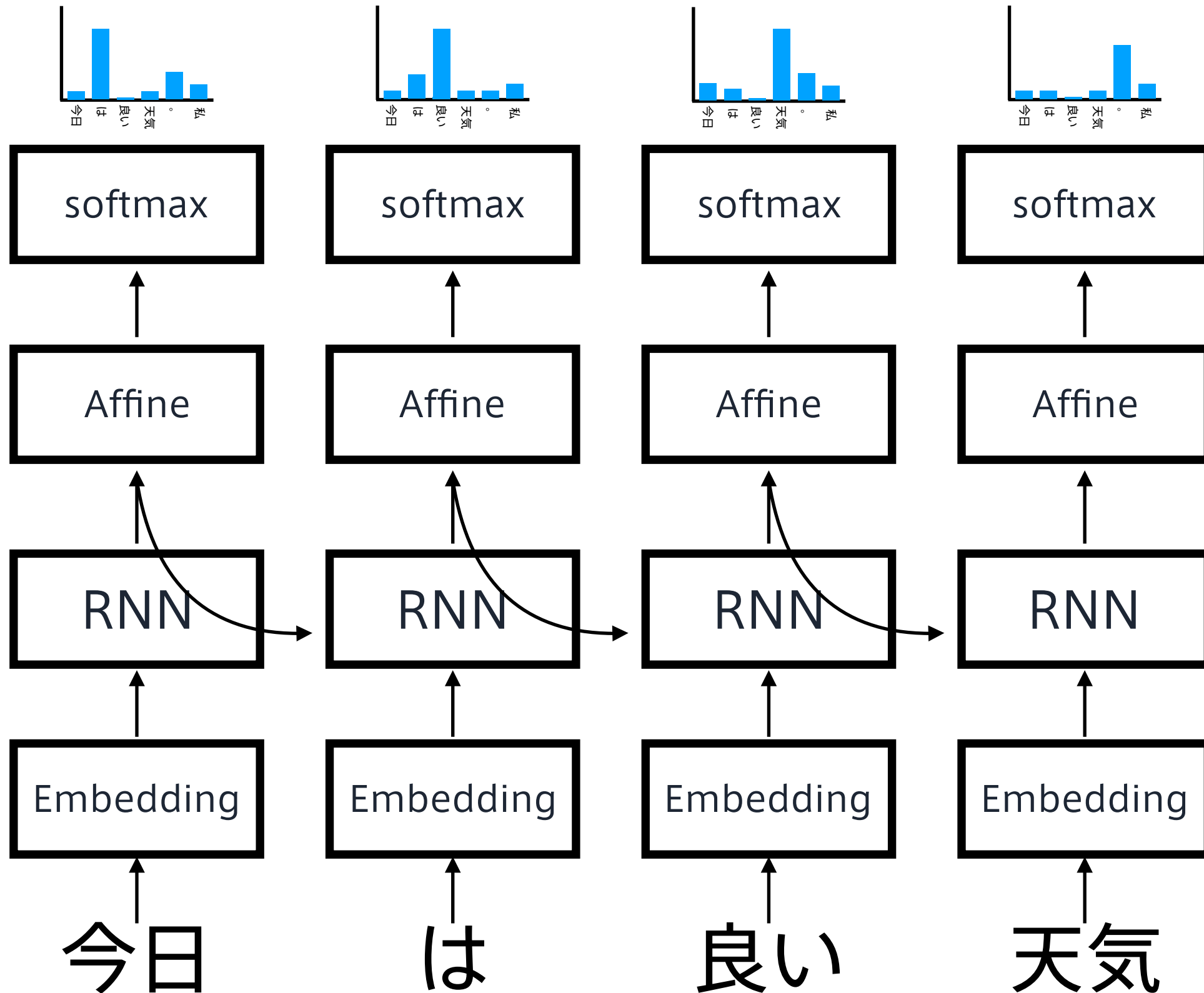
時刻t-1までの情報で、時刻tの事後確率を求めることが目標

→これで同時確率が計算できる



RNN x 言語モデル

各地点で次にどの単語が来れば自然(事後確率最大)かを出力できる



RNN x 言語モデル

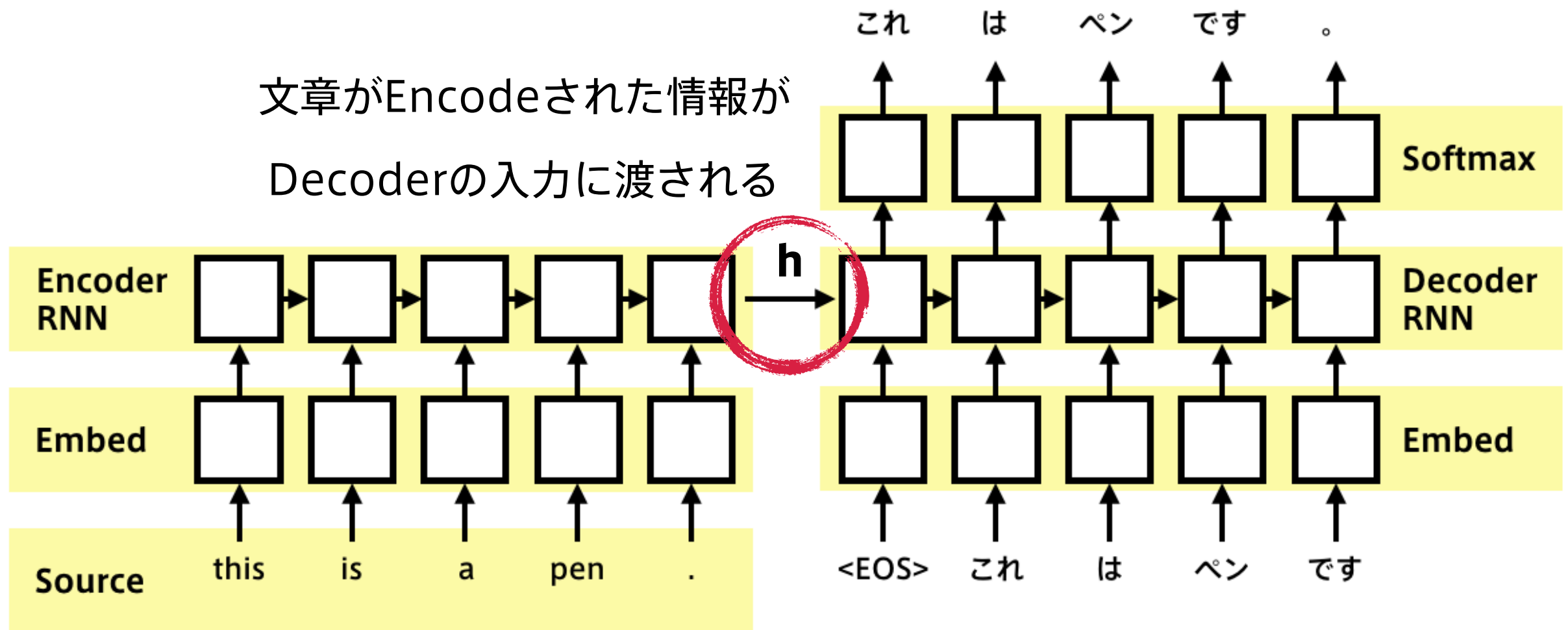
まとめると

- RNNは系列情報を内部状態に変換することができる
- 文章の各単語が現れる際の同時確率は、事後確率で分解できる
 - したがって、事後確率を求めることがRNNの目標になる
- 言語モデルを再現するようにRNNの重みが学習されていれば、ある時点の次の単語を予測することができる
 - 先頭単語を与えれば文章を生成することも可能

Seq2seq

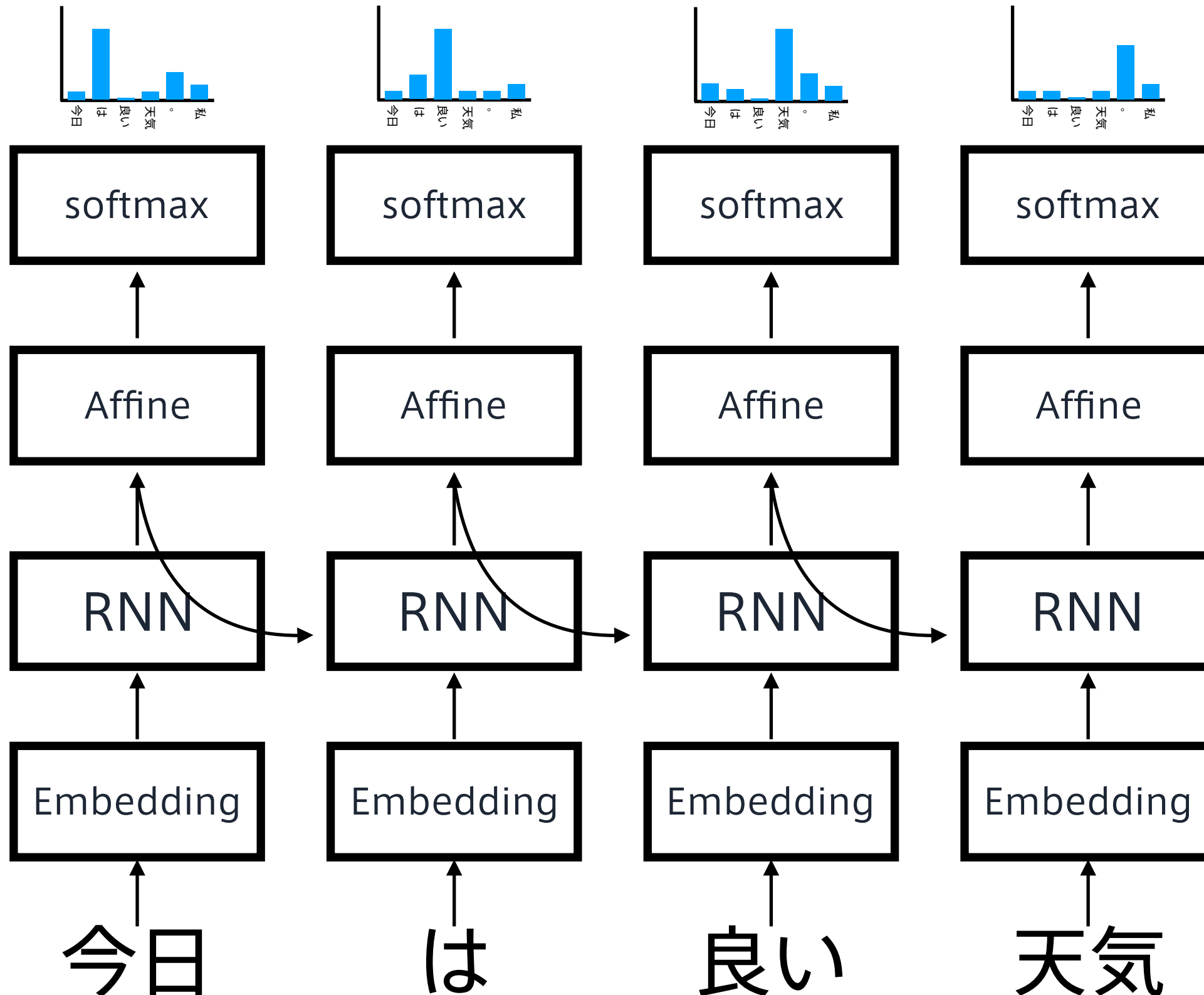
EncoderからDecoderに渡される内部状態ベクトルが鍵

文章がEncodeされた情報が
Decoderの入力に渡される



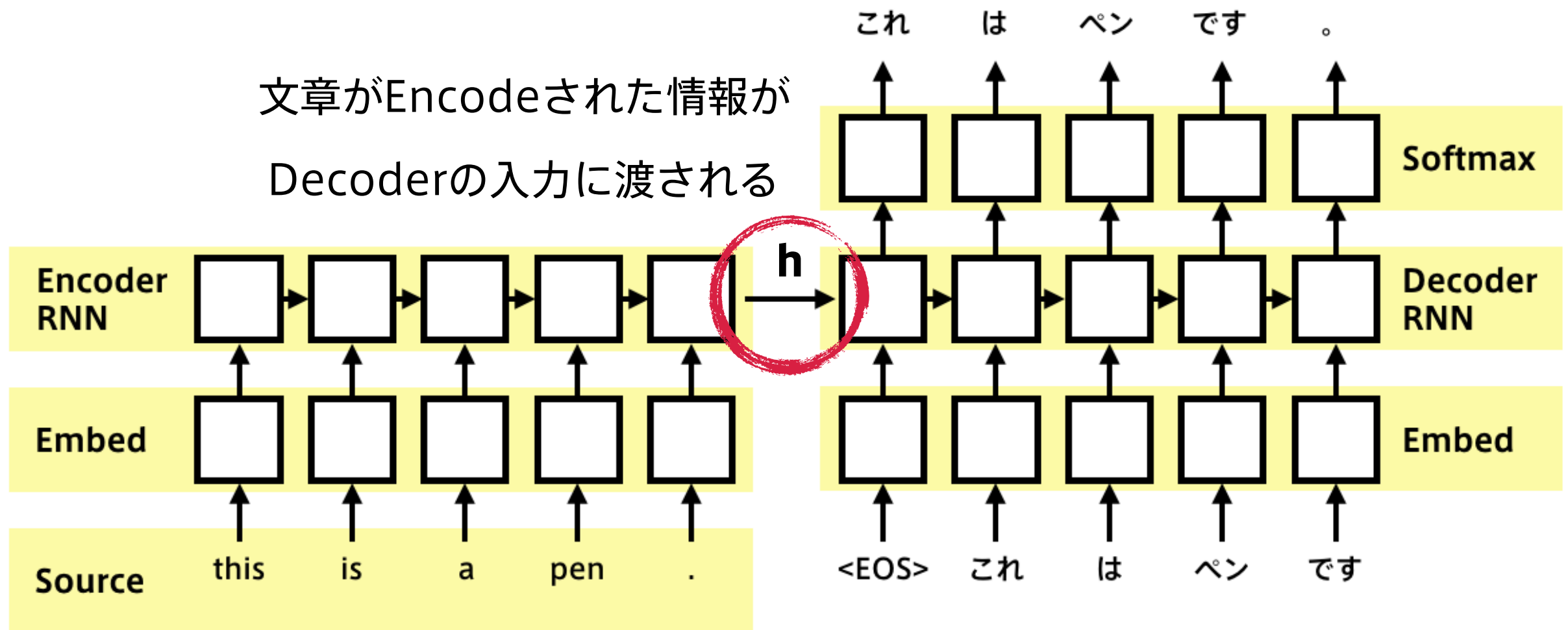
RNN x 言語モデル

各地点で次にどの単語が来れば自然(事後確率最大)かを出力できる



Seq2seq

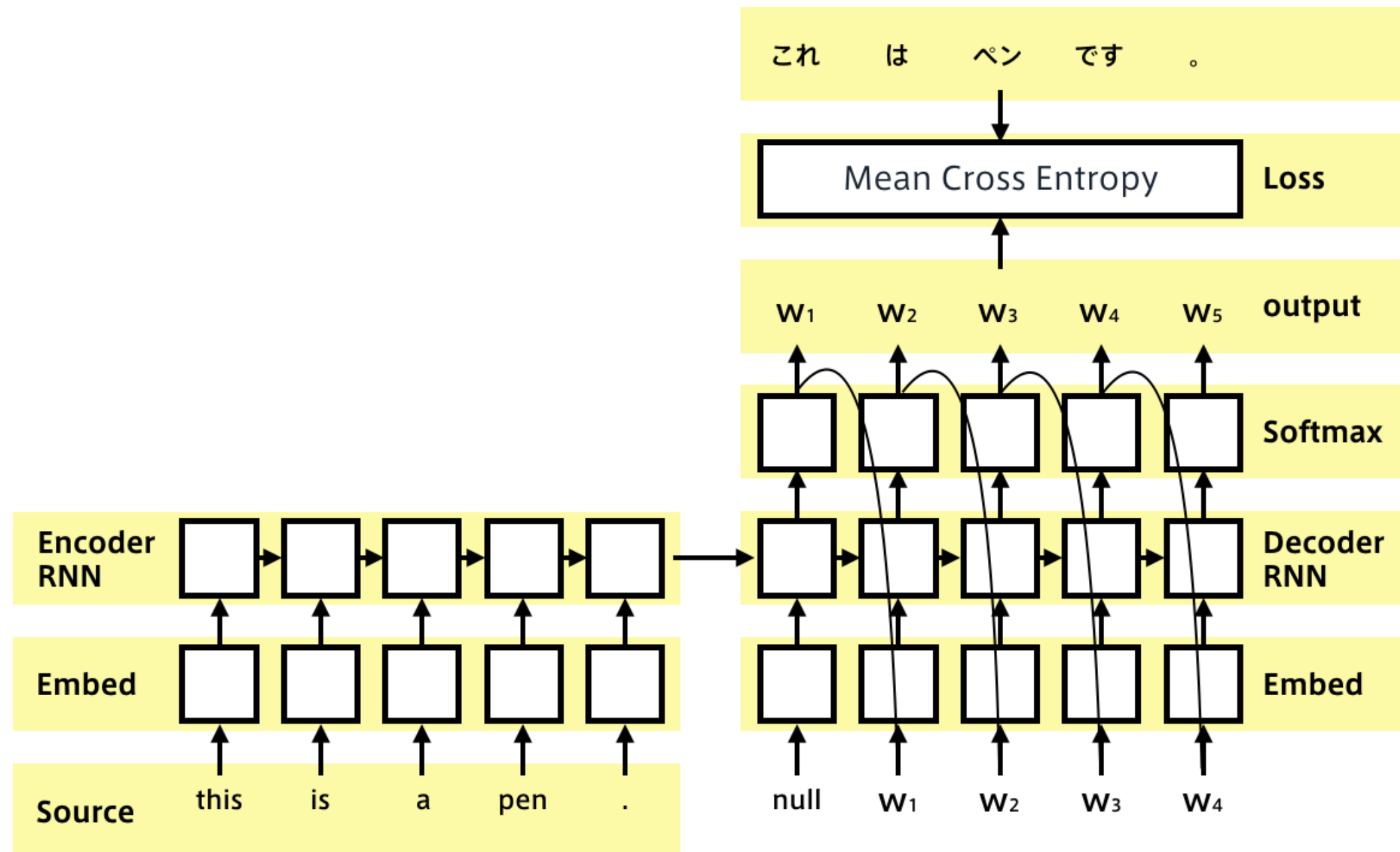
EncoderからDecoderに渡される内部状態ベクトルが鍵



Decoder側の構造は言語モデルRNNとほぼ同じだが
隠れ状態の初期値にEncoder側の内部状態を受け取る

Seq2seq

EncoderからDecoderに渡される内部状態ベクトルが鍵

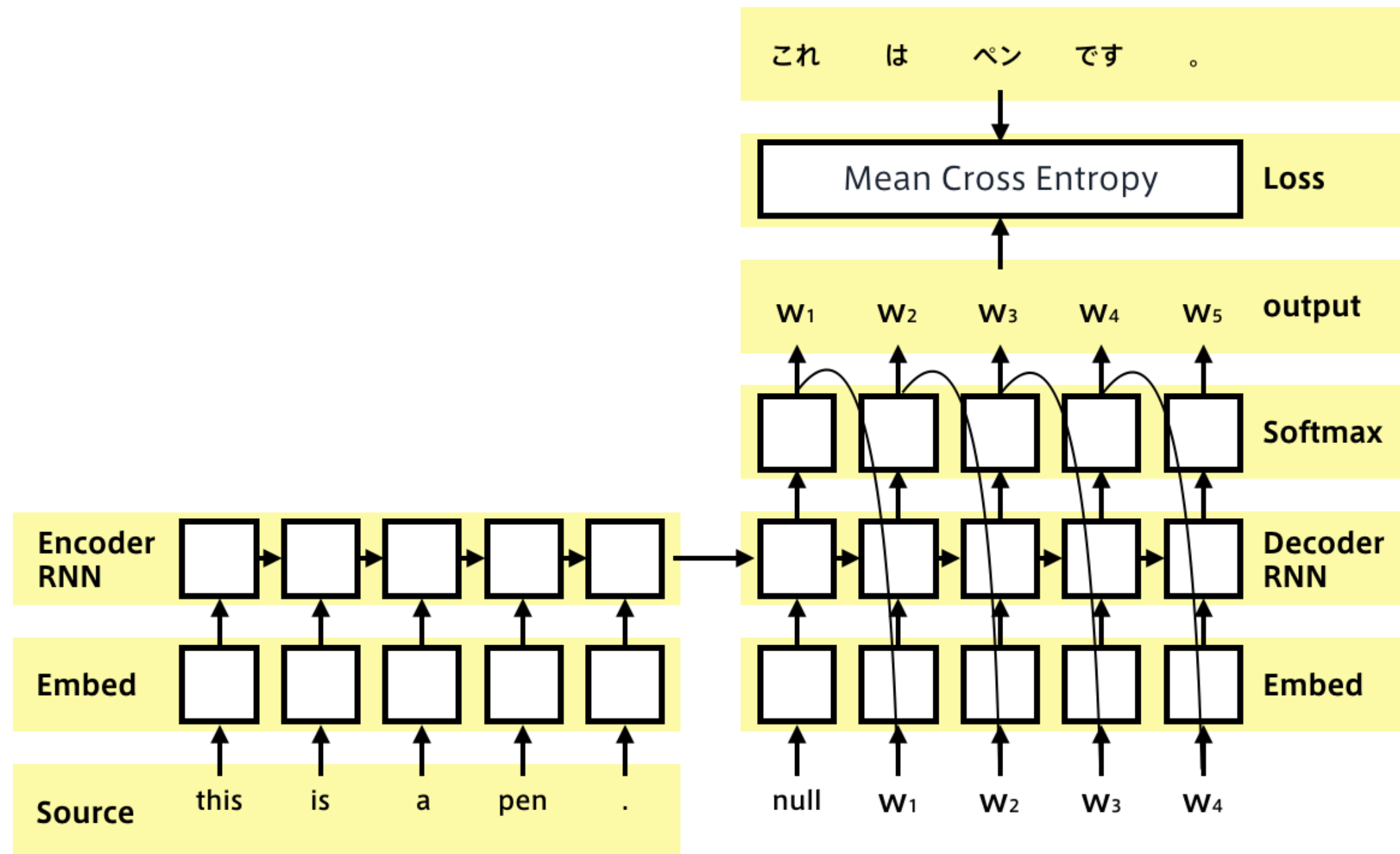


Decoderのoutput側に正解を当てれば教師あり学習が
End2endで行える

<introduction/> <implementation>

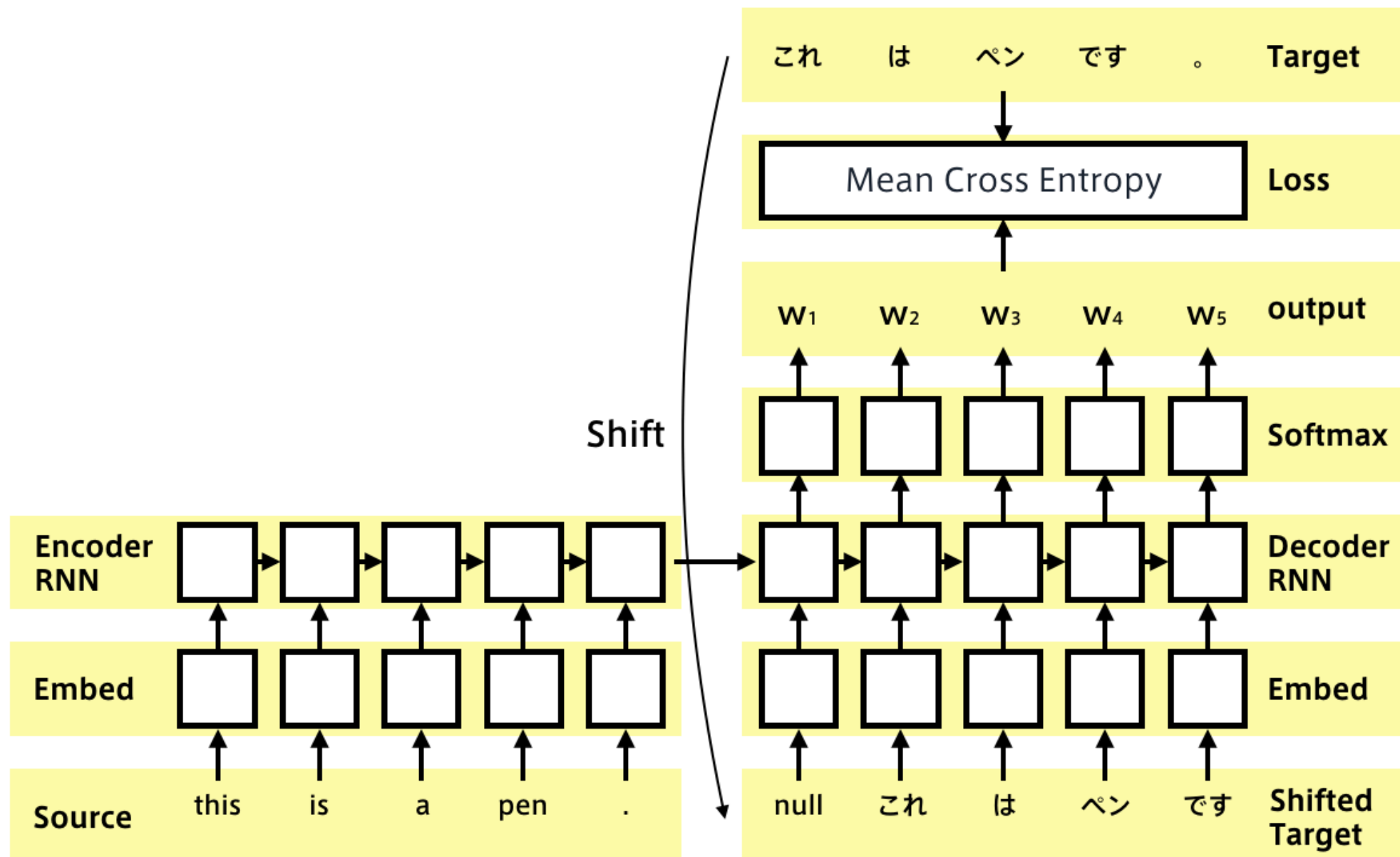
Teacher Forcing

Q. このモデルで起こる問題は？



Teacher Forcing

正解ラベルを直接Decoderの入力にする



Teacher Forcing

Q. このモデルで起こる問題は？

