

RNNと自然言語処理

RNNの基礎からAttention～BERTの概要まで

スカイマインド株式会社

本橋 和貴

自己紹介

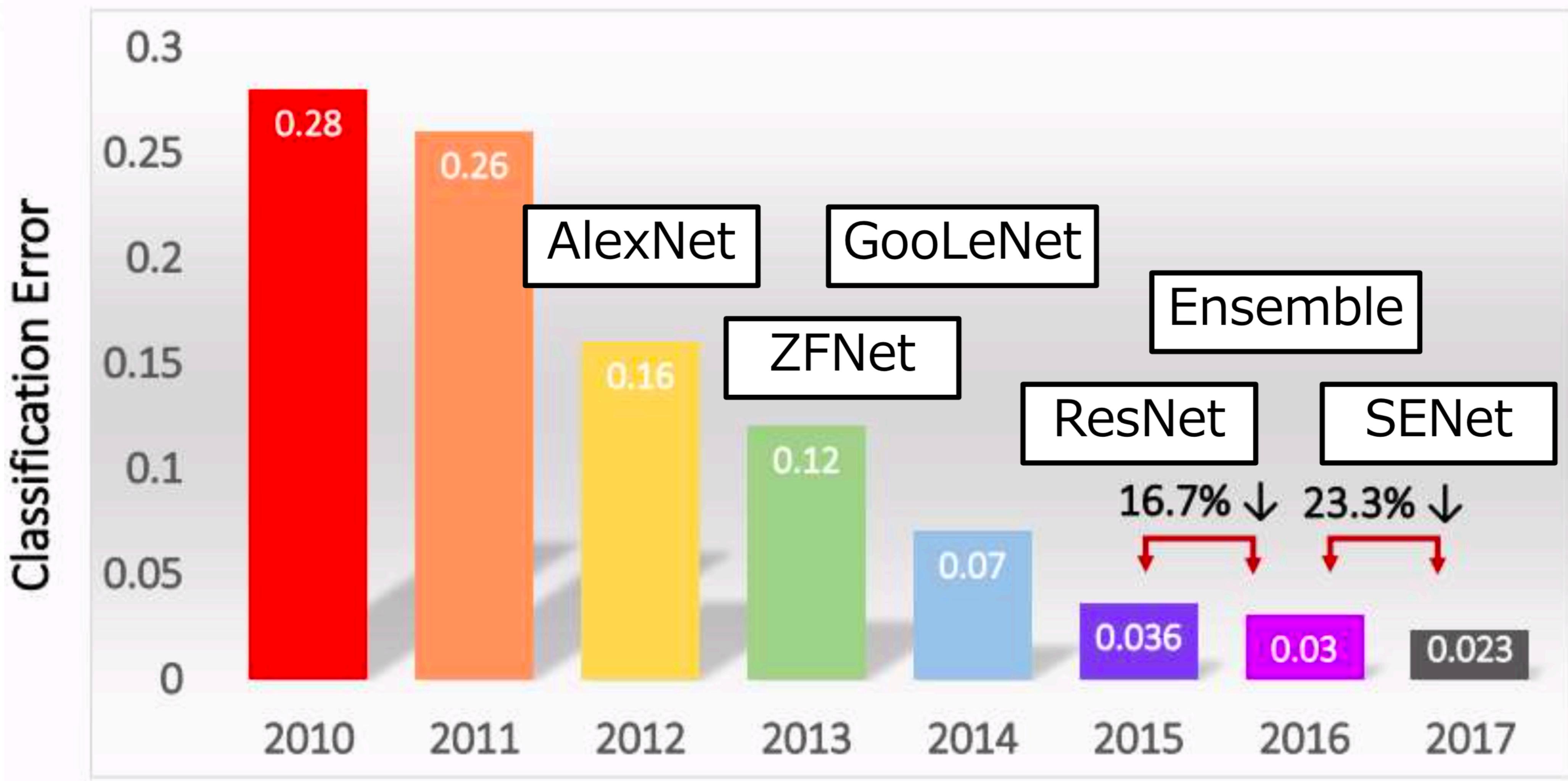
▶ 本橋 和貴 @kmotohas

- スカイマインド株式会社
 - Deep Learning Engineer (前職ではDL+ROS)
- 素粒子物理学実験 (LHC-ATLAS実験) 出身
 - 博士 (理学)
- 好きな定理 : ネーターの定理



系に連続的な**対称性**がある場合はそれに対応する**保存則**が存在すると述べる定理である。(Wikipedia)

ImageNet コンペの優勝モデル



ImageNet データセットと畳み込みニューラルネット

Java sparrow, Java finch, ricebird, Padda oryzivora

Small finch-like Indonesian weaverbird that frequents rice fields

877 pictures 29.19% Popularity Percentile Wordnet IDs

Numbers in brackets: (the number of synsets in the subtree).

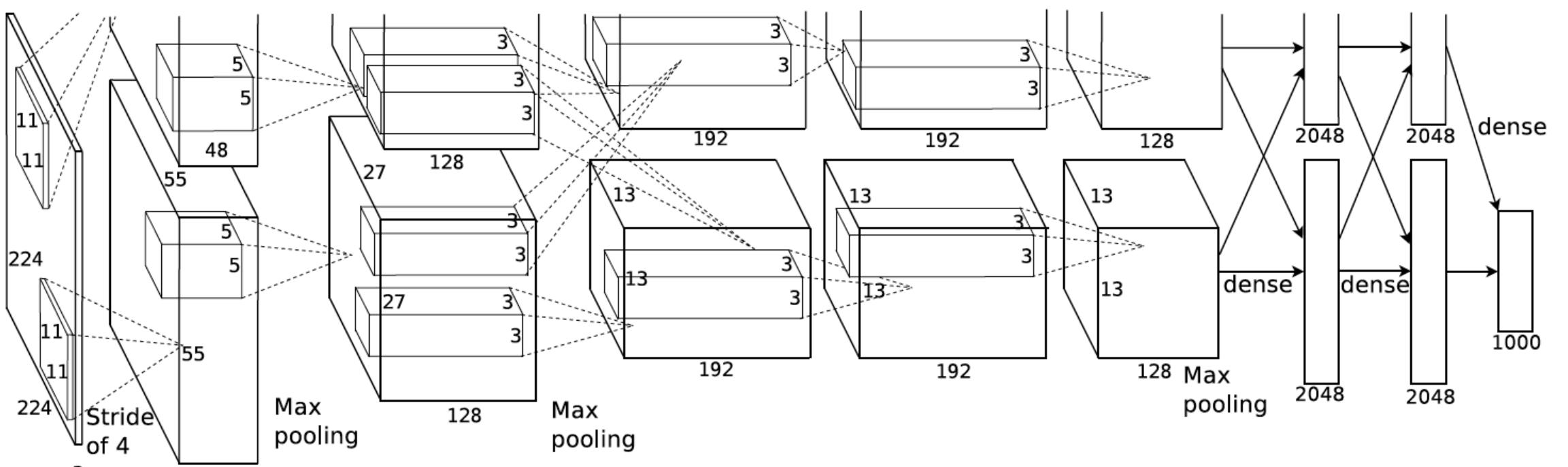
ImageNet 2011 Fall Release (32326)

- plant, flora, plant life (4486)
- geological formation, formation (11)
- natural object (1112)
- sport, athletics (176)
- artifact, artefact (10504)
- fungus (308)
- person, individual, someone, some
- animal, animate being, beast, brute (766)
- invertebrate (766)
- homeotherm, homiotherm, hor
- work animal (4)
- darter (0)
- survivor (0)
- range animal (0)
- creepy-crawly (0)
- domestic animal, domesticated
- molter, moulter (0)
- varmint, varment (0)
- mutant (0)
- critter (0)
- game (47)
- young, offspring (45)
- poikilotherm, ectotherm (0)
- herbivore (0)
- peeper (0)
- pest (1)
- female (4)
- insectivore (0)

Treemap Visualization Images of the Synset Downloads

*Images of children synsets are not included. All images shown are thumbnails. Images may be subject to copyright.

Prev 1 2 3 4 5 6 7 8 9 10 ... 25 26 Next



<https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637>

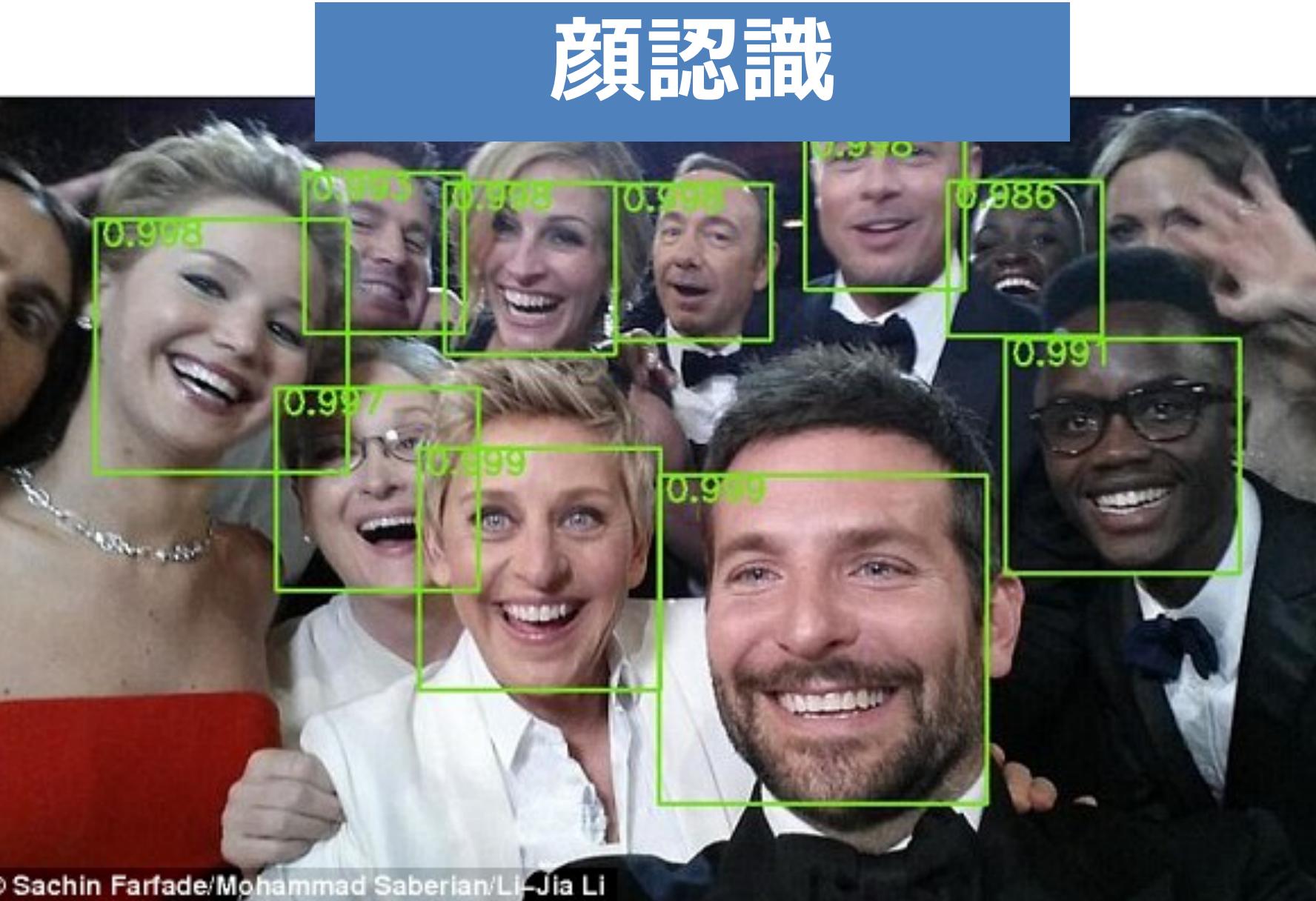
ImageNetの大量画像で訓練済みのニューラルネットワークを
特徴抽出機として別の画像認識系タスクに再利用可能

skymind | 画像認識のアプリケーション

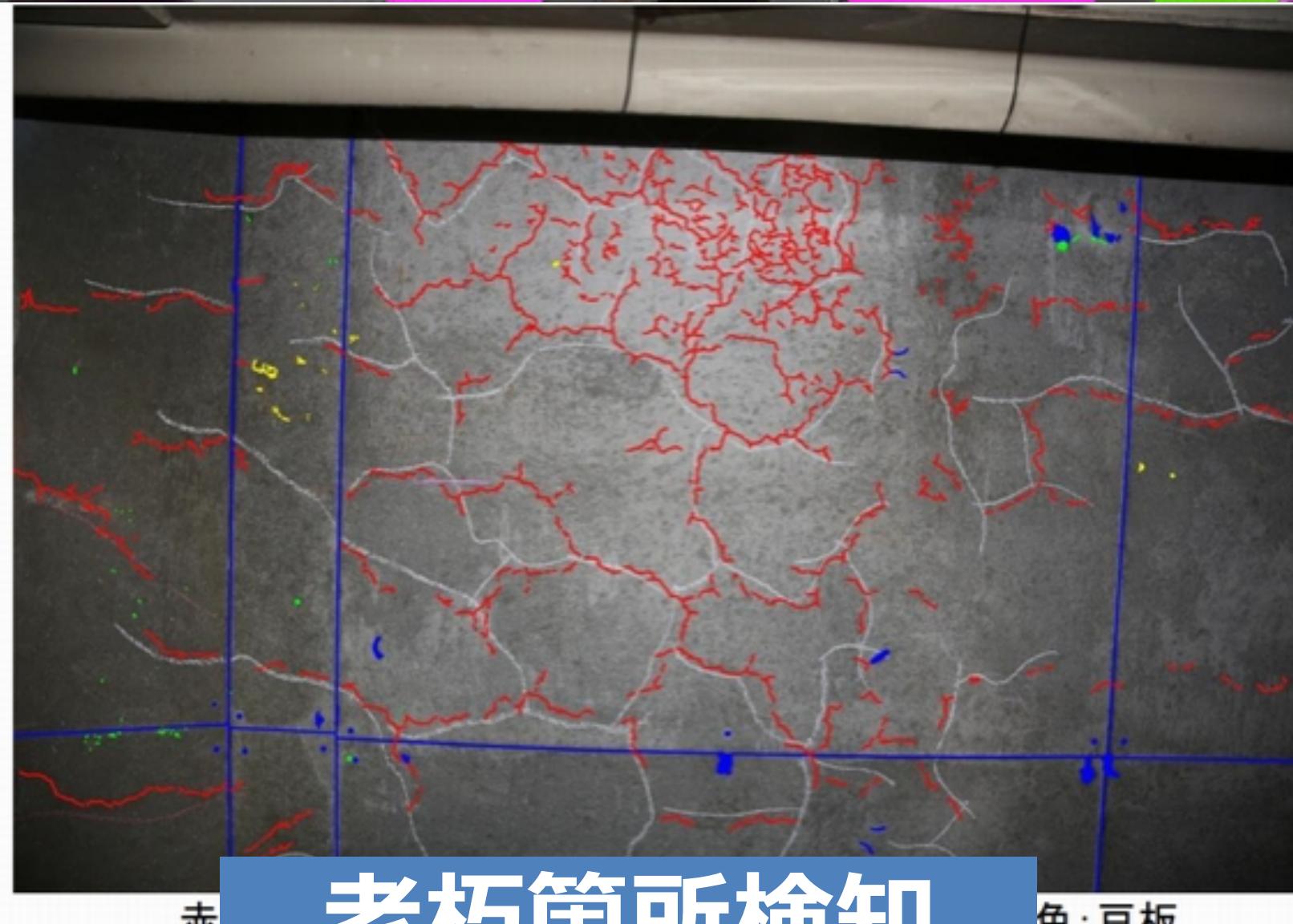
一般物体認識



顔認識



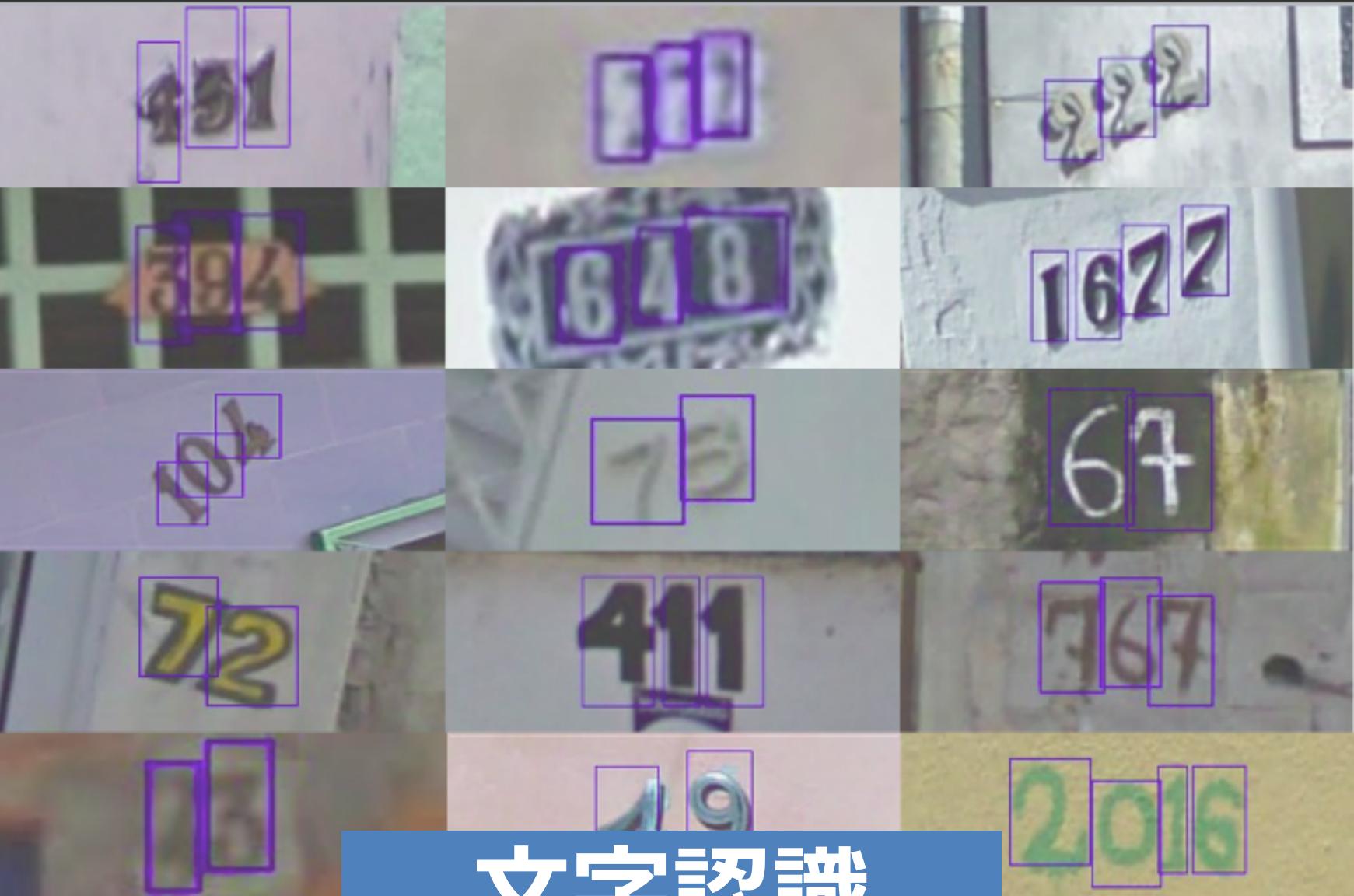
年齢推定



老朽箇所検知

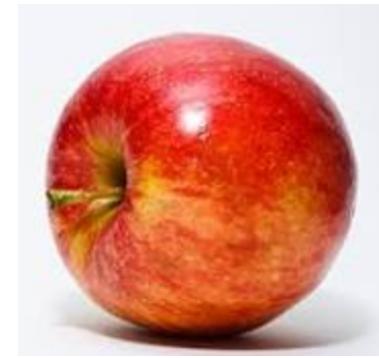


セグメンテーション



文字認識

コンピュータービジョンと自然言語処理



Computer
Vision

Natural
Language
Processing

鳥

りんご

物理世界

記号世界



物理的な光が解析対象

概念的な言葉が解析対象

BERT：自然言語処理用の汎用特徴抽出機

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language

GLEU (8種のNLPタスク、スコアは平均)

Rank	Name	Model	URL	Score
1	Jacob Devlin	BERT: 24-layers, 1024-hidden, 1		80.4
2	Alec Radford	Singletask Pretrain Transformer OpenAI Transformer		72.8
3	Samuel Bowman	BiLSTM+ELMo+Attn ELMo		70.5

SQuAD (質疑応答タスク)

Rank	Model	EM	F1
	Human Performance Stanford University (Rajpurkar et al. '16)	82.304	91.221
1 Oct 05, 2018	BERT (ensemble) Google AI Language https://arxiv.org/abs/1810.04805	87.433	93.160
2 Sep 26, 2018	nlnet (ensemble) Microsoft Research Asia	85.954	91.677

言語モデルを事前学習した
エンコーダを質疑応答や
タグ付け等のNLPタスクに
転移学習する研究やでっ

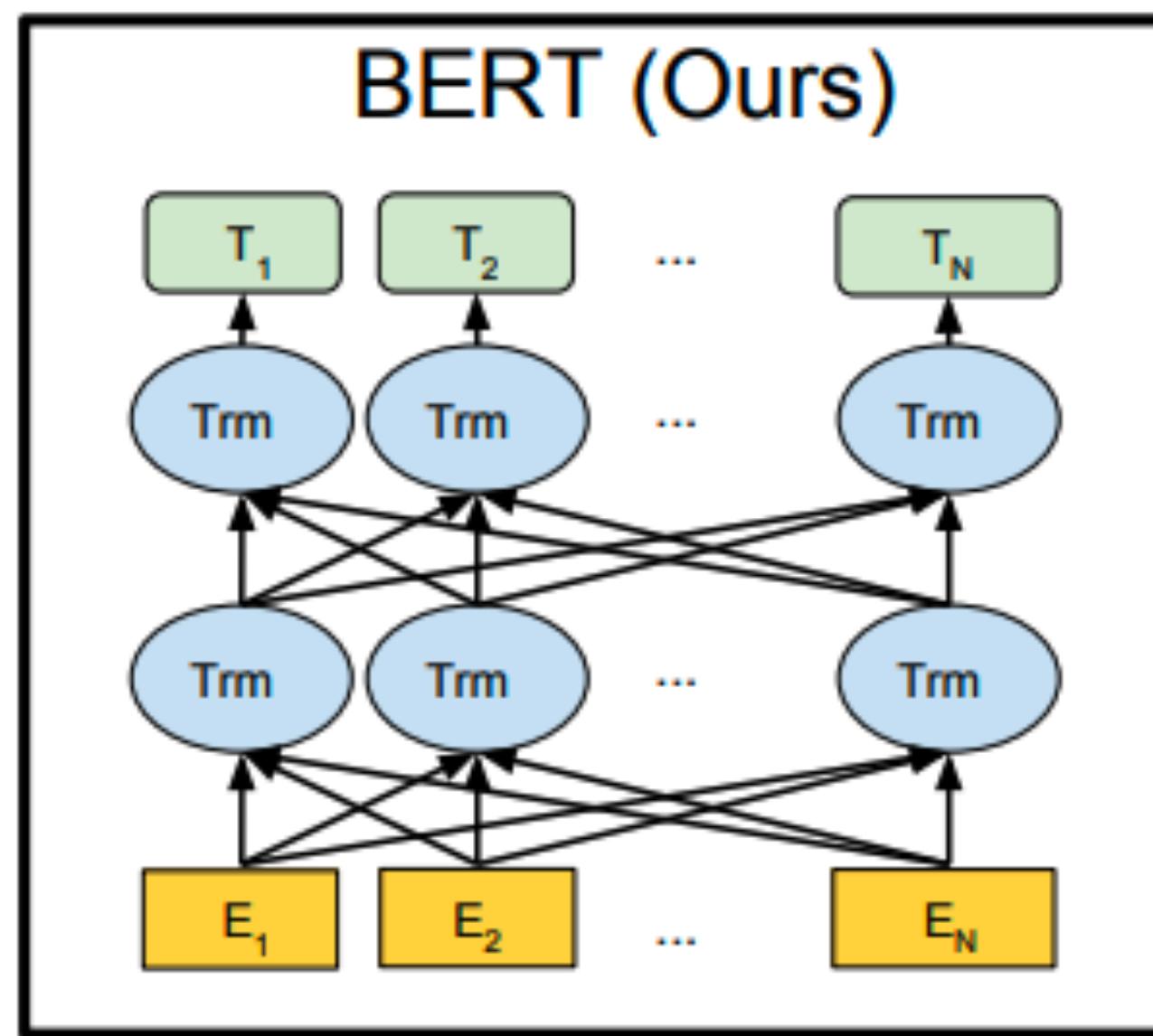
ぶっちぎりのSOTAやん！
こんなん勝てんわっ！



(引用元) https://twitter.com/_Ryobot/status/1050925881894400000/

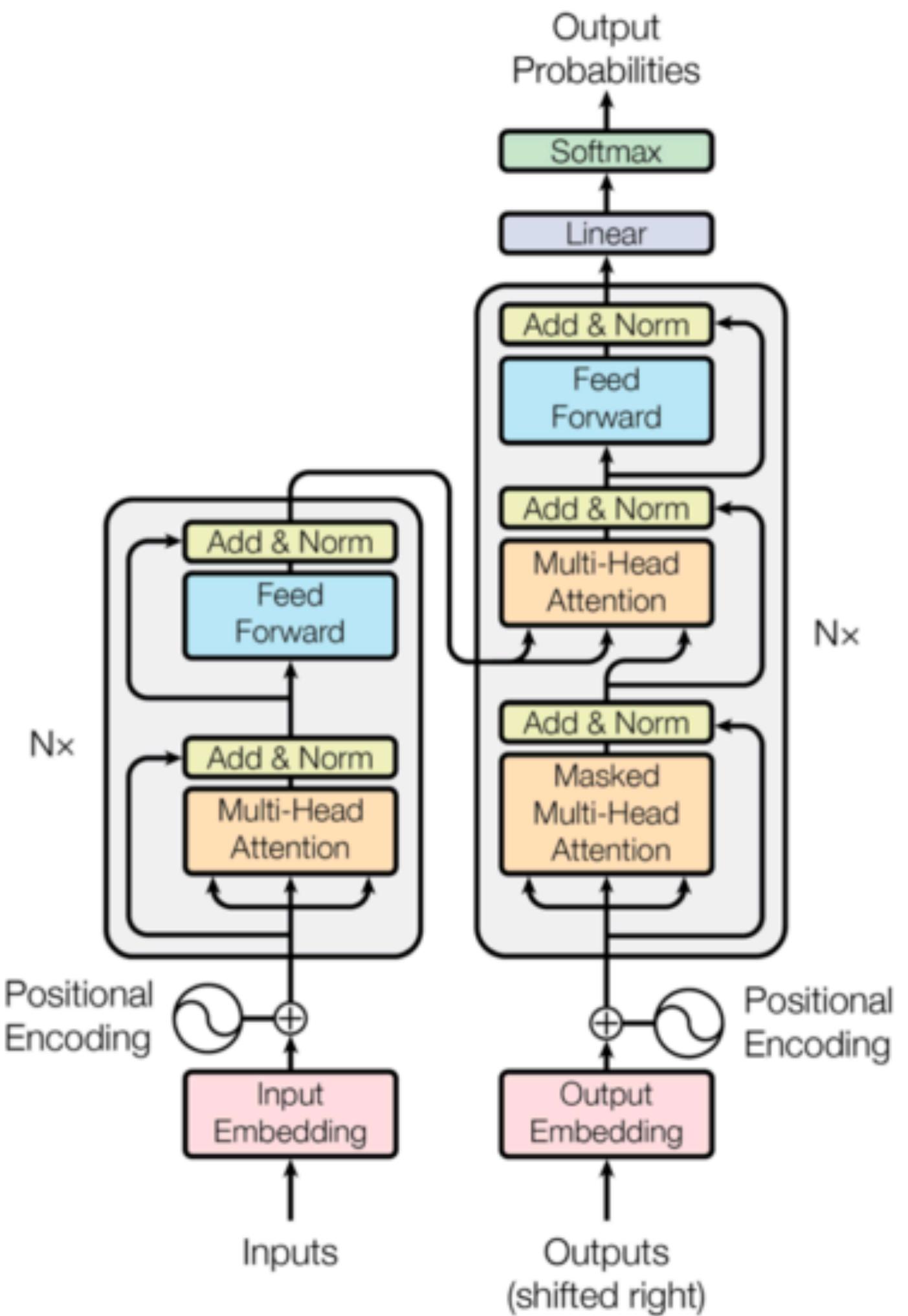
BERTとは？

- ▶ Bidirectional Encoder Representations from Transformers
 - 双方向のTransformerを用いて言語表現をエンコードするモデル



Transformerとは？

- ▶ Googleのニューラル機械翻訳モデル
 - RNNを使わずAttention機構のみ利用



というわけでBERTの雰囲気の理解が目標

► Agenda

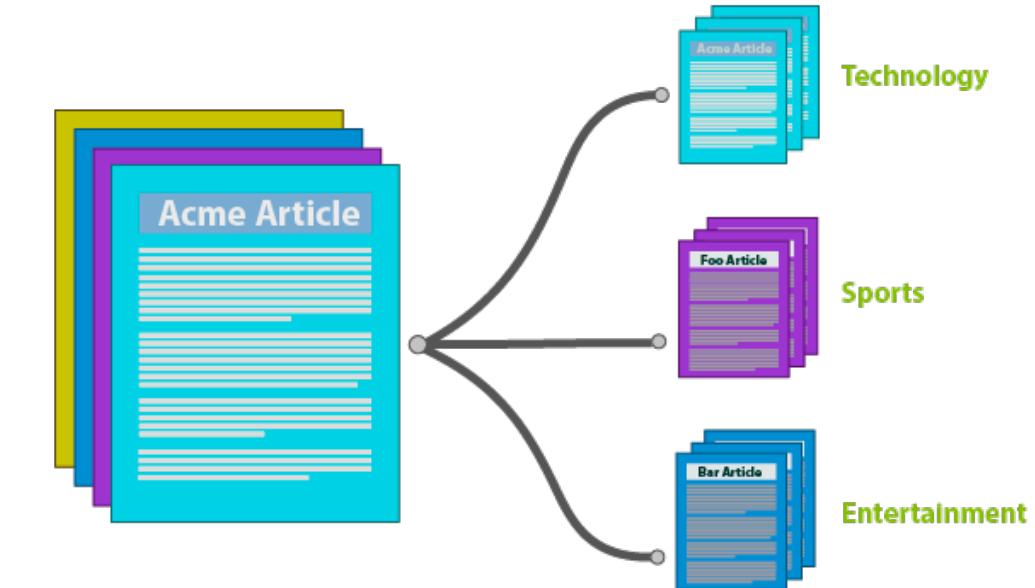
- 自然言語処理とRNN
 - Attention機構
 - Transformer/BERTの概要



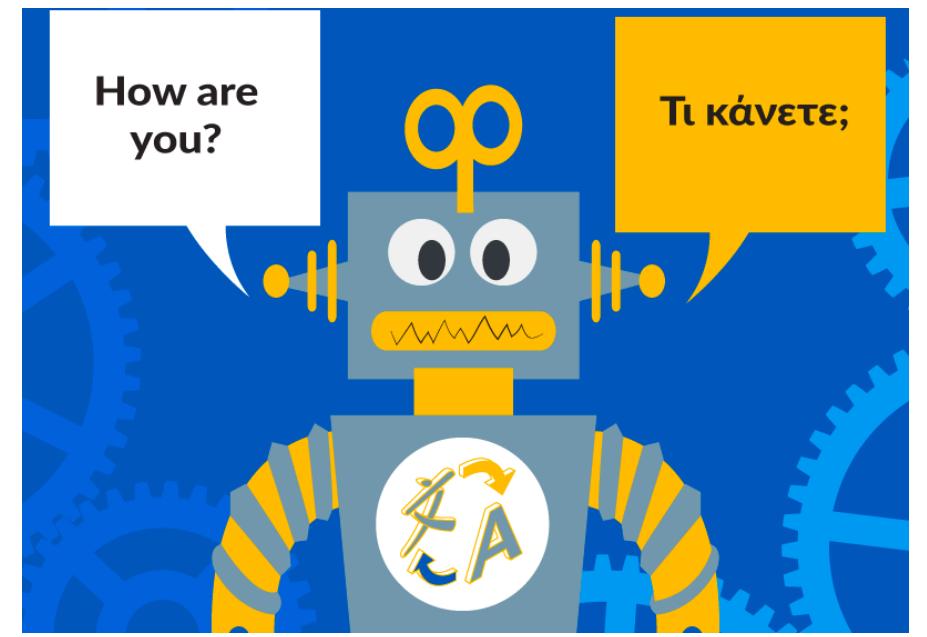
<https://robomind.co.jp/aivskodomotati3/>

自然言語処理の主要な応用タスク

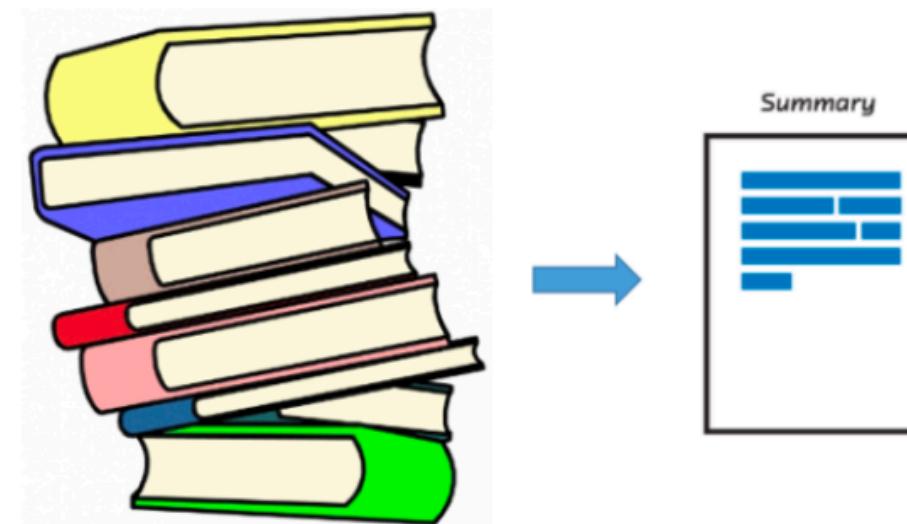
- ▶ 文書分類 (document categorization)
- ▶ 機械翻訳 (machine translation)
- ▶ 文章要約 (text summarization)
- ▶ 質問応答 (question answering)



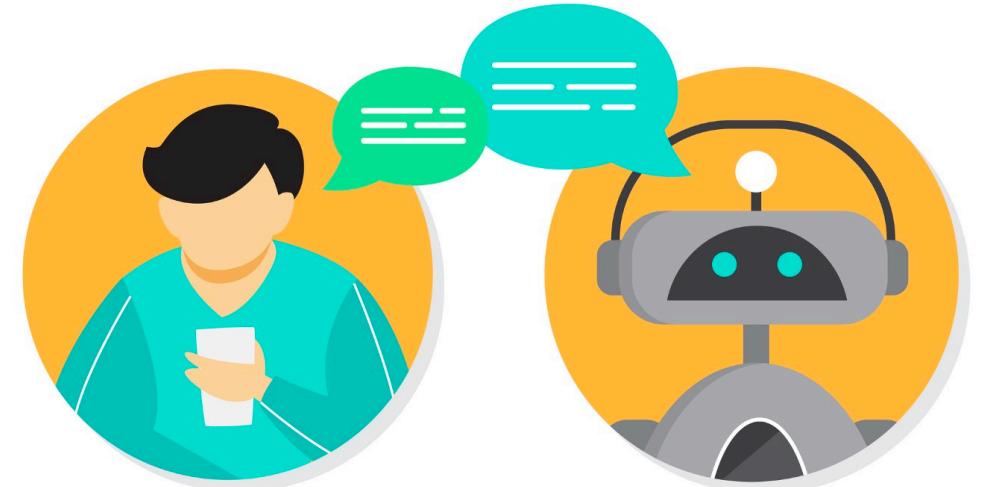
<https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a>



<https://livesalesman.com/why-you-cant-depend-on-machine-translation-for-multilingual-customer-support/>

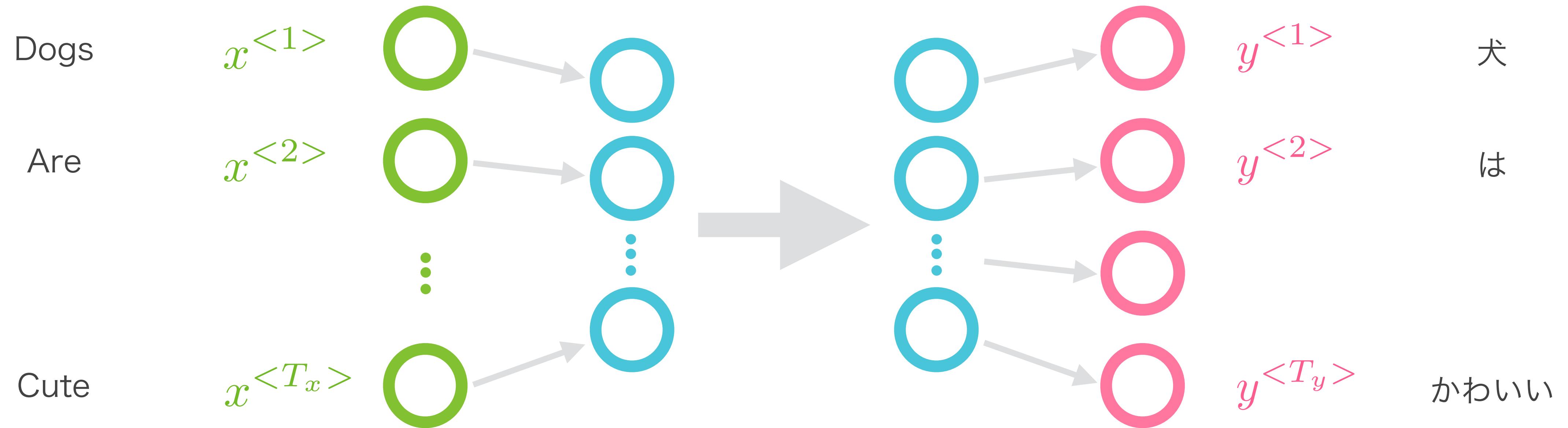


<https://towardsdatascience.com/text-summarization-using-deep-learning-6e379ed2e89c>



<https://towardsdatascience.com/automatic-question-answering-ac7593432842>

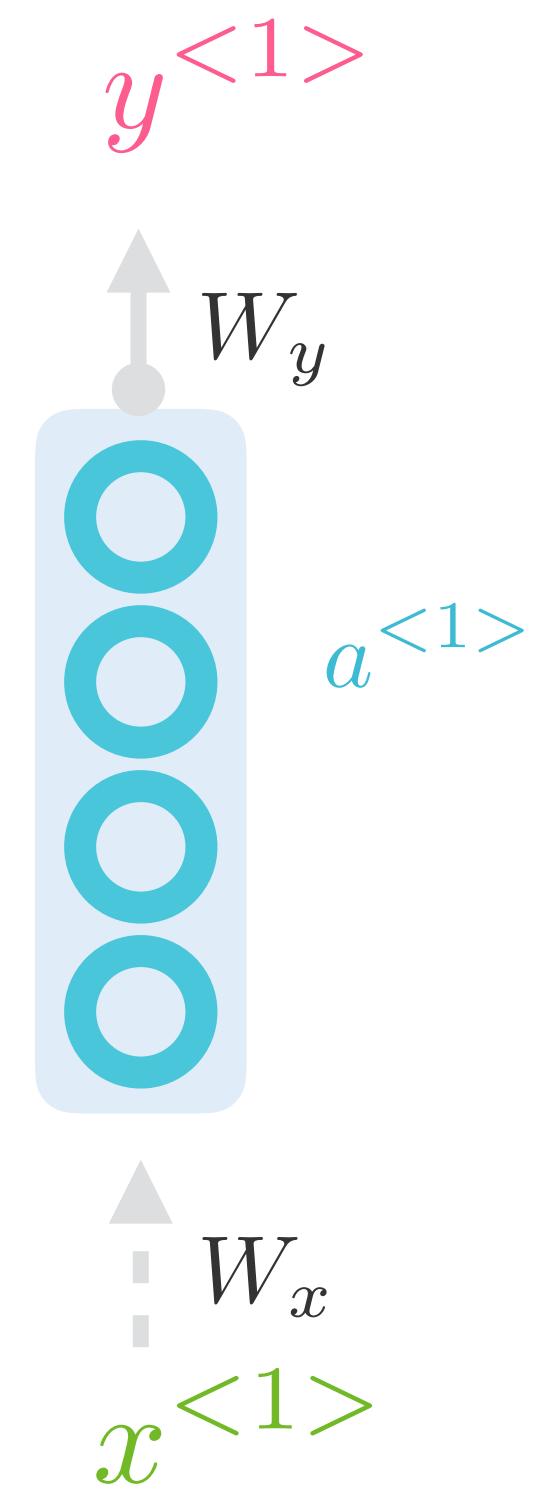
普通の全結合ニューラルネットワークの課題



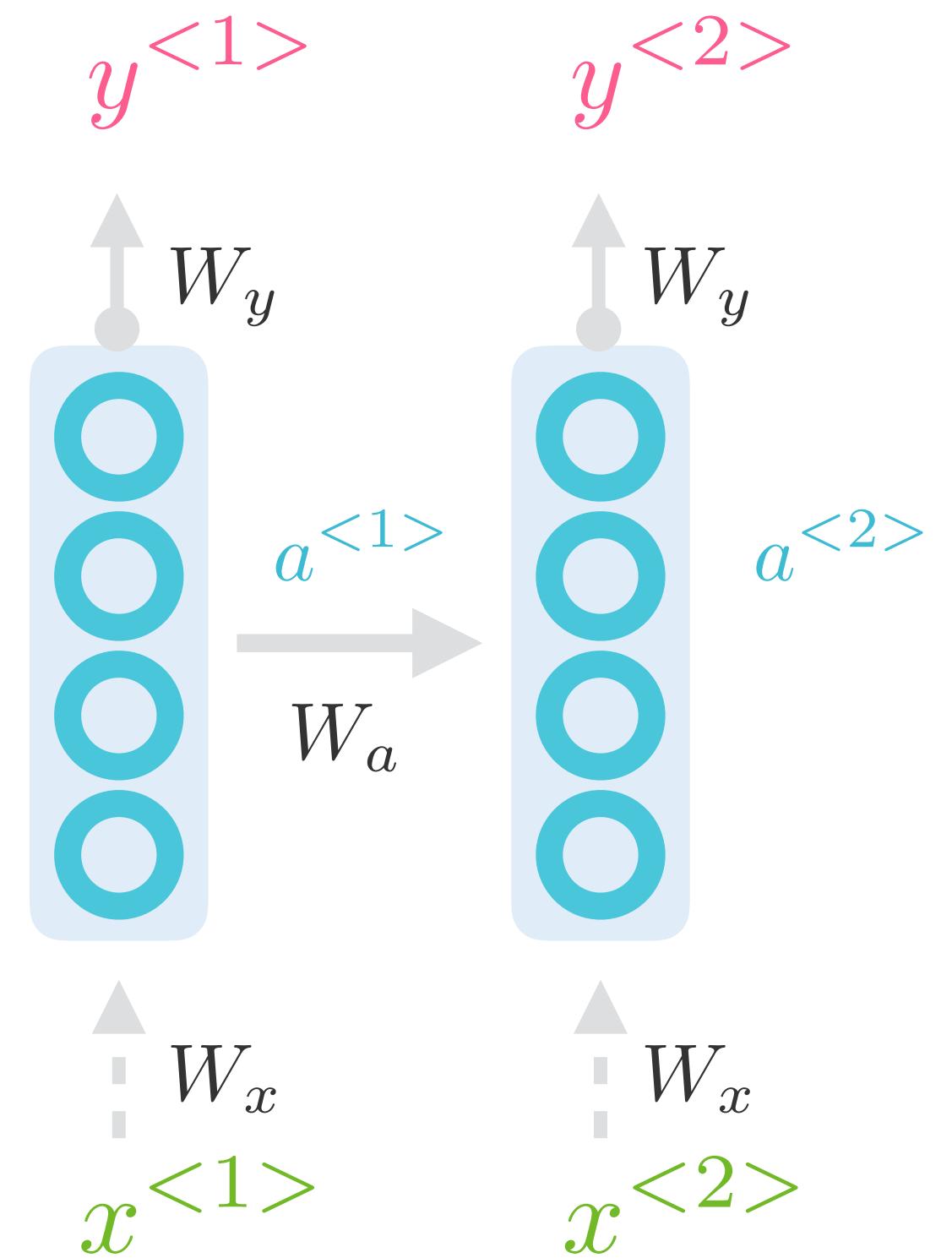
課題

- 一般に、入力の長さ (T_x) も出力の長さ (T_y) もセンテンスごとに異なる
- ある単語の特徴量を学習しても、単語の出現位置が変わると対応できない

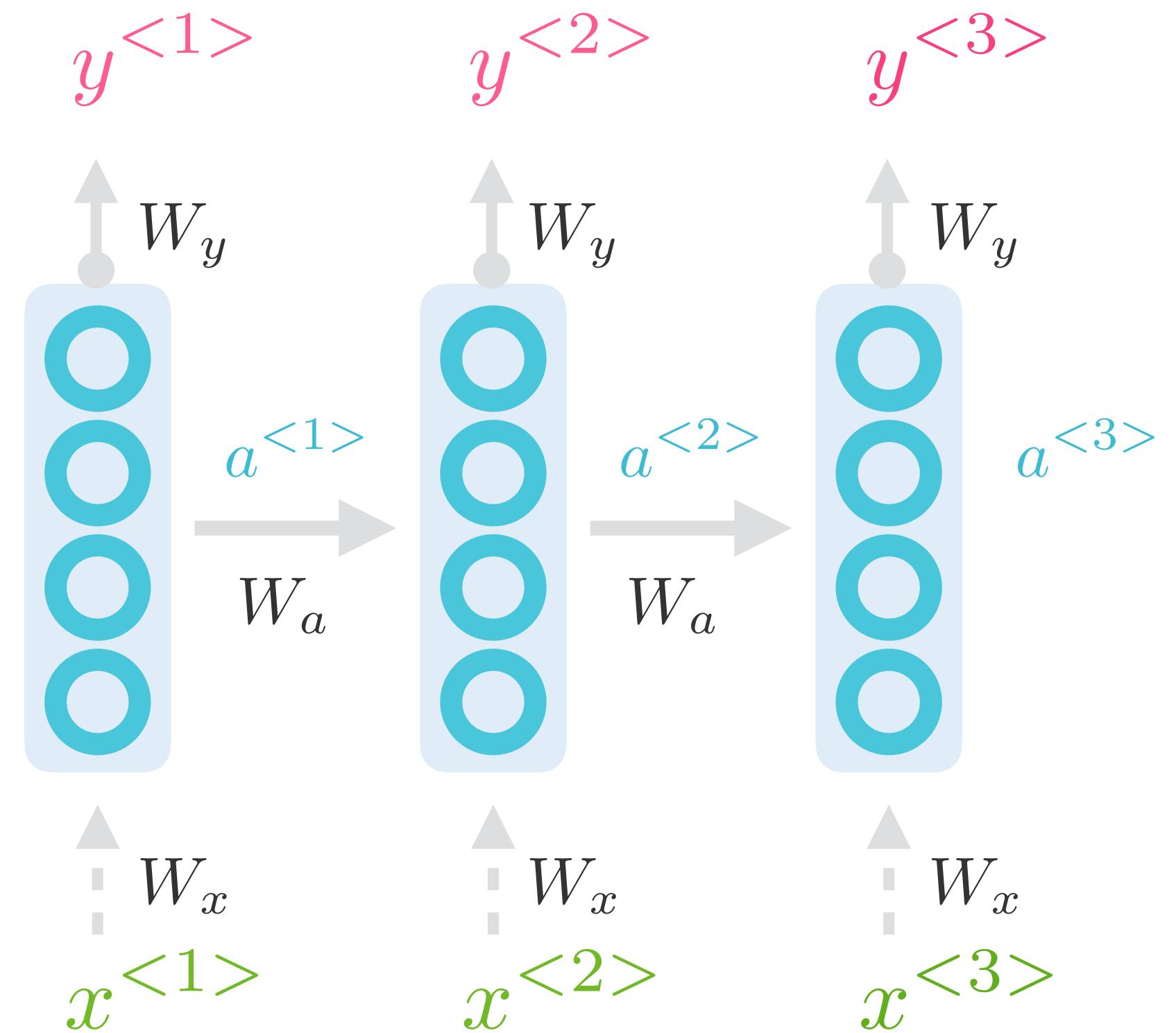
RNN (Recurrent Neural Network)



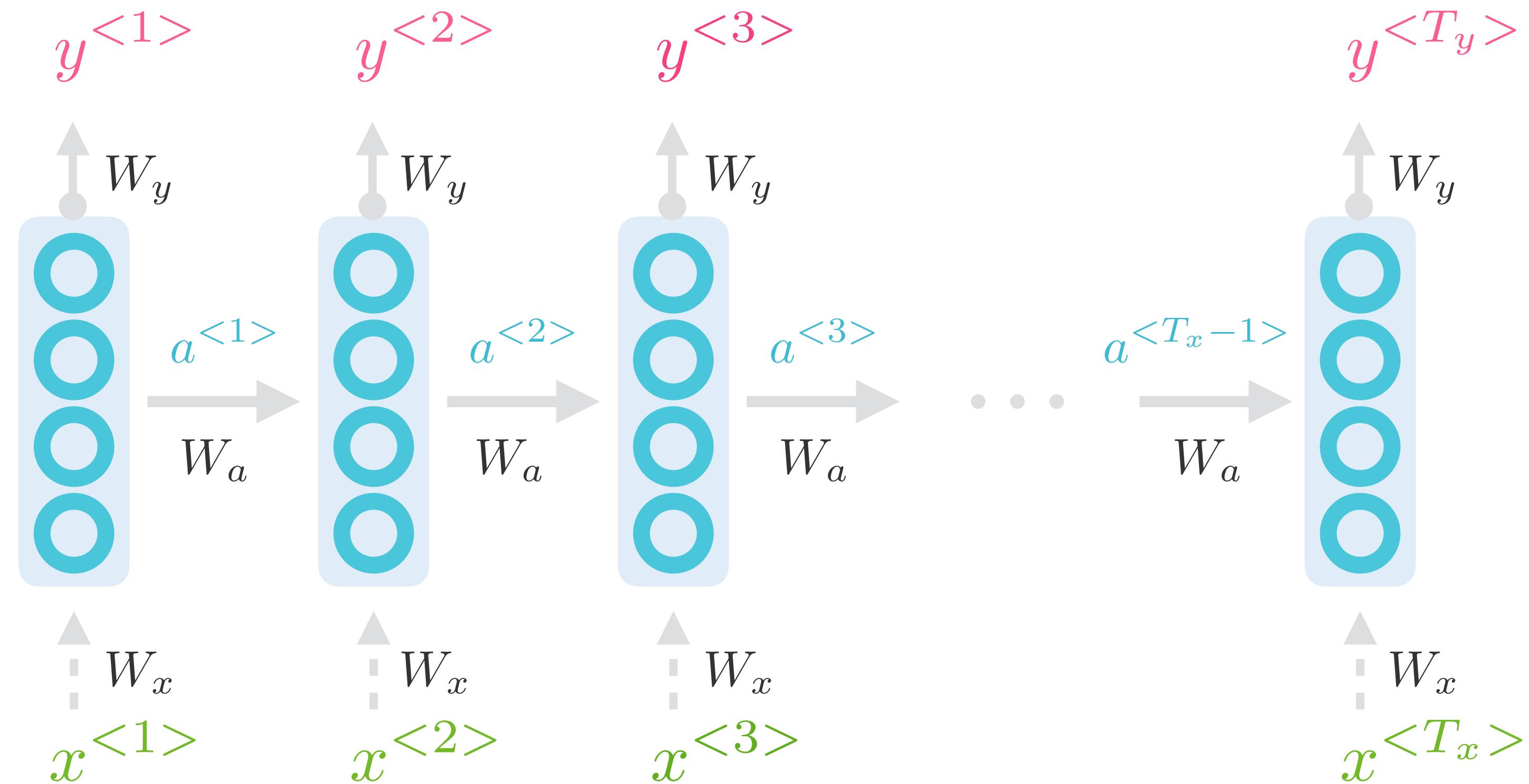
RNN (Recurrent Neural Network)



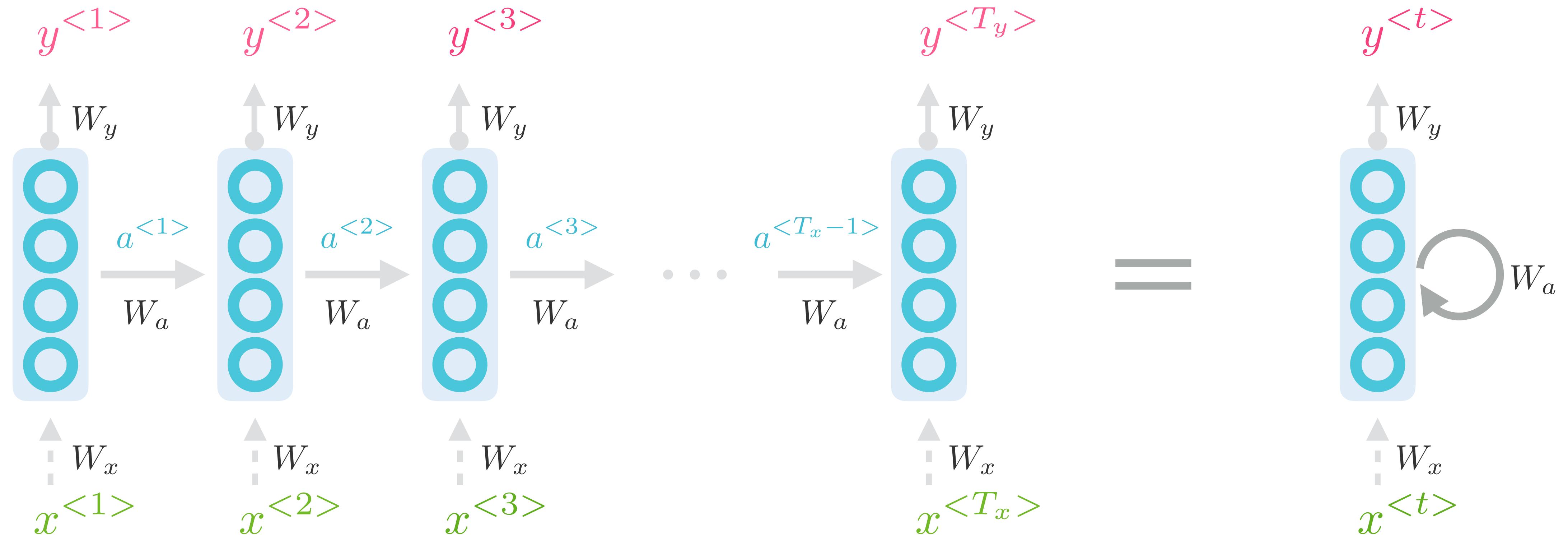
RNN (Recurrent Neural Network)



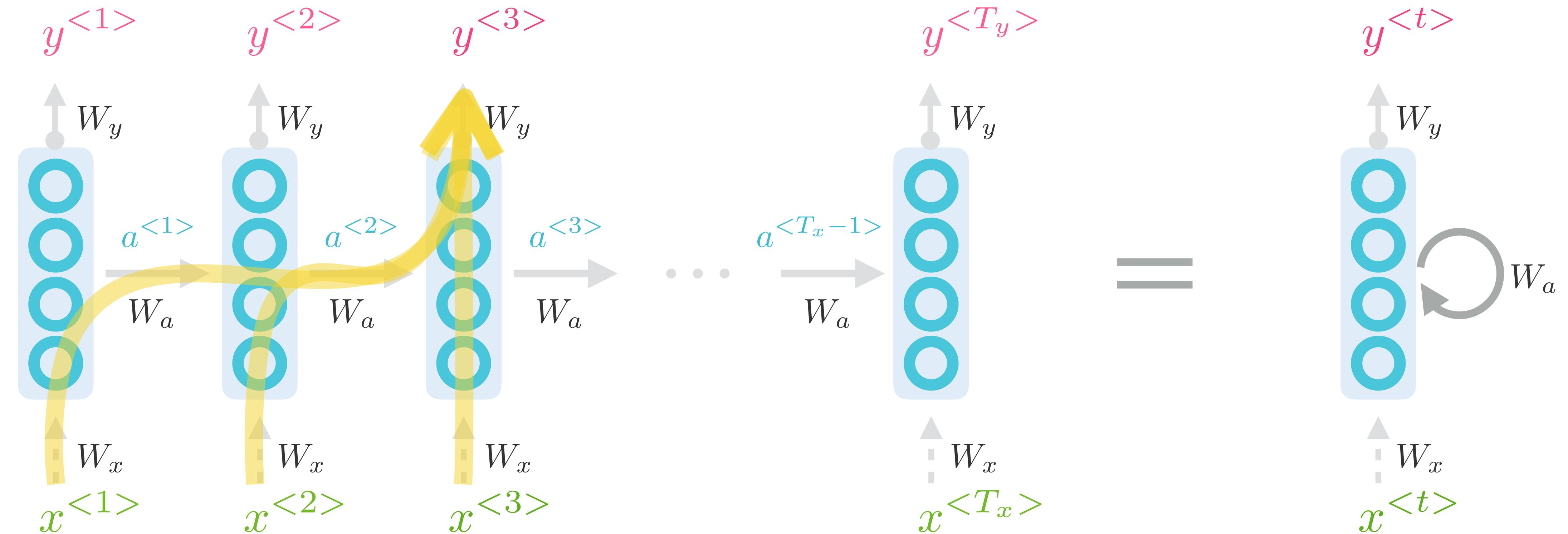
RNN (Recurrent Neural Network)



RNN (Recurrent Neural Network)

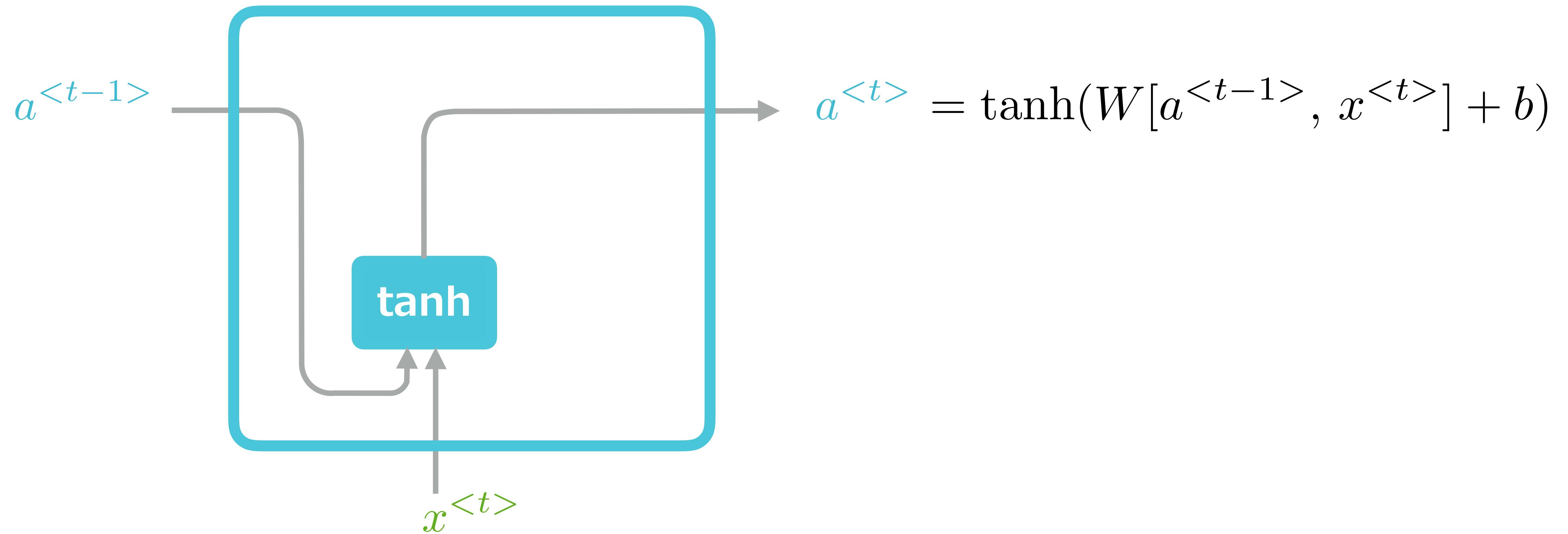


RNN (Recurrent Neural Network)

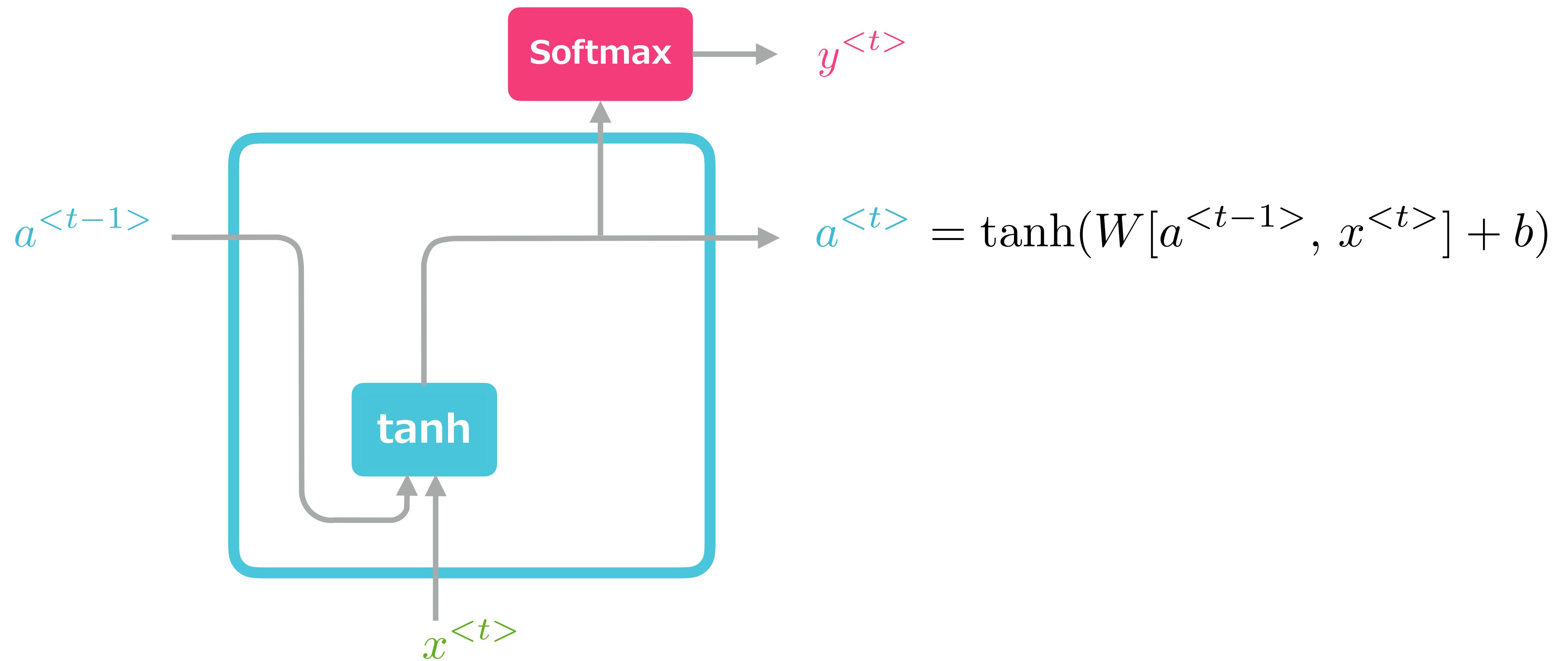


入力の時系列データを抽象的な文脈情報 (context) としてエンコード

RNN ユニット

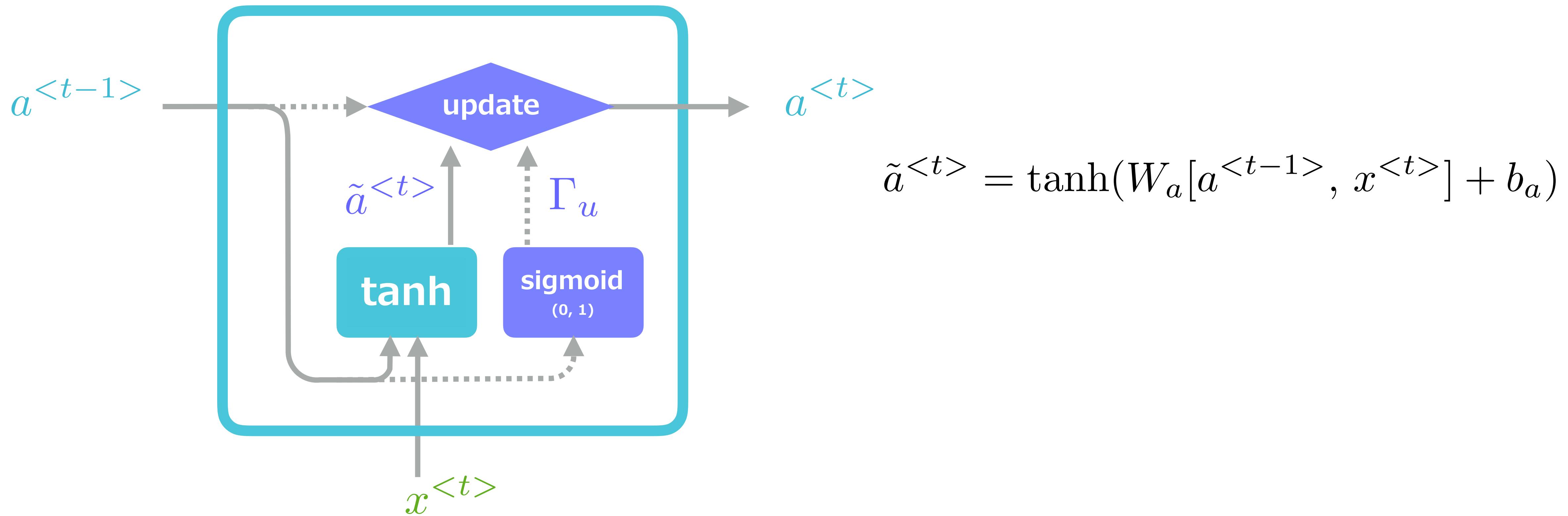


RNN ユニット



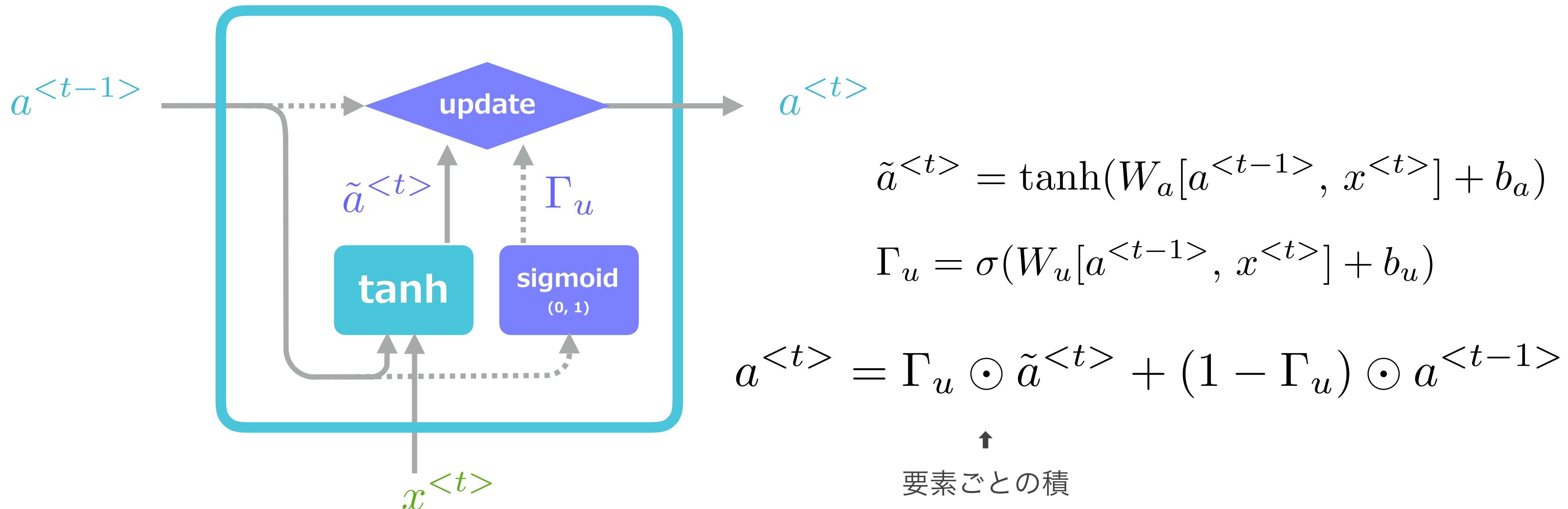
《simplified》 GRU (Gated Recurrent Unit) [2014]

- The **cat**, which already ate ⋯, [was] full
- The **cats**, which already ate ⋯, [were] full



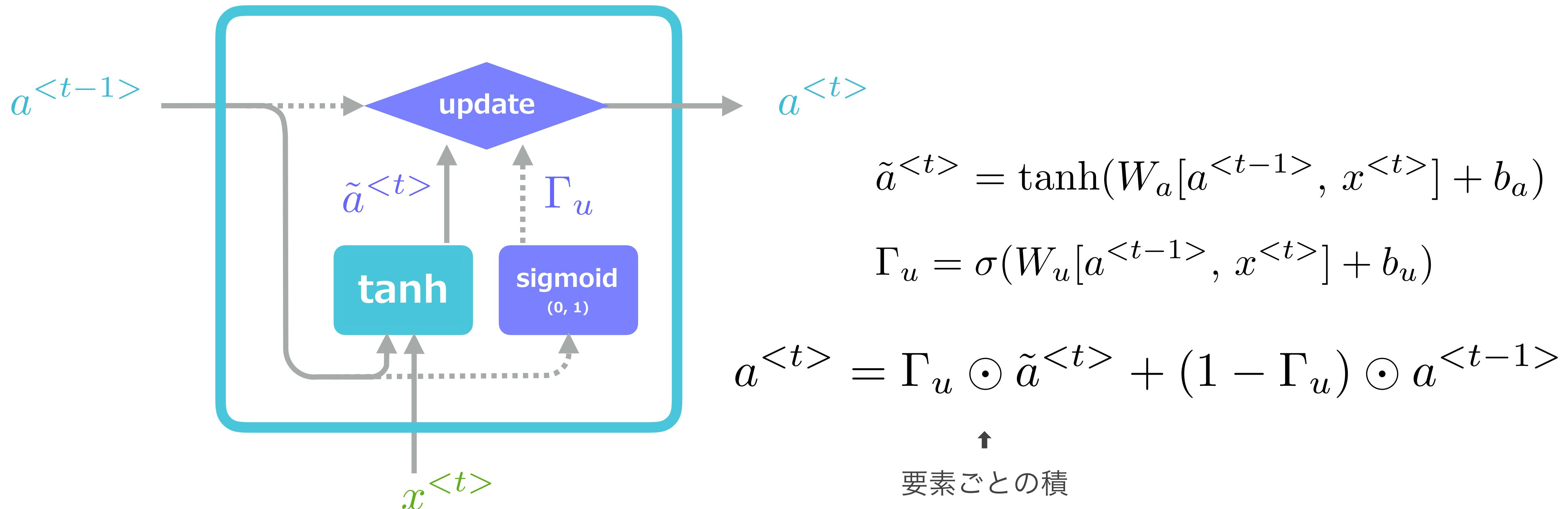
《simplified》 GRU (Gated Recurrent Unit) [2014]

- The **cat**, which already ate ⋯, [was] full
- The **cats**, which already ate ⋯, [were] full



《simplified》 GRU (Gated Recurrent Unit) [2014]

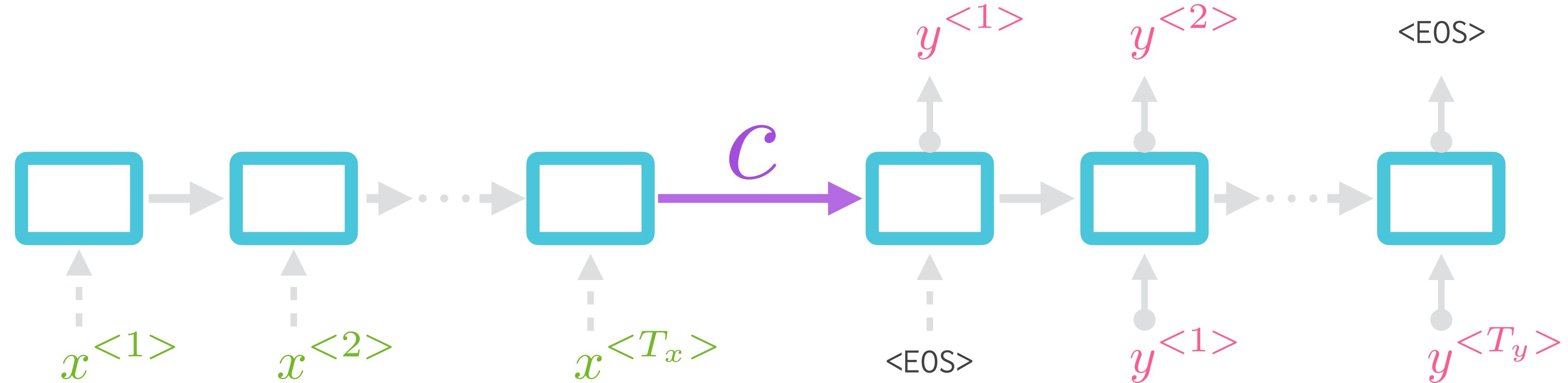
- The **cat**, which already ate ⋯, [was] full
- The **cats**, which already ate ⋯, [were] full



* 本当はアップデートゲートだけでなく、リセットゲートという別のゲートもある

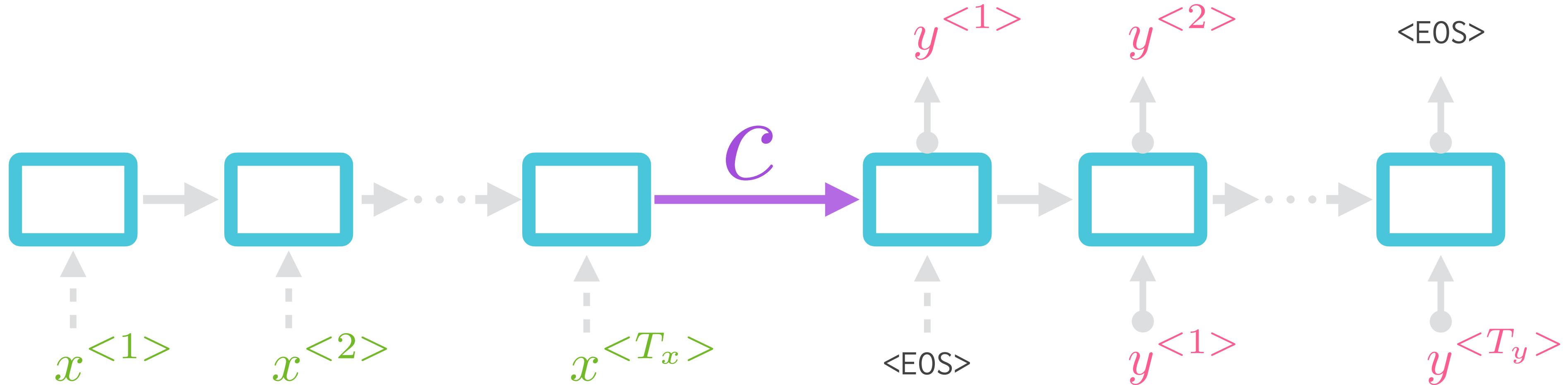
* LSTMもゲートの種類と数が違うだけで似たような構造 (入力ゲート、出力ゲート、忘却ゲート)

seq2seq (Encoder-Decoderモデル) [2014]



- ・入力文を **RNN (LSTM/GRU)** で抽象的な"コンテキスト"にエンコードし、それからデコードして出力文を生成
- ・機械翻訳や対話などのタスクでよく用いられる

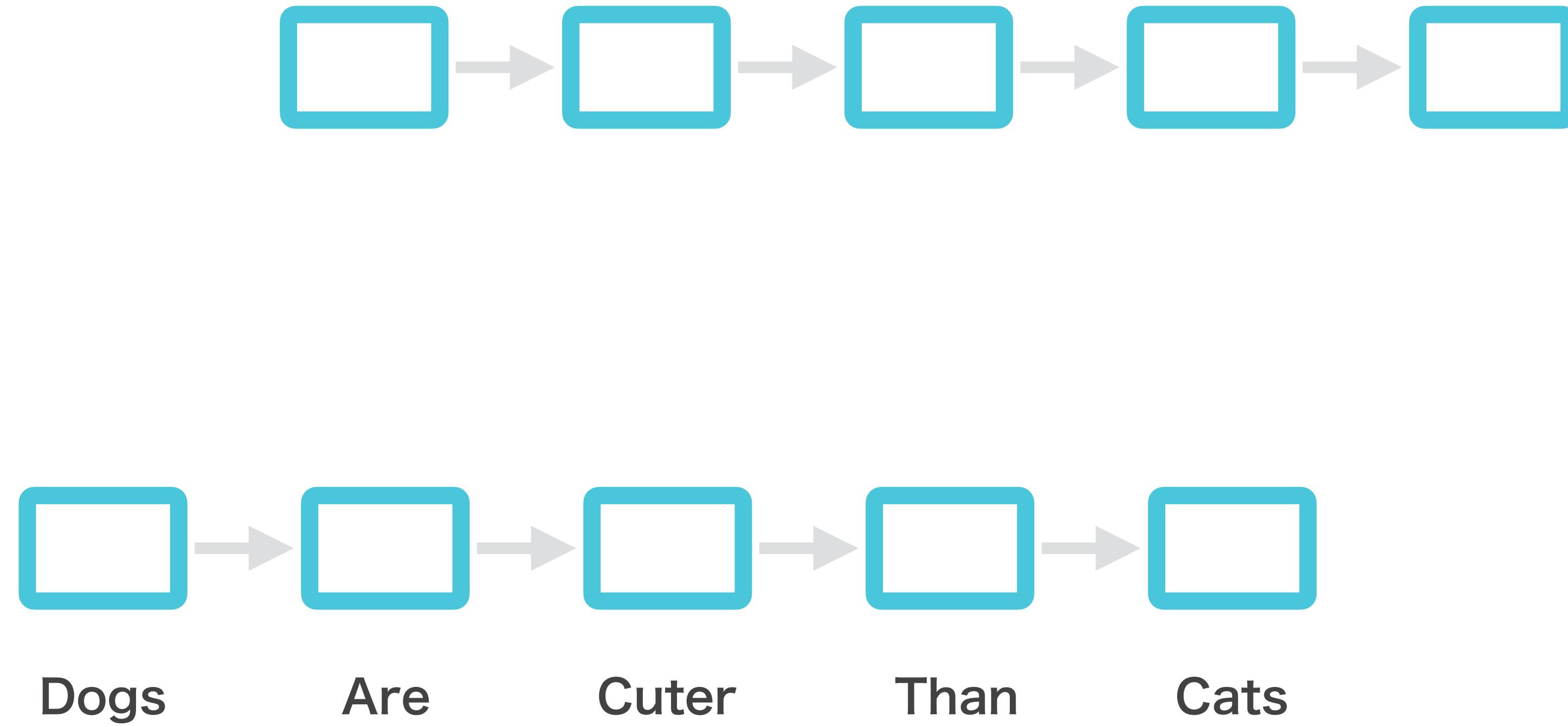
seq2seq (Encoder-Decoderモデル) [2014]



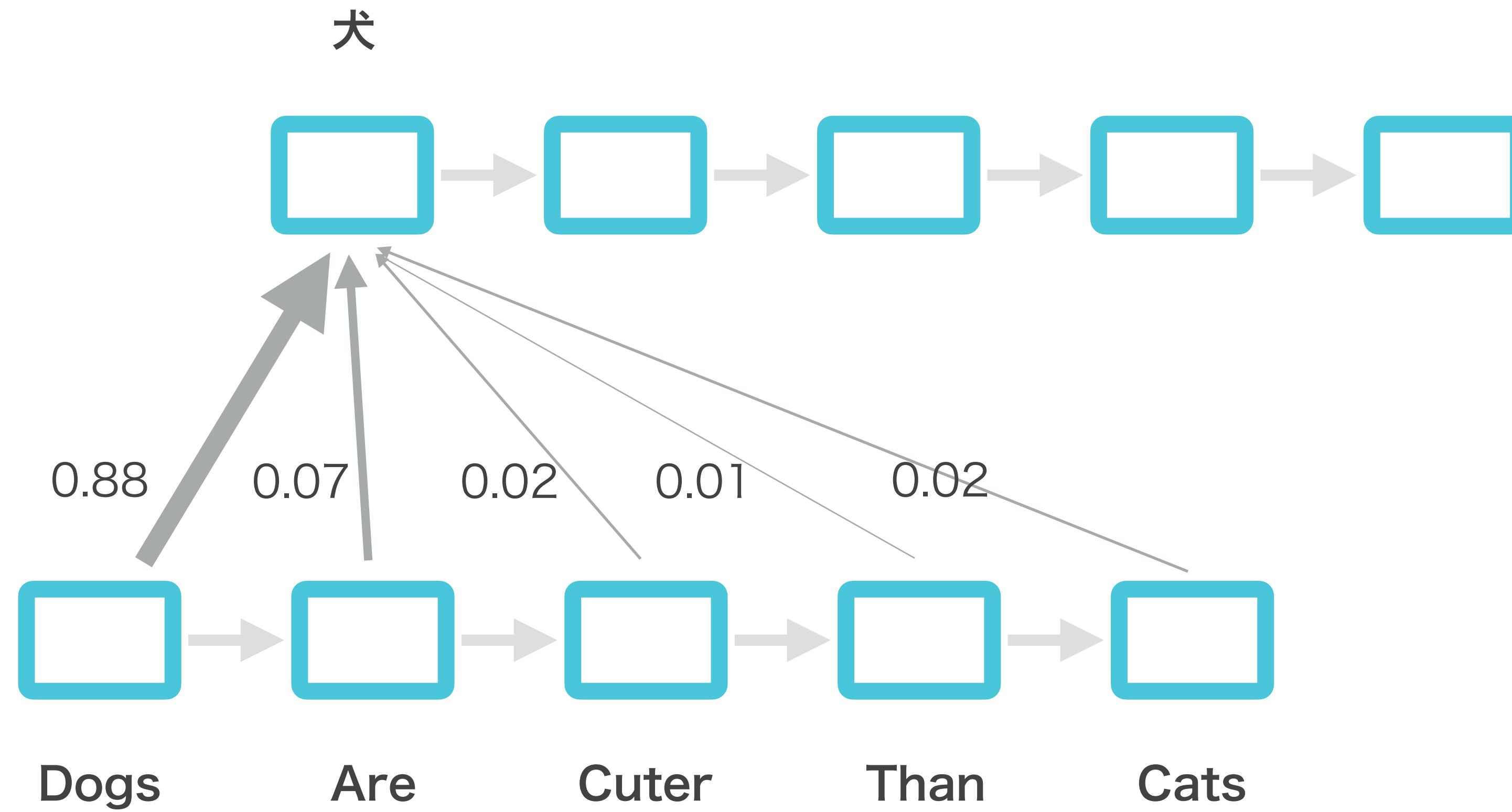
- ・入力文を **RNN (LSTM/GRU)** で抽象的な”コンテキスト”にエンコードし、それからデコードして出力文を生成
- ・機械翻訳や対話などのタスクでよく用いられる
- ・ただし、入力文が長くなると”コンテキスト”に入力文の情報を十分に記憶できなくなってしまう

→ **Attention (注意/注視/注目) 機構の導入 [2015?]**

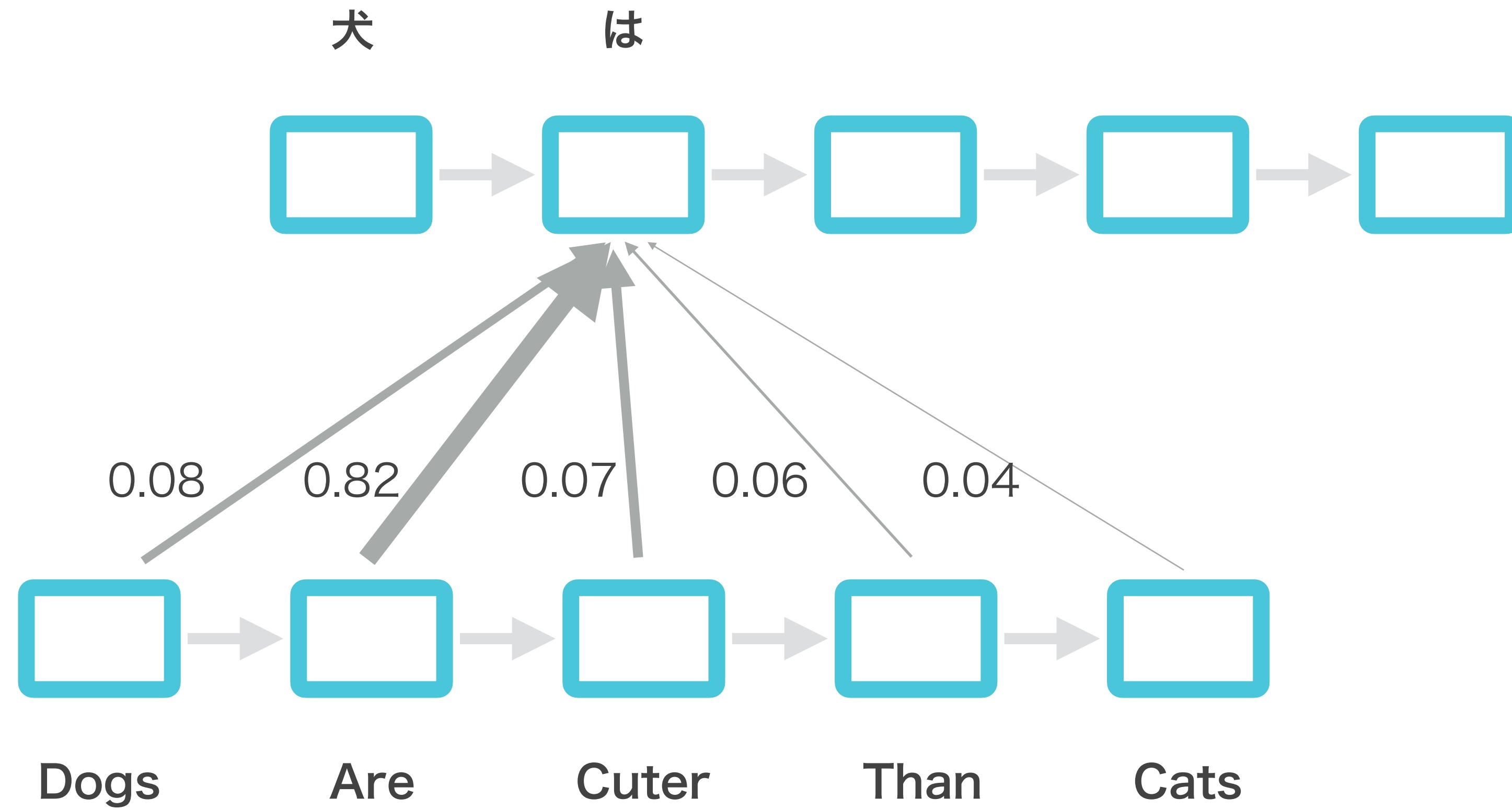
Attention機構のイメージ



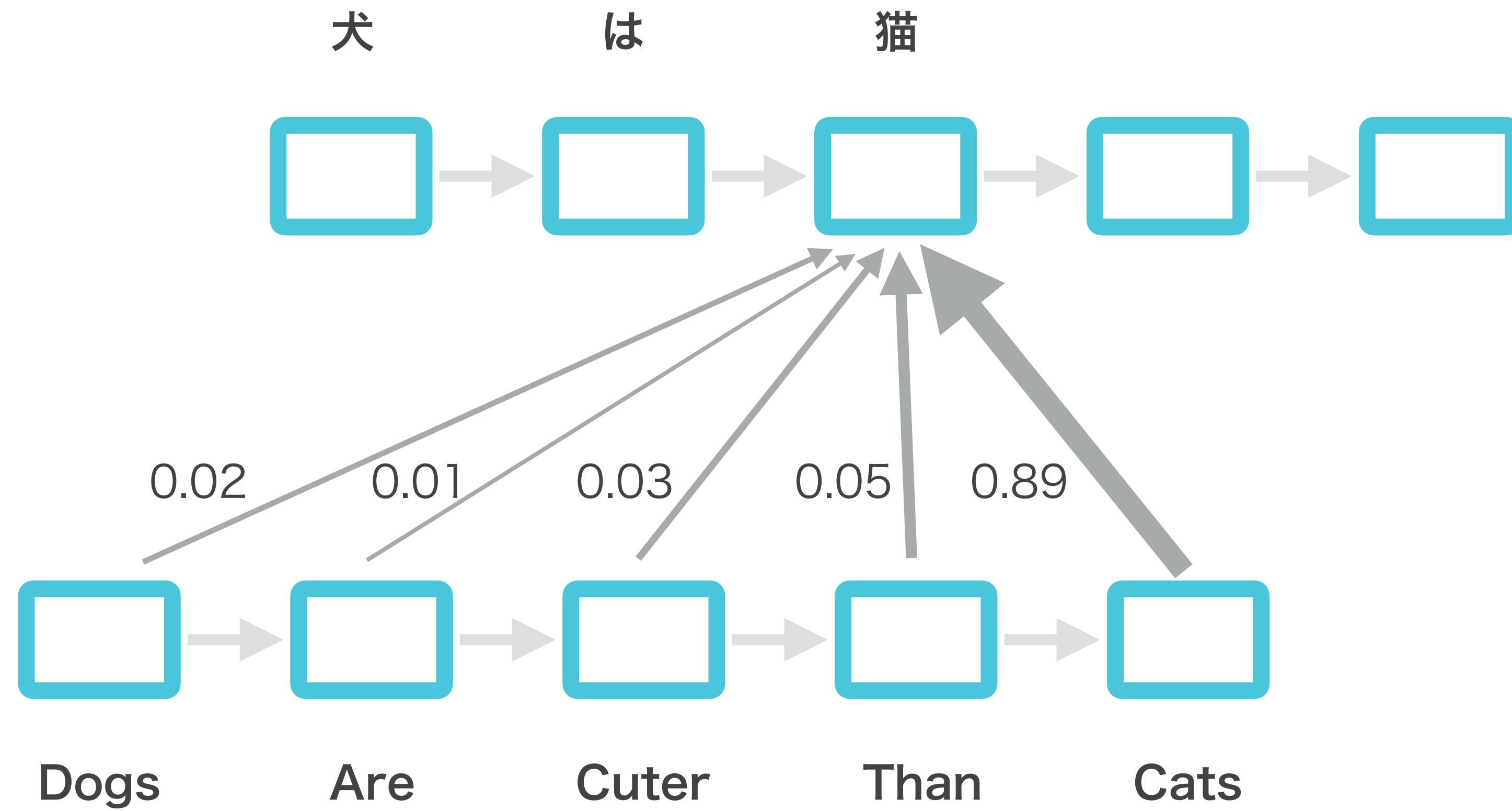
Attention機構のイメージ



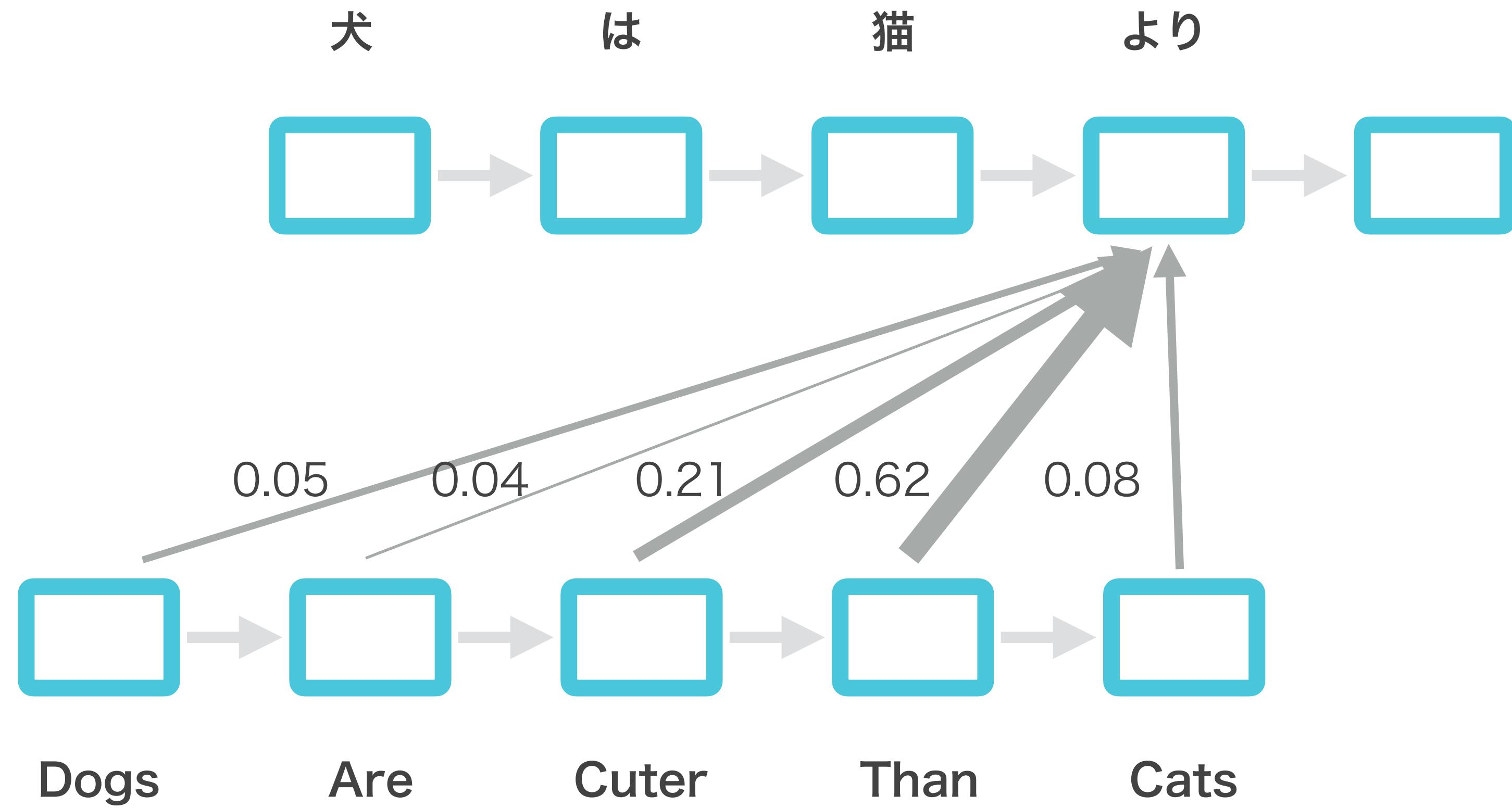
Attention機構のイメージ



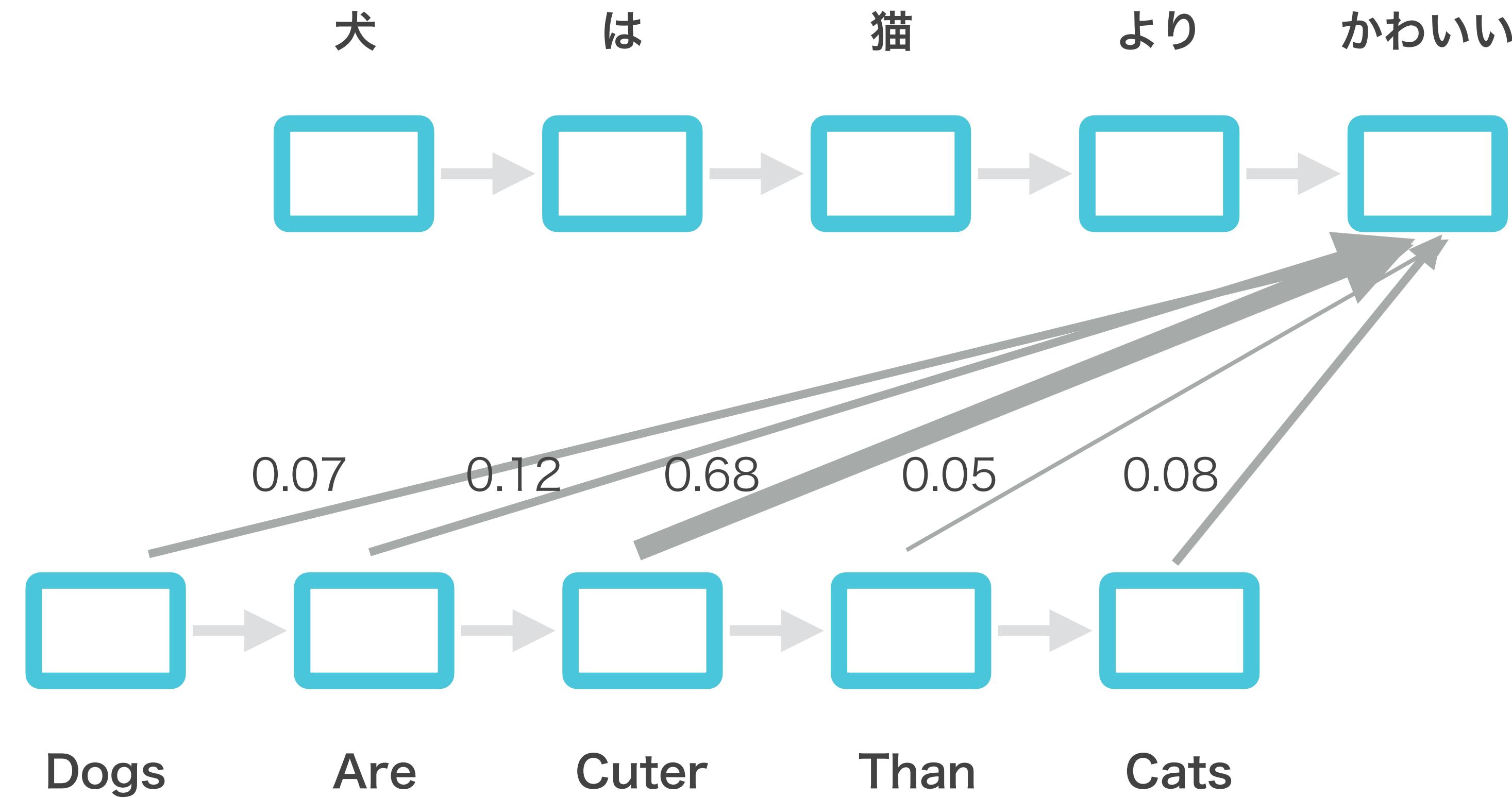
Attention機構のイメージ



Attention機構のイメージ



Attention機構のイメージ



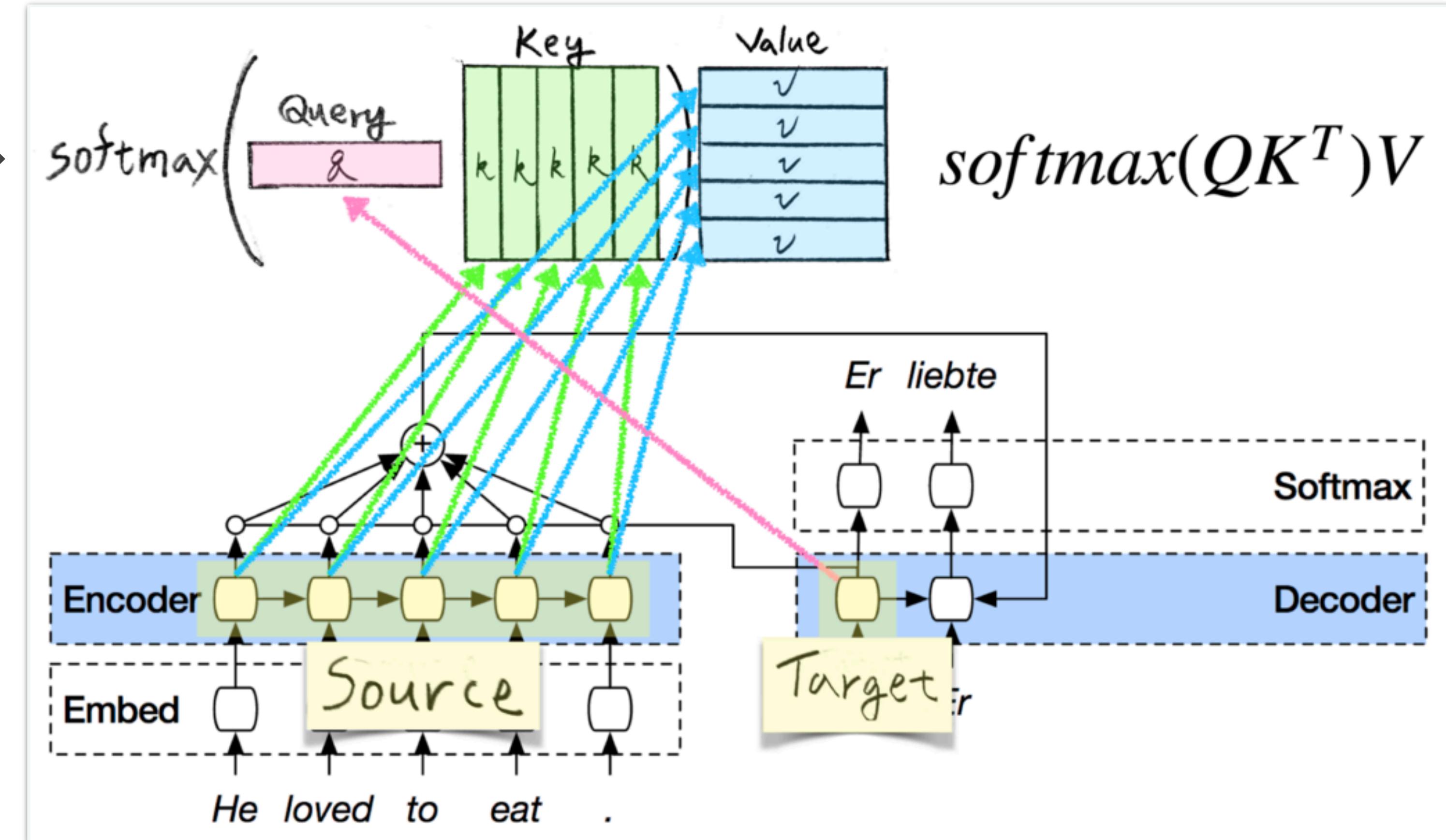
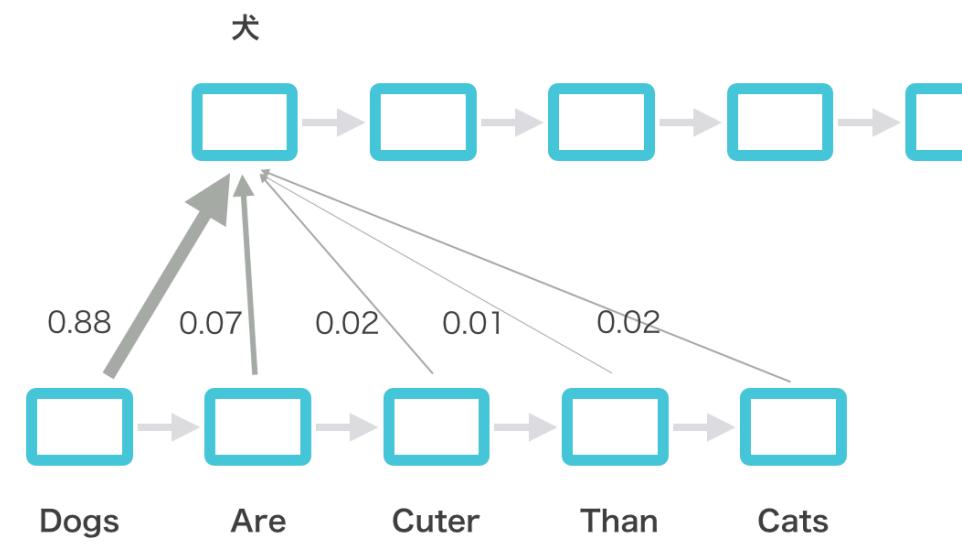
人間が英文翻訳するのと同じような注目の仕方

Attentionの計算式（数ある定義のひとつ）

decoderとencoderの中間状態の類似度を計算

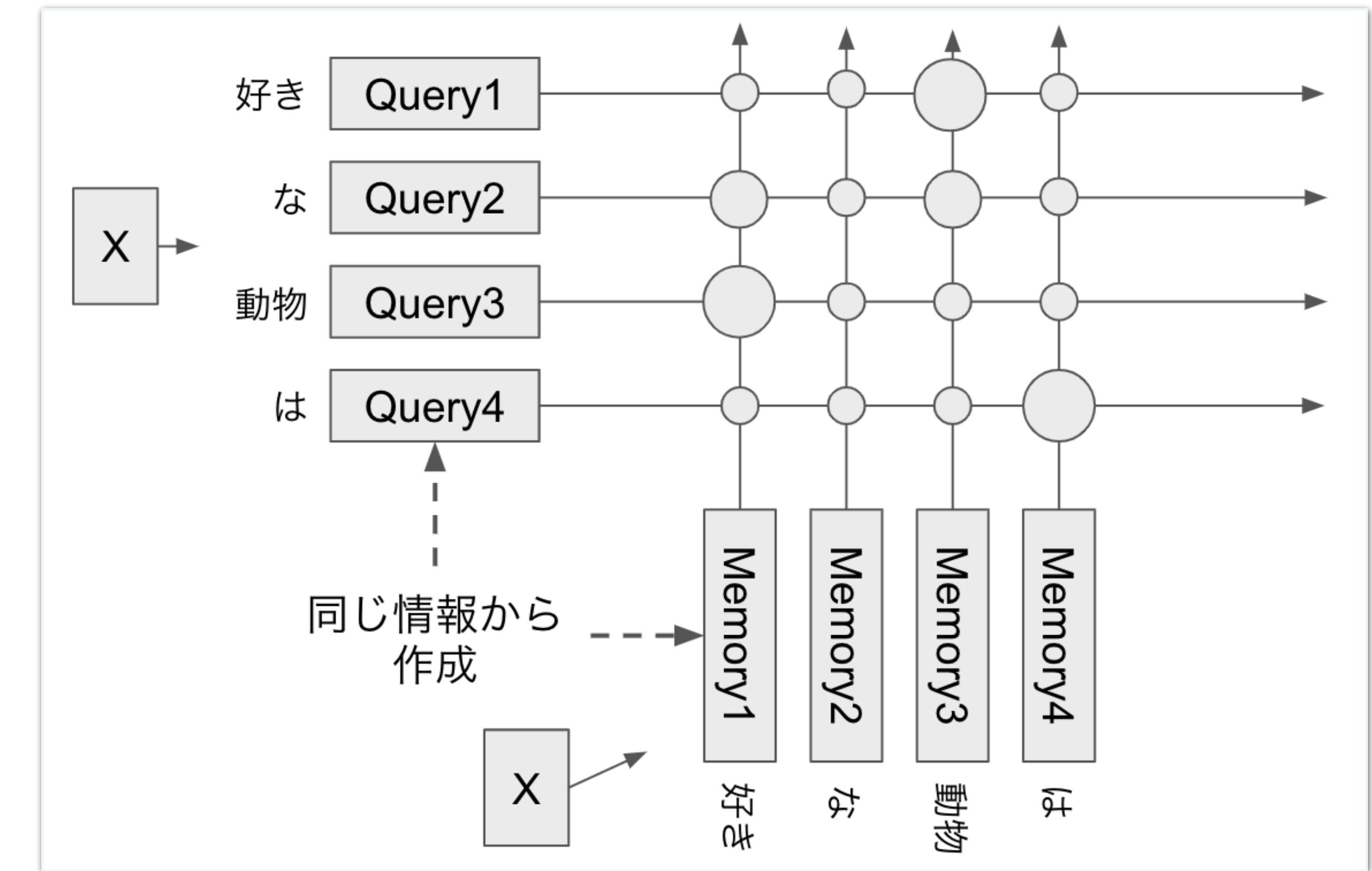
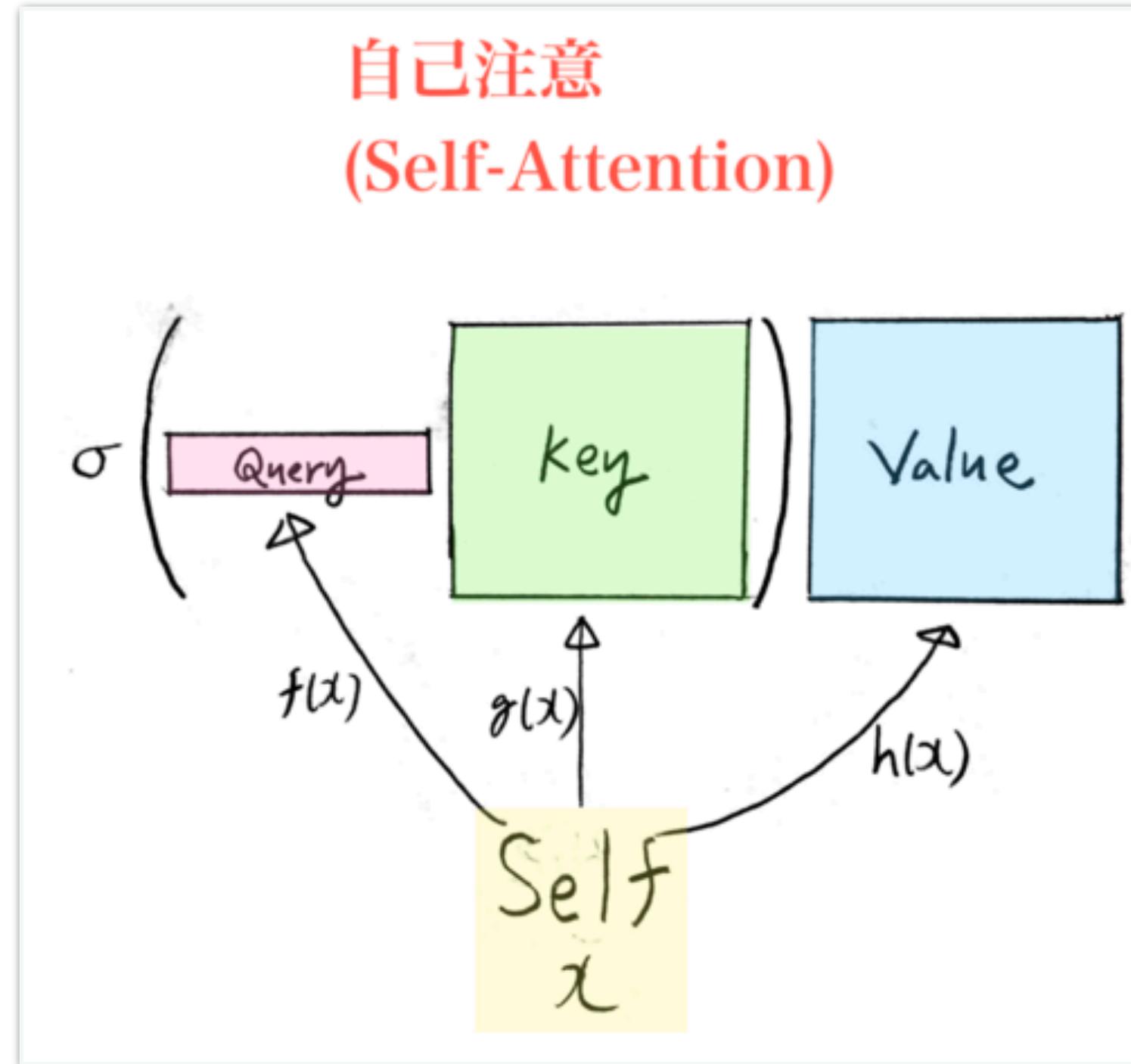


softmaxで合計1の重みに変換 →



<http://deeplearning.hatenablog.com/entry/transformer>

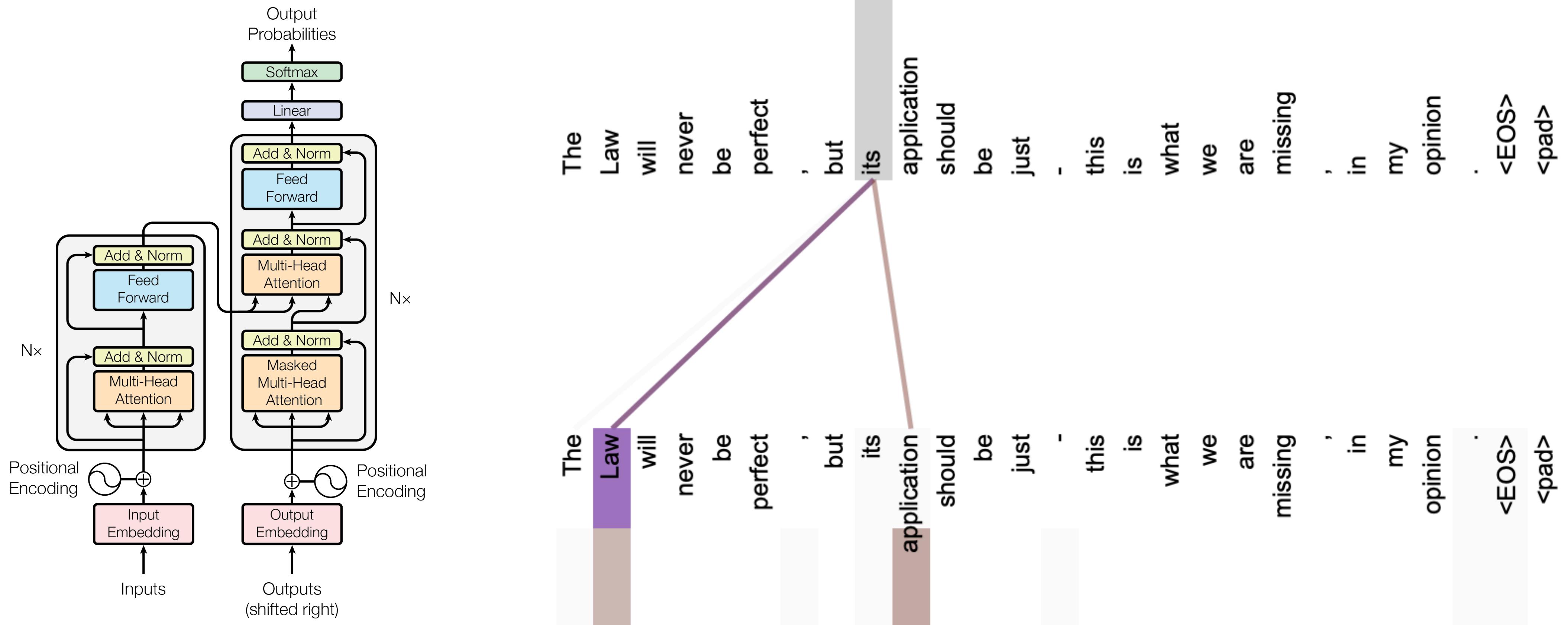
Self-Attention



<http://deeplearning.hatenablog.com/entry/transformer>

<https://qiita.com/Changkengo/items/b9afe7a3341b47dced30>

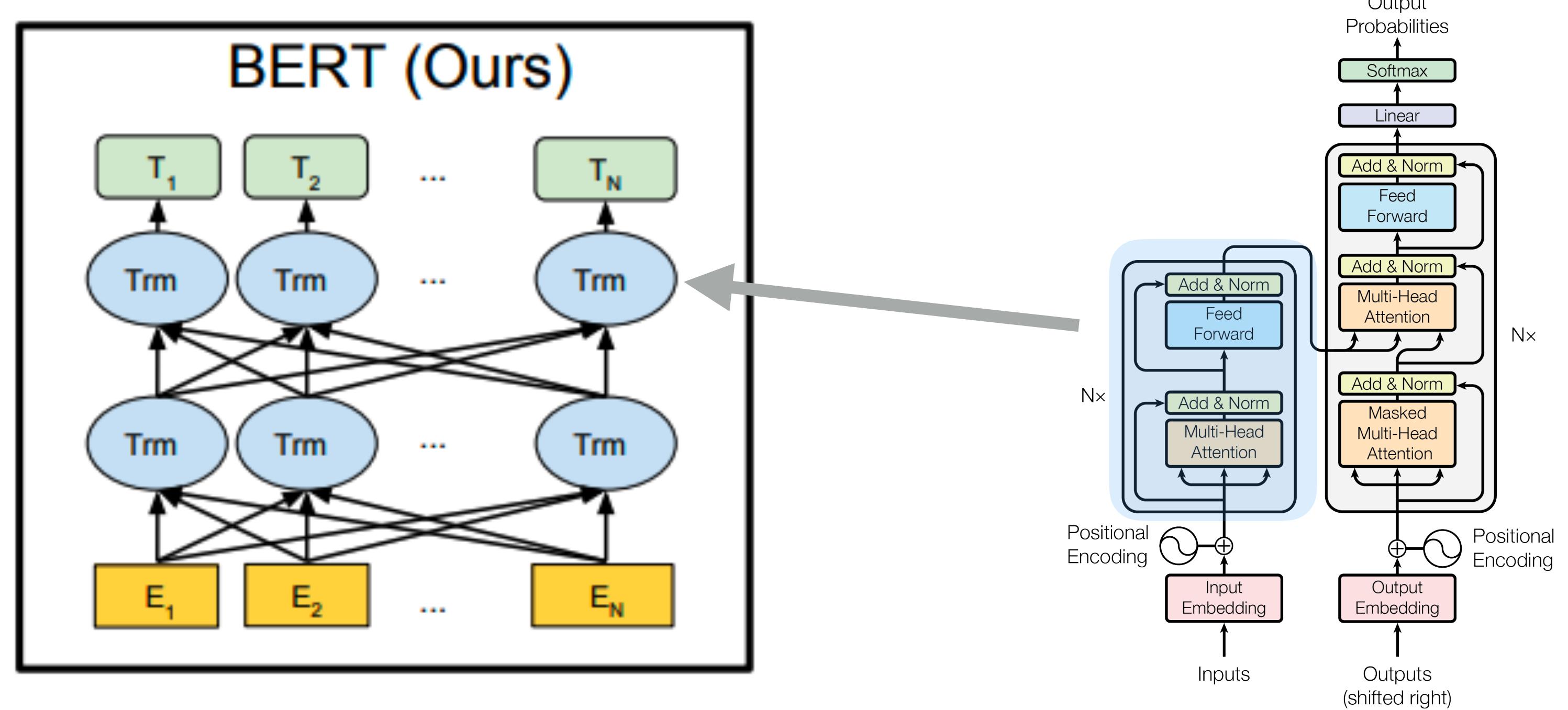
RNN使わずにSelf-Attentionだけでも文脈表現を抽出可能



Attention is All You Need (Transformer) [2017]: <https://arxiv.org/abs/1706.03762>

BERT [2018]

- ▶ Bidirectional Encoder Representations from Transformers
 - 双方向のTransformerを用いて言語表現をエンコードするモデル



BERTを表現抽出器として利用

神田 の 鬼金棒 の ラーメン は 美味しい



大規模データセットで訓練済みのBERT

英語・中国語・多言語対応版、日本語版*

*1GPUだと30日分の訓練



機械翻訳モデル

対話モデル

感情推定モデル

“Ramen noodles at Kikanbo in Kanda is delicious” 「へえ、この後行こうかな！」

ポジティブ

多数のベンチマークタスクでState-Of-The-Artの性能

タスク	概要	前SOTA	BERT
GLUE	8種の言語理解タスク	75.2	81.9
1. MNLI	2入力文の含意/矛盾/中立を判定	82.1	86.7
2. QQP	2質問文が意味的に等価か判定	70.3	72.1
3. QNLI	SQuADの改変。陳述文が質問文の解答を含むか判定	88.1	91.1
4. SST-2	映画レビューの入力文のネガポジを判定	91.3	94.9
5. CoLA	入力文が言語的に正しいか判定	45.4	60.5
6. STS-B	ニュース見出しの2入力文の意味的類似性をスコア付け	80.0	86.5
7. MRPC	ニュース記事の2入力文の意味的等価性を判定	82.3	89.3
8. RTE	2入力文の含意を判定	56.0	70.1
SQuAD	質疑応答タスク。陳述文から質問文の解答を抽出	91.7	93.2
CoNLL	固有表現抽出タスク。単語に人物/組織/位置のタグ付け	92.6	92.8
SWAG	入力文に後続する文を4つの候補文から選択	59.2	86.3

https://twitter.com/_Ryobot/status/1050925881894400000

SWAG (The Situations With Adversarial Generations dataset)

A girl is going across a set up monkey bars. She

(少女が雲梯を進んでいます。彼女は)

(i) jumps up across the monkey bars.

(雲梯で飛び上ります)

(ii) struggles onto the bars to grab her head.

(頭を掴むために雲梯でもがいています)

(iii) gets to the end and stands on a wooden plank.

(終わりまで行き、木の板の上に立ちます)

(iv) jumps up and does a back flip.

(飛び上がり、宙返りをします)

参考) <https://it-mint.com/2018/11/06/datasets-of-deep-learning-bert-1603.html>

穴埋めクイズと隣接文クイズを解いて学習

Input A: [CLS] the man went to [MASK] store [SEP]

Input B: he bought a gallon [MASK] milk [SEP]

Label: **IsNext**

-

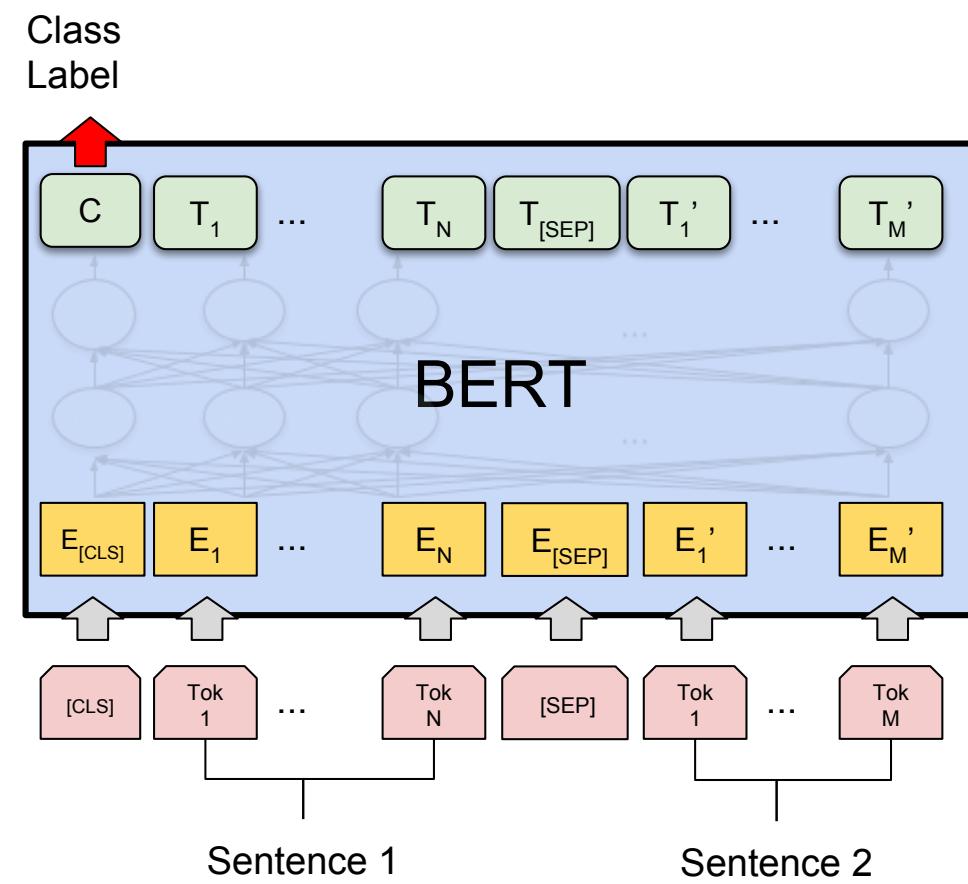
Input A: [CLS] the man went to [MASK] store [SEP]

Input B: penguin [MASK] are flight ##less birds [SEP]

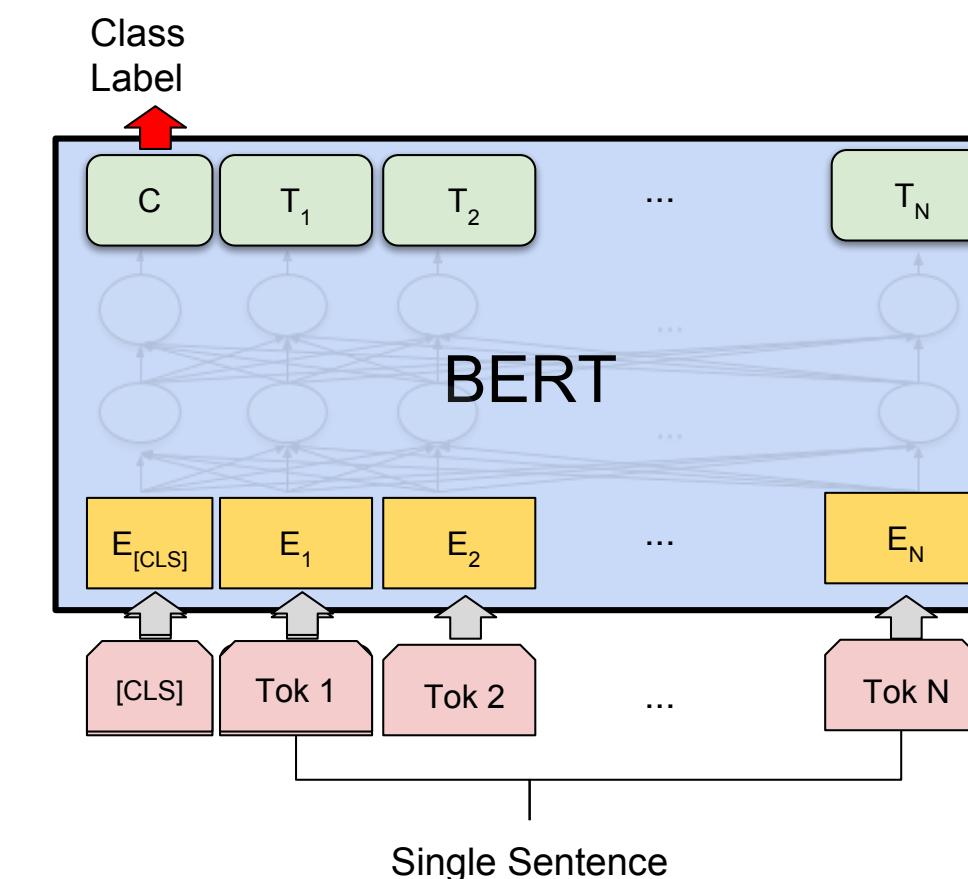
Label: **NotNext**

- ▶ [CLS]: 隣接文クイズ開始の印、[MASK]: 単語の置き換え、[SEP]: センテンスの末尾、##: 接尾辞

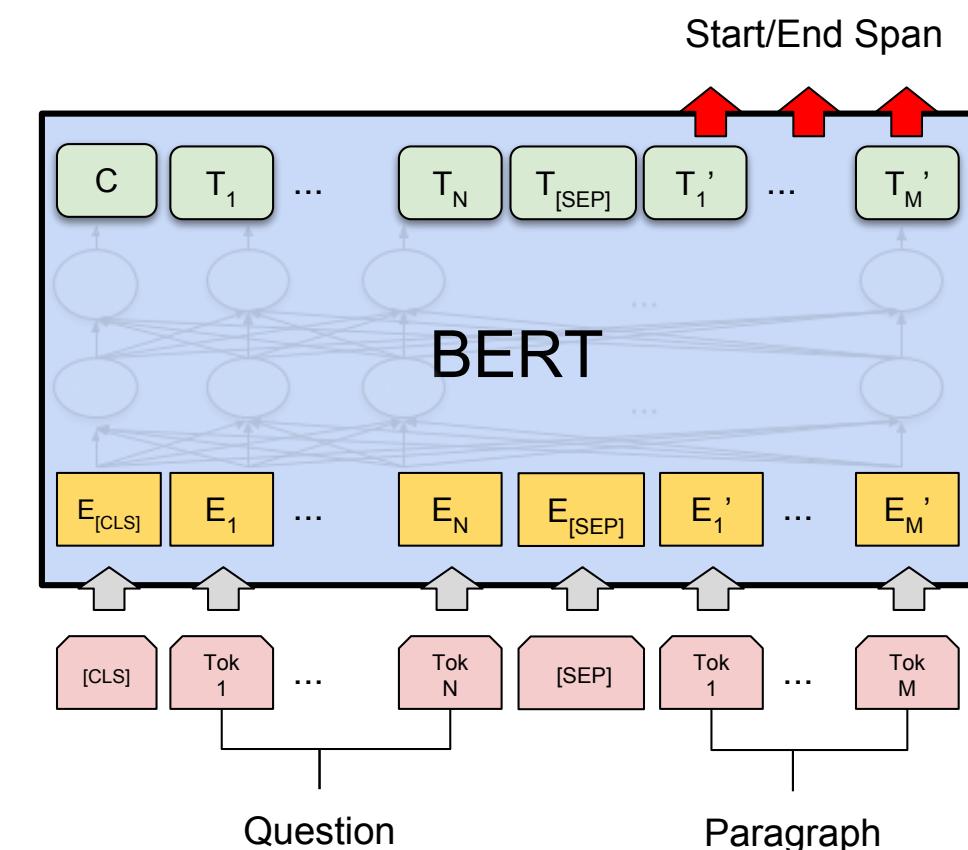
特定のタスク用にシンプルな出力層を追加する



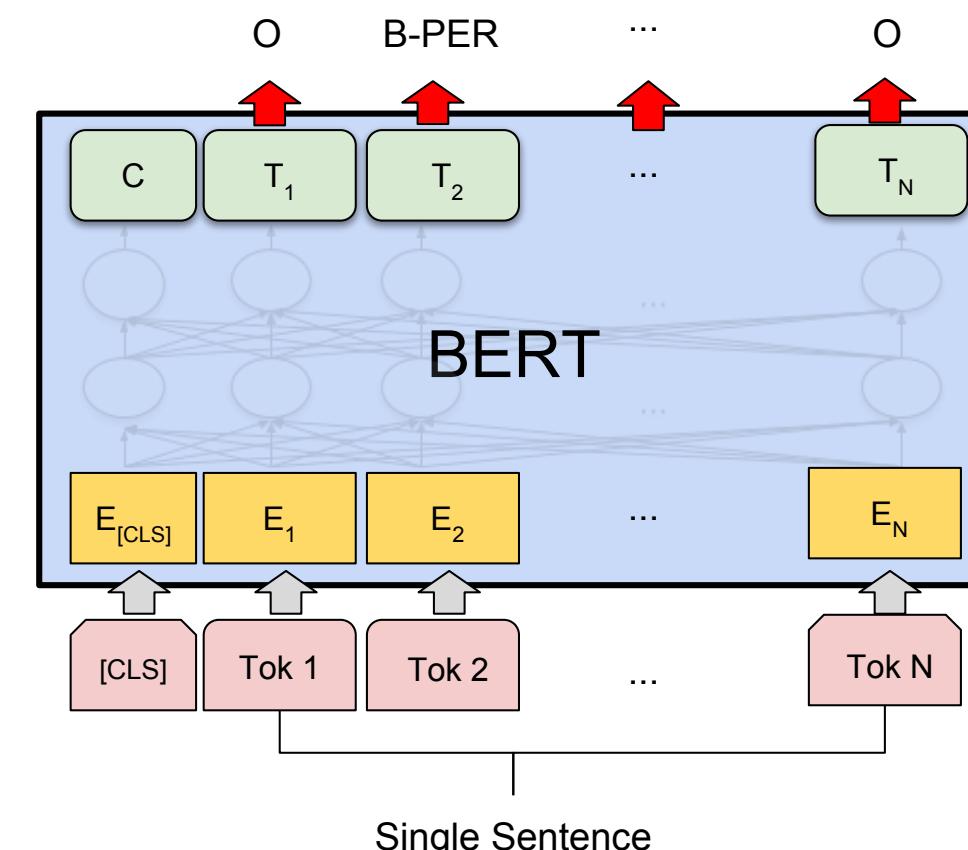
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



(b) Single Sentence Classification Tasks:
SST-2, CoLA



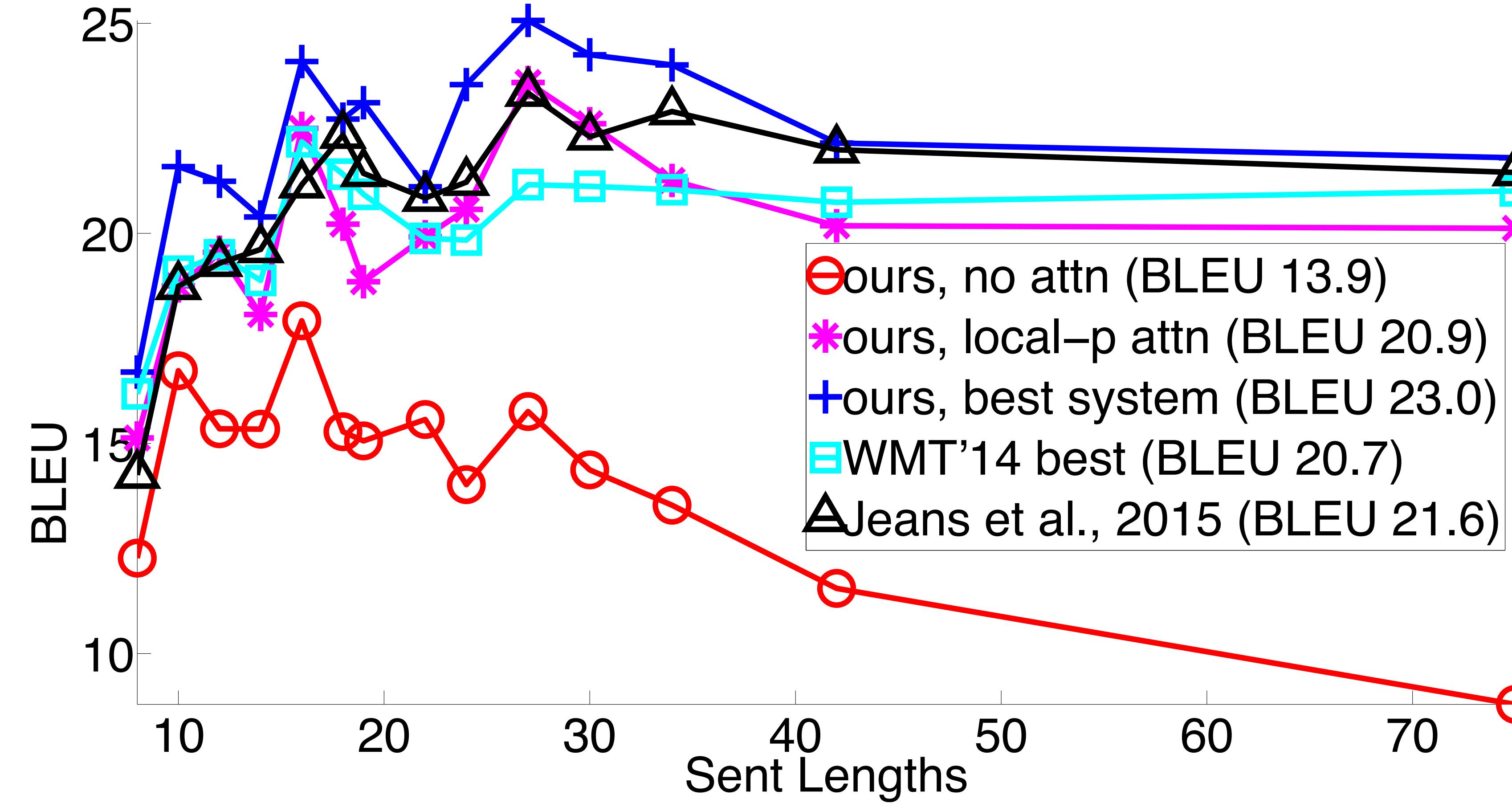
(c) Question Answering Tasks:
SQuAD v1.1



(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

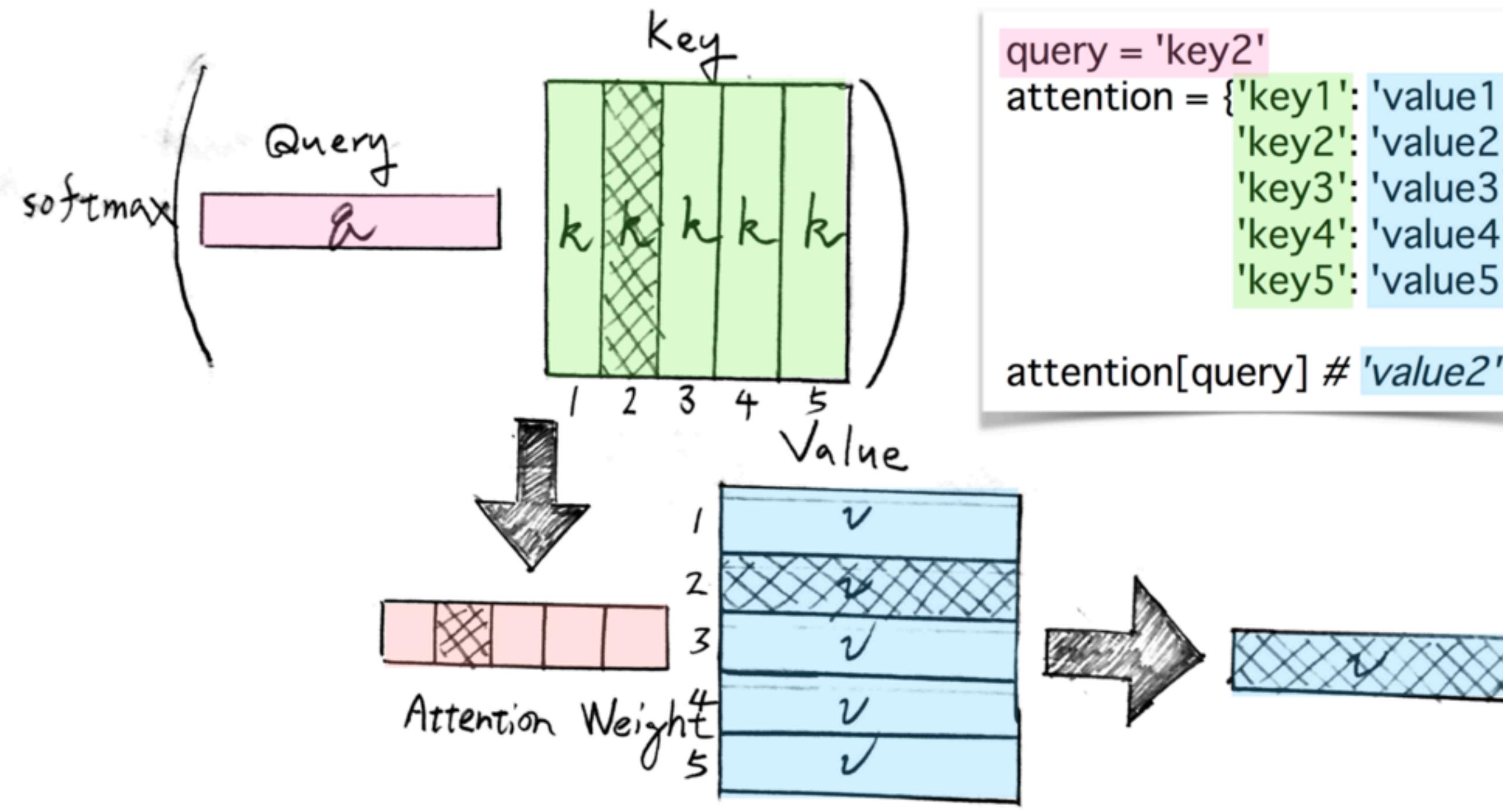
まとめ

- ▶ 時系列データには全結合よりリカレント
- ▶ ゲートを追加して単純なRNNの欠点を解消
 - LSTM, GRUなど
- ▶ 文脈を一箇所にまとめずAttentionで時間の関数に
- ▶ Self-AttentionはRNNなしの強力な文脈表現抽出器に
- ▶ BERTは自然言語処理のドーピング剤（？）



Attentionの解釈

注意は辞書オブジェクト
(Attention is a Dictionary object)



<http://deeplearning.hatenablog.com/entry/transformer>

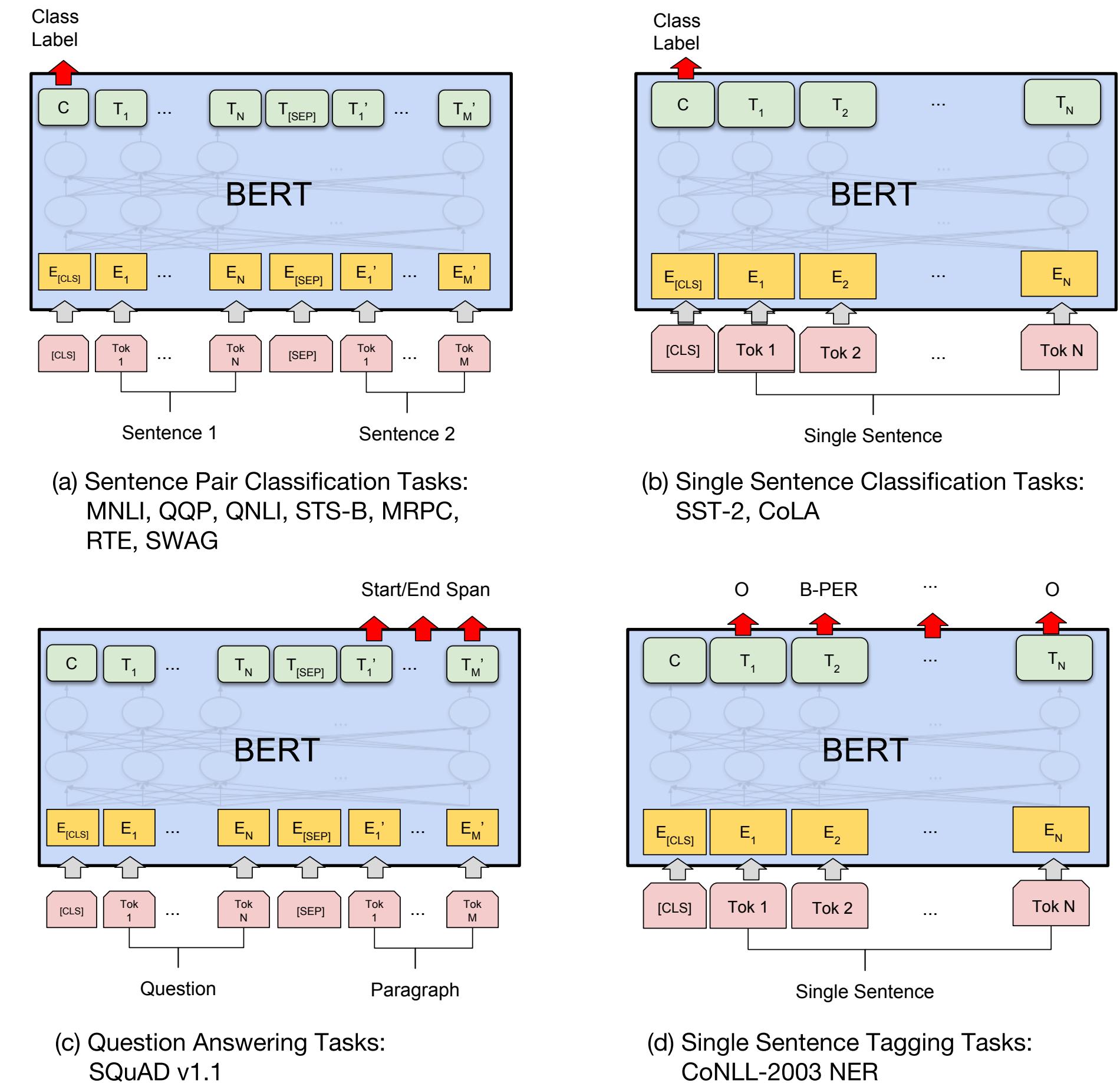


Figure 3: Our task specific models are formed by incorporating BERT with one additional output layer, so a minimal number of parameters need to be learned from scratch. Among the tasks, (a) and (b) are sequence-level tasks while (c) and (d) are token-level tasks. In the figure, E represents the input embedding, T_i represents the contextual representation of token i , [CLS] is the special symbol for classification output, and [SEP] is the special symbol to separate non-consecutive token sequences.