深層学習と自然言語処理

東京大学大学院
情報理工学系研究科
鶴岡 慶雅

概要

- ニューラルネットワーク
  - 多層ニューラルネットワーク
  - リカレントニューラルネットワーク
  - 畳み込みニューラルネットワーク

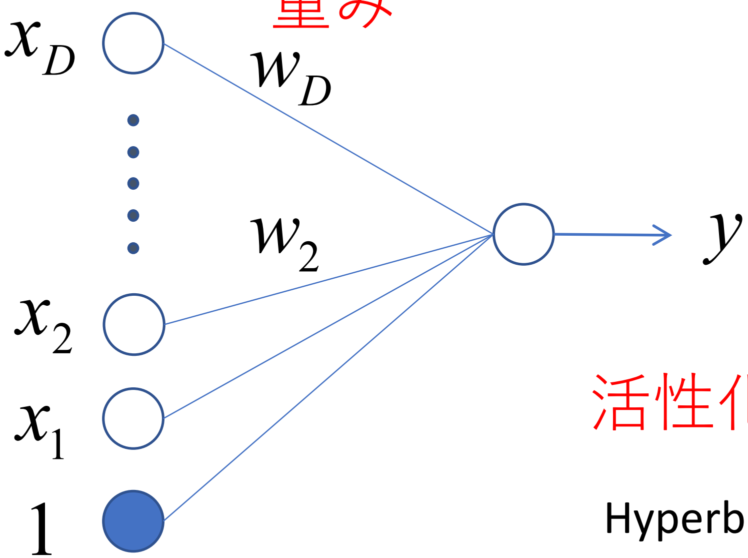
- 自然言語処理
  - 基盤技術
    - 言語モデル、品詞タグ付け、チャンキング、固有表現認識、構文解析
  - 応用
    - 機械翻訳、対話、要約、説明文生成、質問応答、プログラム生成

ニューラルネットワーク

- ニューロン

重み

入力

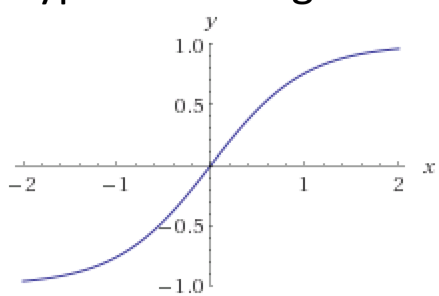


入力の線形和に非線形な活性化関数を適用

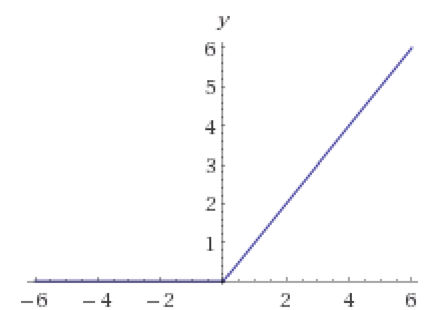
$$y = f\left(\sum_{i=0}^{D} w_i x_i\right)$$

活性化関数

Hyperbolic tangent

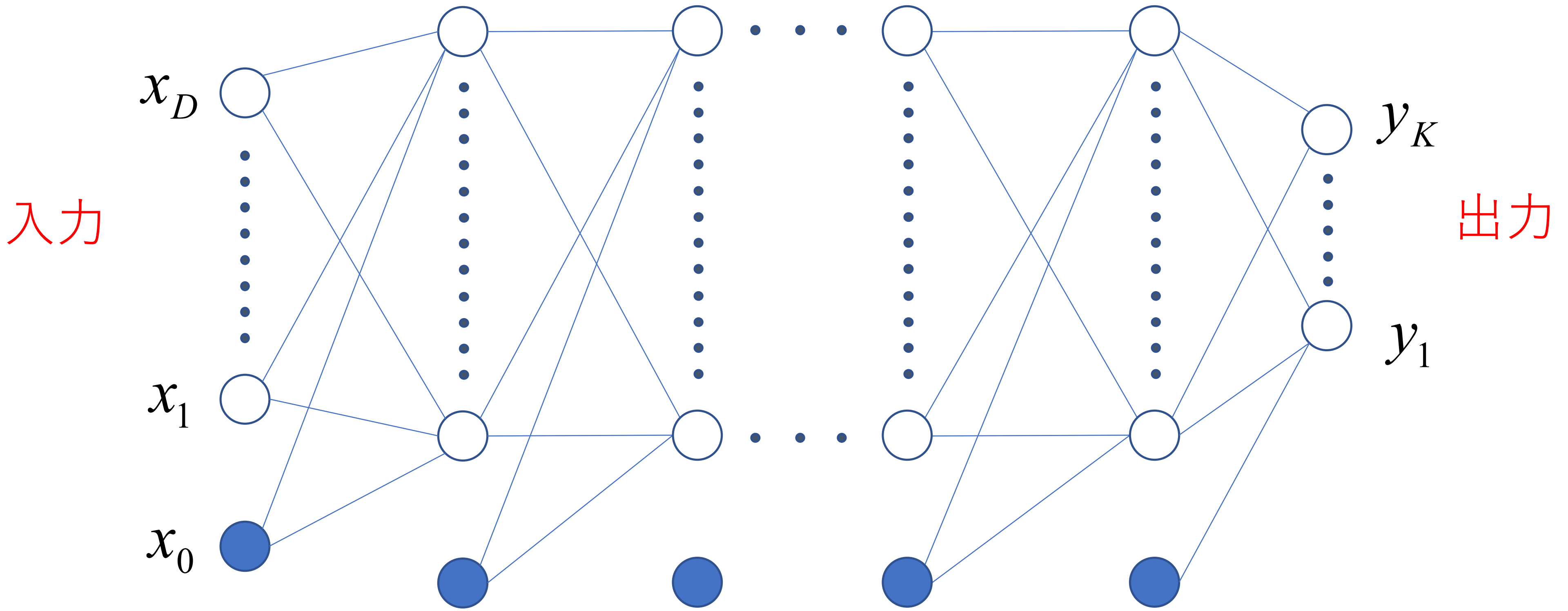


ReLU (Rectified Linear Unit)



多層ニューラルネットワーク

- 多数の入出力のペアから入出力関係を学習

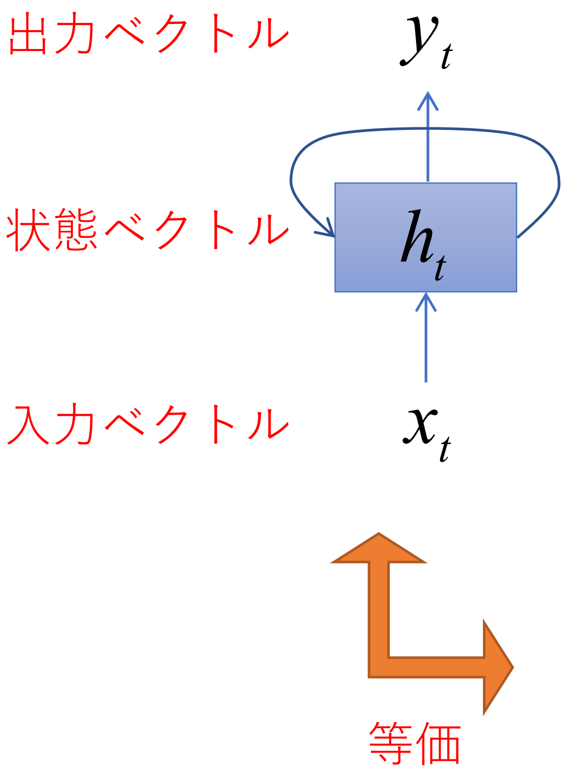


入出力の次元は固定

→　不定形な構造を持つ入出力は扱いにくい

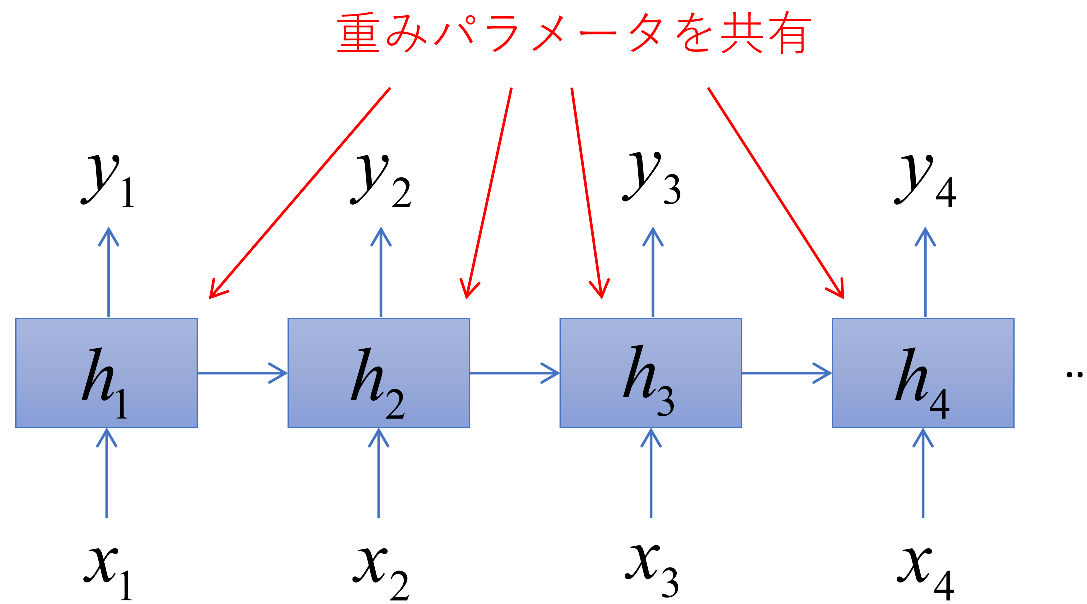
リカレントニューラルネットワーク (Recurrent Neural Network, RNN)

- 任意の長さの系列を扱うことができる



$$h_t = sigmoid\left(W^{\mathrm{hx}} x_t + W^{\mathrm{hh}} h_{t-1}\right)$$

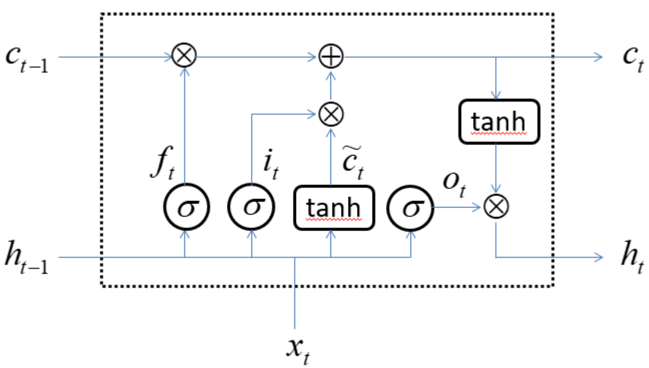
$$y_t = W^{\mathrm{yh}} h_t$$



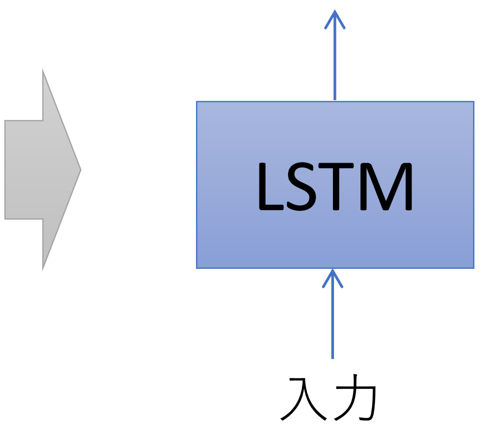
LSTM (Long Short-Term Memory)

- 単純な RNN の問題点
  - 勾配消失問題
  - 長距離の依存関係をとらえられない

- Long Short-Term Memory (LSTM)



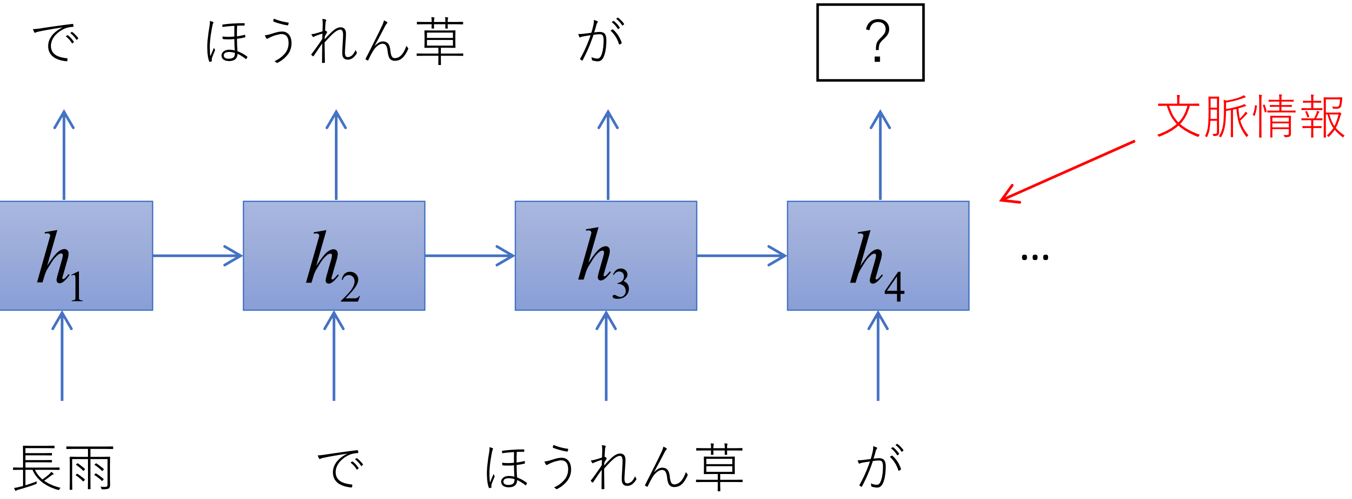
$$i_t = \sigma\left(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}\right)$$
$$f_t = \sigma\left(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}\right)$$
$$o_t = \sigma\left(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}\right)$$
$$\tilde{c}_t = \tanh\left(W^{(\tilde{c})}x_t + U^{(\tilde{c})}h_{t-1} + b^{(\tilde{c})}\right)$$
$$c_t = i_t \circ \tilde{c}_t + f_t \circ c_{t-1}$$
$$h_t = o_t \circ \tanh(c_t)$$



RNNと自然言語処理

- 自然言語処理では文字や単語の系列を扱う
  - 言語モデル、品詞タグ付け、固有表現認識、機械翻訳、etc.

- 例）言語モデル
  - 次の単語を予測



品詞タグ付け

- 文中の各単語に品詞情報を付与

Paul Krugman , a professor at Princeton University, was
NNP NNP , DT NN IN NNP NNP , VBD
awarded the Nobel Prize in Economics on Monday.
VBN DT NNP NNP IN NNP IN NNP

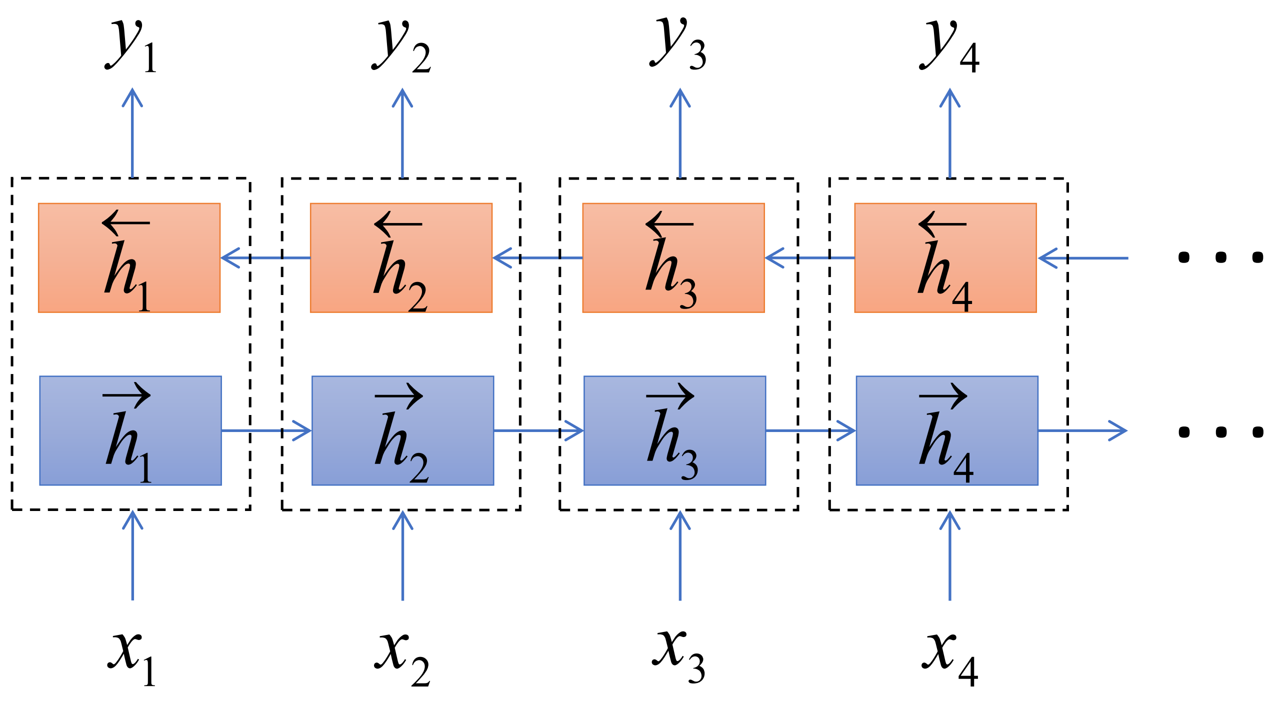
- 品詞タグ

NN: 名詞
NNP: 固有名詞
DT: 限定詞
:

IN: 前置詞
VBD: 動詞（過去形）
VBN: 動詞（過去分詞）
:

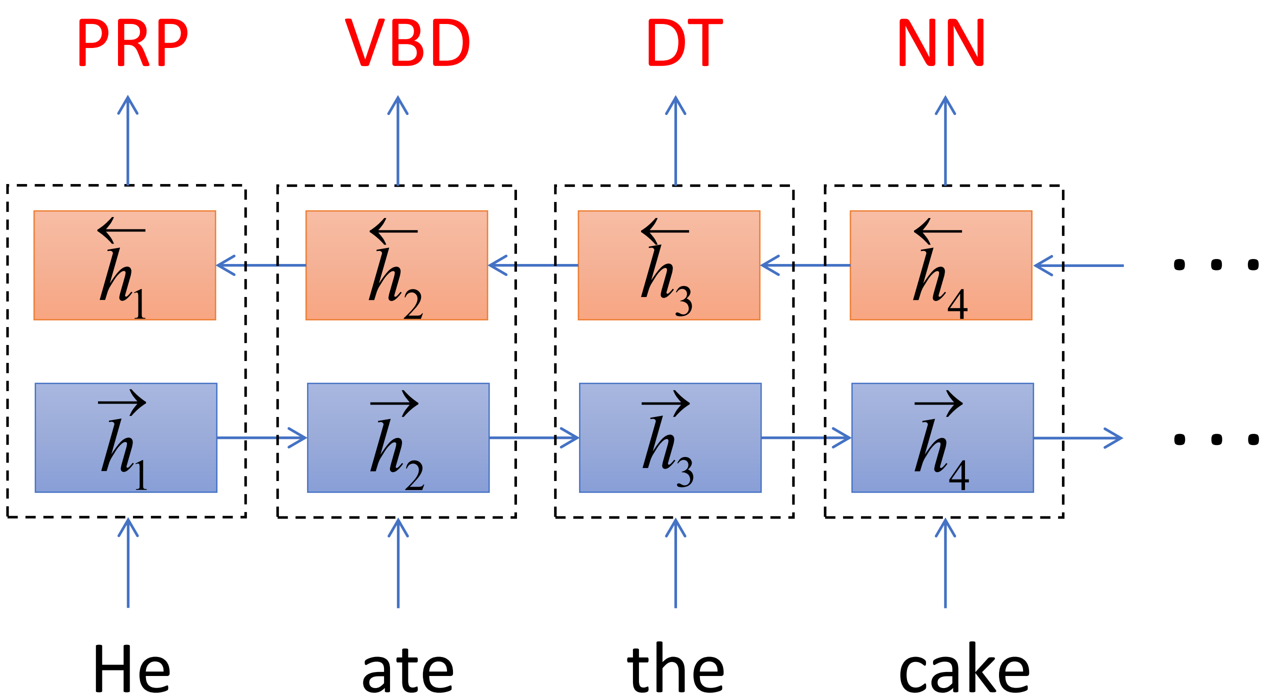
Bidirectional LSTM (BiLSTM)

- ２つのRNN（順方向と逆方向）
  - 左右両方向の文脈情報を捉えられる



BiLSTMによる品詞タグ付け

- 学習データ
  - Wall Street Journal コーパス：約40,000文



→ CRF 層を追加 (Huang et al., 2015; Ma and Hovy, 2016)

チャンキング(shallow parsing)

He reckons the current account deficit will narrow to

NP VP NP VP PP

only # 1.8 billion in September .

NP PP NP

- 文をフラットな句に分割
- 再帰的な分割は行わない

チャンキング (shallow parsing)

He	*reckons*	*the*	*current*	*account*	*deficit*	*will*	*narrow*	*to*
B_{NP}	B_{VP}	B_{NP}	I_{NP}	I_{NP}	I_{NP}	B_{VP}	I_{VP}	B_{PP}

only	*#*	*1.8*	*billion*	*in*	*September*	*.*
B_{NP}	I_{NP}	I_{NP}	I_{NP}	B_{PP}	B_{NP}	O

- 各単語に対するタグ付けの問題に変換できる
    - B: チャンクの先頭
    - I: チャンクの中（先頭以外）
    - O: チャンクの外

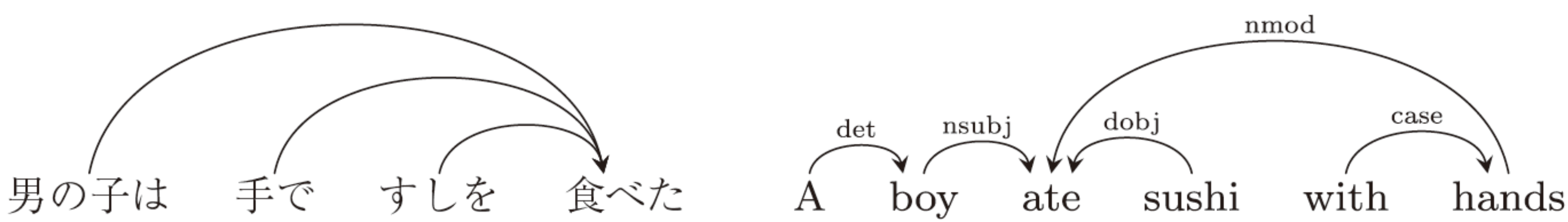
固有表現認識 (named entity recognition)

The peri-kappa B site mediates human immunodeficiency
DNA virus
virus type 2 enhancer activation in monocytes …
cell_type

- 文中の固有表現を認識
- チャンキングと同様、系列ラベリングの問題として処理できる

構文解析

- 依存構造（dependency structure）
  - 文中の単語間の関係をグラフで表す
    - 係り元（dependent）から係り先（head, 主辞）へのエッジ（矢印を逆向きに描くことも多い）
  - 係り受け構造とも呼ばれる



Shift-Reduce法による依存造解析

- Shift-Reduce 法
  - バッファー（buffer）
    - 解析前の単語列を格納
  - スタック（stack）
    - 解析後の依存構造を格納
  - アクション（action）
    - Shift
      - バッファーの先頭の単語をスタックに移動
    - Reduce
      - スタックトップの２つの単語の間にエッジを生成

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	**STACK**	**BUFFER**
Shift **ReduceL** **ReduceR**		I saw a dog with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	I	saw a dog with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	I saw	a dog with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift **ReduceL** **ReduceR**	saw	a dog with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	saw a	dog with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift **ReduceL** **ReduceR**	saw a dog	with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	saw dog	with eyebrows

Shift-Reduce 法

I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift **ReduceL** **ReduceR**	saw dog with	eyebrows

Shift-Reduce 法

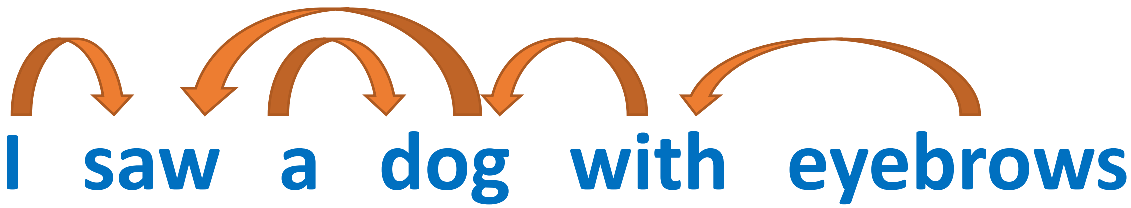
I saw a dog with eyebrows

OPERATION	STACK	BUFFER
Shift **ReduceL** **ReduceR**	saw dog with eyebrows	

Shift-Reduce 法

OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	saw dog with	

Shift-Reduce 法

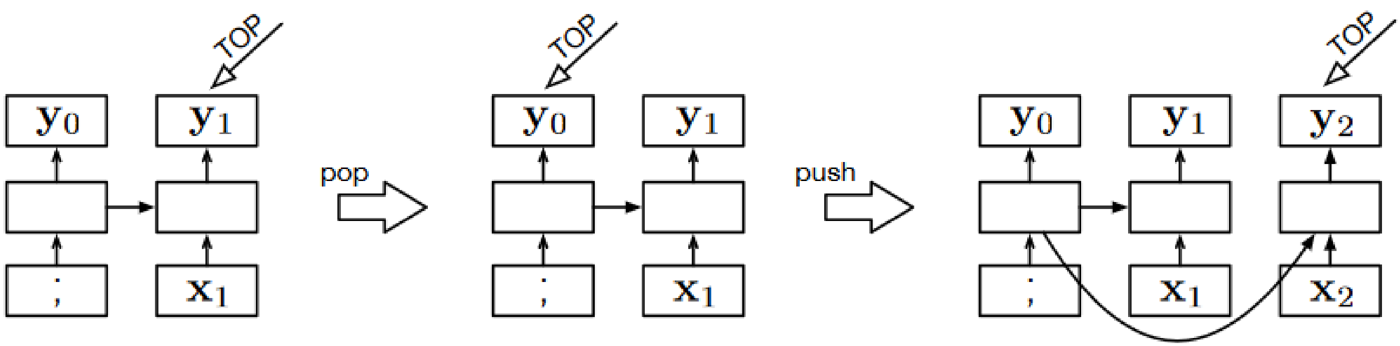


OPERATION	STACK	BUFFER
Shift ReduceL ReduceR	saw dog	

Shift-Reduce 法

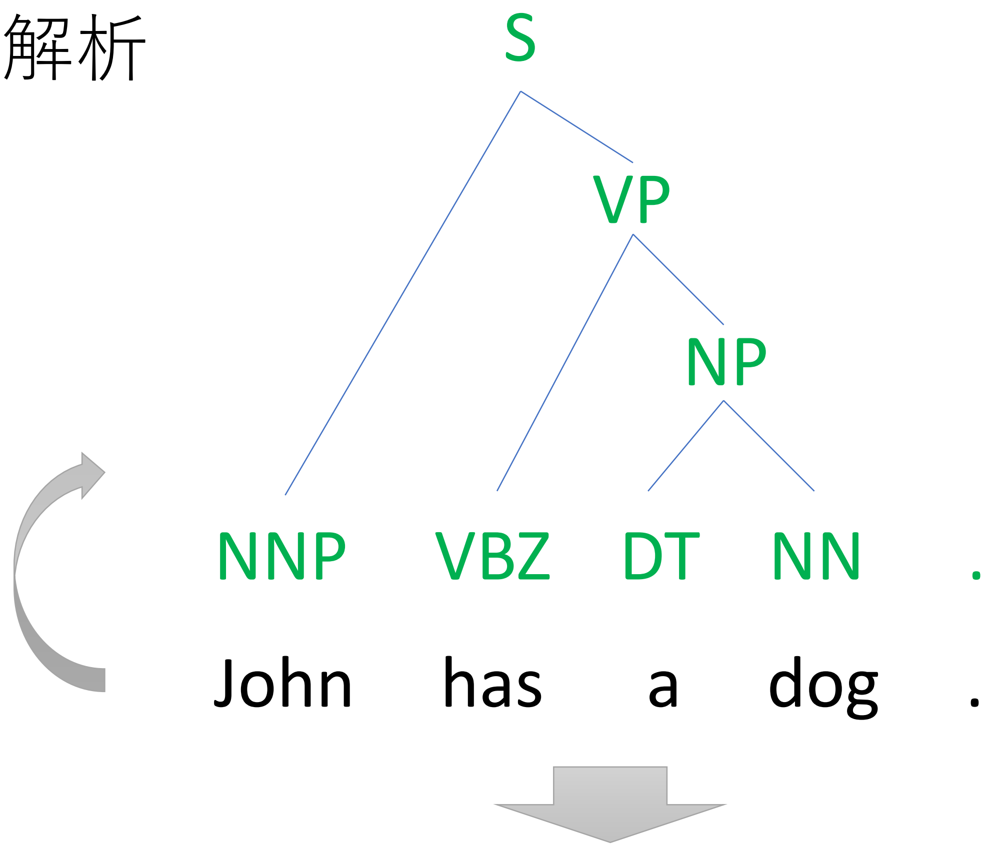
- 学習モデル
  - Stack LSTM (Dyer et al., 2015)
  - FFNN + ビーム探索 + Early Update (Andor et al., 2016)
  - BiLSTM (Kiperwasser & Goldberg, 2016)

- Stack LSTM (Dyer et al., 2015)



構文解析

- 句構造解析



(S (NP NNP)$_{NP}$ (VP VBZ (NP DT NN)$_{NP}$)$_{VP}$.)$_{S}$

Recurrent Neural Network Grammar (RNNG) (Dyer et al., 2016)

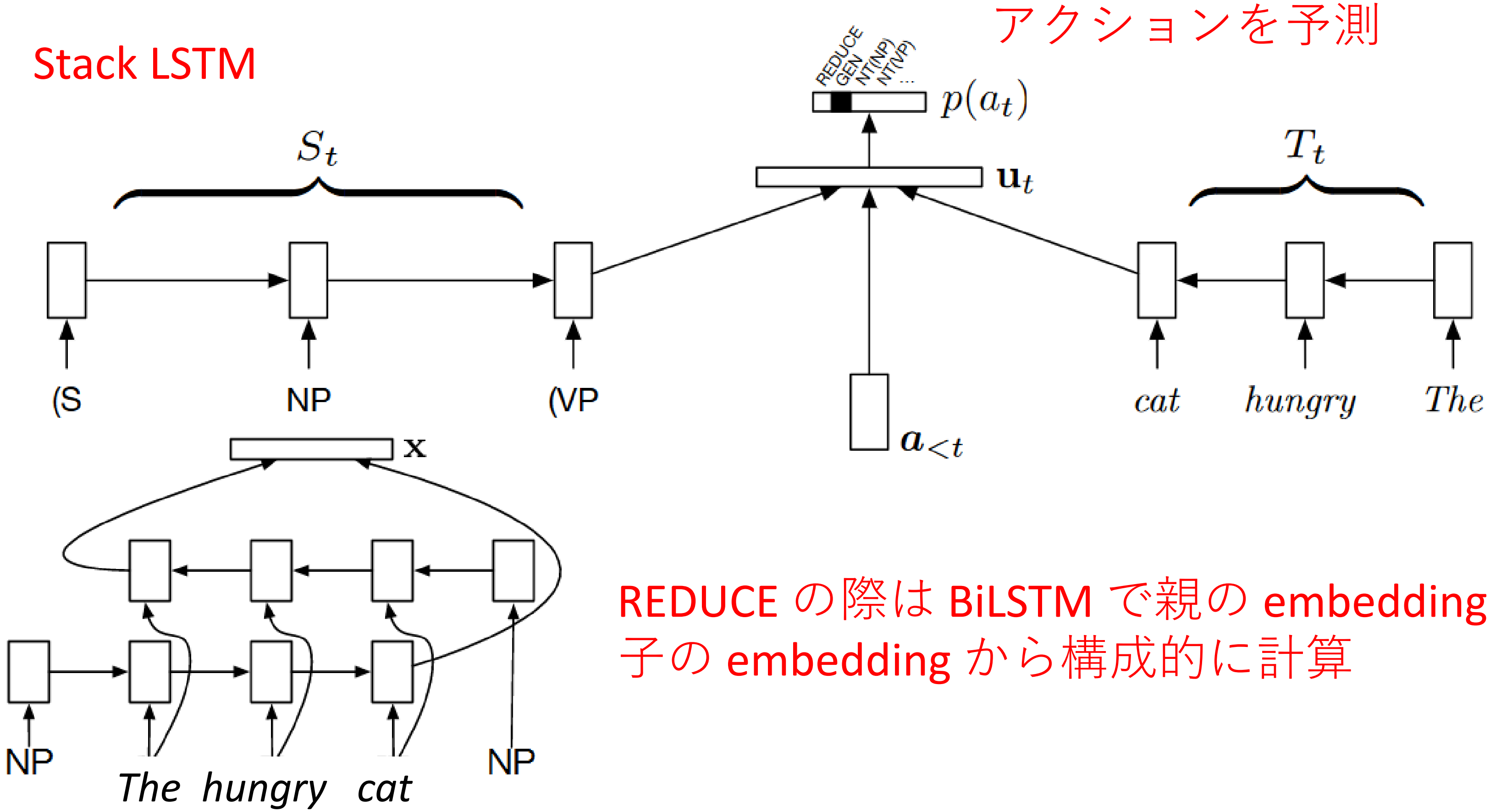
入力文: *The hungry cat meows .*

	Stack	**Buffer**	**Action**
0		*The* | *hungry* | *cat* | *meows* | *.*	NT(S)
1	(S	*The* | *hungry* | *cat* | *meows* | *.*	NT(NP)
2	(S | (NP	*The* | *hungry* | *cat* | *meows* | *.*	SHIFT
3	(S | (NP | *The*	*hungry* | *cat* | *meows* | *.*	SHIFT
4	(S | (NP | *The hungry*	*cat* | *meows* | *.*	SHIFT
5	(S | (NP | *The hungry* | *cat*	*meows* | *.*	REDUCE
6	(S | (NP | *The hungry* | *cat*)	*meows* | *.*	NT(VP)
7	(S | (NP | *The hungry* | *cat*) (VP	*meows* | *.*	SHIFT
8	(S | (NP | *The hungry* | *cat*) (VP *meows*	*.*	REDUCE
9	(S | (NP | *The hungry* | *cat*) (VP *meows*)	*.*	SHIFT
10	(S | (NP | *The hungry* | *cat*) (VP *meows*) | *.*		REDUCE
11	(S | (NP | *The hungry* | *cat*) (VP *meows*) | *.*)		

Recurrent Neural Network Grammar (RNNG) (Dyer et al., 2016)

Stack LSTM

アクションを予測



REDUCE の際は BiLSTM で親の embedding を子の embedding から構成的に計算

構文解析の精度

- 依存構造解析

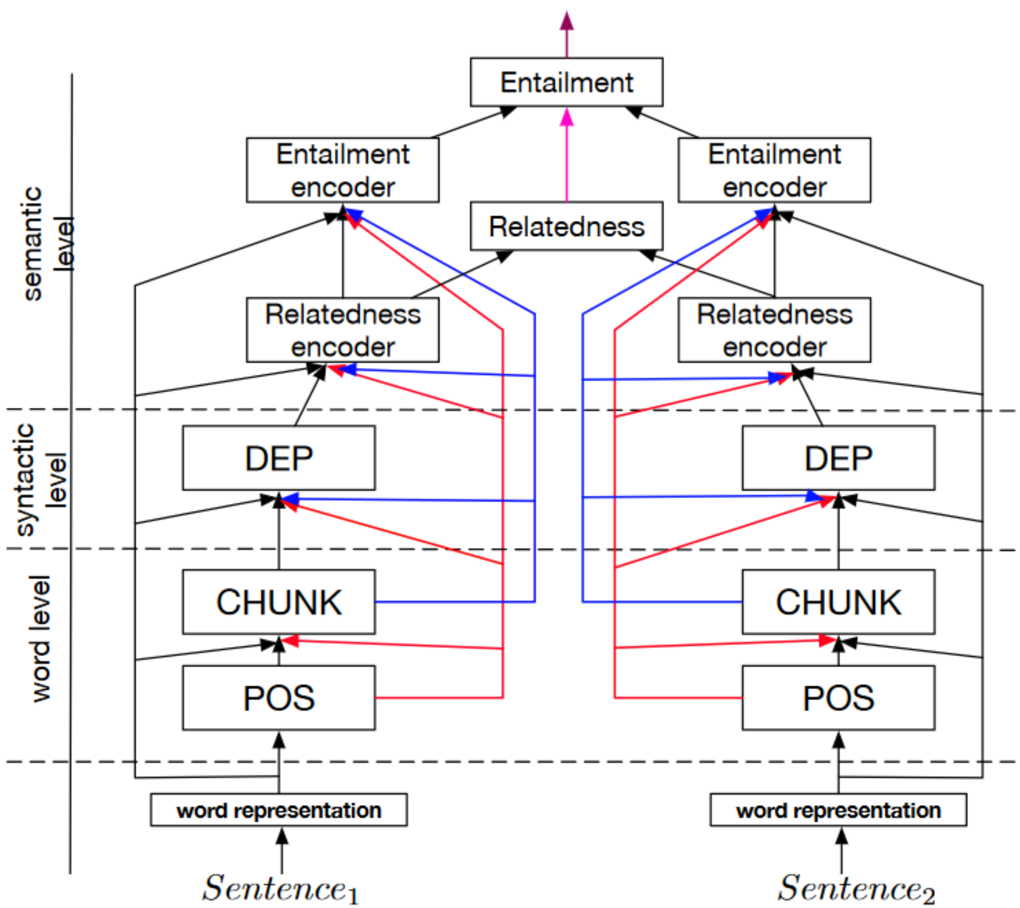
パーザー	UAS	LAS
Turbo parser (Martins et al., 2013)	92.9	90.6
SyntaxNet (Andor et al., 2016)	94.6	92.8
Deep Biaffine Attention (Dozat and Manning, 2016)	95.4	93.8
Recurrent Neural Network Grammar (Dyer et al. 2016)	95.8	94.6

- 句構造解析

パーザー	F1 score
Symbol-refined CFG (Petrov and Klein, 2007)	90.1
Bayesian Symbol-refined TSG (Shindo et al. 2012)	91.1
Recurrent Neural Network Grammar (Dyer et al. 2016)	93.6

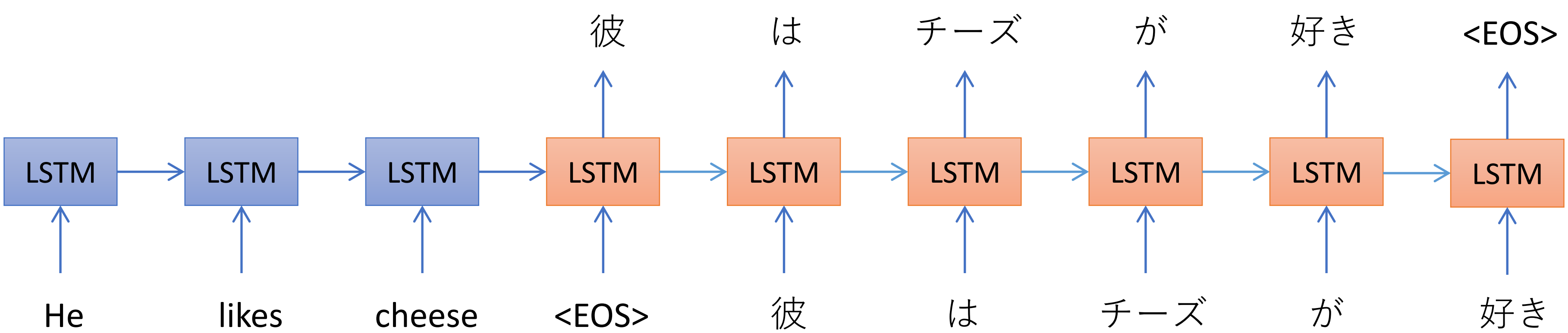
マルチタスク学習による解析

- Joint Many-Task Model (Hashimoto et al., 2016)
  - ５つのタスクを同時学習



ニューラル機械翻訳

- エンコーダ・デコーダモデル (Sutskever et al., 2014)
  - Encoder RNN
    - 翻訳元の文を読み込み、実数値ベクトルに変換
  - Decoder RNN
    - 実数値ベクトルから翻訳先言語の文を生成



アテンション (Bahdanau et al., 2015)

- 翻訳先の各単語を選択する際に、翻訳元の文中の各単語の隠れ状態の情報を利用

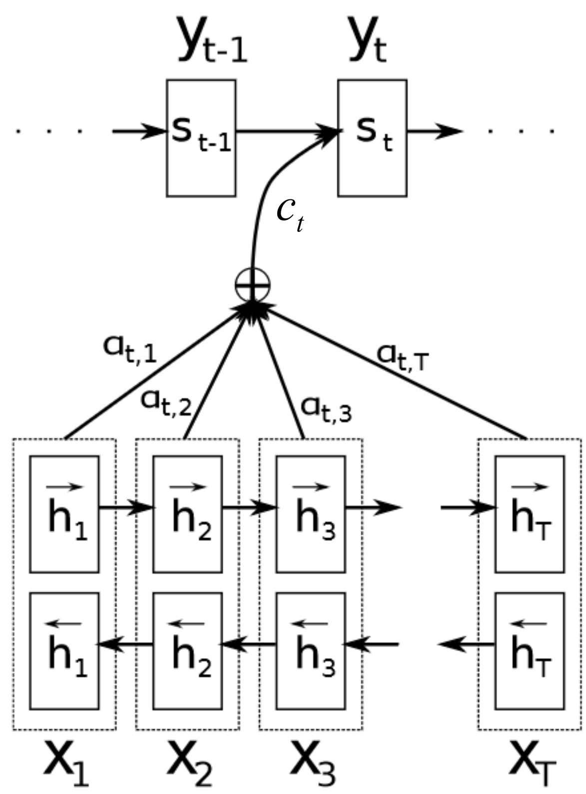
翻訳元の各単語の隠れ状態の加重平均

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j$$

重み（すべて足すと１）

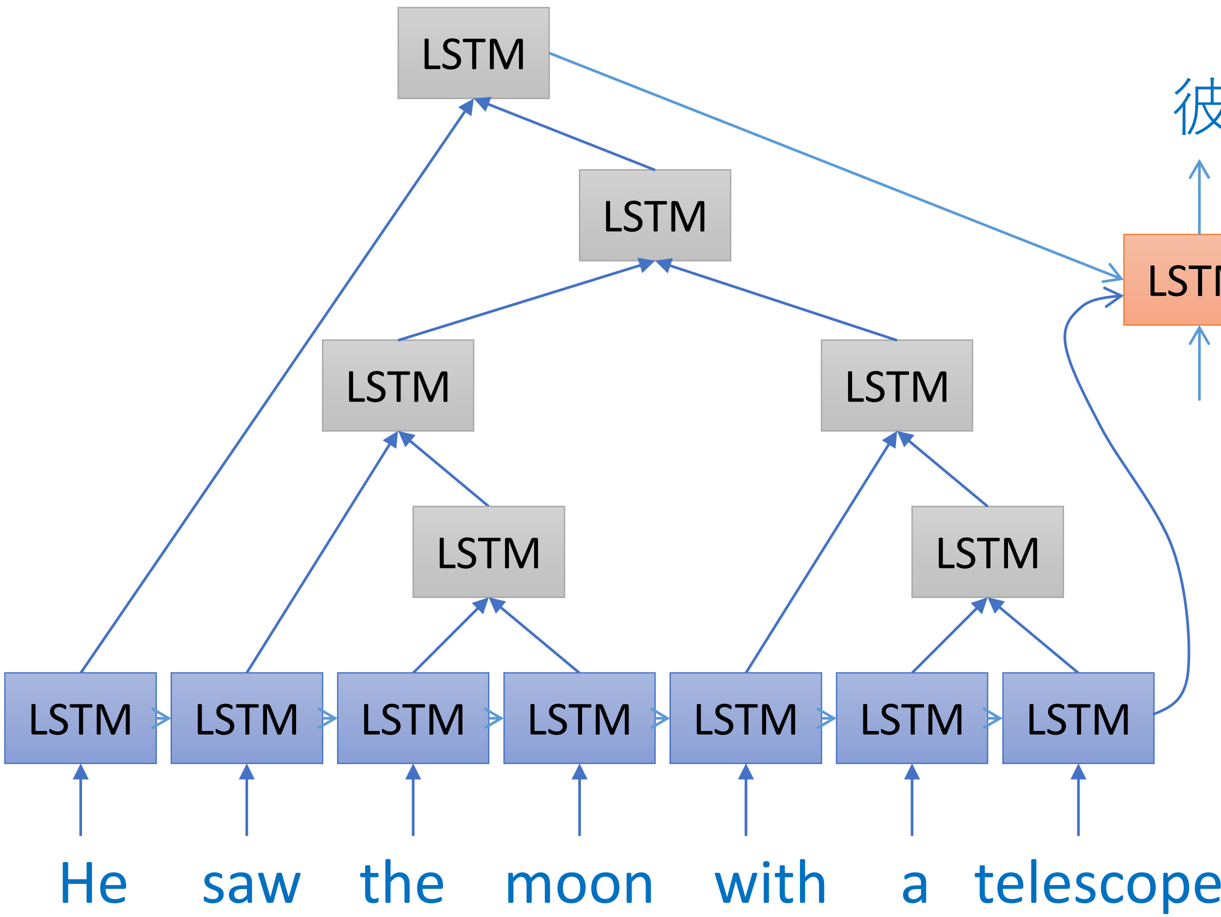
$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})}$$

$$e_{ij} = FeedForwardNN(s_{i-1}, h_j)$$



Tree-to-sequence モデル

- 入力文の句構造を利用 (Eriguchi et al. 2016)



翻訳例

In information technology and electron field, the application of nanotechnology to next generation semiconductors, high-density information record technology, miniature integrated circuit elements, electric power saving displays using carbon nano-tube, etc. can be expected.

情報 技術 と 電子 分野 で は， 次 世代 半 導体 へ の ナノテク ノロジー の 応用， 高 密度 情報 記録 技術， 超 小型 集積 回 路 要素， カーボン ナノ チューブ を 用い た 省 電力 表示 な ど が 期待 でき る 。

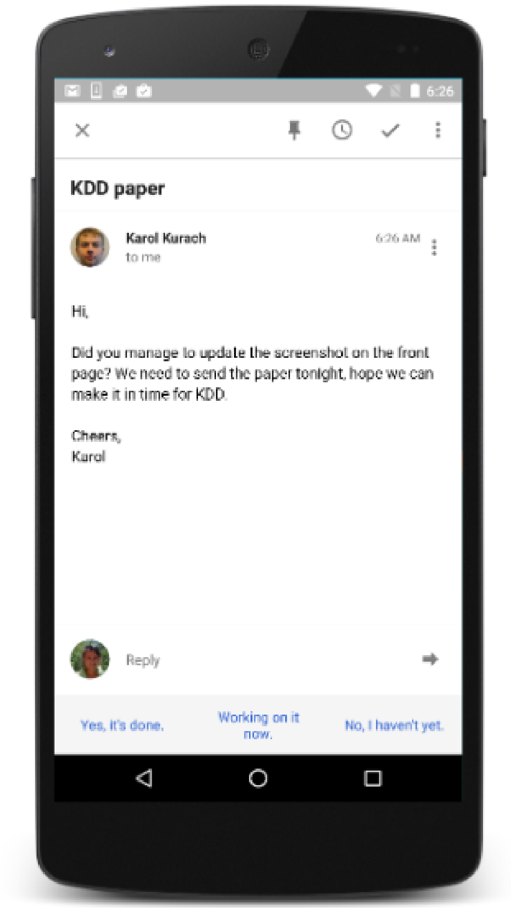
ニューラル会話モデル (Vinyals & Le, 2015)

- エンコーダ・デコーダモデルで会話文を学習
  - 学習データ
    - 映画の中での会話　6200万文
- 学習したモデルと人間との会話の例

Human: who is skywalker ?
Machine: he is a hero .
Human: who is bill clinton ?
Machine: he 's a billionaire .
Human: is sky blue or black ?
Machine: blue .
Human: does a cat have a tail ?
Machine: yes .

Human: does a cat have a wing ?
Machine: no
Human: can a cat fly ?
Machine: no .
Human: how many legs does a cat have ?
Machine: four , i think .
Human: how many legs does a spider have ?
Machine: three , i think .

Eメールの返事の自動生成 (Kannan et al., 2016)



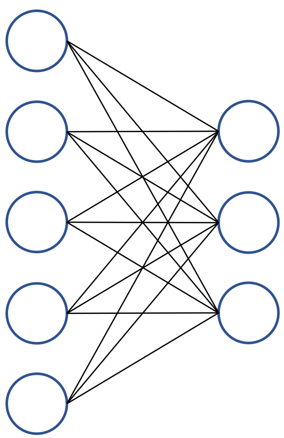
Query	Top generated responses
Hi, I thought it would be great for us to sit down and chat. I am free Tuesday and Wenesday. Can you do either of those days? Thanks! –Alice	I can do Tuesday. I can do Wednesday. How about Tuesday? I can do Tuesday! I can do Tuesday. What time works for you? I can do Wednesday! I can do Tuesday or Wednesday. How about Wednesday? I can do Wednesday. What time works for you? I can do either.

Table 1: Generated response examples.

Kannan et al., Smart Reply: Automated Response Suggestion for Email, KDD 2016

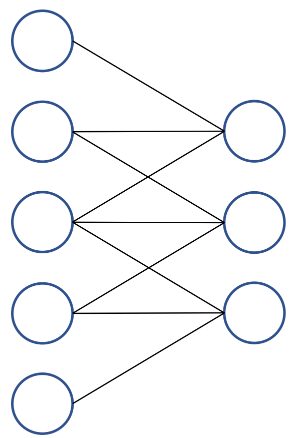
畳み込みニューラルネットワーク（Convolutional Neural Network, CNN）

- 全結合



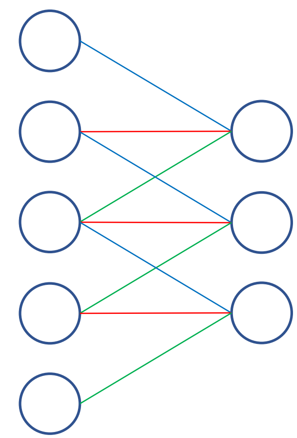
パラメータ数
5 x 3 = 15

- 局所的結合



パラメータ数
3 x 3 = 9

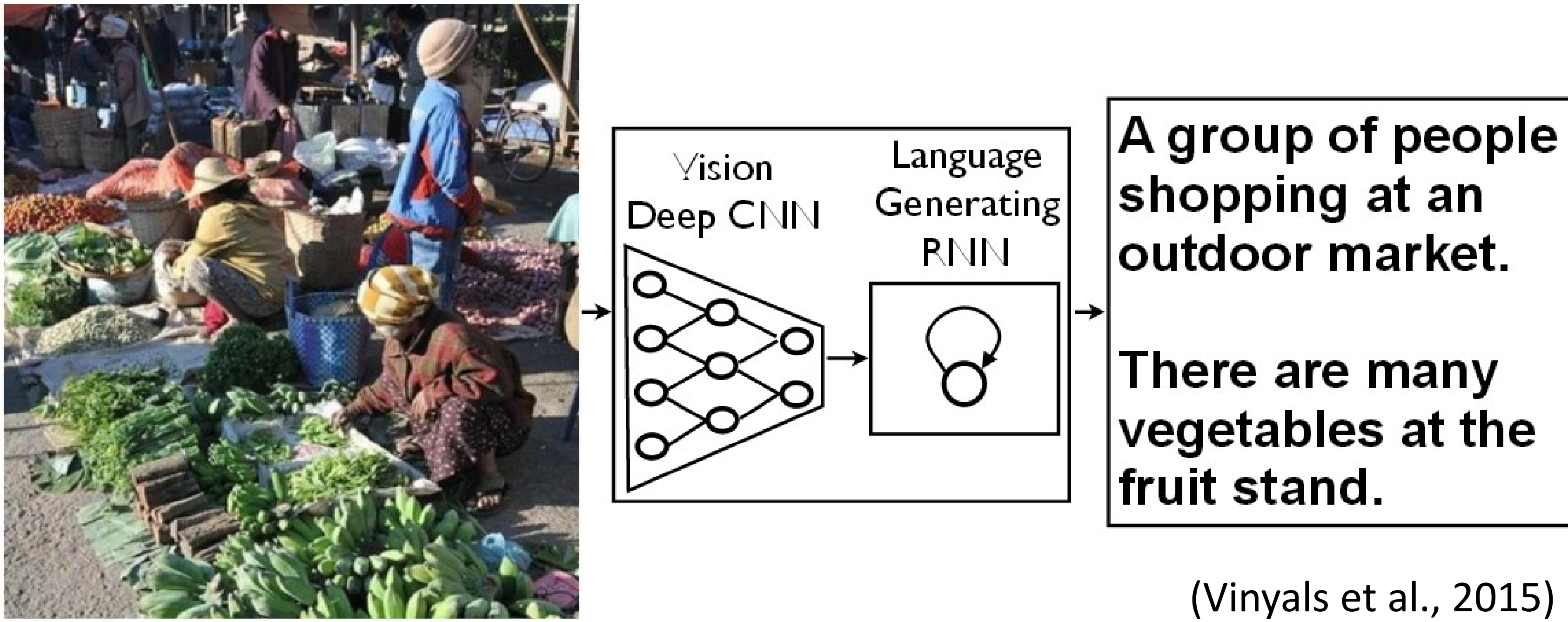
- パラメータ共有



パラメータ数
3

パラメータ数を減らすことにより過学習を回避
画像認識、テキスト分類などに有効

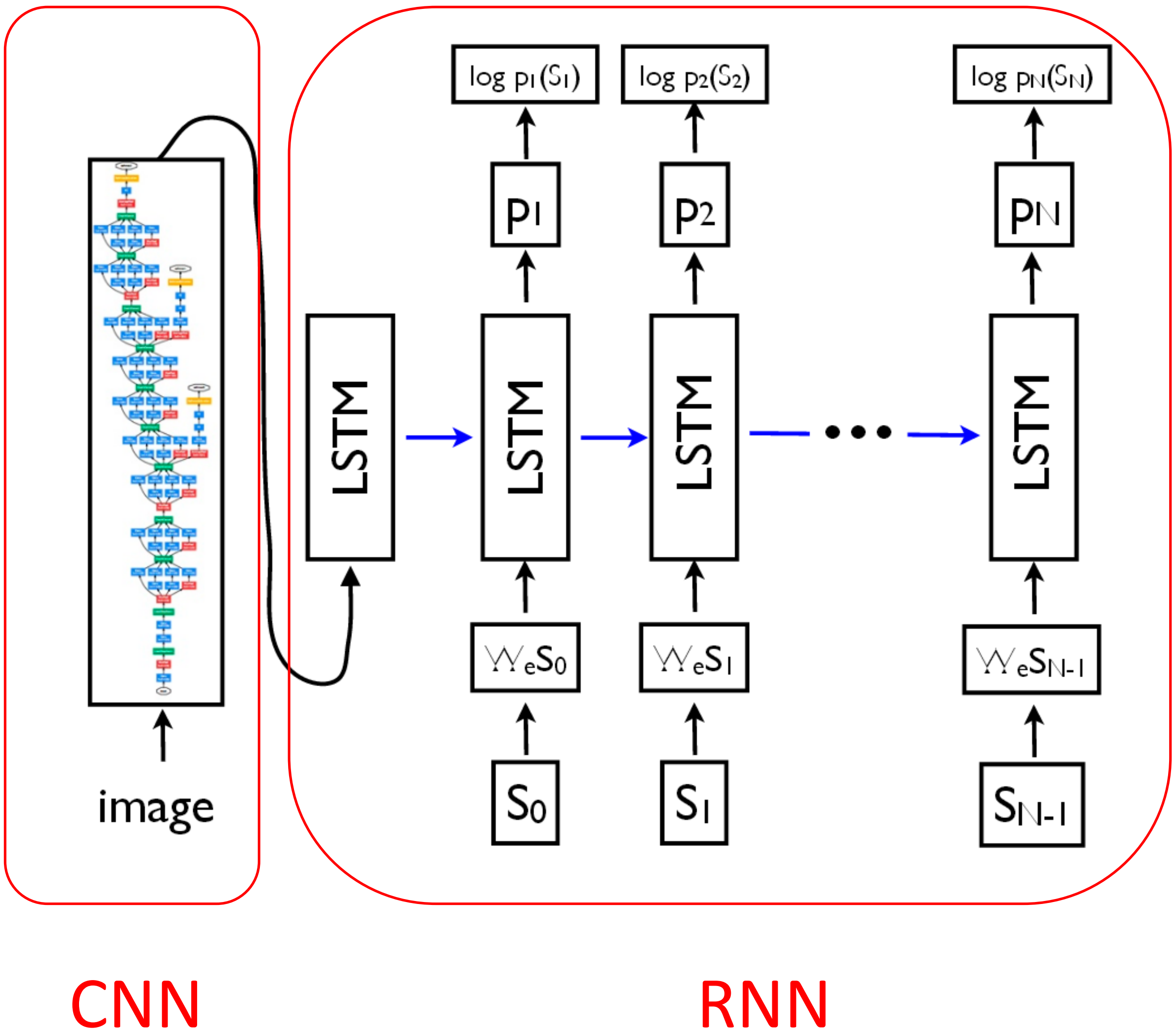
画像の説明文の生成



(Vinyals et al., 2015)

1. 大量のラベル付き画像で画像認識CNNを学習
2. 説明文付きの画像で言語生成RNNを学習

画像の説明文の生成



(Vinyals et al., 2015)

説明文生成例 (Vinyals et al., 2015)

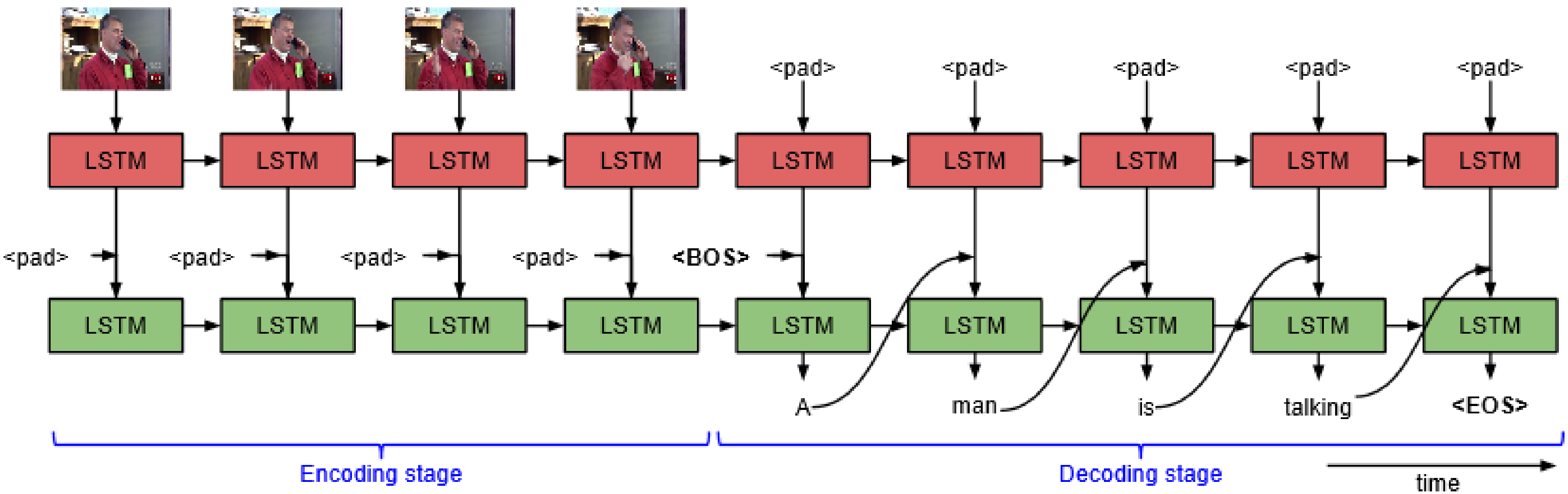
A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A group of young people playing a game of frisbee.

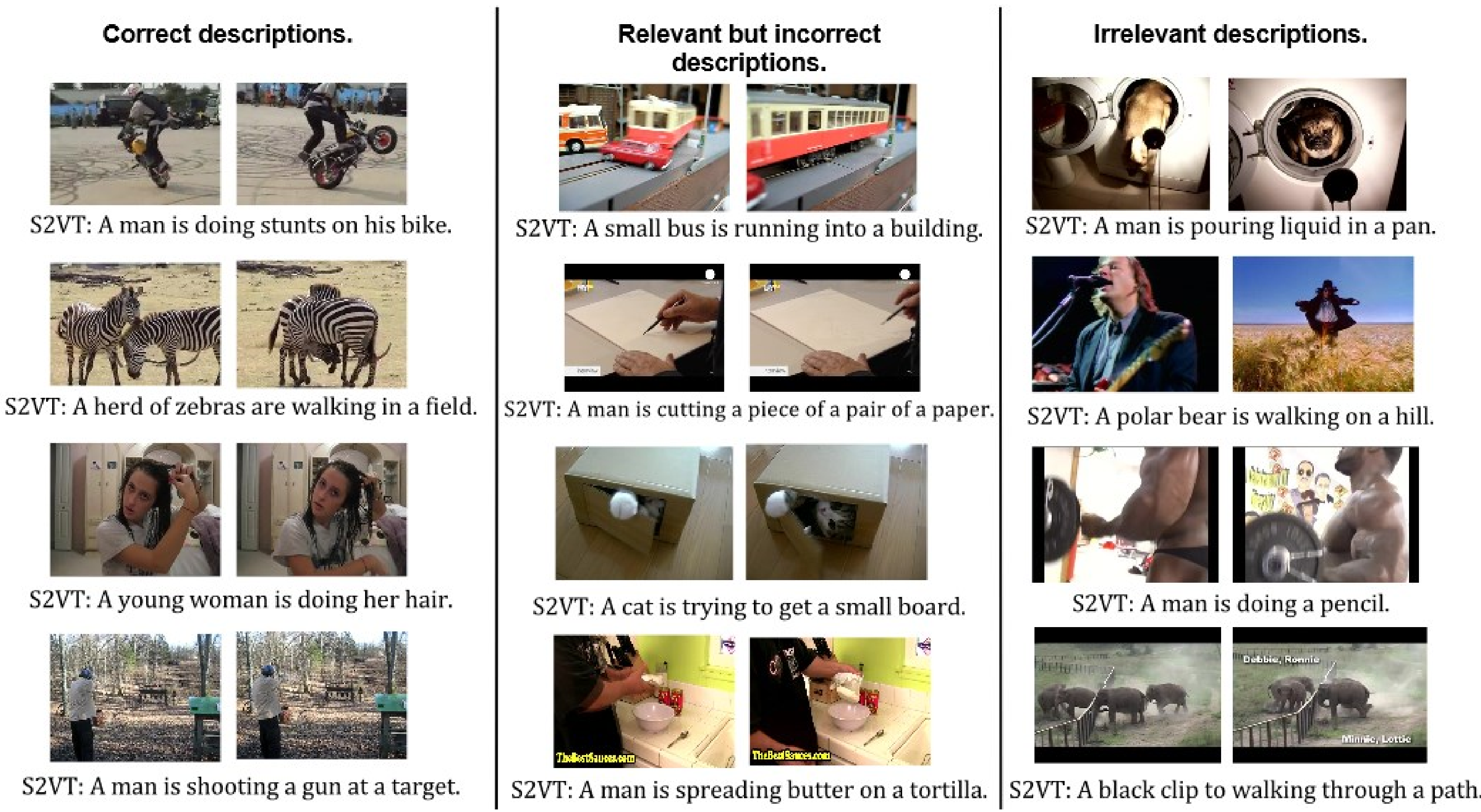
Two hockey players are fighting over the puck.

動画の説明文の生成



Venugopalan et al., Sequence to Sequence – Video to Text, ICCV 2015

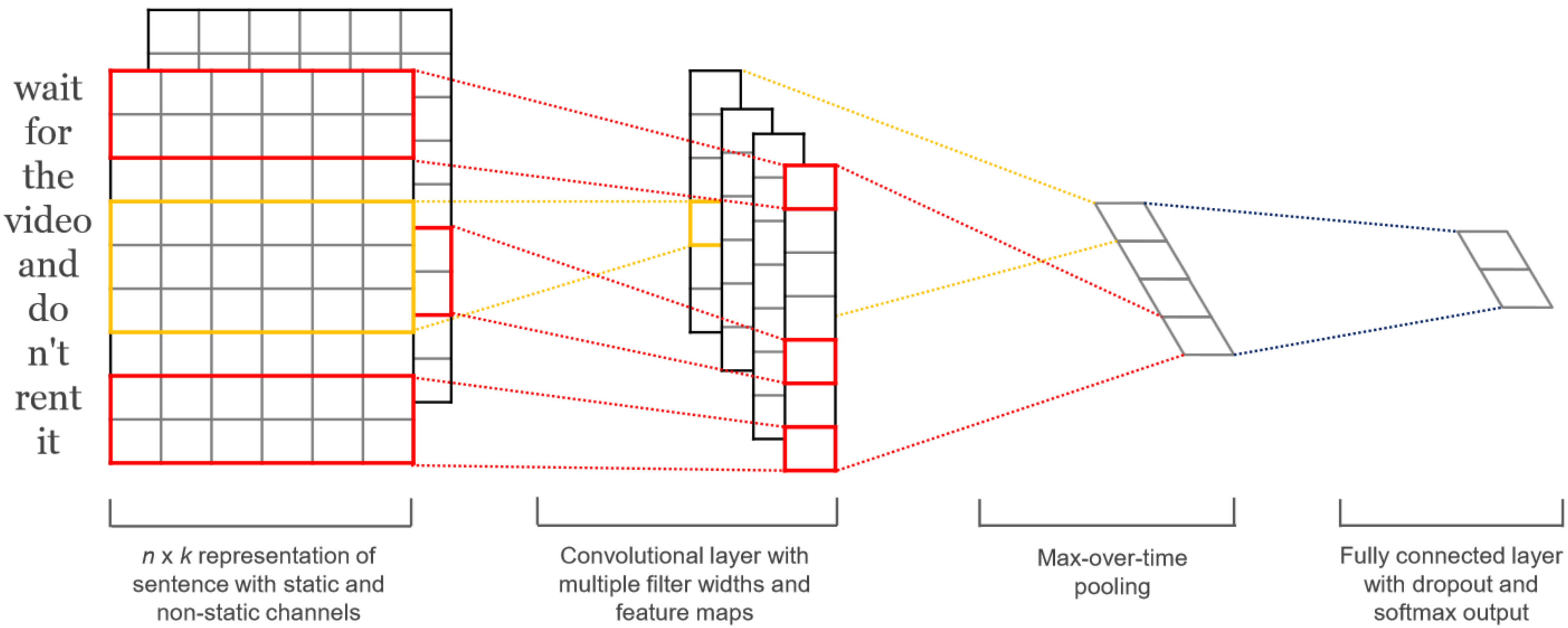
動画の説明文の生成



Venugopalan et al., Sequence to Sequence – Video to Text, ICCV 2015

CNN による文分類 (Kim, 2014)

- 単語 n-gram（的なものを）をソフトに検出



SQuAD
(The Sanford Question Answering Dataset)

Rajpurkar et al. (2016)

- Wikipedia記事に関するQAデータセット
- 大規模
  - 500記事、10万QAペア
  - クラウドソーシングによって作成
- 質問の答えは文書中の単語列

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

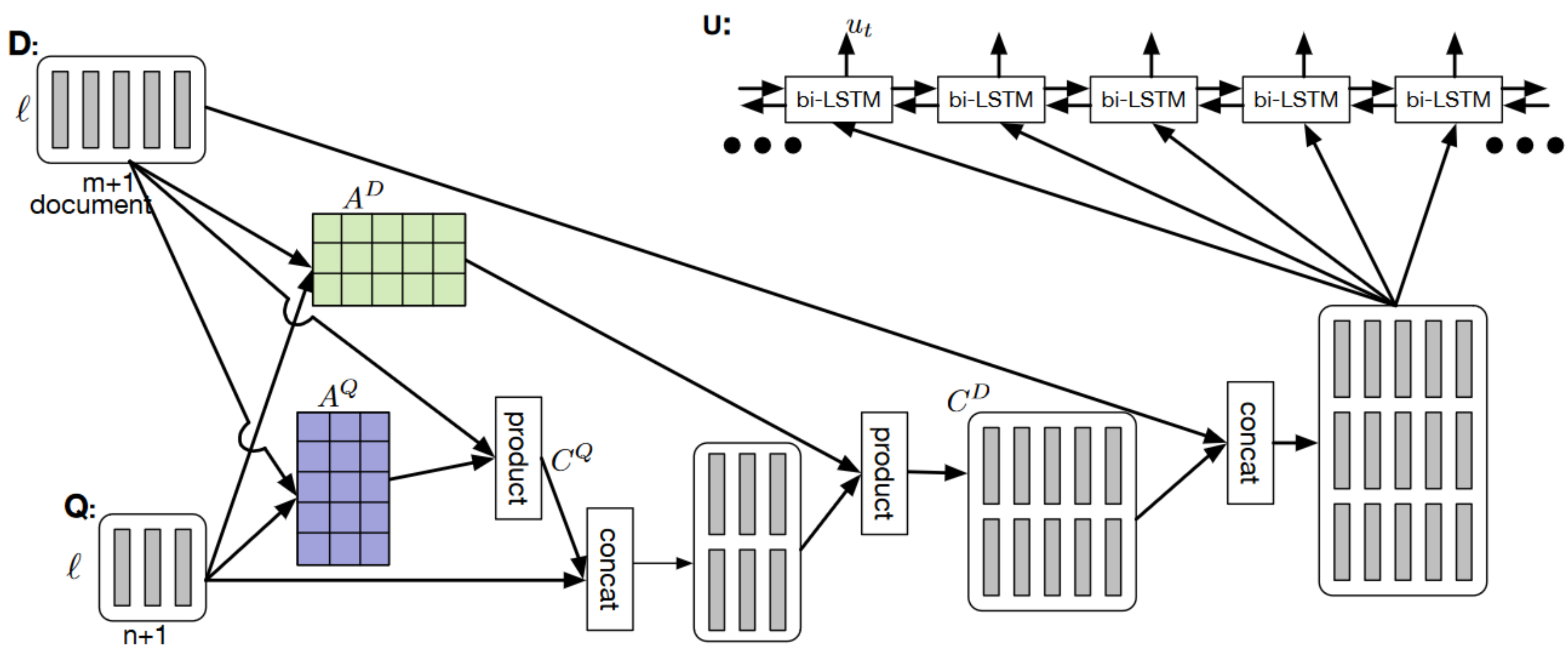
What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Dynamic Coattention Networks (Xiong et al., 2016)

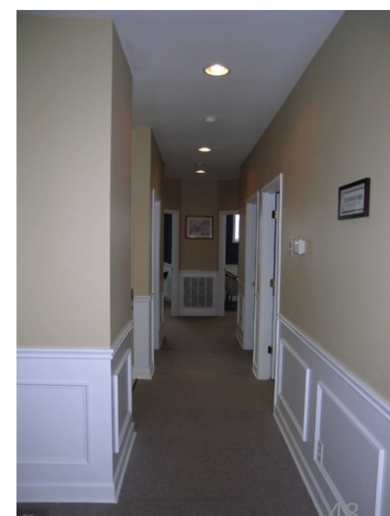
- 文書と質問の中の各単語を互いの類似度で重みづけ



推論を必要とする質問応答（QA）

文書

Mary got the football there.
John moved to the bedroom.
Sandra went back to the kitchen.
Mary travelled to the hallway.
John got the football there.
John went to the hallway.
John put down the football.
Mary went to the garden.

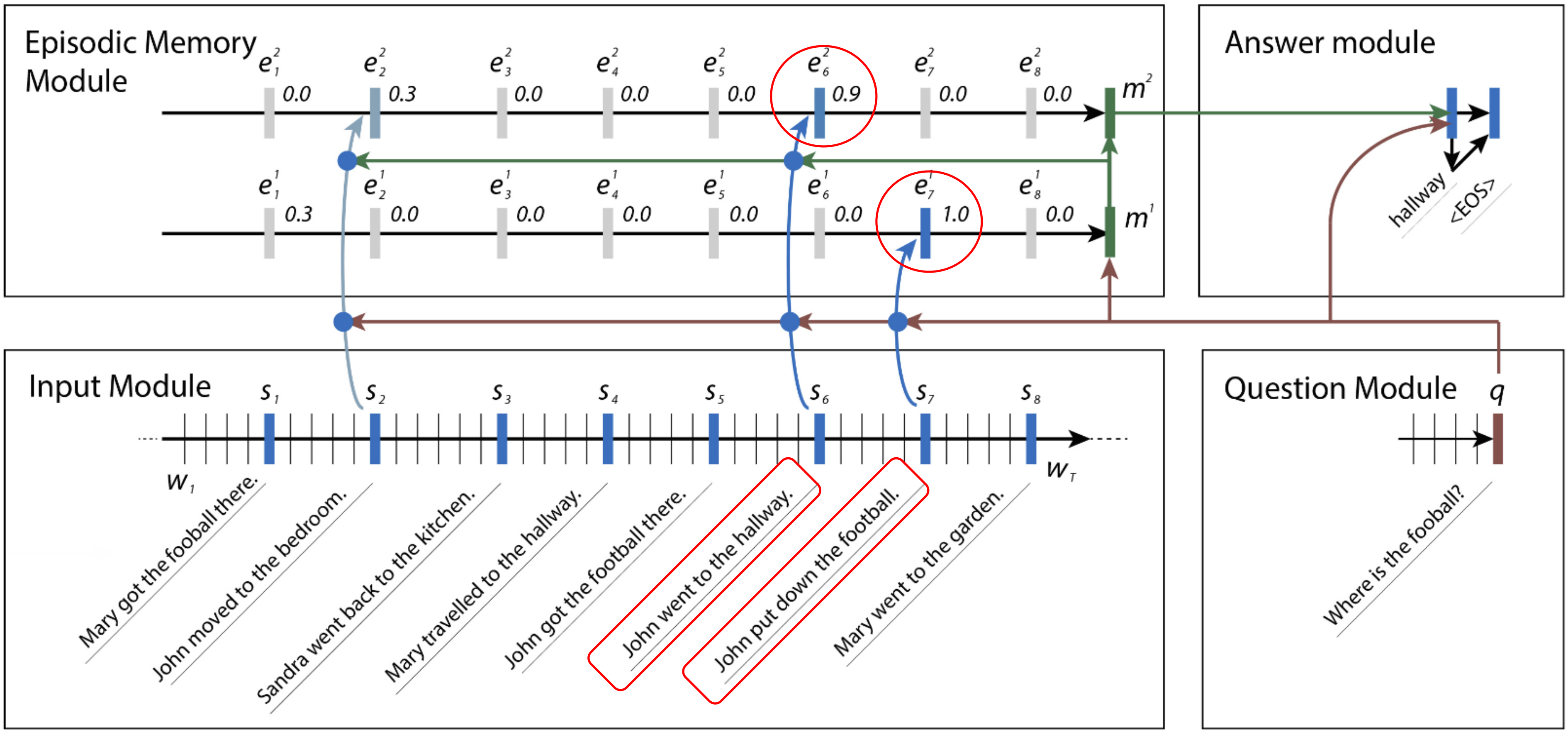


質問

Where is the football?

Dynamic Memory Networks (Kumar et al., 2016)

- 答えを導出するために必要な文を順次推定



文書要約

Original Text (truncated): lagos, nigeria (cnn) a day after winning nigeria's presidency, ***muhammadu buhari*** told cnn's christiane amanpour that **he plans to aggressively fight corruption that has long plagued nigeria** an…

…

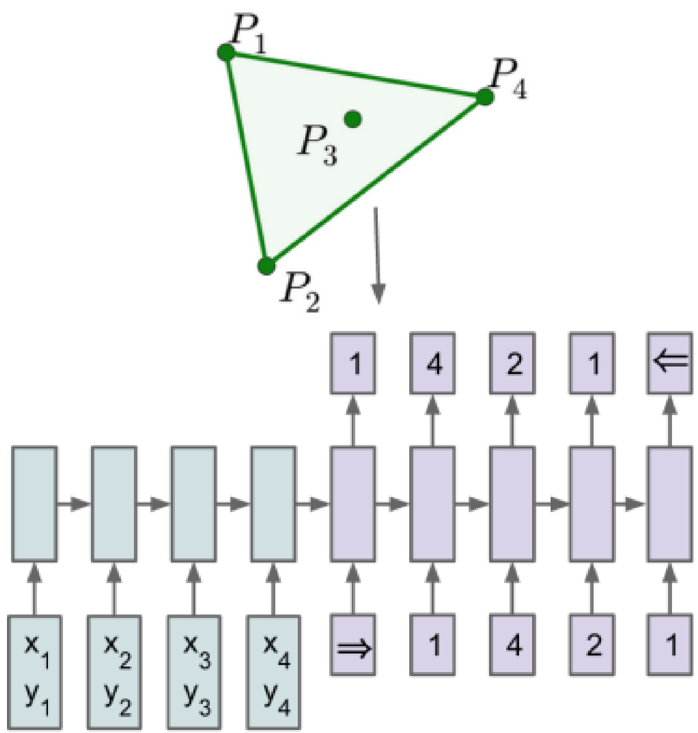
about 2 million votes, according to nigeria's independent national electoral commission. **the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.**

生成された要約： *muhammadu buhari* says he plans to aggressively fight corruption that has long plagued nigeria. he says his administration is confident it will be able to thwart criminals. the win comes after a long history of military rule, coups and botched attempts at democracy in africa's most populous nation.

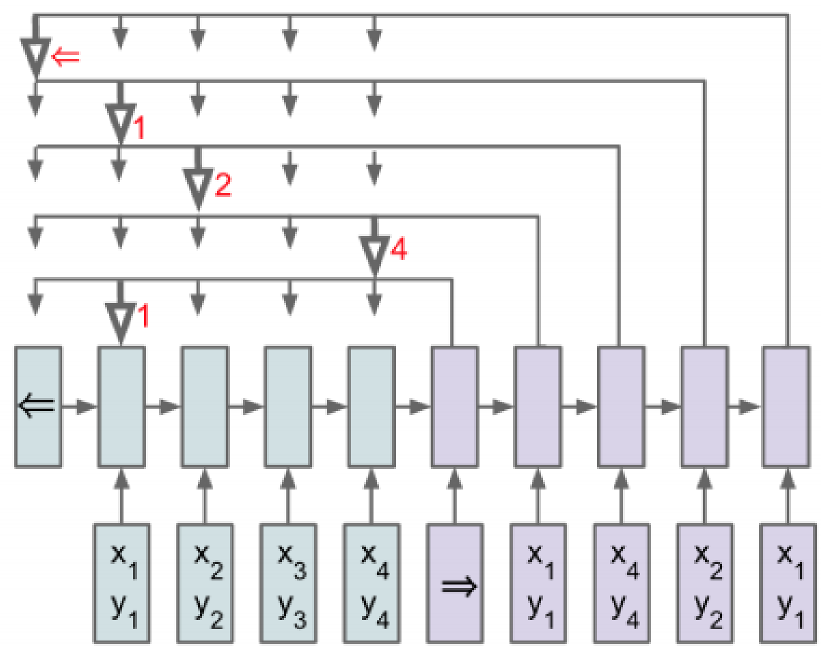
(See et al., 2017)

Pointer networks (Vinyals et al., 2015)

- 入力系列中の要素を指すポインタの系列を出力



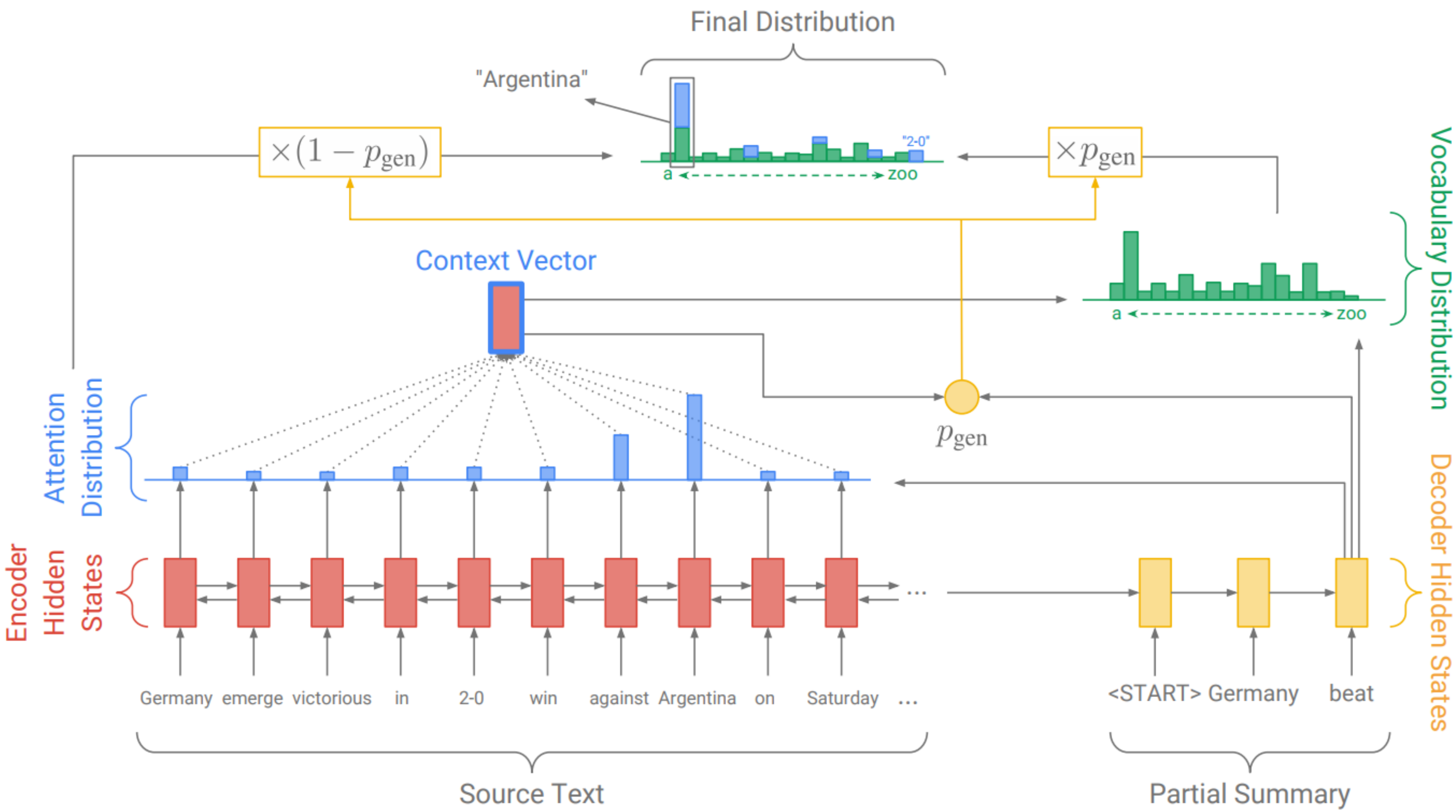
(a) Sequence-to-Sequence



(b) Ptr-Net

Pointer-generator model (See et al., 2017)

- ポインタを併用して要約文の単語を生成



プログラム自動生成

```
class DivineFavor(SpellCard):
    def __init__(self):
        super().__init__("Divine Favor", 3,
    CHARACTER_CLASS.PALADIN, CARD_RARITY.RARE)

    def use(self, player, game):
        super().use(player, game)
        difference = len(game.other_player.hand)
       - len(player.hand)
        for i in range(0, difference):
            player.draw()
```

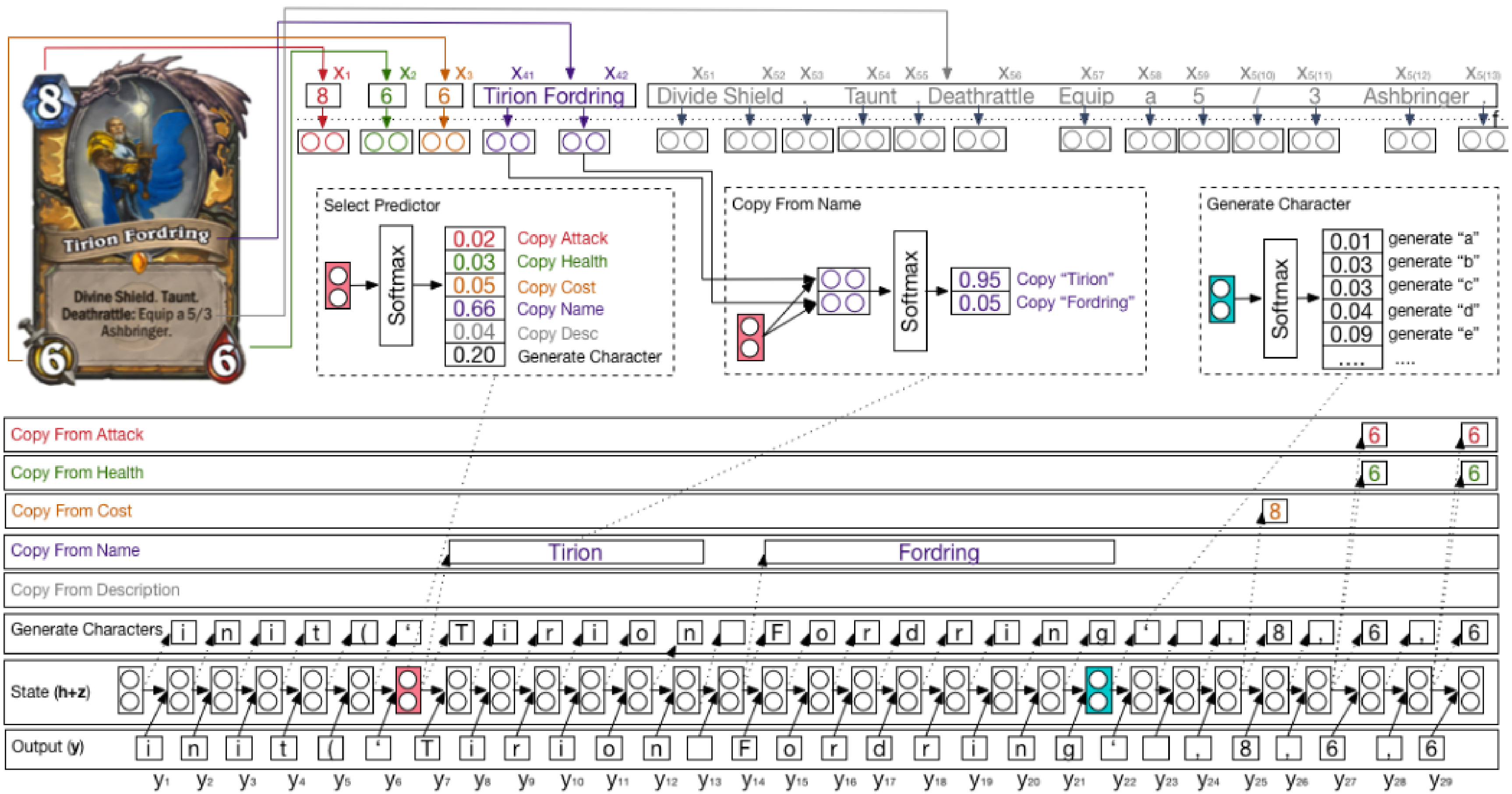
Magic the Gathering

Hearthstone

Ling et al., Latent Predictor Networks for Code Generation, ACL 2016

Latent Predictor Networks (Ling et al., 2016)

- 構造を持つ入力へのポインタも併用



生成例


```
class MadderBomber(MinionCard):       BLEU = 100.0
  def __init__(self):
    super().__init__("Madder Bomber", 5,
    CHARACTER_CLASS.ALL, CARD_RARITY.RARE,
    battlecry=Battlecry(Damage(1),
    CharacterSelector(players=BothPlayer(),
       picker= RandomPicker(6))))

def create_minion(self, player):§
  return Minion(5, 4)§
```


```
class Preparation(SpellCard):              BLEU = 64.2
  def __init__(self):
    super().__init__("Preparation", 0,
    CHARACTER_CLASS.ROGUE, CARD_RARITY.EPIC,
    target_func=hearthbreaker.targeting.find_minion_spell_target)

  def use(self, player, game):
    super().use(player, game)
    self.target.change_attack(3)

                player.add_aura(AuraUntil(ManaChange(-3),
                CardSelector(condition=IsSpell()), SpellCast()))
```

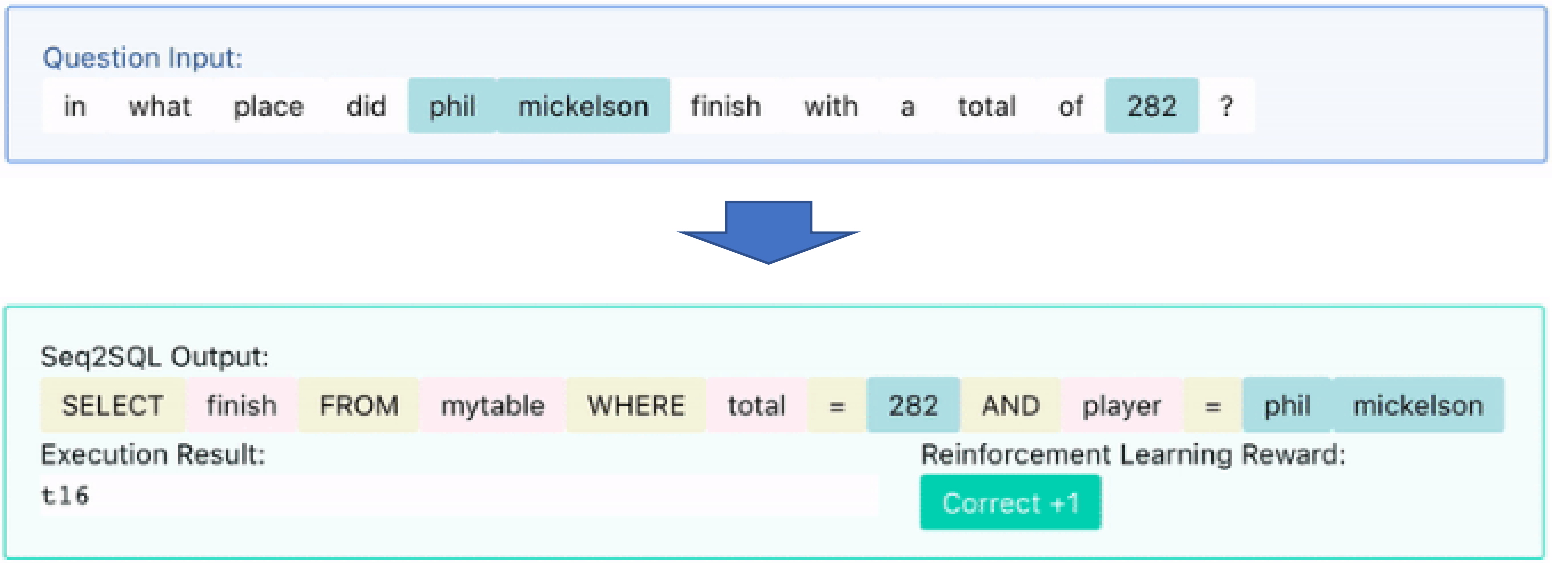
Ling et al., Latent Predictor Networks for Code Generation, ACL 2016

Text to Python Code (Yin and Neubig, 2017)

input	*join app_config.path and string 'locale' into a file path, substitute it for localedir.*
pred.	`localedir = os.path.join(app_config.path, 'locale')` ✓
input	*self.plural is an lambda function with an argument n, which returns result of boolean expression n not equal to integer 1*
pred.	`self.plural = lambda n: len(n)` ✗
ref.	`self.plural = lambda n: int(n!=1)`

Yin and Neubig, A Syntactic Neural Model for General-Purpose Code Generation, ACL 2017

Seq2SQL (Zhong et al., 2017)



実行

データベース

Database Table: mytable

player	country	year (s) won	total	to par	finish
Tiger Woods	United States	1999 , 2000	270	−18	1
Shaun Micheel	United States	2003	275	−13	2
Phil Mickelson	United States	2005	282	−6	T16
David Toms	United States	2001	282	−6	T16
Davis Love III	United States	1997	286	−2	T34
Rich Beem	United States	2002	291	+3	T49
Bob Tway	United States	1986	296	+8	T65

SQL Vocabulary:
SELECT WHERE FROM AND MAX MIN COUNT SUM AVG = > <

- 教師付き学習

  ＋
- 強化学習
  - 方策勾配法により、クエリの実行結果を報酬として学習

https://einstein.ai/research/how-to-talk-to-your-database

まとめ

- ニューラルネットワーク
  - リカレントニューラルネットワーク
  - 畳み込みニューラルネットワーク
- エンコーダー・デコーダーモデル
  - 系列から系列への変換
- 複雑なタスクを簡単なアーキテクチャで実現
  - レゴブロック (?) の組み合わせ、End-to-end 学習
- 大幅な性能向上
  - 構文解析、機械翻訳、質問応答、etc