2016/01/18

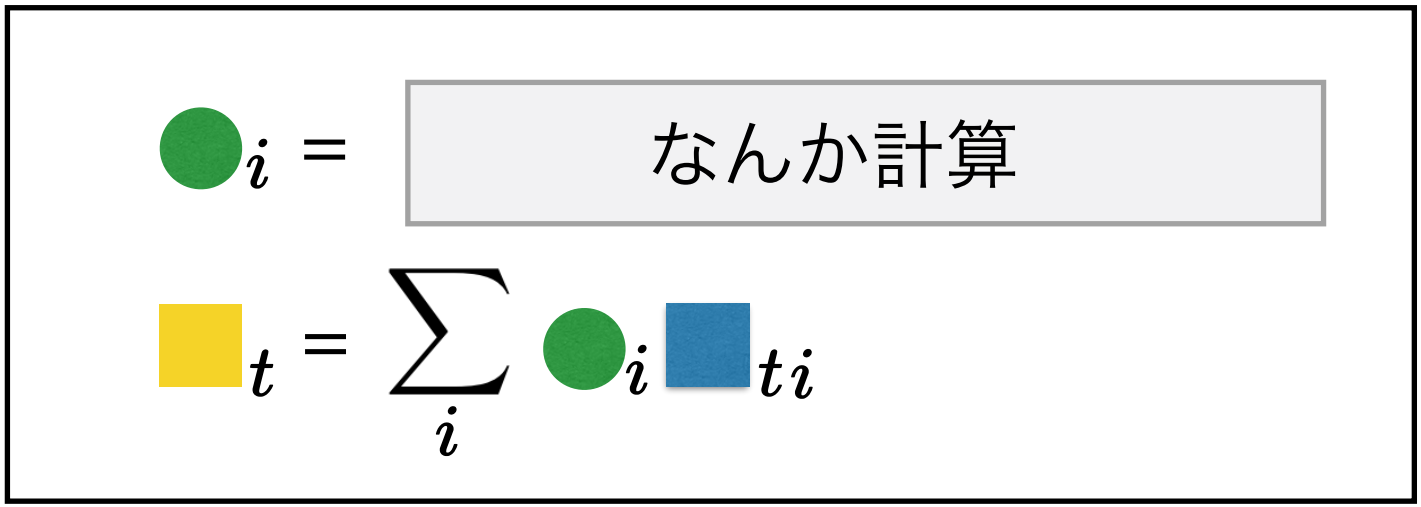
最近のDeep Learning界隈における

Attention 事情

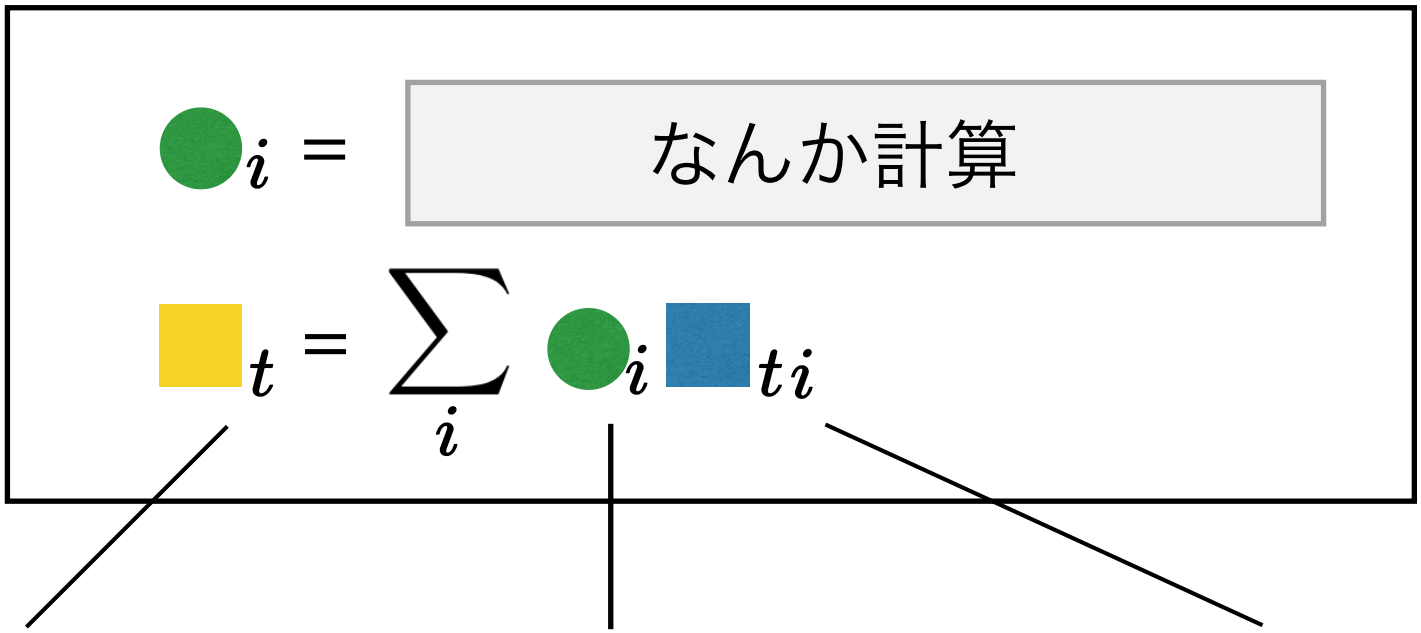
neural network with attention: survey

Yuta Kikuchi
@kiyukuta

今日はこんな形，よく見ることになります



今日はこんな形，よく見ることになります

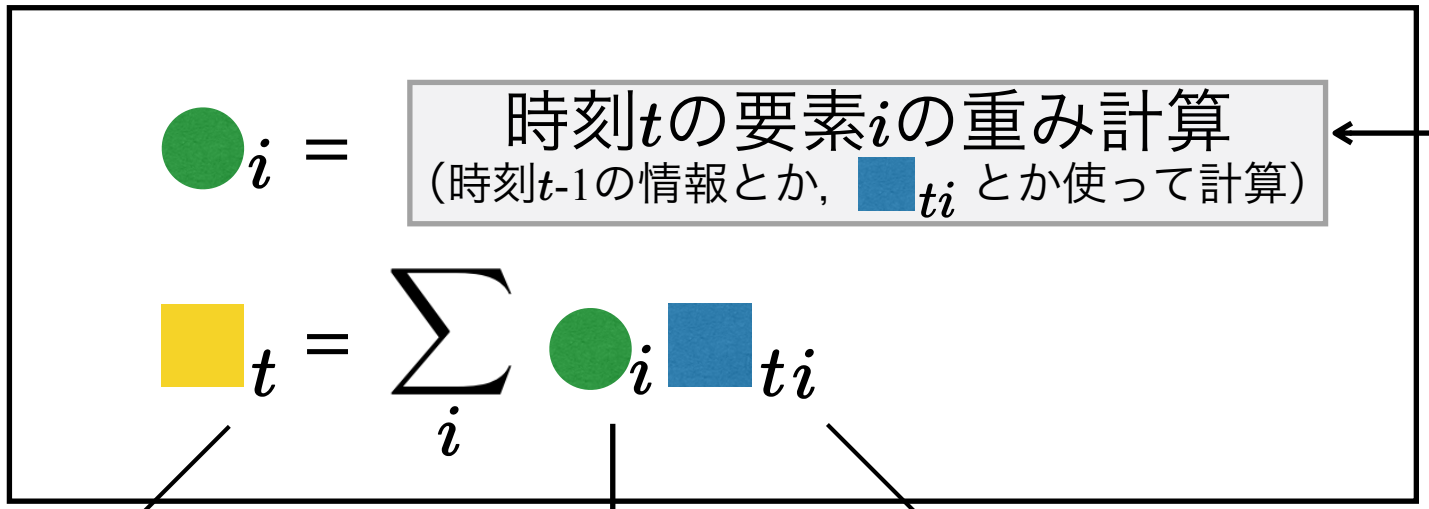


最終的に時刻tで
使われる情報

要素iの重み
(よくsoftmaxされている)

要素i
（専ら隠れ層ベクトル）

今日はこんな形，よく見ることになります



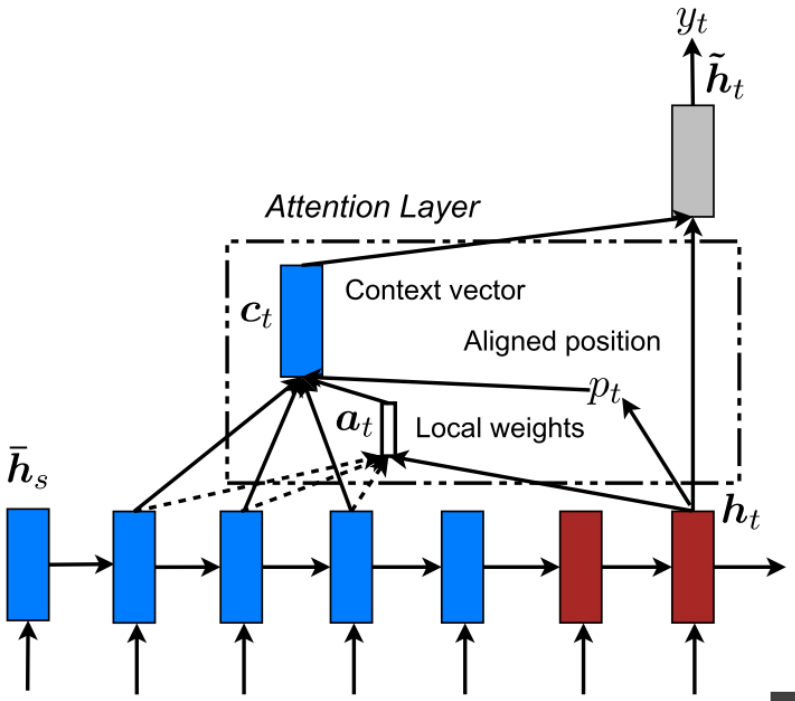
最終的に時刻tで
使われる情報

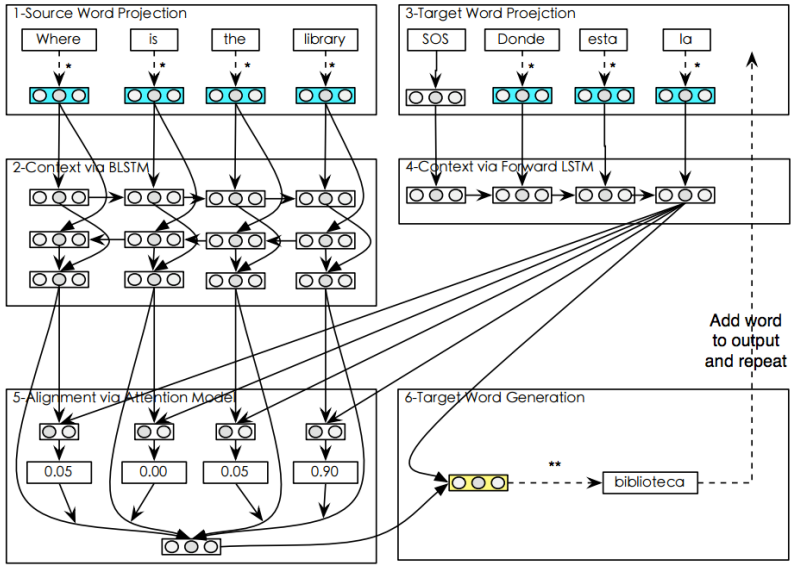
要素iの重み

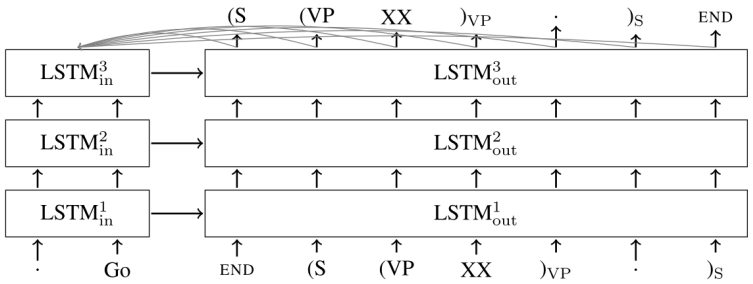
要素i
（専ら隠れ層ベクトル）

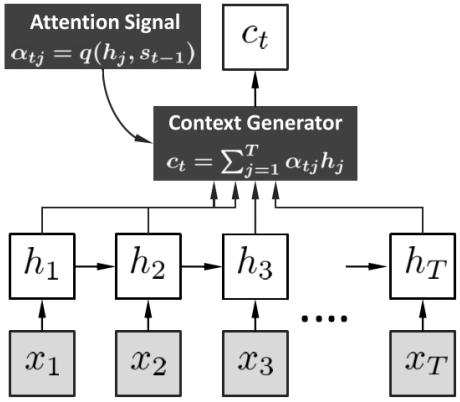
使う情報は今回の隠れ層だったり前回の隠れ層だったり
計算も隠れ層同士の内積だったりMLPを用意したり

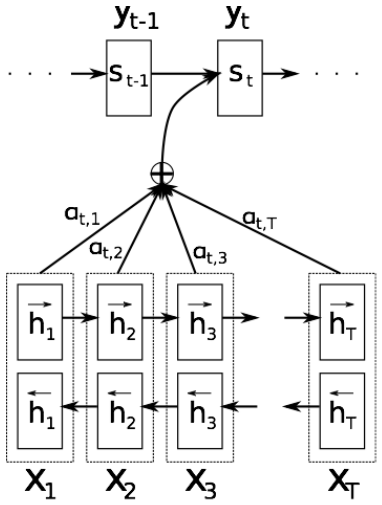
Attentionという同じ技術の似たような適用方法を複数の著者が「各々の思い思いの図」で書いてます





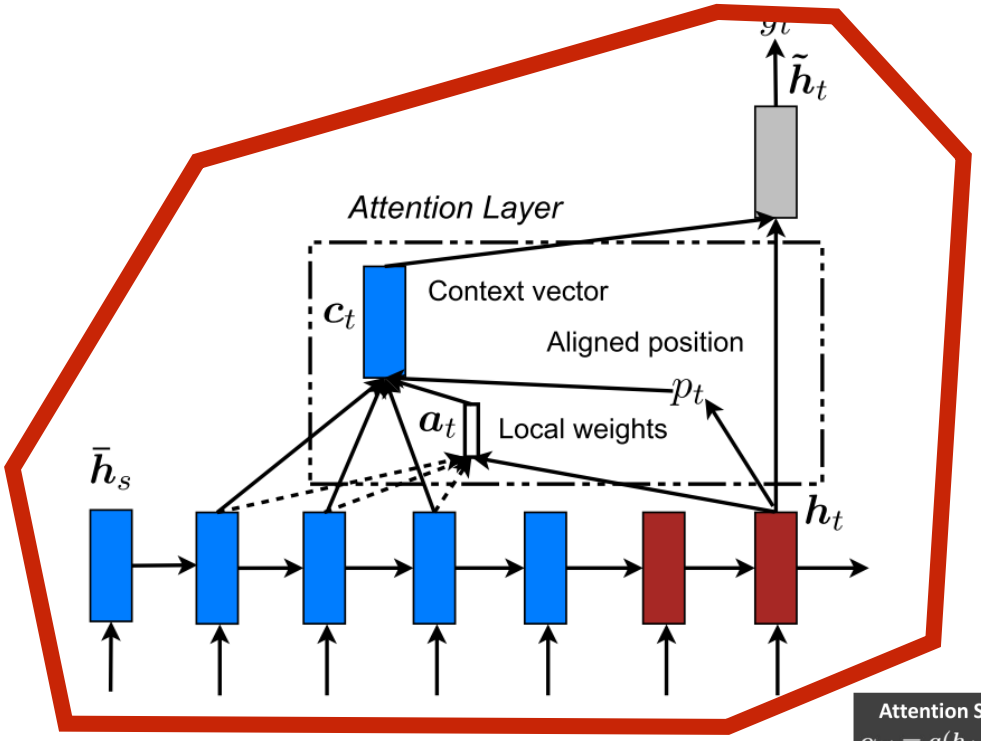


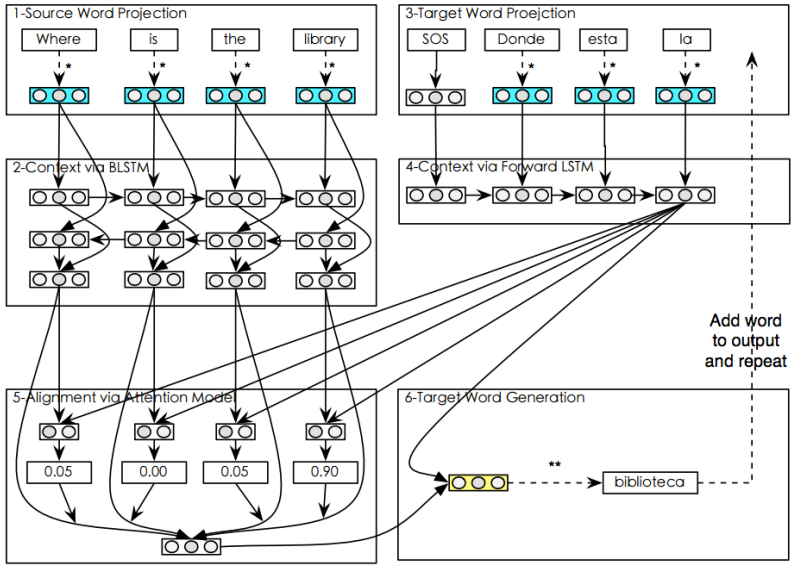


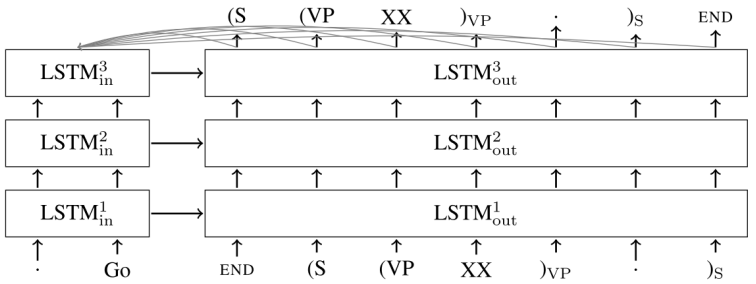


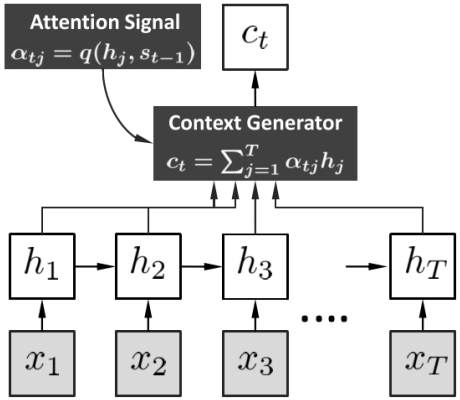
この中で（attention的な意味で）仲間はずれは？？？？

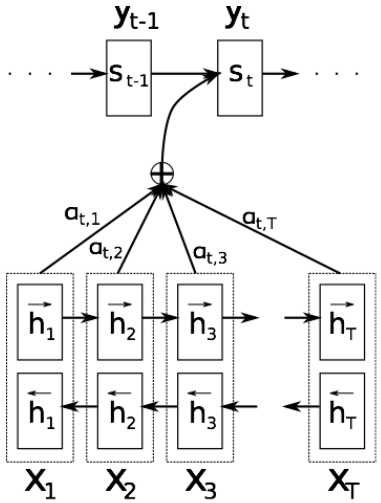
Attentionという同じ技術の似たような適用方法を複数の著者が「各々の思い思いの図」で書いてます











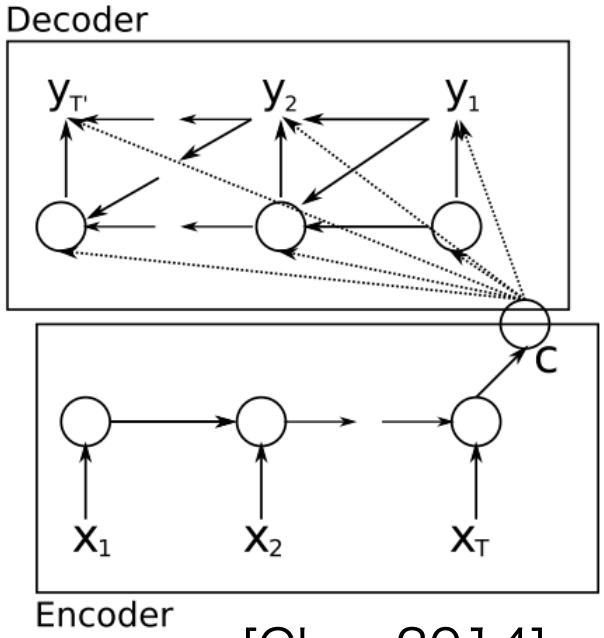
この中で（attention的な意味で）仲間はずれは？？？？

└(細かいところはちがったり)

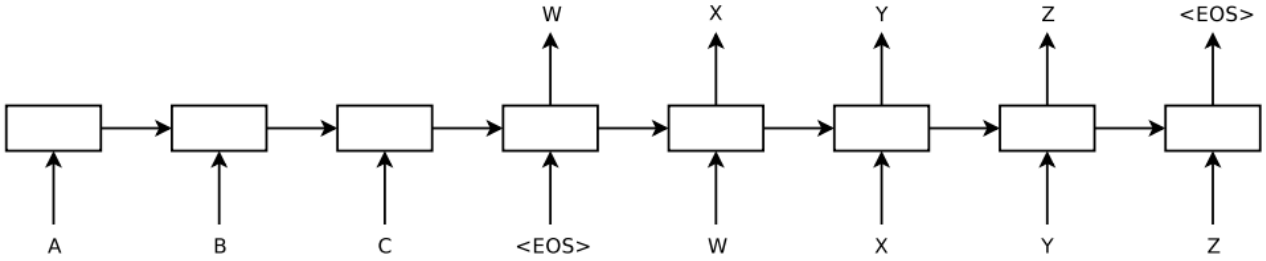
Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- **Background3: Encoder-Decoder approach (aka. sequence to sequence approach)**
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

Encoder-Decoder approach



[Cho+2014]

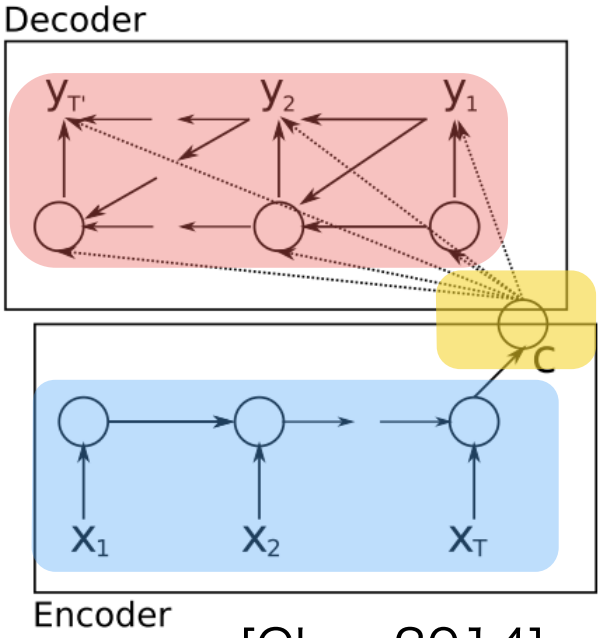


[Sutskever+2014]

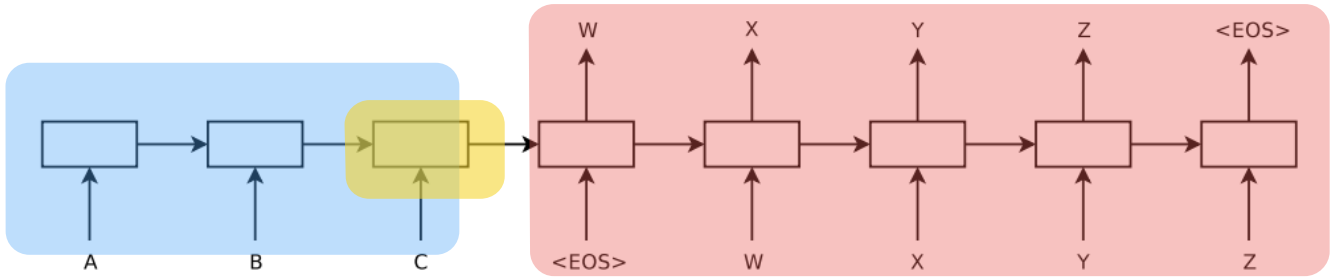
入力側(Encoder)と出力側(Decoder)それぞれに**RNN**を用意

Encoderによって入力系列を中間ノードに変換し，その情報を元に

Decoderが系列を出力



[Cho+2014]

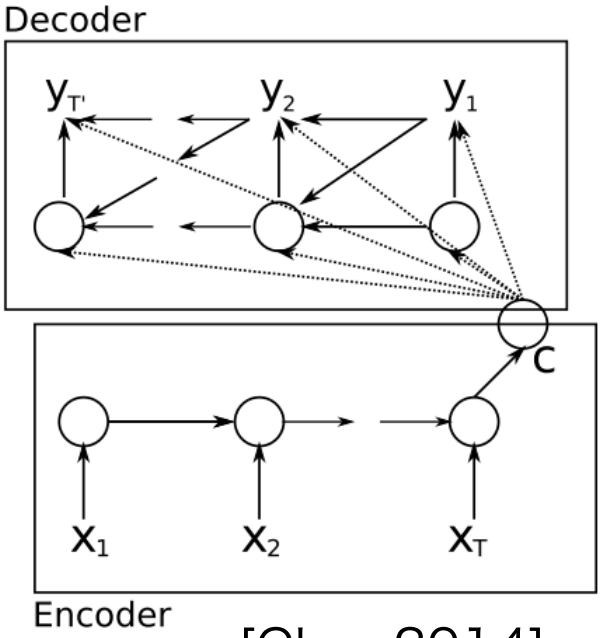


[Sutskever+2014]

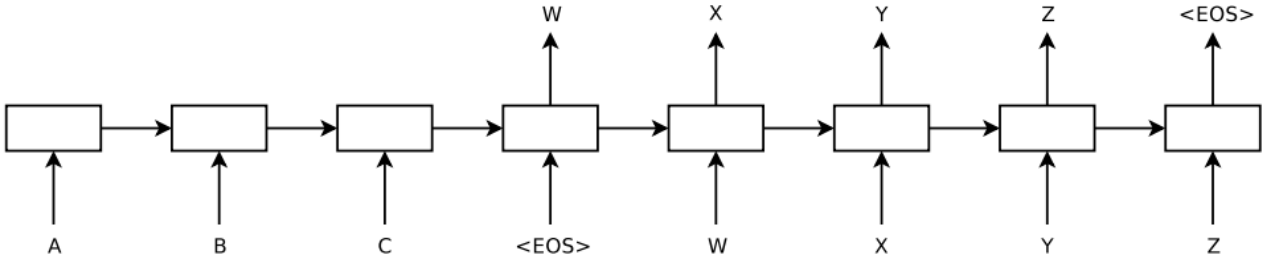
入力側(Encoder)と出力側(Decoder)それぞれに**RNN**を用意

Encoderによって入力系列を中間ノードに変換し，その情報を元に

Decoderが系列を出力



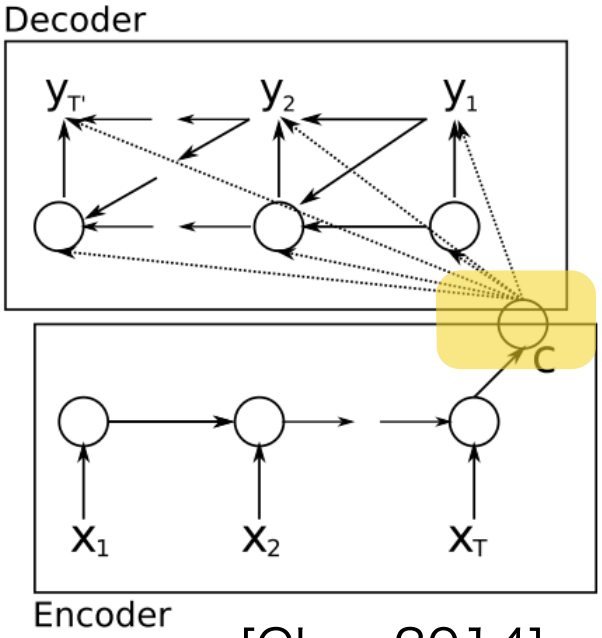
[Cho+2014]



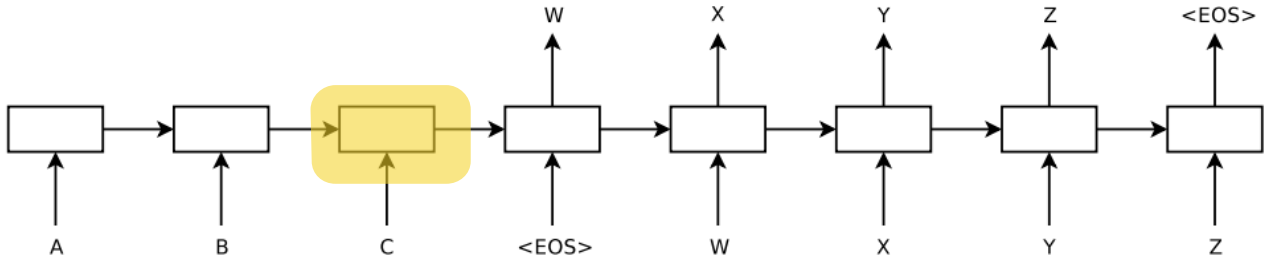
[Sutskever+2014]

機械翻訳やキャプション生成(encoder=CNN)で話題になり
様々なタスクでの適用例が報告されている[YYYYYYY+201X]

そしてAttentionの出現

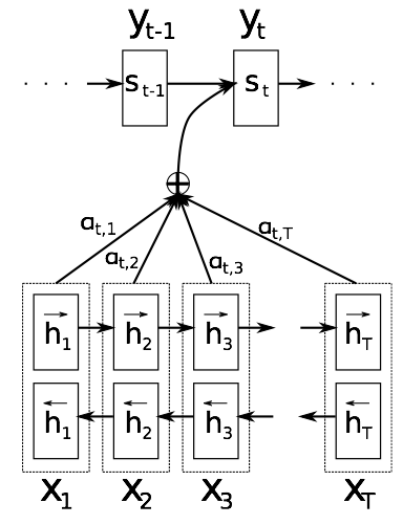


[Cho+2014]



[Sutskever+2014]

Although most of the previousworks (see, e.g., Cho et al., 2014a; Sutskever et al., 2014; Kalchbrenner and Blunsom, 2013) used to encode a variable-length input sentence into a fixed-length vector, **it is not necessary,** and even it may be beneficial to have a variable-length vector, as we will show later



[Bahdanau+2014]

Agenda

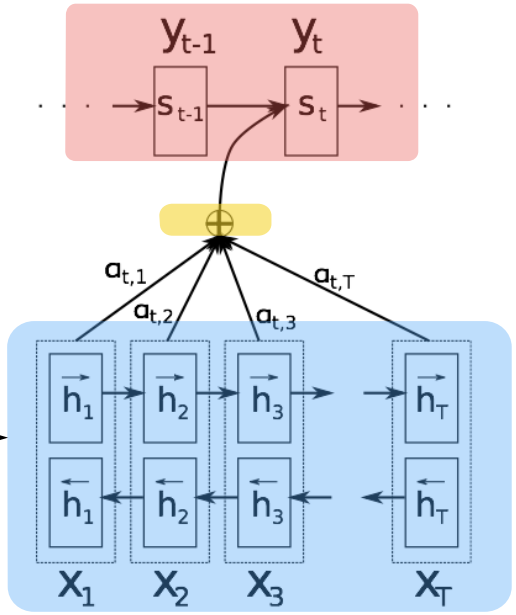
- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (aka. sequence to sequence approach)
- **Attention mechanism and its variants**
  - **Global attention**
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

Attention mechanism

(global attention

for convenience)

Although most of the previousworks (see, e.g., Cho et al., 2014a; Sutskever et al., 2014; Kalchbrenner and Blunsom, 2013) used to **encode a variable-length input sentence into a fixed-length vector, it is not necessary,** and even it may be beneficial to have a variable-length vector, as we will show later



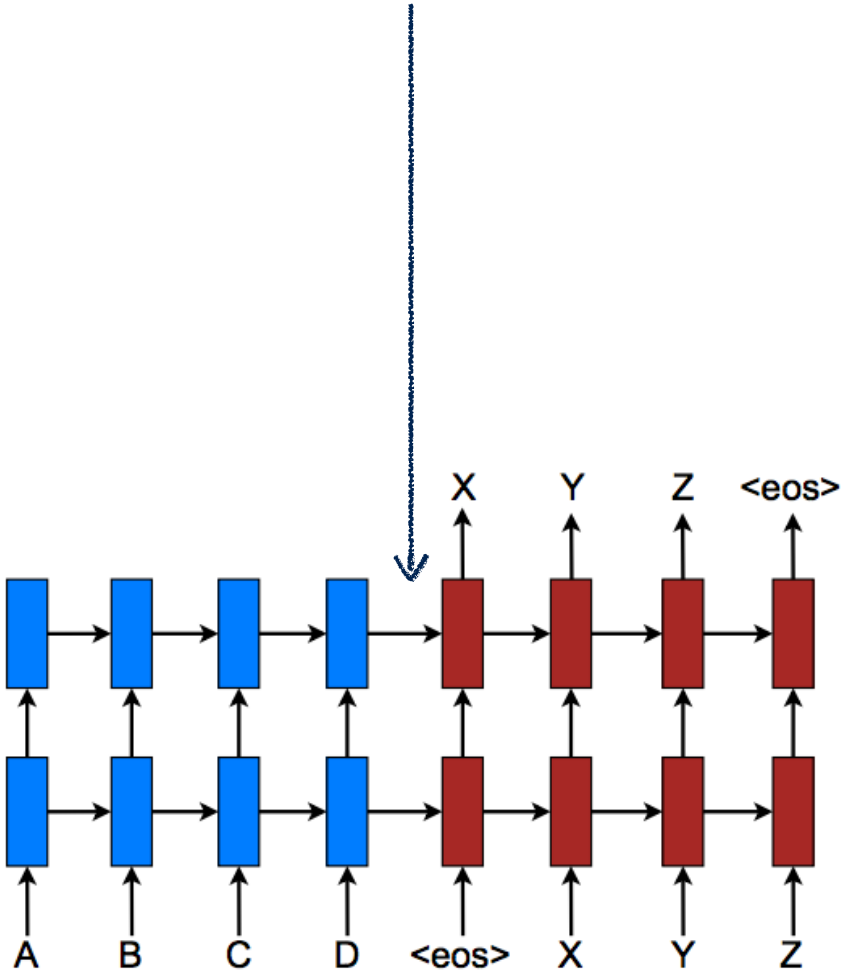
[Bahdanau+2014]

入力系列をシンプルなRNN Encoderによって一つの固定長ベクトルに詰め込むのではなく，Encode中の各隠れ層をとっておいて，decodeの各ステップでそれらを使う（どう使うかはもうすぐ説明）

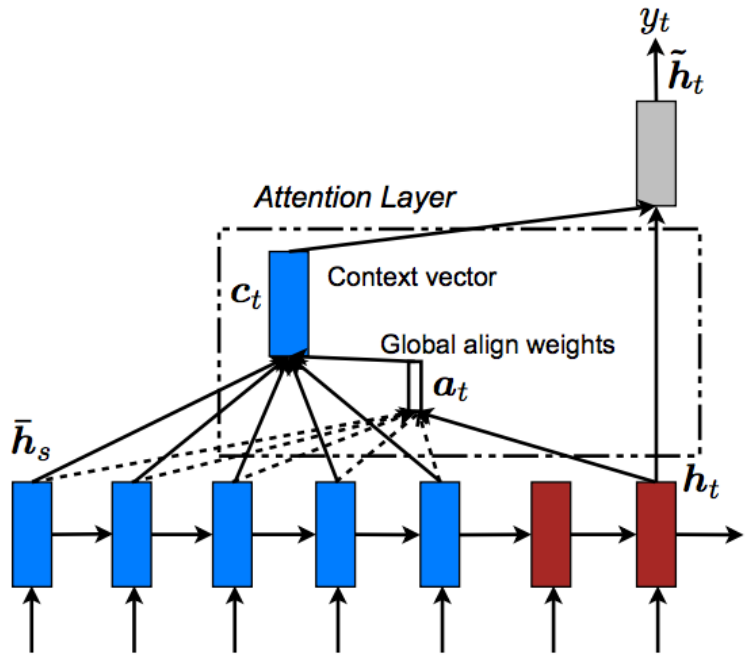
最終的に一つのベクトルとしてdecoderに渡すのは同じだが，渡すベクトルが文脈（e.g. これまでの自身の出力）に応じて動的に変わる

Figures from [Luong+2015] for comparison

Enc-decがEncoderが入力系列を一つのベクトルに圧縮
Decoderが出力時の初期値として使う



simple enc-dec

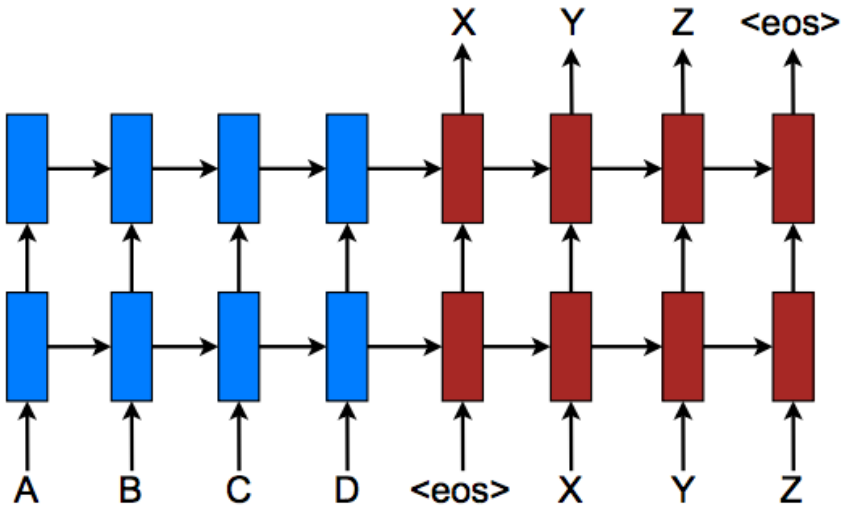


enc-dec + attention

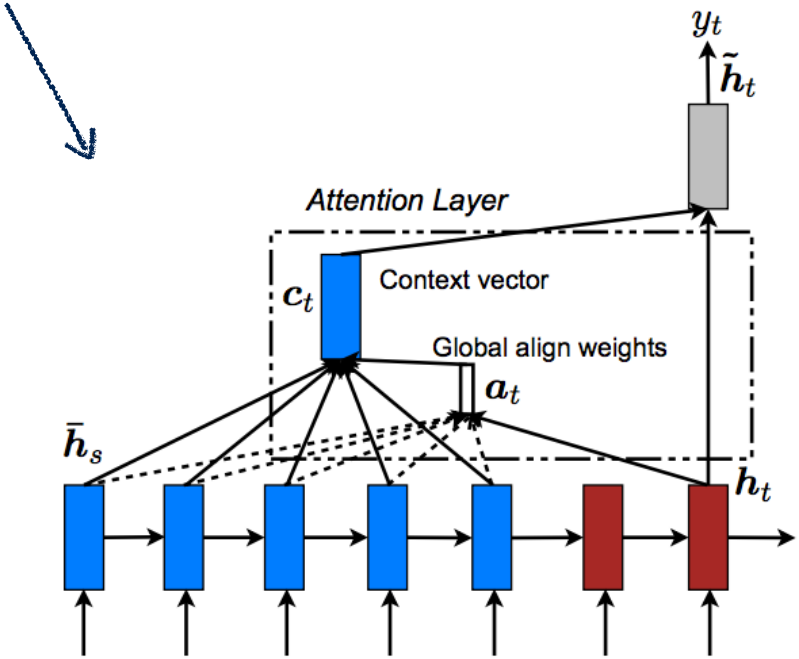
Figures from [Luong+2015] for comparison

Enc-decがEncoderが入力系列を一つのベクトルに圧縮
Decoderが出力時の初期値として使う

Attentionでは出力ステップtで，その時の隠れ層h_tを使い
入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_t
を使って出力y_tを予測



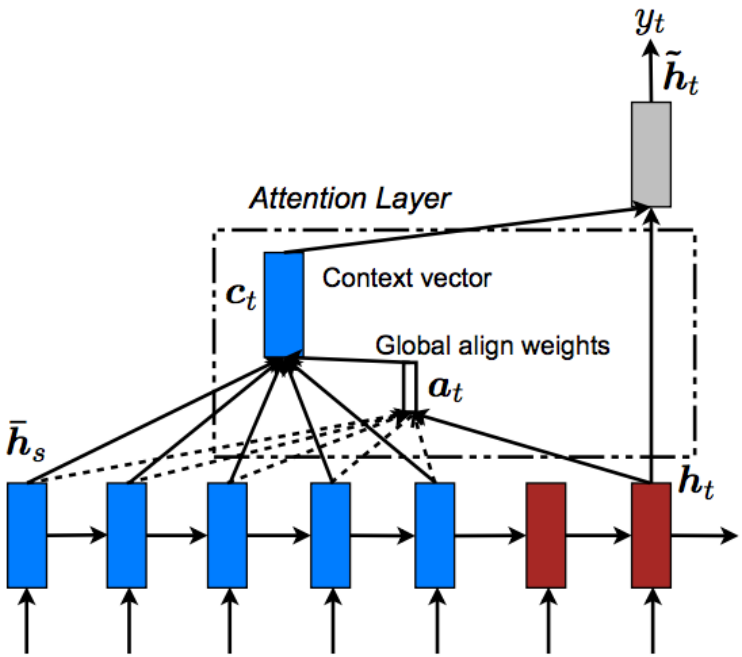
enc-dec



attention

Figures from [Luong+2015] for comparison

Attentionでは出力ステップtで，その時の隠れ層h_tを使い入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_tを使って出力y_tを予測



$$a_t(s) = \frac{\exp\left(\text{score}(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(\text{score}(h_t, \bar{h}_{s'})\right)}$$

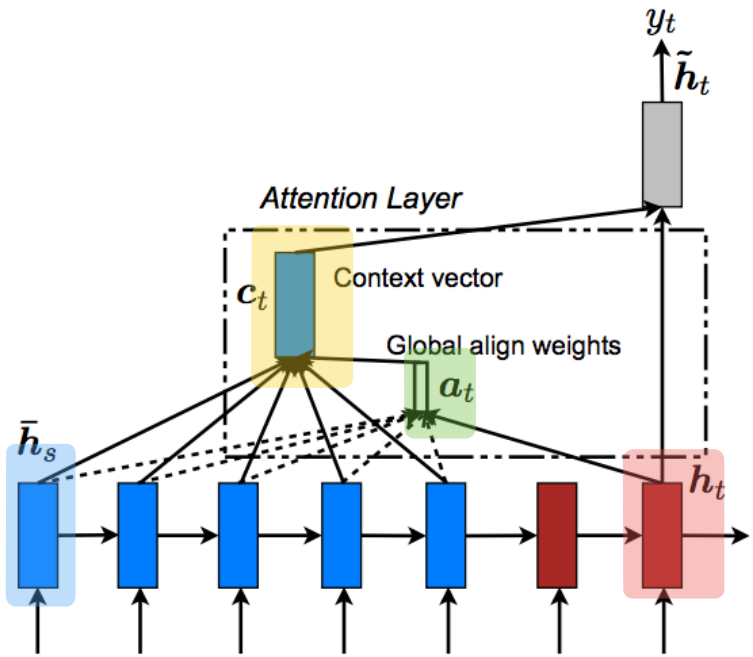
$$c_t = \sum_s a_t(s)\bar{h}_s$$

$$\tilde{h}_t = \tanh(W_c[c_t; h_t])$$

$$p(y_t|y_{<t}, x) = \text{softmax}(W_s\tilde{h}_t)$$

Figures from [Luong+2015] for comparison

Attentionでは出力ステップtで，その時の隠れ層h_tを使い入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_tを使って出力y_tを予測



$$a_t(s) = \frac{\exp\left(\text{score}(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(\text{score}(h_t, \bar{h}_{s'})\right)}$$

$$c_t = \sum_s a_t(s)\bar{h}_s$$

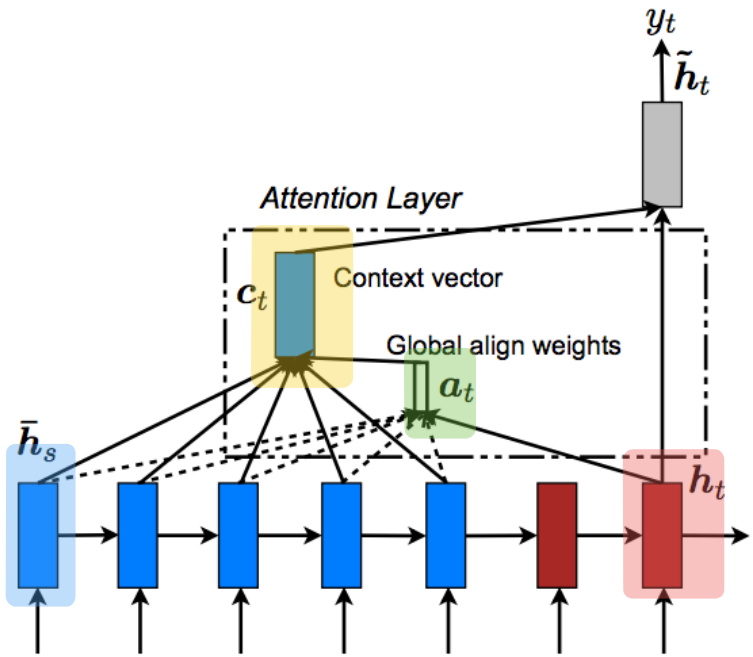
$$\tilde{h}_t = \tanh(W_c[c_t; h_t])$$

$$p(y_t|y_{<t}, x) = \text{softmax}(W_s\tilde{h}_t)$$

ここは単語出力用

Figures from [Luong+2015] for comparison

Attentionでは出力ステップtで，その時の隠れ層h_tを使い入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_tを使って出力y_tを予測



$$a_t(s) = \frac{\exp\left(\text{score}(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(\text{score}(h_t, \bar{h}_{s'})\right)}$$

$$c_t = \sum_s a_t(s)\bar{h}_s$$

$\tilde{h}_t = \tanh(W_c[c_t; h_t])$

$p(y_t$

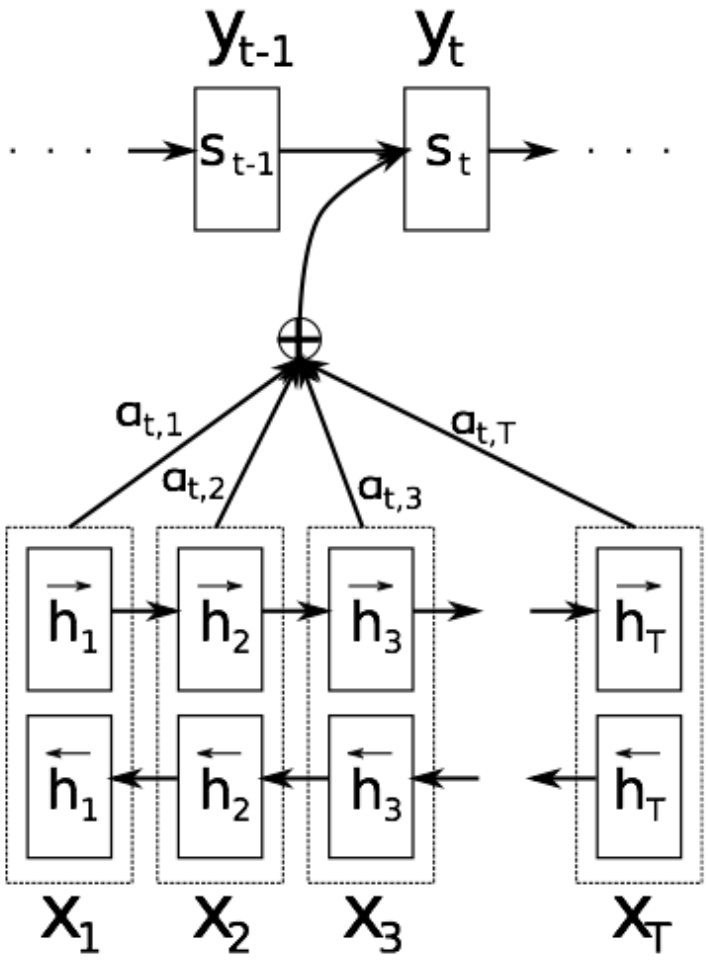
● $_i$ = 時刻tの要素iの重み計算
（時刻t-1の情報とか，■ $_{ti}$ とか使って計算）

■ $_t = \sum_i$ ● $_i$ ■ $_{ti}$

冒頭で見たやつだ！！

Figures from [Luong+2015] for comparison

Attentionでは出力ステップtで，その時の隠れ層h_tを使い入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_tを使って出力y_tを予測



$$a_t(s) = \frac{\exp\left(\text{score}(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(\text{score}(h_t, \bar{h}_{s'})\right)}$$

$$c_t = \sum_s a_t(s)\bar{h}_s$$

$p(y_t$

$\tilde{h}_t = \tanh(W_c[c_t; h_t])$

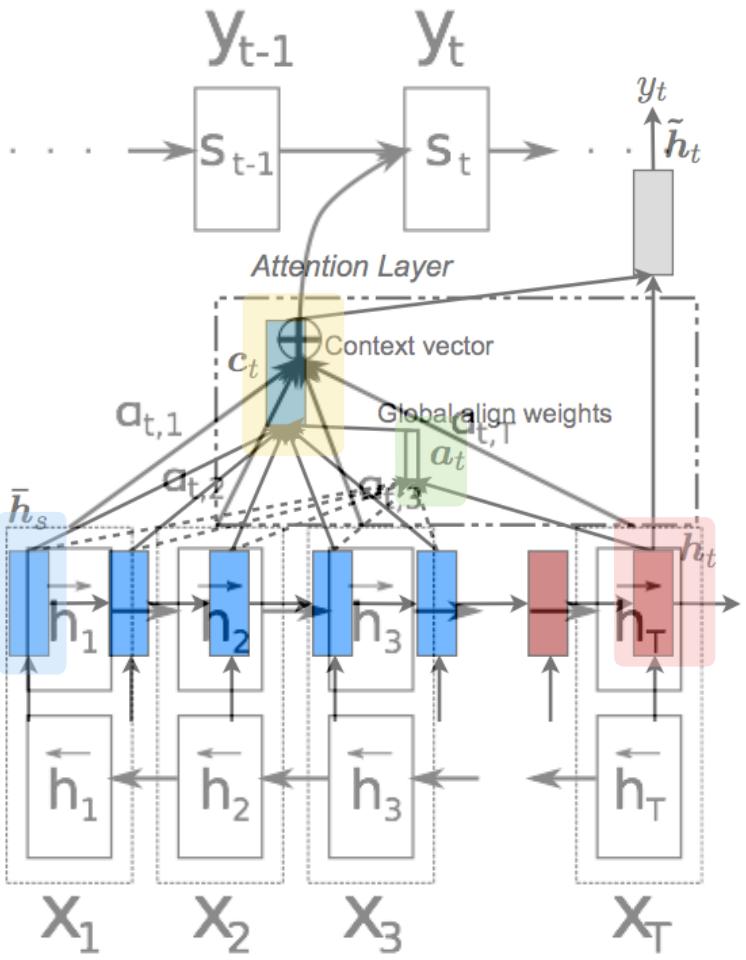
$\blacksquare_i$ = 時刻tの要素iの重み計算
（時刻t-1の情報とか，$\blacksquare_{ti}$ とか使って計算）

$$\blacksquare_t = \sum_i \blacksquare_i \blacksquare_{ti}$$

冒頭で見たやつだ！！

Figures from [Luong+2015] for comparison

Attentionでは出力ステップtで，その時の隠れ層h_tを使い入力側の各隠れ層を荷重a_tで加重平均した文脈ベクトルc_tを使って出力y_tを予測



$$a_t(s) = \frac{\exp\left(\text{score}(h_t, \bar{h}_s)\right)}{\sum_{s'} \exp\left(\text{score}(h_t, \bar{h}_{s'})\right)}$$

$$c_t = \sum_s a_t(s)\bar{h}_s$$

$$\tilde{h}_t = \tanh(W_c[c_t; h_t])$$

$$p(y_t$$

i = 時刻tの要素iの重み計算
（時刻t-1の情報とか，ti とか使って計算）

$$t = \sum_i i\, ti$$

冒頭で見たやつだ！！

Describing Multimedia Content using
Attention-based Encoder–Decoder Networks [Cho+2015]
の言葉を借りつつ

まとめると

Attentionは入力と出力のアライメントを学習するもの

ソフトが多いけどハード(今日扱わない)もあるよ！

シンプルなEnc-decと比較してAttention導入の利点は

- 扱う入力側の情報を動的に変化させることができる
- 解釈が容易になる（アライメントとして可視化できる）

[Bahdanau+2014]以前のAttentionっぽいものは？

画像処理分野で単純に”attention”とかでたどると
本物のヒトの注視点データを用いた研究とか出てきた
（もっと以前にあるかは未調査）

Learning where to Attend with Deep Architectures for Image Tracking [Denil+2011]

テキストだと[Graves2013]に出てくる
“soft window” が関係深そう

ちょっとタスクが面白いので紹介

Generating Sequences With Recurrent Neural Networks [Graves2013]

§4 Handwritten Prediction （前置き）

手書き文字のストロークから**次のペンの動き**を予想

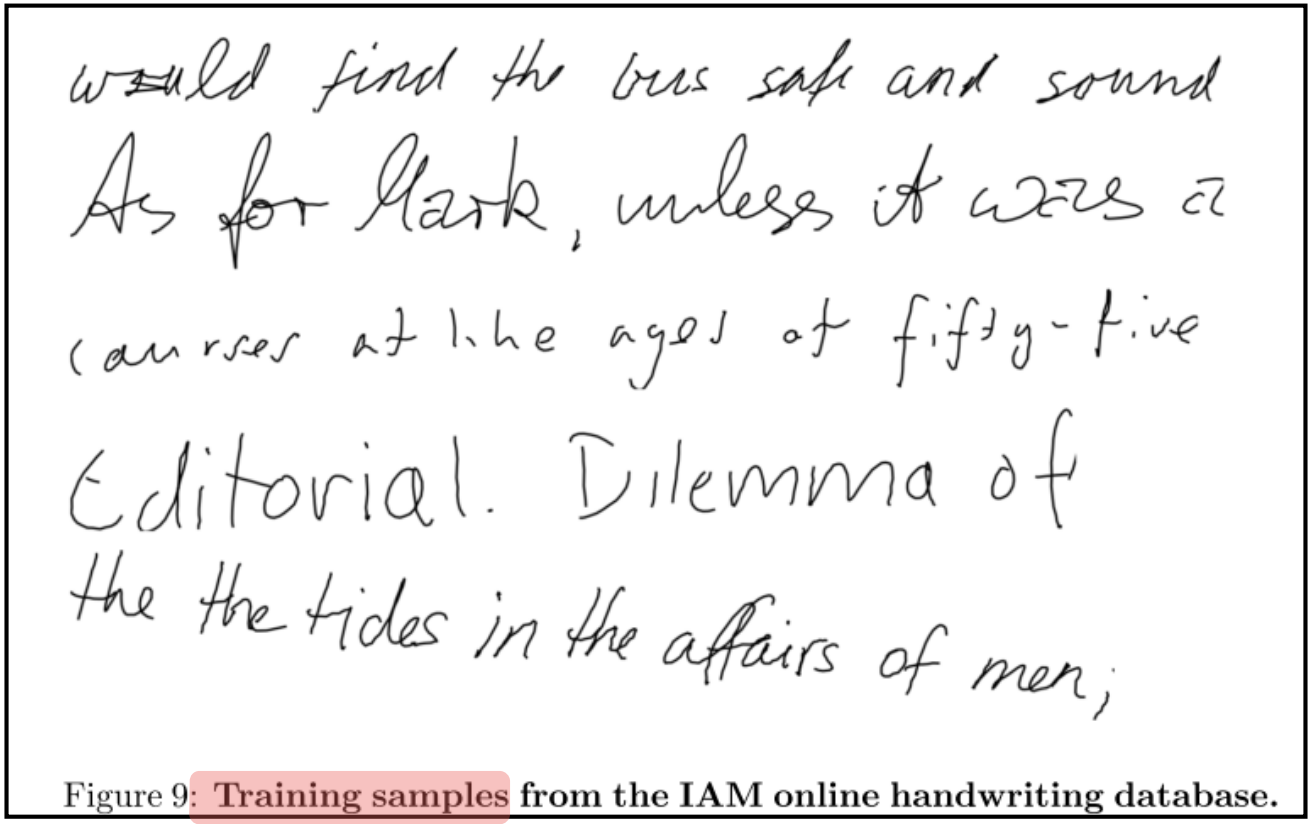


Figure 9: Training samples from the IAM online handwriting database.

えっこんなの生成できるのk....て人手かよ汚すぎだなおい！

x座標　　y座標　　ペンが接地中or not

NNへの入力: $x_t \in \mathbb{R} \times \mathbb{R} \times \{0, 1\}$

NNの出力: $y_t = \left(e_t, \{\pi_t^j, \mu_t^j, \sigma_t^j, \rho_t^j\}_{j=1}^M\right)$

ペンが離れるか否か　↓の混合ガウス分布の**パラメータ**

$$\Pr(x_{t+1}|y_t) = \sum_{j=1}^{M} \pi_t^j \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-Z}{2(1-\rho^2)}\right] \begin{cases} e_t & \text{if } (x_{t+1})_3 = 1 \\ 1 - e_t & \text{otherwise} \end{cases}$$

$$Z = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2}$$

Pr(次のペン位置 | NNが予測した混合ガウス分布)

これが大きくなるような**パラメータ**を出力するよう学習

本題

§5 Handwritten Synthesis

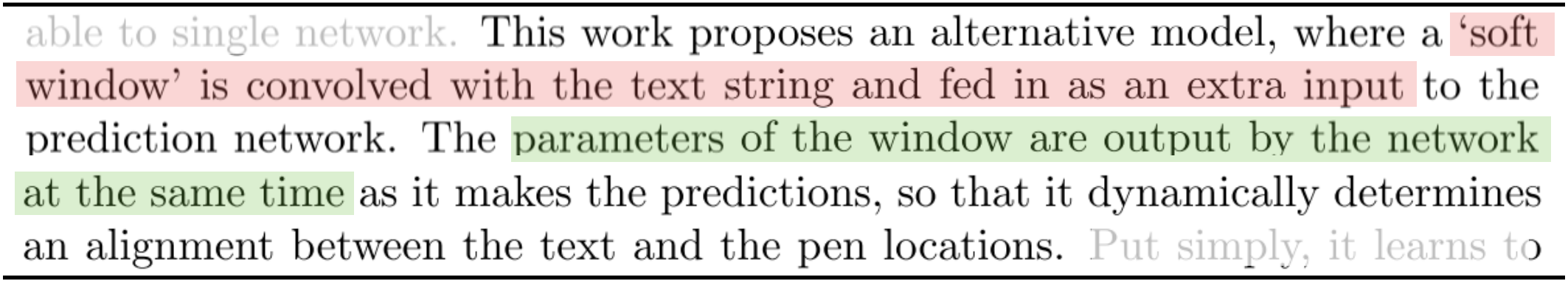
文字列からその手書き文字画像を生成

入出力は§4と同じだが文脈情報としてクエリ文字列を使う

特徴

- 入力と出力の”長さ”がかなり異なる
- 文字と手描きストローク間のアライメント情報はない

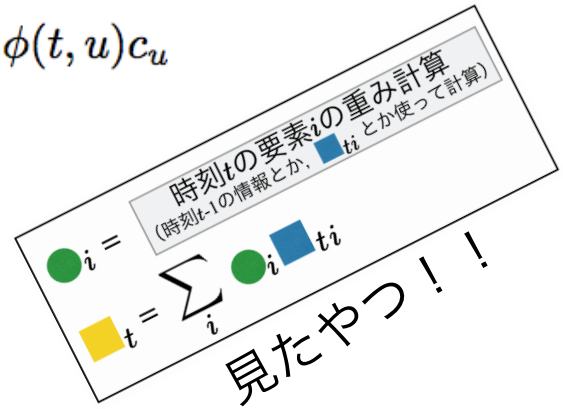
↓



[Graves2013]

$$\phi(t,u) = \sum_{k=1}^{K} \alpha_t^k \exp\left(-\beta_t^k \left(\kappa_t^k - u\right)^2\right)$$

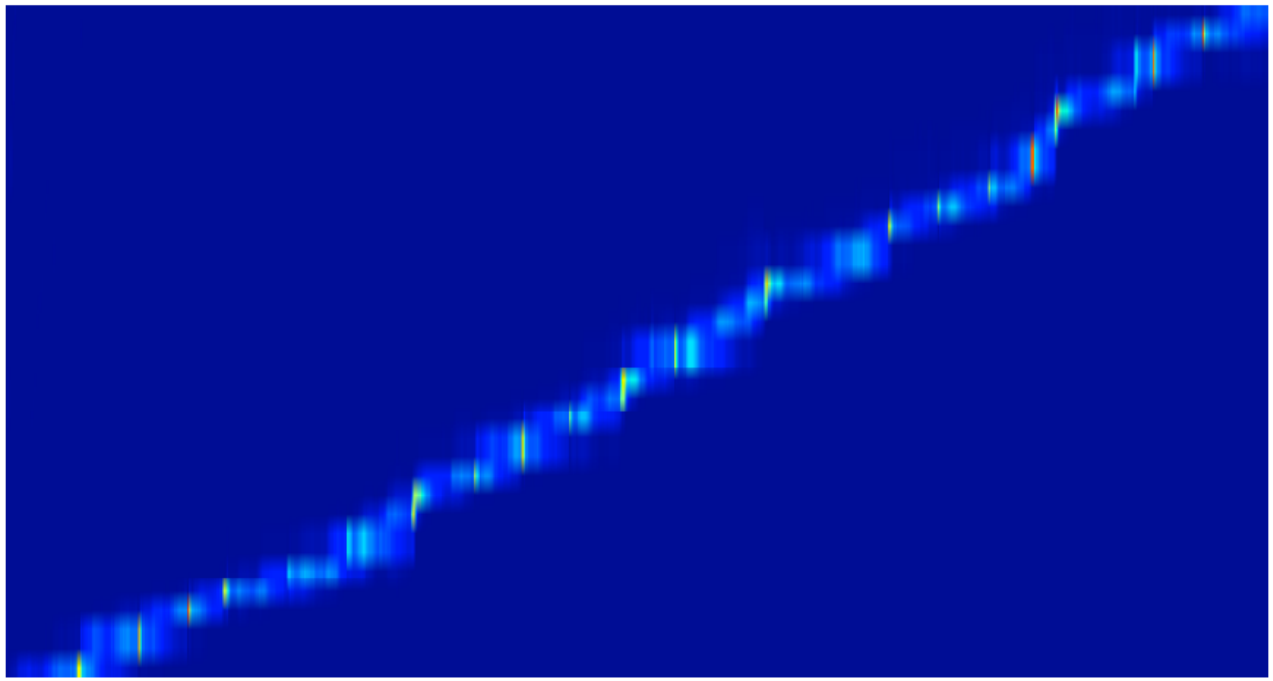
$$w_t = \sum_{u=1}^{U} \phi(t,u) c_u$$



ステップtにおける入力文字uの重み $\phi(t,u)$

$$(\hat{\alpha}_t, \hat{\beta}_t, \hat{\kappa}_t) = W_{h^1 p} h_t^1 + b_p$$
$$\alpha_t = \exp\left(\hat{\alpha}_t\right)$$
$$\beta_t = \exp\left(\hat{\beta}_t\right)$$
$$\kappa_t = \kappa_{t-1} + \exp\left(\hat{\kappa}_t\right)$$

ガウス関数のパラメータは
その時の隠れ層hで決定



Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Gobal attention

ここまで

  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - **Global attention**

ここまで

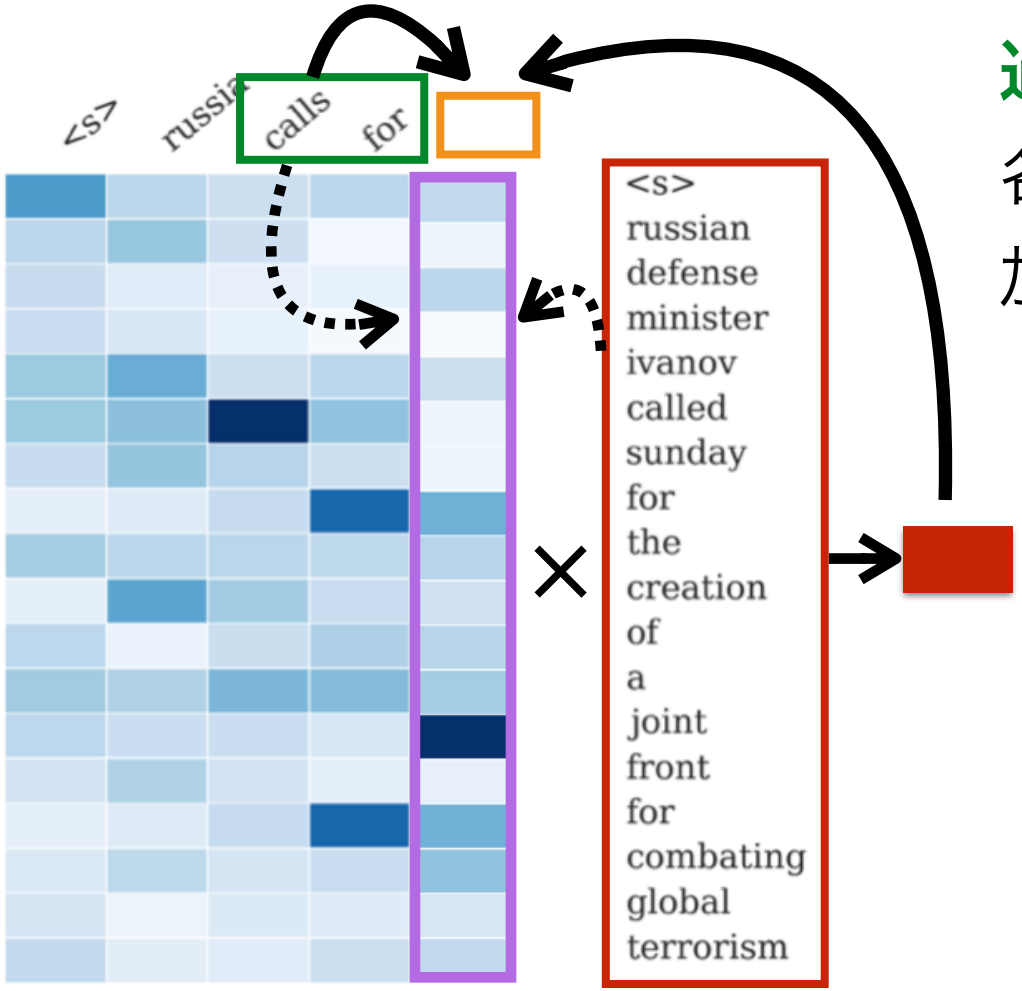
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

翻訳以外のタスクでは？

Global attentionを使ったその他の論文たち
(数スライドずつ)

A Neural Attention Model for Abstractive Sentence Summarization [Rush+2015]

- 文の要約 (not 文書の要約)
- [Bahdanau+2014]を簡素化したネットワーク構造



過去の自分の出力と**入力文**から
各入力単語の**重み**を計算し
加重平均ベクトルを**予測**に使う

bag-of-word embeddings の加重平均

詳しくはこちら！

http://www.slideshare.net/yutakikuchi927/a-neural-attention-model-for-sentence-summarization

Model	DUC-2004 ROUGE-1	DUC-2004 ROUGE-2	DUC-2004 ROUGE-L	Gigaword ROUGE-1	Gigaword ROUGE-2	Gigaword ROUGE-L	Ext. %
IR	11.06	1.67	9.67	16.91	5.55	15.58	29.2
PREFIX	22.43	6.49	19.65	23.14	8.25	21.73	100
COMPRESS	19.77	4.02	17.30	19.63	5.13	18.28	100
W&L	22	6	17	-	-	-	-
TOPIARY	25.12	6.46	20.12	-	-	-	-
MOSES+	26.50	8.13	22.85	28.77	12.10	26.44	70.5
ABS	26.55	7.06	22.05	30.88	12.22	27.77	85.4
ABS+	28.18	8.49	23.81	31.00	12.65	28.34	91.5
REFERENCE	29.21	8.38	24.46	-	-	-	45.6

良い言い換え(I:入力，A:システムの出力)

I(4): australian foreign minister stephen smith sunday congratulated new zealand 's new prime minister-elect john key as he praised ousted leader helen clark as a " gutsy " and respected politician .

A: australian foreign minister congratulates new nz pm after election

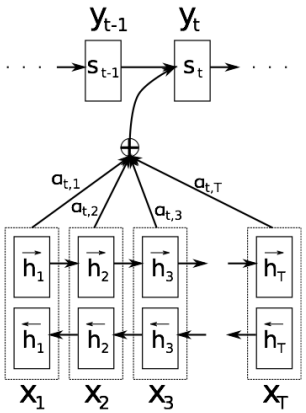
失敗な言い換え(I:入力，A:システムの出力)

I(11): russia 's gas and oil giant gazprom and us oil major chevron have set up a joint venture based in resource-rich northwestern siberia , the interfax news agency reported thursday quoting gazprom officials .

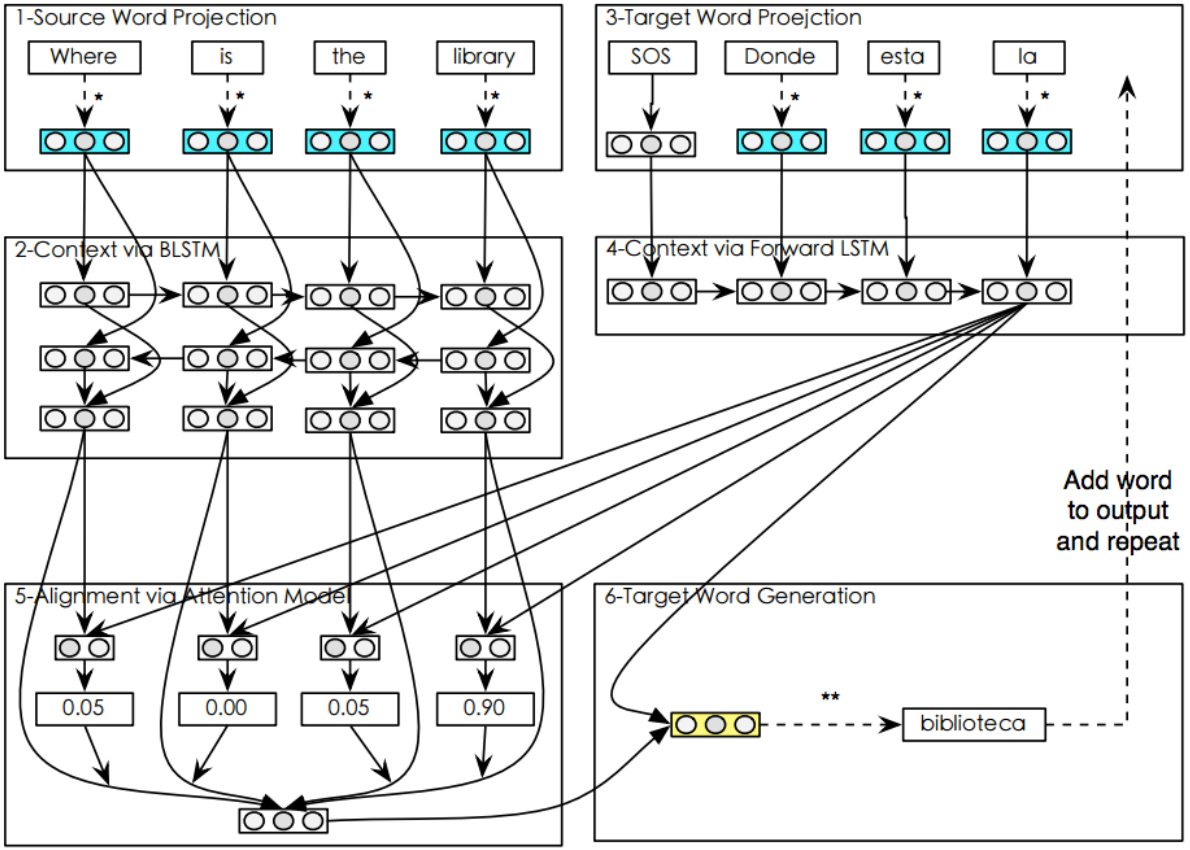
A: russian oil giant chevron set up siberia joint venture

Character-based Neural Machine Translation [Ling+2015]

[Bahdanau+2014]を文字レベルに拡張



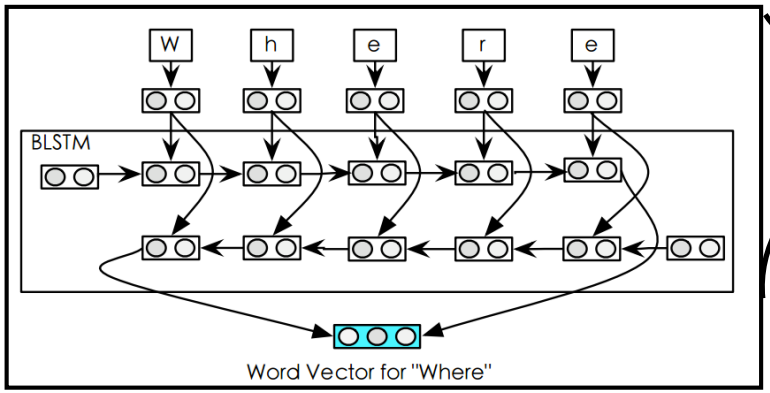
[Bahdanau+2014]



attention

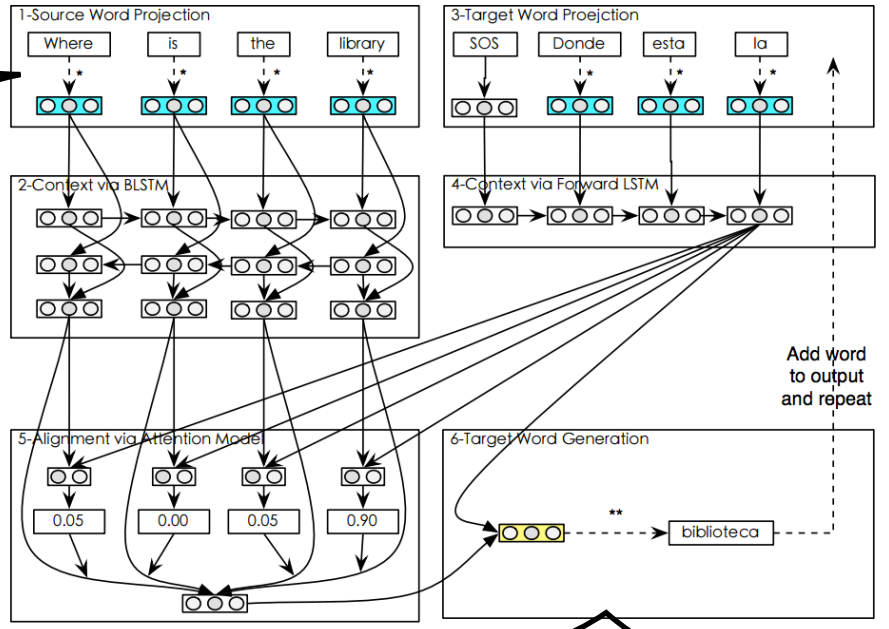
Character-based Neural Machine Translation [Ling+2015]

[Bahdanau+2014]を文字レベルに拡張



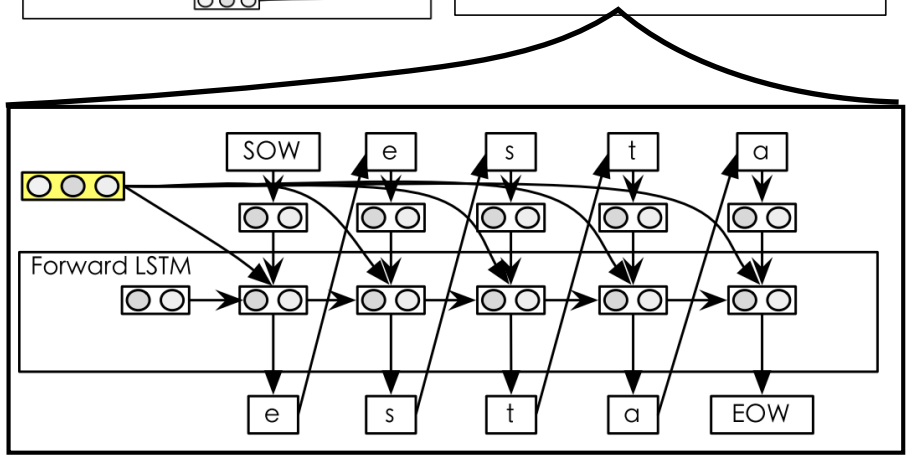
文字列をベクトルに

[Bahdanau+2014]



BLEUはwordとくらべて微増だが，過去のcharベース手法がwordを超えられなかったことを考えるとすごいとの主張

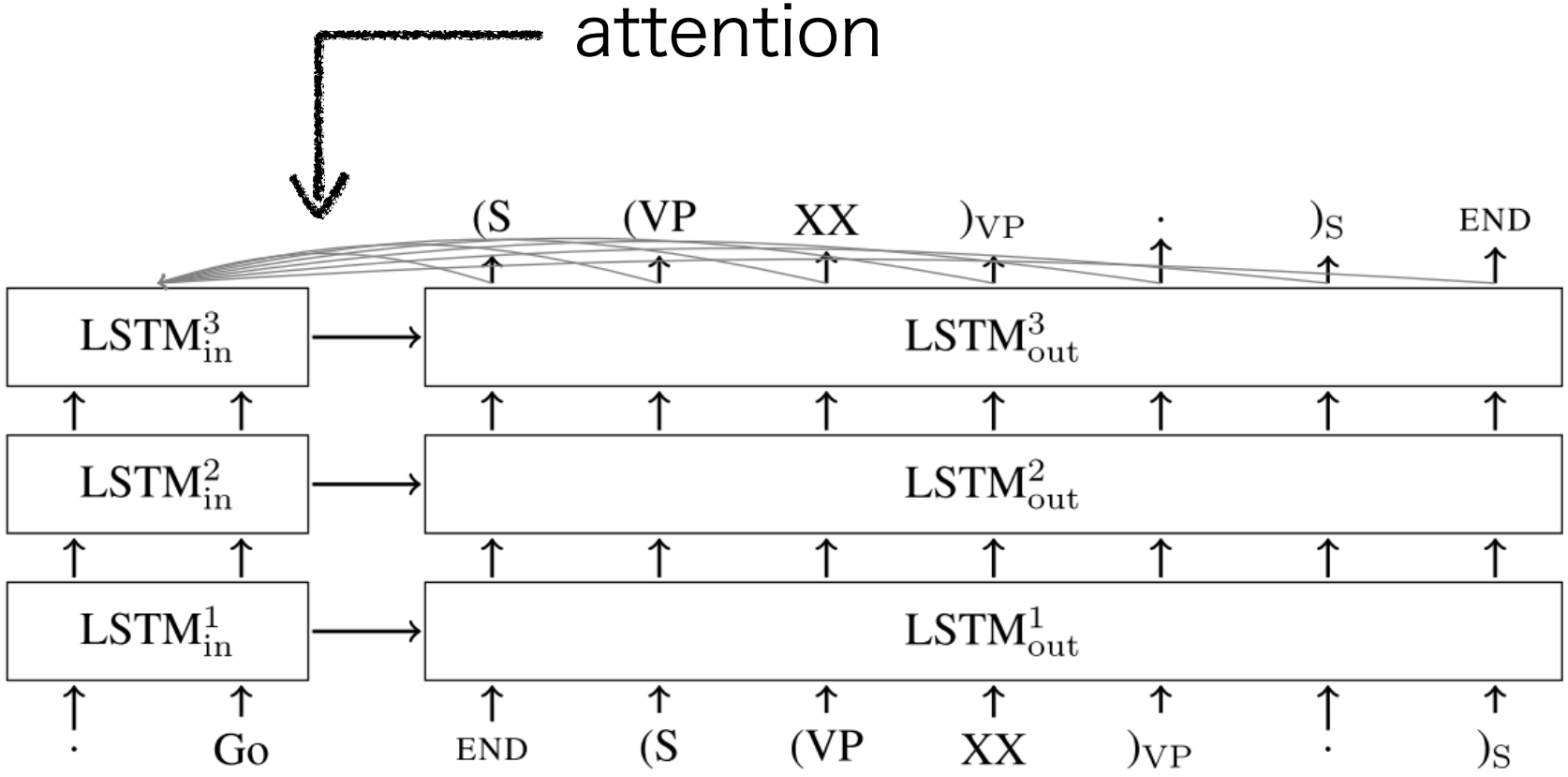
Attentionという意味での拡張はないが文字ベース翻訳知らない勢としては面白い事例もあったので興味あれば原稿を



文字列の生成

Grammar as a Foreign Language [Vinyals+2014]

入力文が文，出力が句構造の構文木

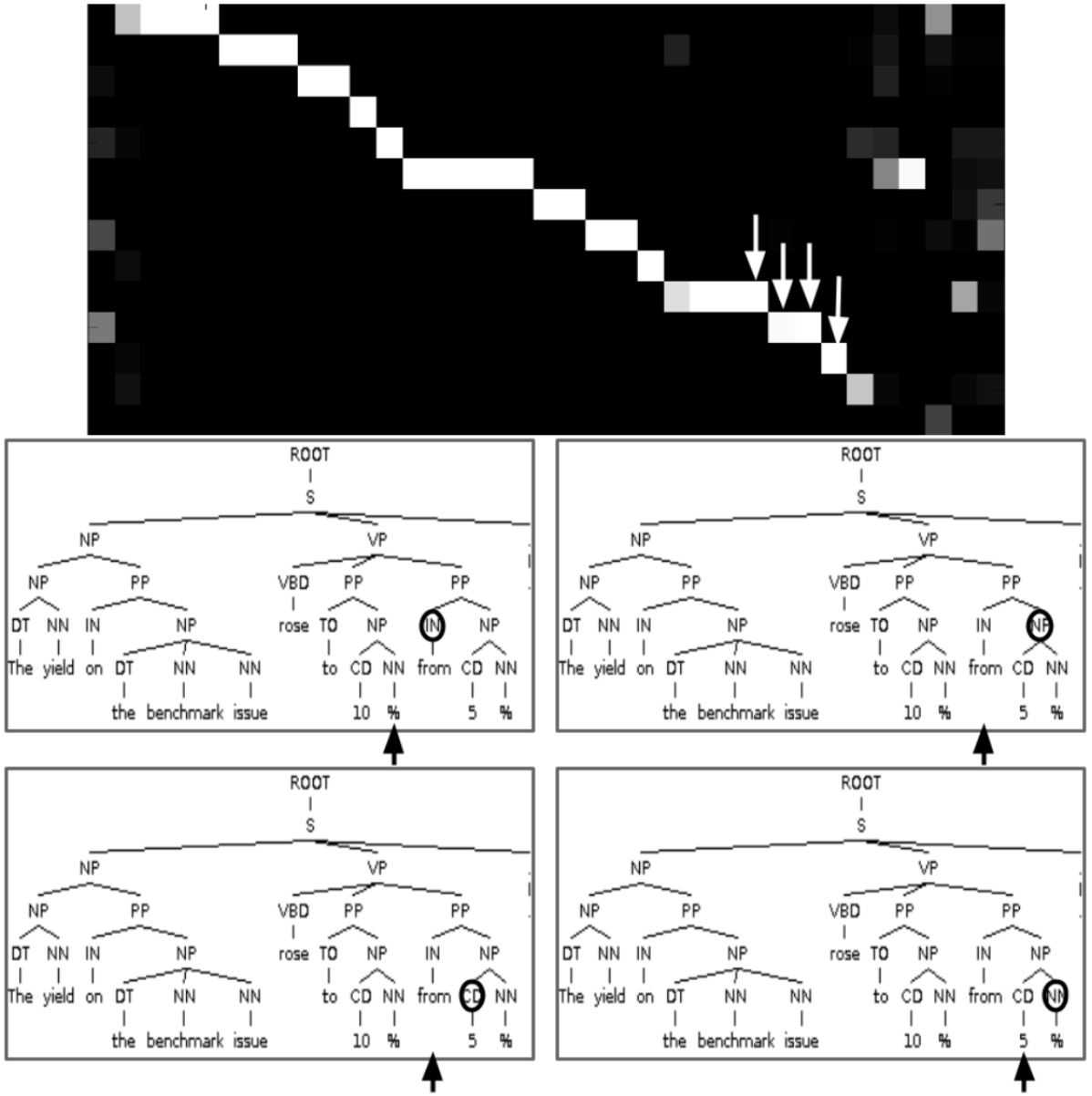


「Go.」という文での例．入力を逆順で入れる

Parser	Training Set	WSJ 22	WSJ 23
baseline LSTM+D	WSJ only	< 70	< 70
LSTM+A+D	WSJ only	88.7	88.3
LSTM+A+D ensemble	WSJ only	90.7	90.5
baseline LSTM	BerkeleyParser corpus	91.0	90.5
LSTM+A	high-confidence corpus	93.3	92.5
LSTM+A ensemble	high-confidence corpus	**93.5**	**92.8**
Petrov et al. (2006) [12]	WSJ only	91.1	90.4
Zhu et al. (2013) [13]	WSJ only	N/A	90.4
Petrov et al. (2010) ensemble [14]	WSJ only	92.5	91.8
Zhu et al. (2013) [13]	semi-supervised	N/A	91.3
Huang & Harper (2009) [15]	semi-supervised	N/A	91.3
McClosky et al. (2006) [16]	semi-supervised	92.4	92.1
Huang & Harper (2010) ensemble [17]	semi-supervised	92.8	92.4

Table 1: F1 scores of various parsers on the development and test set. See text for discussion.

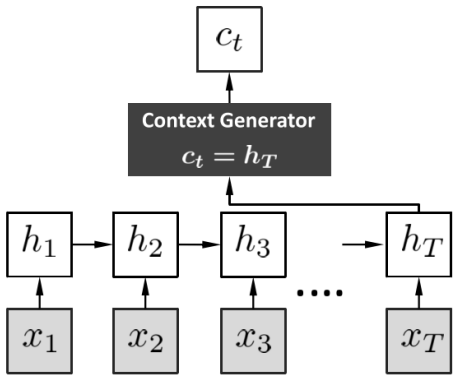
tri-training（BerkeleyとZParの２パーザの出力が一致した文とその木を訓練データとする）によりデータ数増やすと精度が結構あがる（約11million文）


ROOT S NP VP NP PP VBD PP PP DT NN IN NP rose TO NP IN NP The yield on DT NN NN to CD NN from CD NN the benchmark issue 10 % 5 %
ROOT S NP VP NP PP VBD PP PP DT NN IN NP rose TO NP IN NP The yield on DT NN NN to CD NN from CD NN the benchmark issue 10 % 5 %
ROOT S NP VP NP PP VBD PP PP DT NN IN NP rose TO NP IN NP The yield on DT NN NN to CD NN from CD NN the benchmark issue 10 % 5 %
ROOT S NP VP NP PP VBD PP PP DT NN IN NP rose TO NP IN NP The yield on DT NN NN to CD NN from CD NN the benchmark issue 10 % 5 %

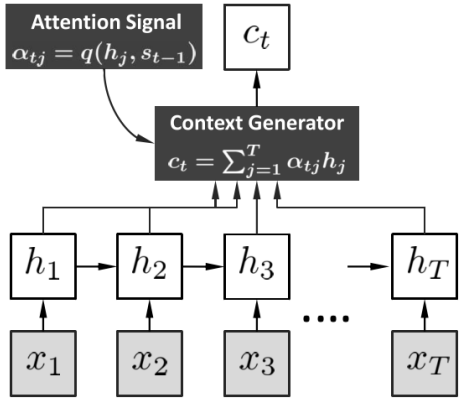
Neural Responding Machine for Short-Text Conversation [Shang+2015]

対話生成（入力された単語列への応対としての単語列の出力）

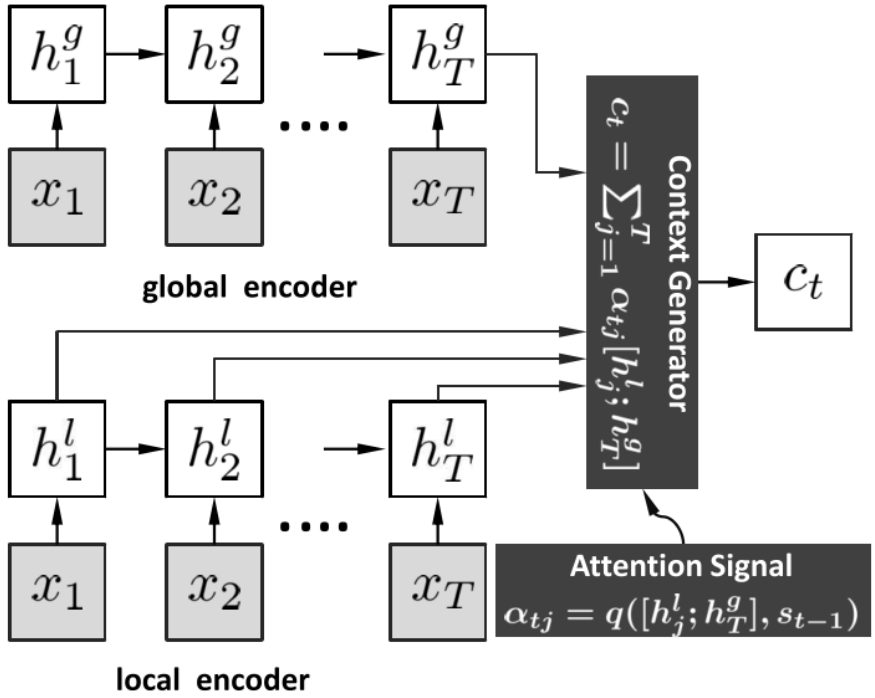
以下，Encoderだけの図（c_tがデコーダに渡される）



普通のenc-dec



attention



Joint

ID	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr.-based
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧… High fever…
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, **CuiDuoLa**	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这是哪儿啊？真美！ Where is it? It is so beautiful!	阳光明媚，心情舒畅。 Such a nice sunny day! I am in a great mood.	这种天气真不是盖的。 It is indeed a very nice weather.	文山啊 出去踏青寻找灵感 哈哈 **WenShan**, let's go out to get some inspiration. Ha! Ha!

Figure 8: Some responses generated by different models (originally in Chinese with their literal English translation), where the words in boldfaces are entity names.

Teaching Machines to Read and Comprehend Karl [Hermann+15]

Original Version	Anonymised Version
Context	
The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...	the *ent381* producer allegedly struck by *ent212* will not press charges against the " *ent153* " host , his lawyer said friday . *ent212* , who hosted one of the most - watched television shows in the world , was dropped by the *ent381* wednesday after an internal investigation by the *ent180* broadcaster found he had subjected producer *ent193* " to an unprovoked physical and verbal attack . " ...
Query	
Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.	producer **X** will not press charges against *ent212* , his lawyer says .
Answer	
Oisin Tymon	*ent193*

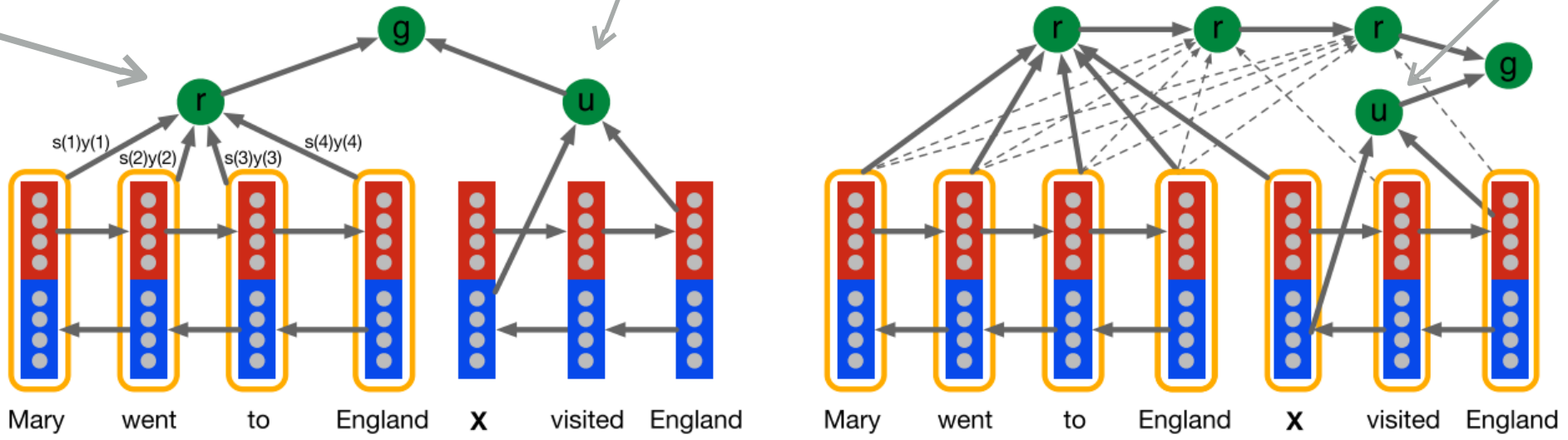
$$p(a|d,q) \propto \exp\left(W(a)g(d,q)\right), \quad \text{s.t. } a \in V,$$

Contextと穴あき（X）Queryが与えられた時,

Xに何が入るかをContextから選択

基本は両方向RNN

Queryベクトル (u)は両方向RNNの最後の隠れ状態を連結したもの



(a) Attentive Reader.

(b) Impatient Reader.

Contextのベクトル(r)はQueryの情報(u)を使ったattention

$$m(t) = \tanh\left(W_{ym} y_d(t) + W_{um} u\right)$$
$$s(t) \propto \exp\left(\mathrm{w}_{ms}^{\intercal} m(t)\right),$$
$$r = y_d s,$$

Queryの各文字でRNNすることでrを作る

各ステップでattention

	CNN		Daily Mail	
	valid	test	valid	test
Maximum frequency	30.5	33.2	25.6	25.5
Exclusive frequency	36.6	39.3	32.7	32.8
Frame-semantic model	36.3	40.2	35.5	35.5
Word distance model	50.5	50.9	56.4	55.5
Deep LSTM Reader	55.0	57.0	63.3	62.2
Uniform Reader	39.0	39.4	34.6	34.4
Attentive Reader	61.6	63.0	**70.5**	**69.0**
Impatient Reader	**61.8**	**63.8**	69.0	68.0

Table 5: Accuracy of all the models and benchmarks on the CNN and Daily Mail datasets. The Uniform Reader baseline sets all of the $m(t)$ parameters to be equal.

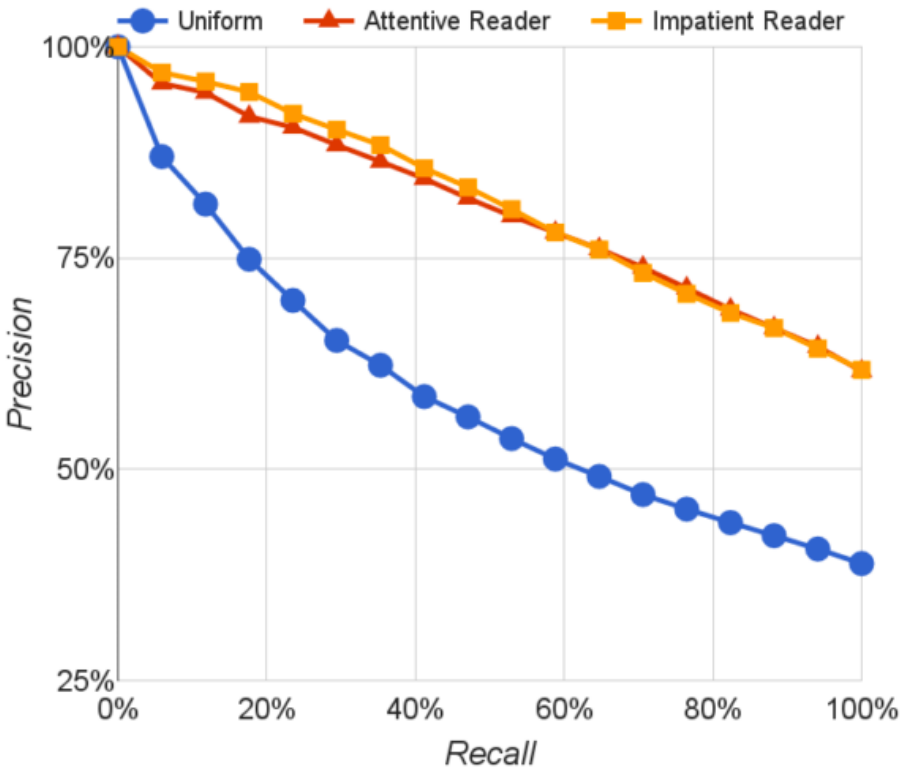


Figure 2: Precision@Recall for the attention models on the CNN validation data.

Attention(前スライドのふたつ)強い

二前前の(a)Attentive Readerのヒートマップ

by *ent423* , *ent261* correspondent updated 9:49 pm et , thu march 19 , 2015 (*ent261*) a *ent114* was killed in a parachute accident in *ent45* , *ent85* , near *ent312* , a *ent119* official told *ent261* on wednesday . he was identified thursday as special warfare operator 3rd class *ent23* , 29 , of *ent187* , *ent265* . `` *ent23* distinguished himself consistently throughout his career . he was the epitome of the quiet professional in all facets of his life , and he leaves an inspiring legacy of natural tenacity and focused

. . .

ent119 identifies deceased sailor as **X** , who leaves behind a wife

正解

by *ent58* , *ent61* updated 11:44 am et , tue march 10 , 2015 (*ent61*) a suicide attacker detonated a car bomb near a police vehicle in the capital of southern *ent29* 's *ent85* on tuesday , killing seven people and injuring 23 others , the province 's deputy governor said . the attack happened at about 6 p.m. in the *ent8* area of *ent67* city , said *ent30* , deputy governor of *ent85* . several children were among the wounded , and the majority of casualties were civilians , *ent30* said . details about the attacker 's identity and motive were n't immediately available .

car bomb detonated near police vehicle in **X** , deputy governor says

間違い

爆破が起きた地理情報（の粒度）に曖昧性があり，
正解は選べなかったけど文脈的には正しいattention

(b)の方の時系列Heat mapもあるので気になる人はsee the paper

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - **Global attention**

ここまで

  - Local attention
  - Pointer networks
  - Attention for image
- Attention techniques
- NN with Memory
  - (end-to-end) memor
  - neural turing machin
  - language modeling

翻訳以外のタスクでは？

- 文要約 [Rush+2015]
- パージング [Vinyals+2014]
- 文字ベース翻訳 [Ling+2015]
- 対話生成 [Shang+2015]
- 穴埋め [Hermann+15]

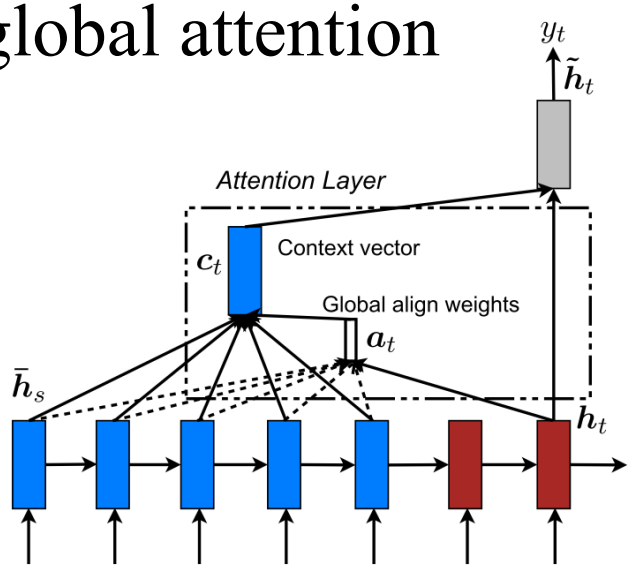
など…

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - **Local attention**
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

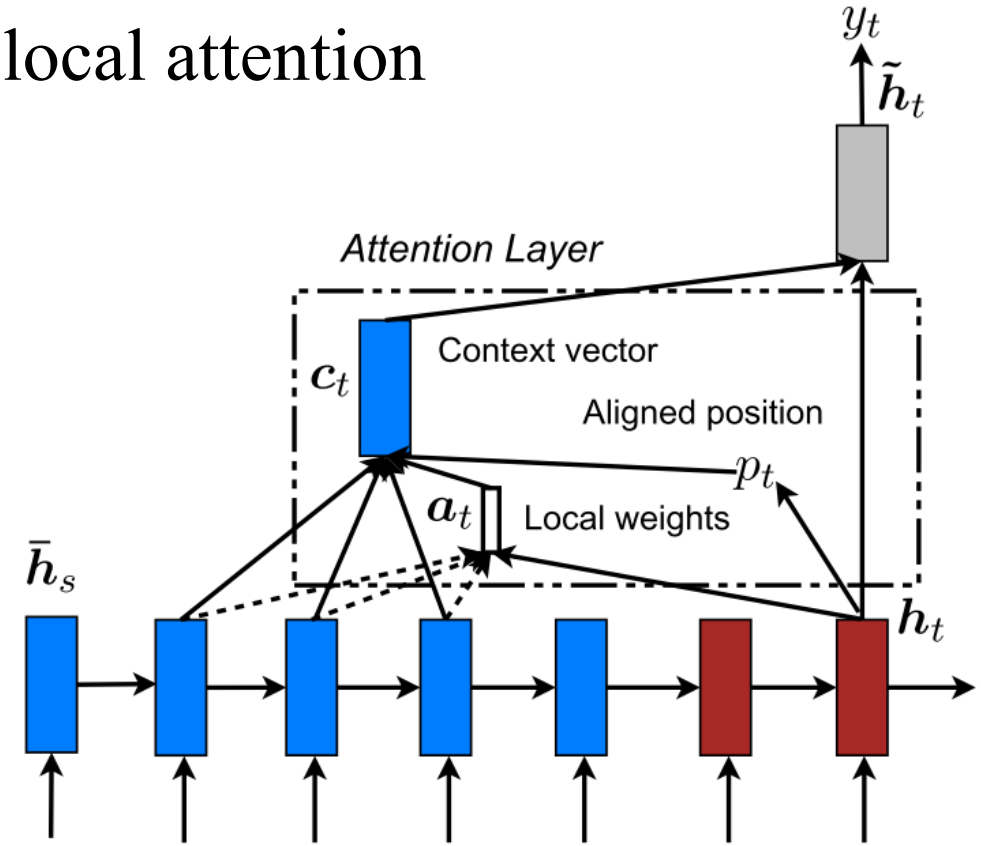
Effective Approaches to Attention-based Neural Machine Translation [Luong+15]

global attention



入力全体を加重平均

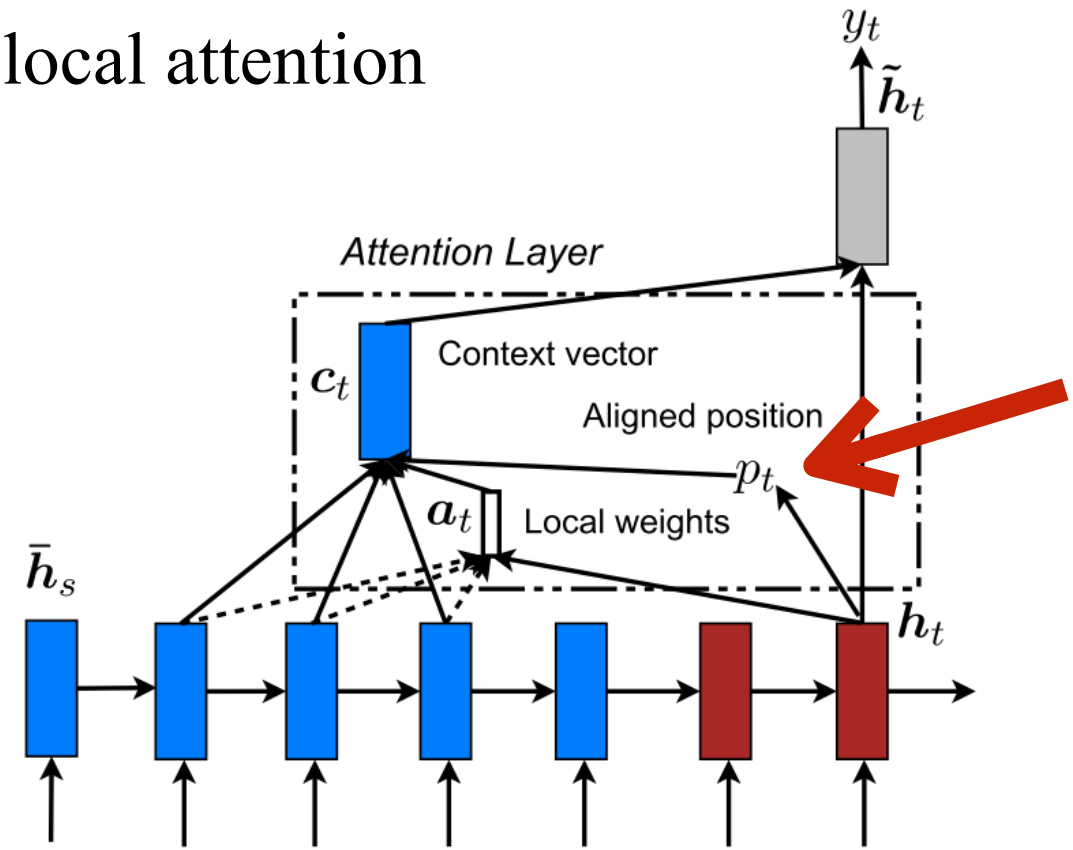
local attention



Neural Machine Translation

位置情報の情報も利用したattention

local attention

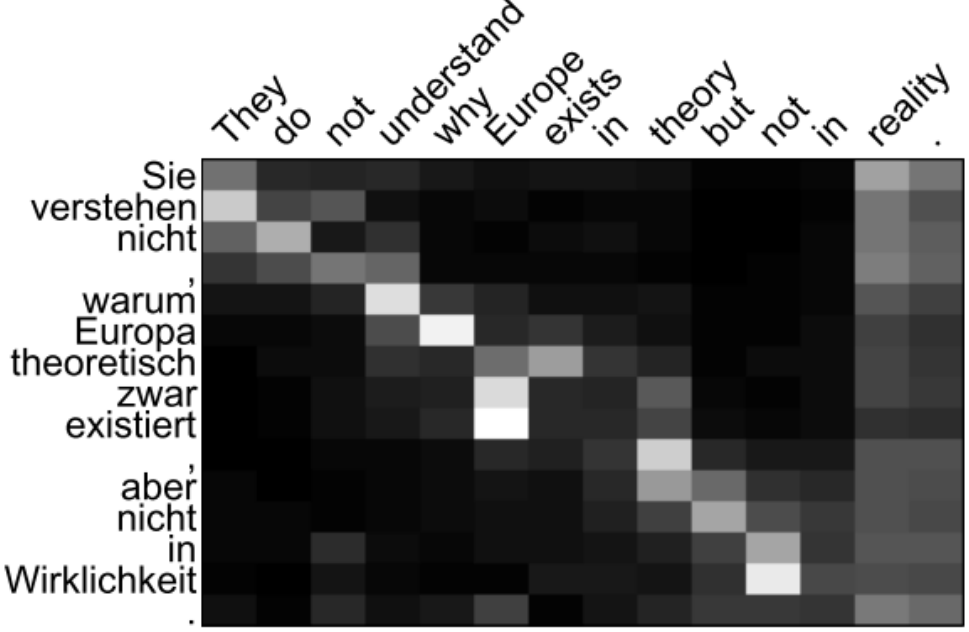


時刻tで注目する位置: $p_t = \underbrace{S}_{\text{文長}} \cdot \underbrace{\text{sigmoid}(\boldsymbol{v}_p^\top \tanh(\boldsymbol{W_p h}_t))}_{0.0\sim1.0}$

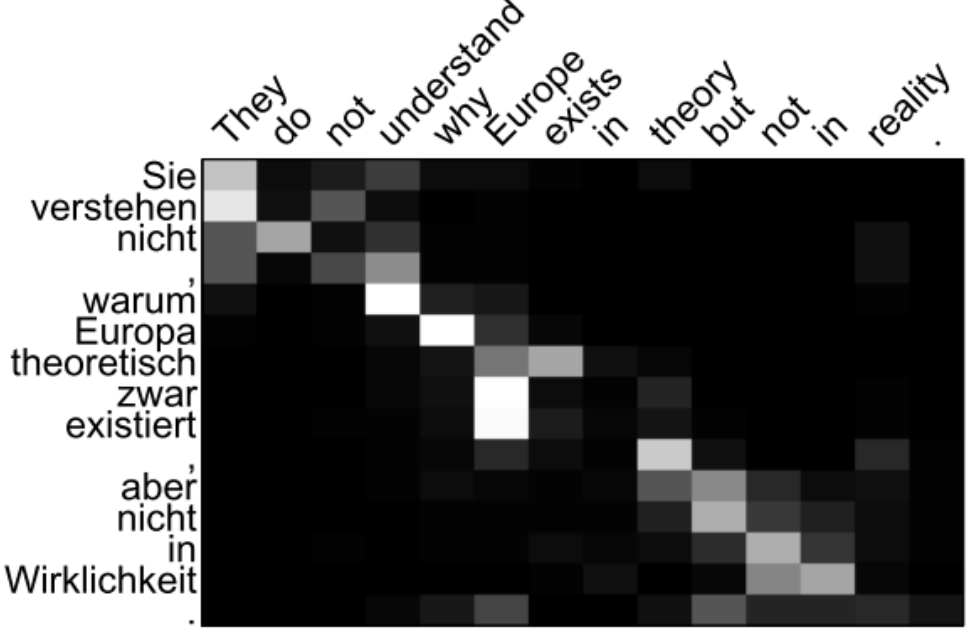
p_tを使ってattention : $\boldsymbol{a}_t(s) = \text{align}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s) \boxed{\exp\left(-\frac{(s-p_t)^2}{2\sigma^2}\right)}$

/

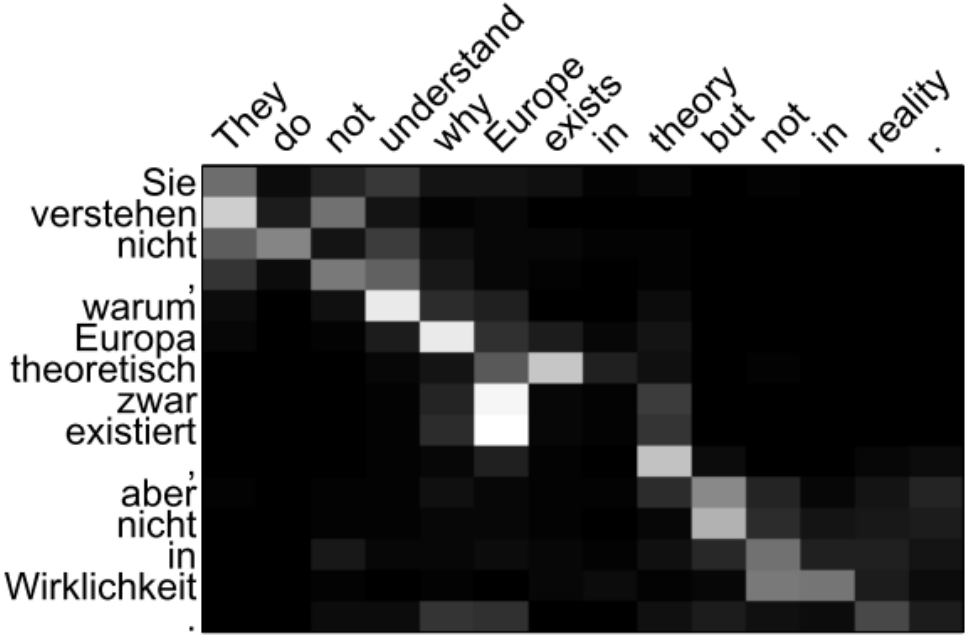
global attentionの時の重みも使う

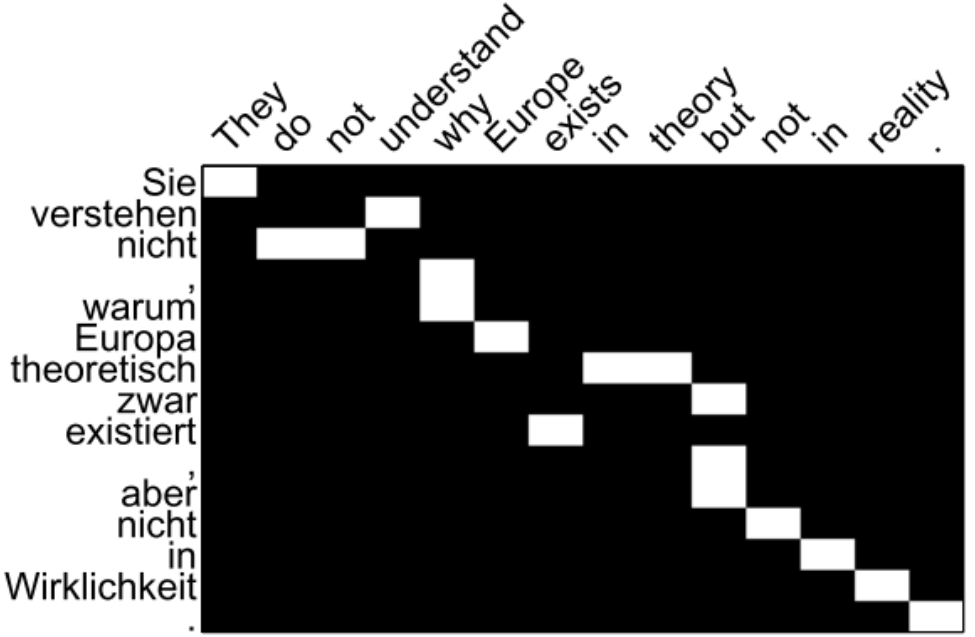


↑ global



↑ これ





↑ gold

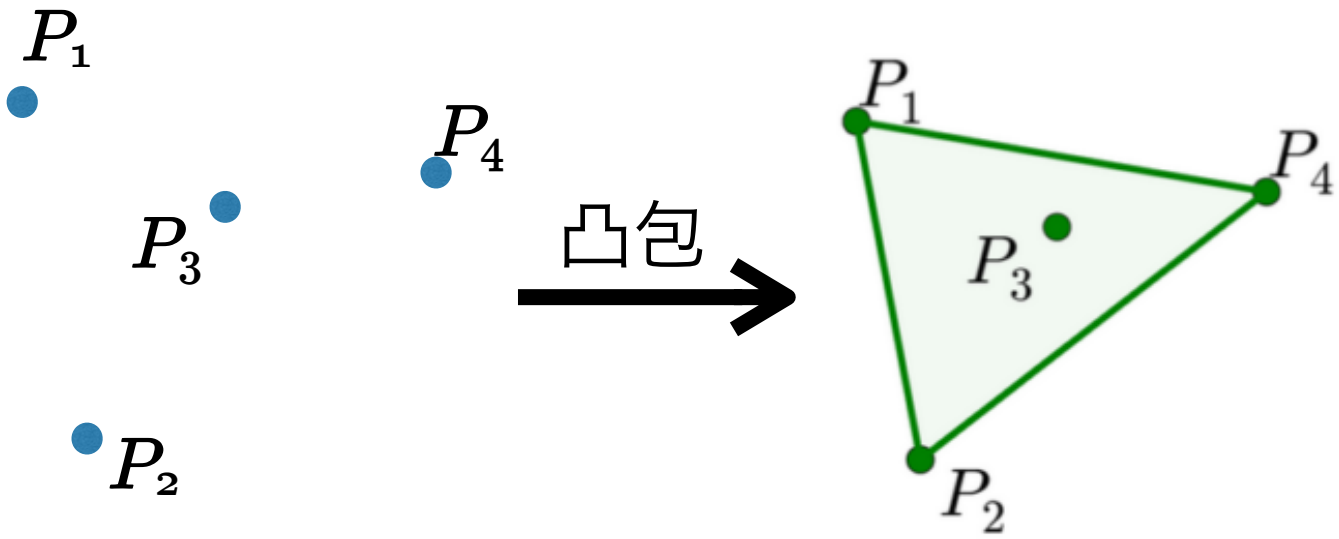
Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - **Pointer networks**
  - Attention for image (image caption generation)
- Attention techniques
- NN with Memory

Pointer Networks [Vinyals+2015]

（Attentionの異なる使いかた）

NNに組合せ最適化問題を解かせる

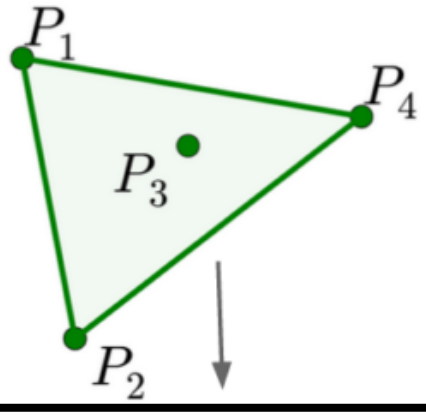


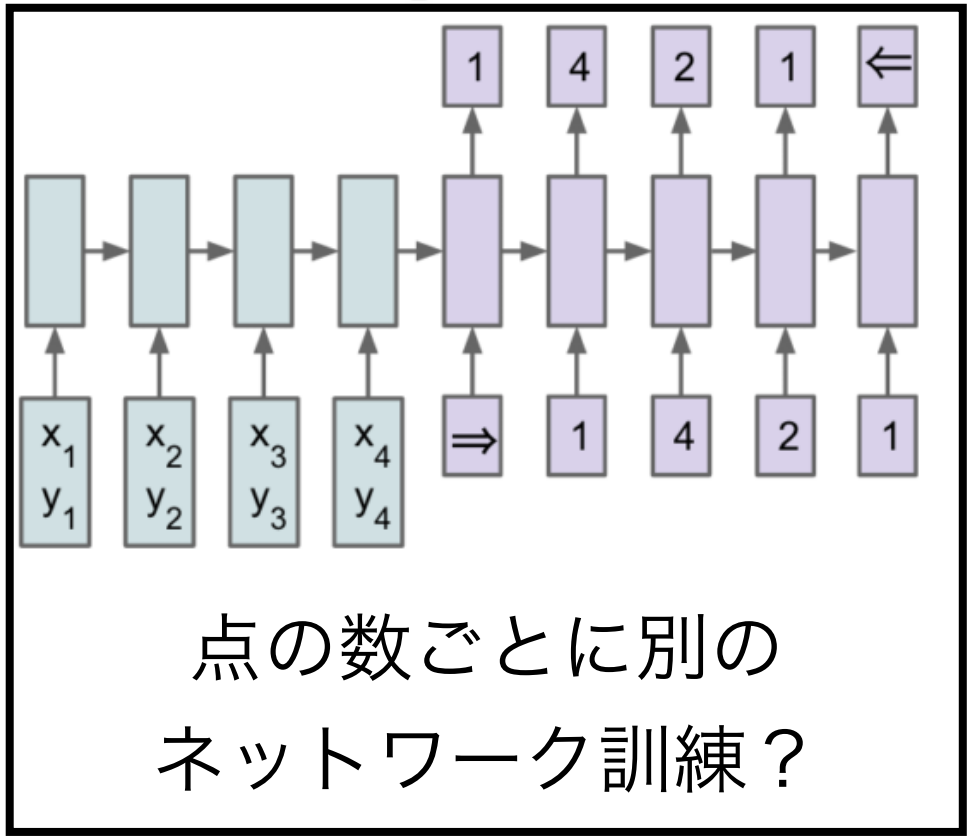
翻訳のように「固定された語彙リストからの選択」ではなく

「入力系列のいずれかそのものを選択する」というNN

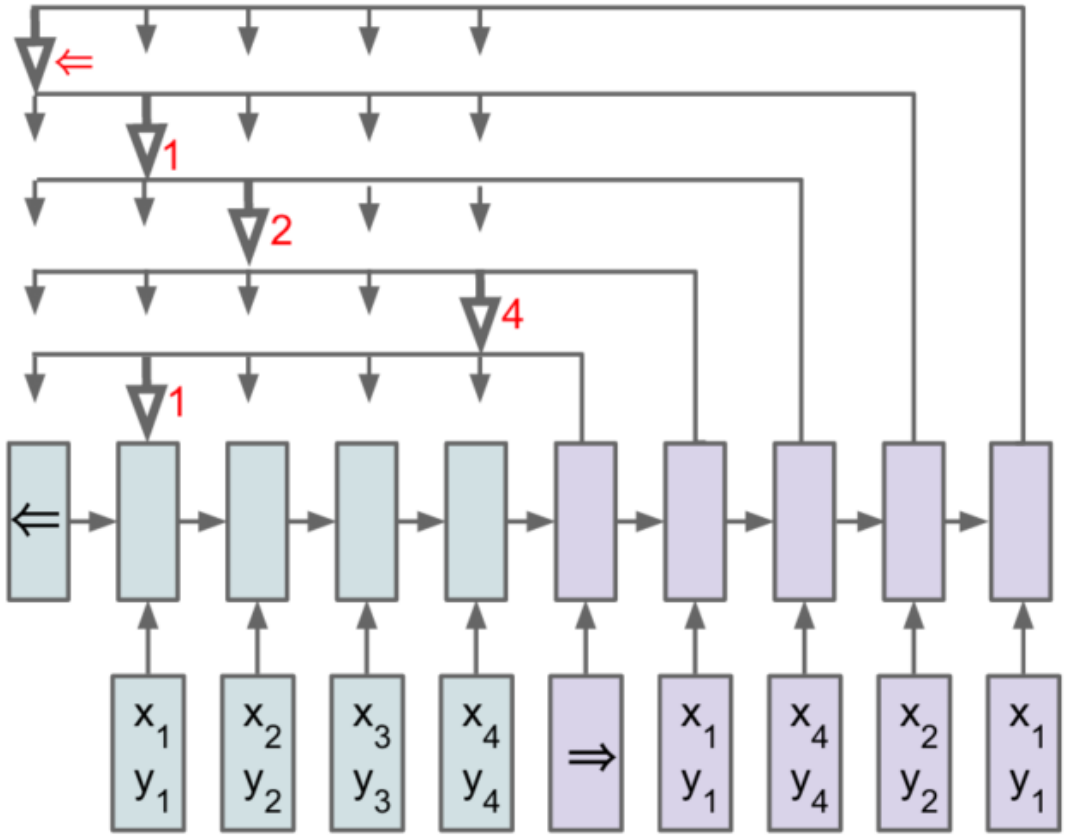
考慮するP_iの数は問題によって違う

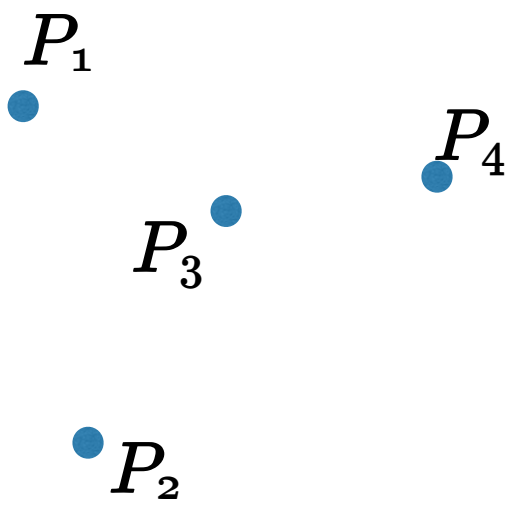
Decoderの出力層(softmax)の次元を固定できない





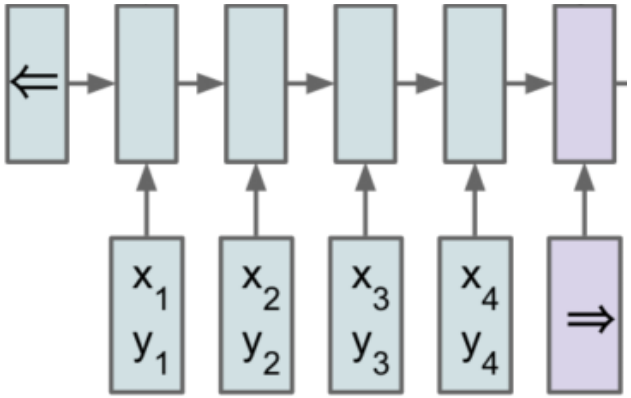
点の数ごとに別の
ネットワーク訓練？



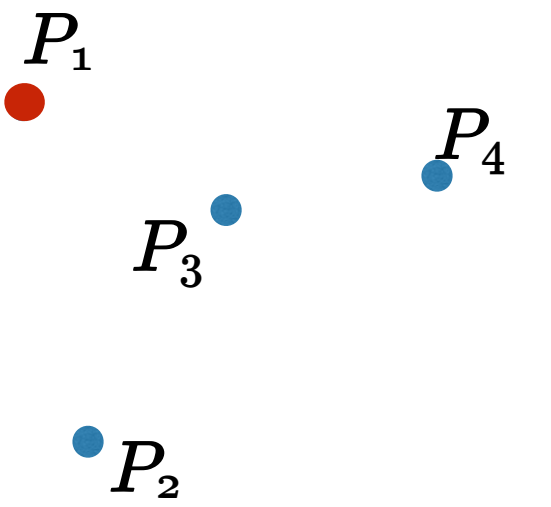


uの分布

これまでのattentionで荷重ベクトルの荷重として使ってきた

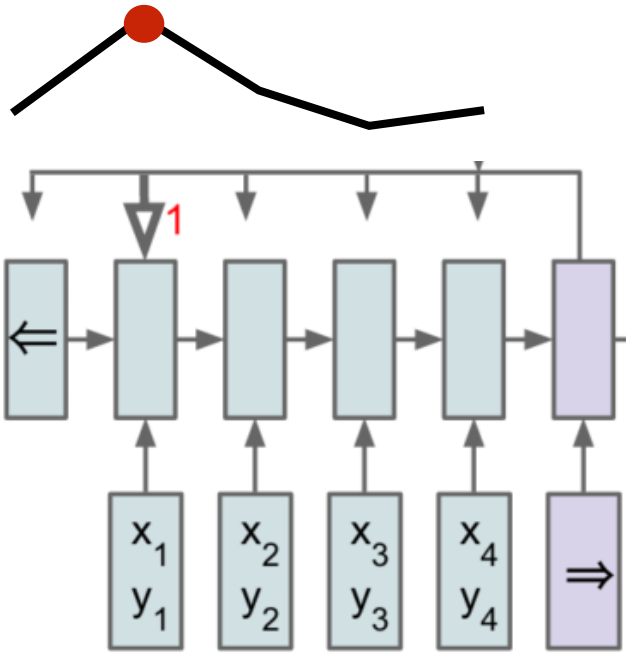


$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$



uの分布

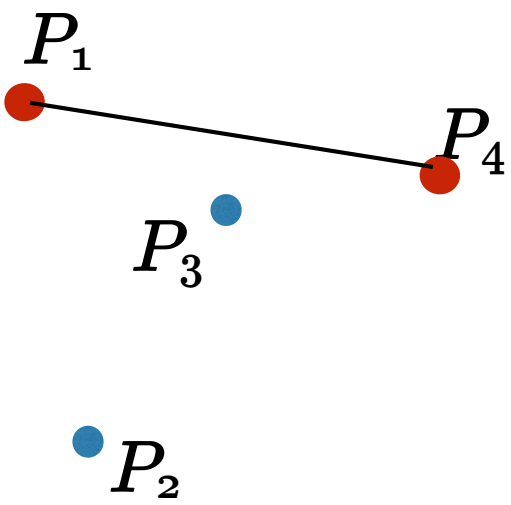
これまでのattentionで荷重ベクトルの荷重として使ってきた

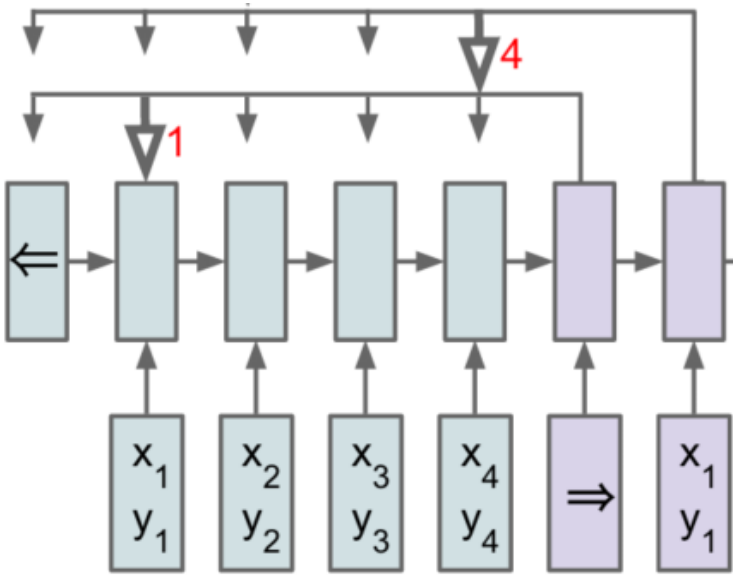


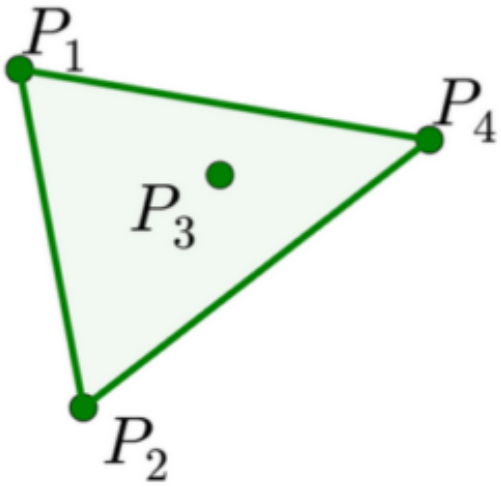
$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i)$$

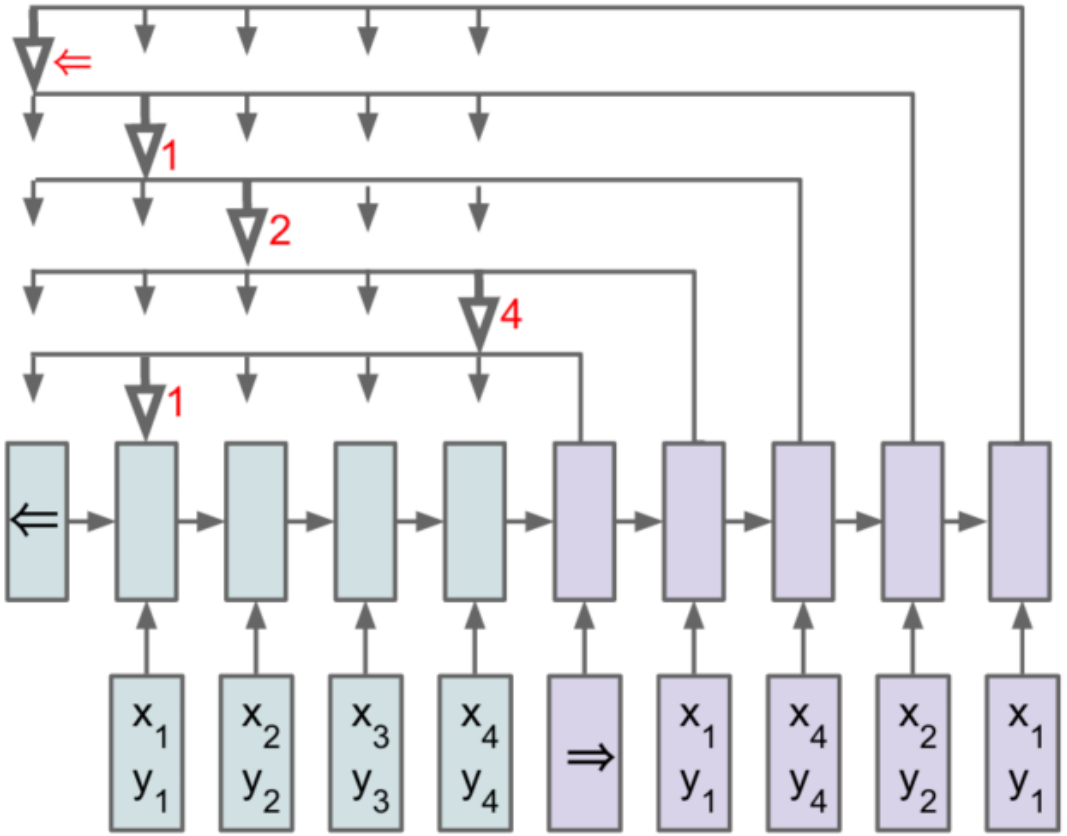
$$p(C_i|C_1, \ldots, C_{i-1}, \mathcal{P}) = \text{softmax}(u^i)$$

attentionの分布をそのまま答えに！

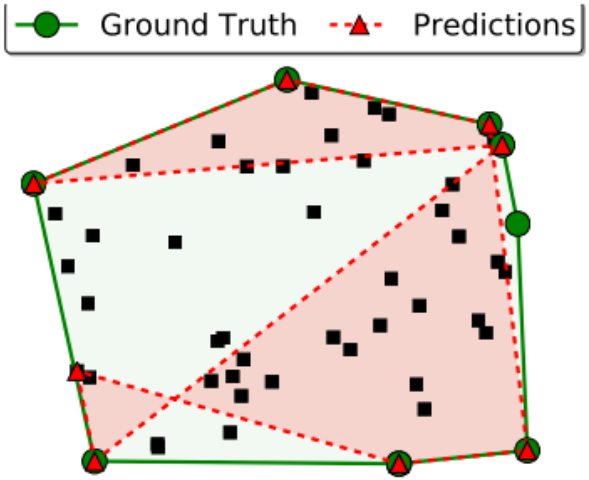

P_1
P_4
P_3
P_2


4
1
⇐
x_1 y_1
x_2 y_2
x_3 y_3
x_4 y_4
⇒
x_1 y_1

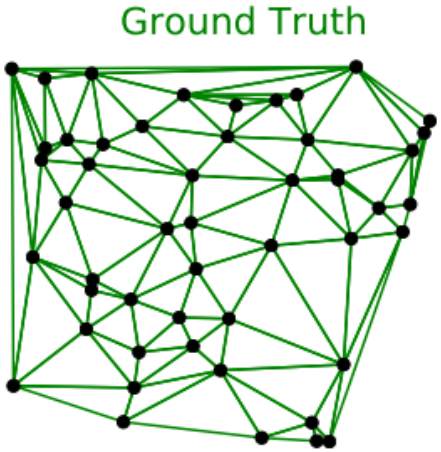

P_1
P_4
P_3
P_2


⇐
1
2
4
1
⇐
x1 y1
x2 y2
x3 y3
x4 y4
⇒
x1 y1
x4 y4
x2 y2
x1 y1

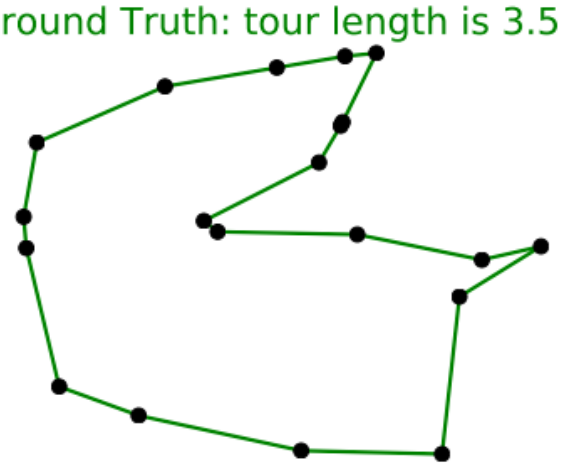
緑が正解，赤が予測



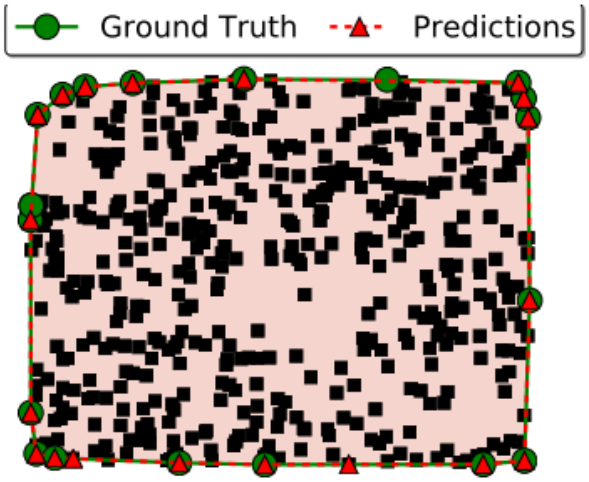
(a) LSTM, m=50, n=50

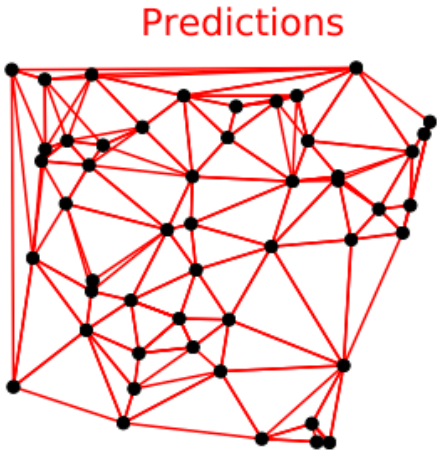


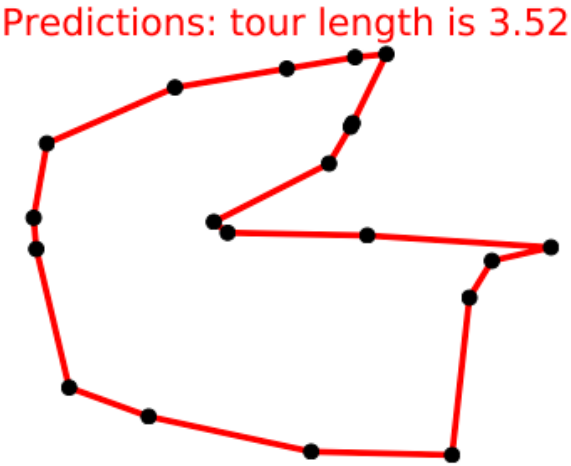
(b) Truth, n=50



(c) Truth, n=20





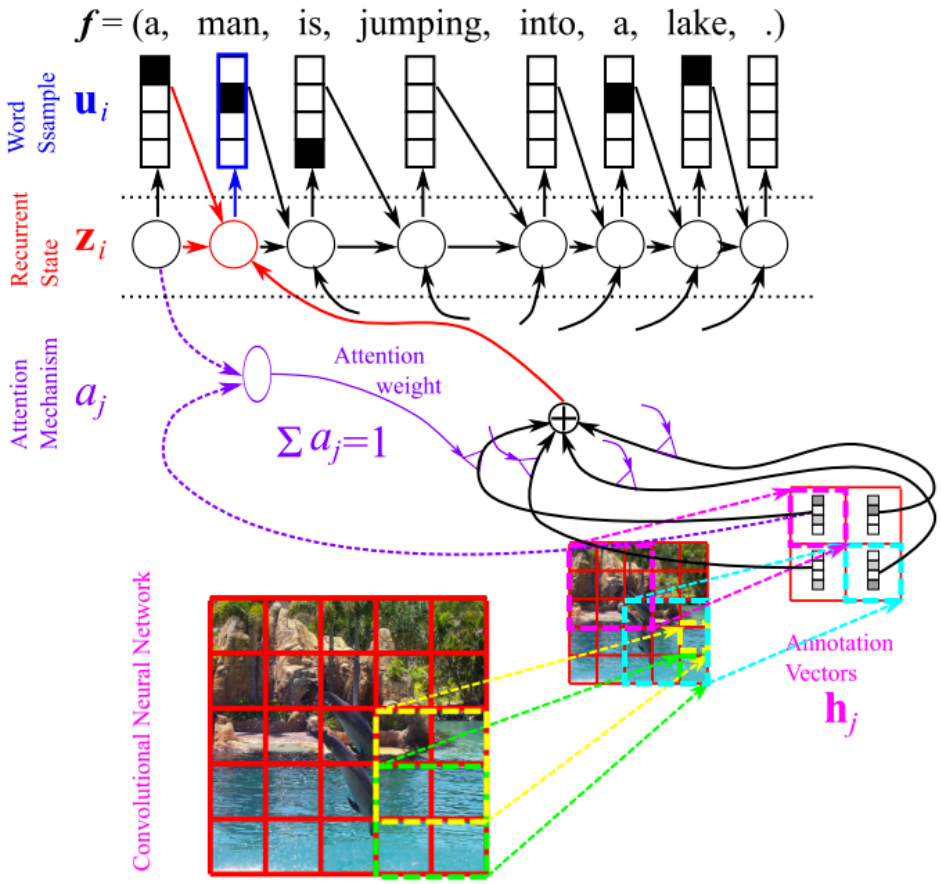


LSTM→ふつうのEnc-dec+attention

Agenda

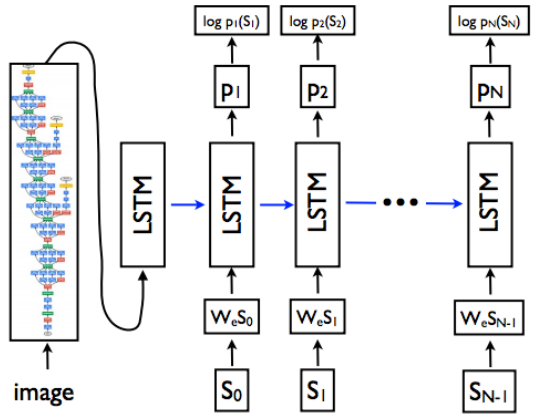
- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - Pointer networks
  - **Attention for image (image caption generation)**
- Attention techniques
- NN with Memory

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention



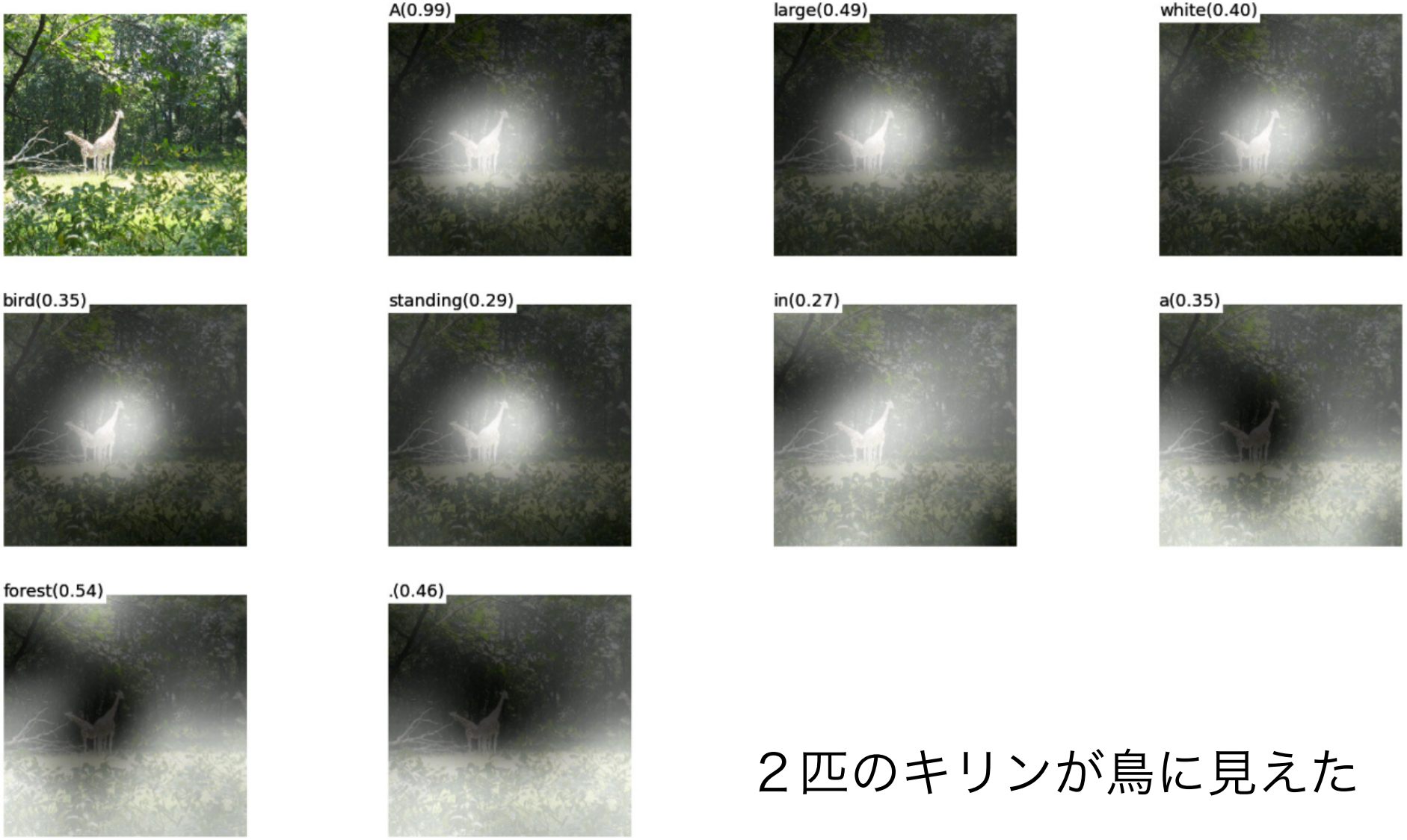
画像から説明文[Xu+2015]

CNNの最後のConv層を使い
画像のどこに注目するかを決める



2014年の終わり話題になった[Vinyals+2015b]
は最終(fully connected)層をdecoderへ渡す

[Xu+2015]からの図



２匹のキリンが鳥に見えた

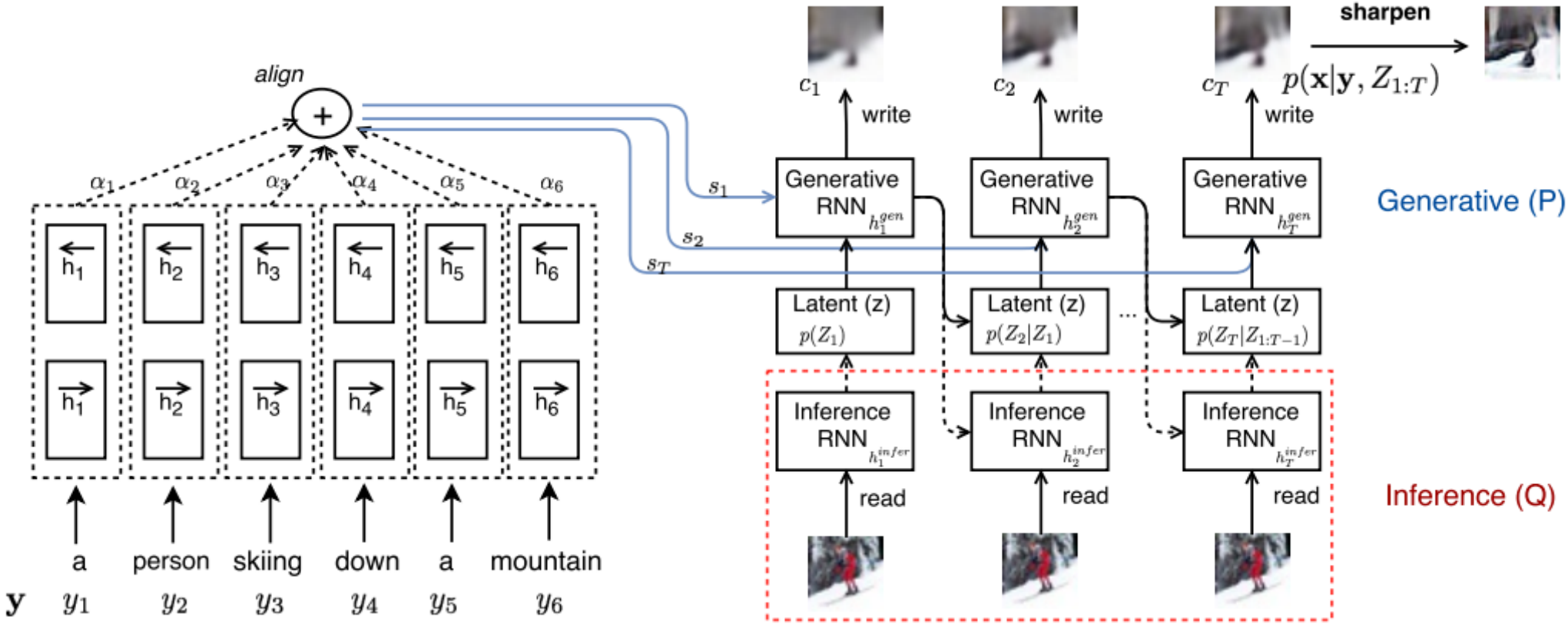
(b) A large white bird standing in a forest.

パーカーの丸いマークを
時計と勘違い

(b) A woman holding a clock in her hand.

Generating Images from Captions with Attention
[Mansimov+2016]

そんなことが可能なのか…？



両方向RNN

DRAW[Gregor+2015]の拡張

実際に生成した画像

A stop sign is flying in blue skies.

A herd of elephants flying in the blue skies.

A toilet seat sits open in the grass field.

A person skiing on sand clad vast desert.

キャプションの単語(色やテーブルにのってるもの)を変えた時の画像の変化

A yellow school bus parked in a parking lot.

A red school bus parked in a parking lot.

A green school bus parked in a parking lot.

A blue school bus parked in a parking lot.

The decadent chocolate desert is on the table.

A bowl of bananas is on the table.

A vintage photo of a cat.

A vintage photo of a dog.

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- **Attention techniques**
- NN with Memory

Attentionの良さを高めるための工夫たち

Doubly Stochastic Attention from [Xu+2015]

ふつうのAttention: 各ステップで入力で荷重の総和取ると1になる
(softmax的な意味で)

→ 入力系列の各要素(i)について時間(t)での総和も1に近づける

$$L_d = -\log(P(\mathbf{y}|\mathbf{x})) + \lambda \sum_i^L (1 - \sum_t^C \alpha_{ti})^2$$

これによって入力全体をまんべんなくAttentionさせる
BLEUの向上にも，定性的な向上にも役立ったとのこと

Attentionの良さを高めるための工夫たち

Gated Attention from [Xu+2015]

Attentionの結果できた加重平均ベクトルを
どのくらい使うかをgateで制御する

$$\phi\left(\{\mathbf{a}_i\}, \{\alpha_i\}\right) = \beta \sum_i^L \alpha_i \mathbf{a}_i$$

$$\beta_t = \sigma(f_\beta(\mathbf{h}_{t-1}))$$

これを入れると，画像中のオブジェクトへの
Attentionがより強調されたっぽい

Attentionの良さを高めるための工夫たち

Weak Supervision from [Ling+2015]

文字ベース機械翻訳のやつ

Attentionに明示的な教師はあまり使われていないが
この論文はIBM model4のアライメント情報を
弱教師として利用

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- **NN with Memory**

最近，LSTMのメモリーブロック的な意味ではない
外部Memoryを使ったNNたち(翻訳の[Meng+2016]など他にもちらほら)

Neural Turing Machines [Graves+2014]

アルゴリズム

End-To-End Memory Networks[Sukhbaatar+2015]

QA

Recurrent Memory Network for Language Modeling[Tran+2016]

言語モデル

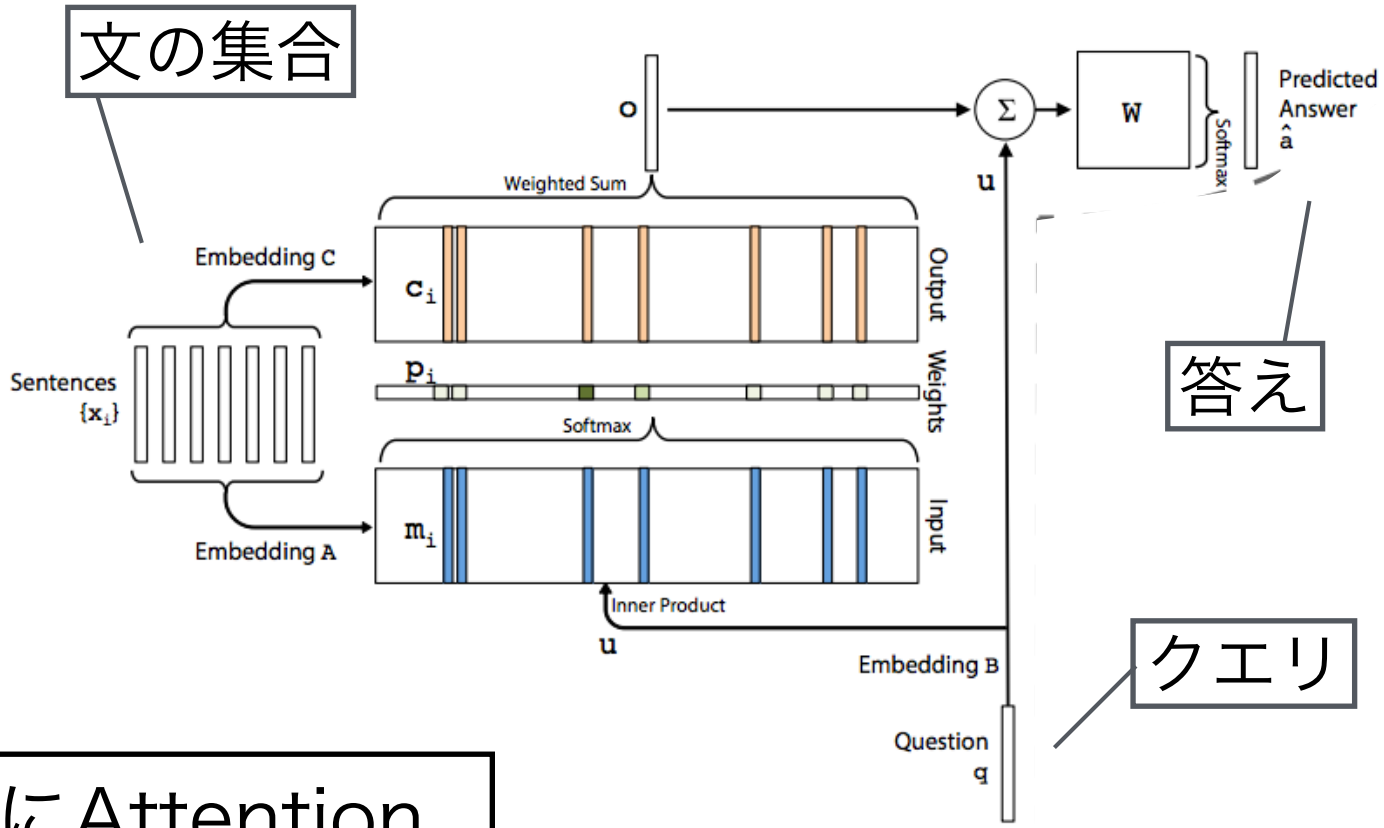
End-To-End Memory Networks[Sukhbaatar+2015]

タスクはQA

```
Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom
```

$$m_i = \sum_j A x_{ij}$$

文は単語ベクトルの和



実際はわりと普通にAttention

$$p_i = \mathrm{Softmax}(u^T m_i)$$

$$o = \sum_i p_i c_i$$

入力のattention荷重を決めるためのLookupTableと，実際に加重平均されるものを別に

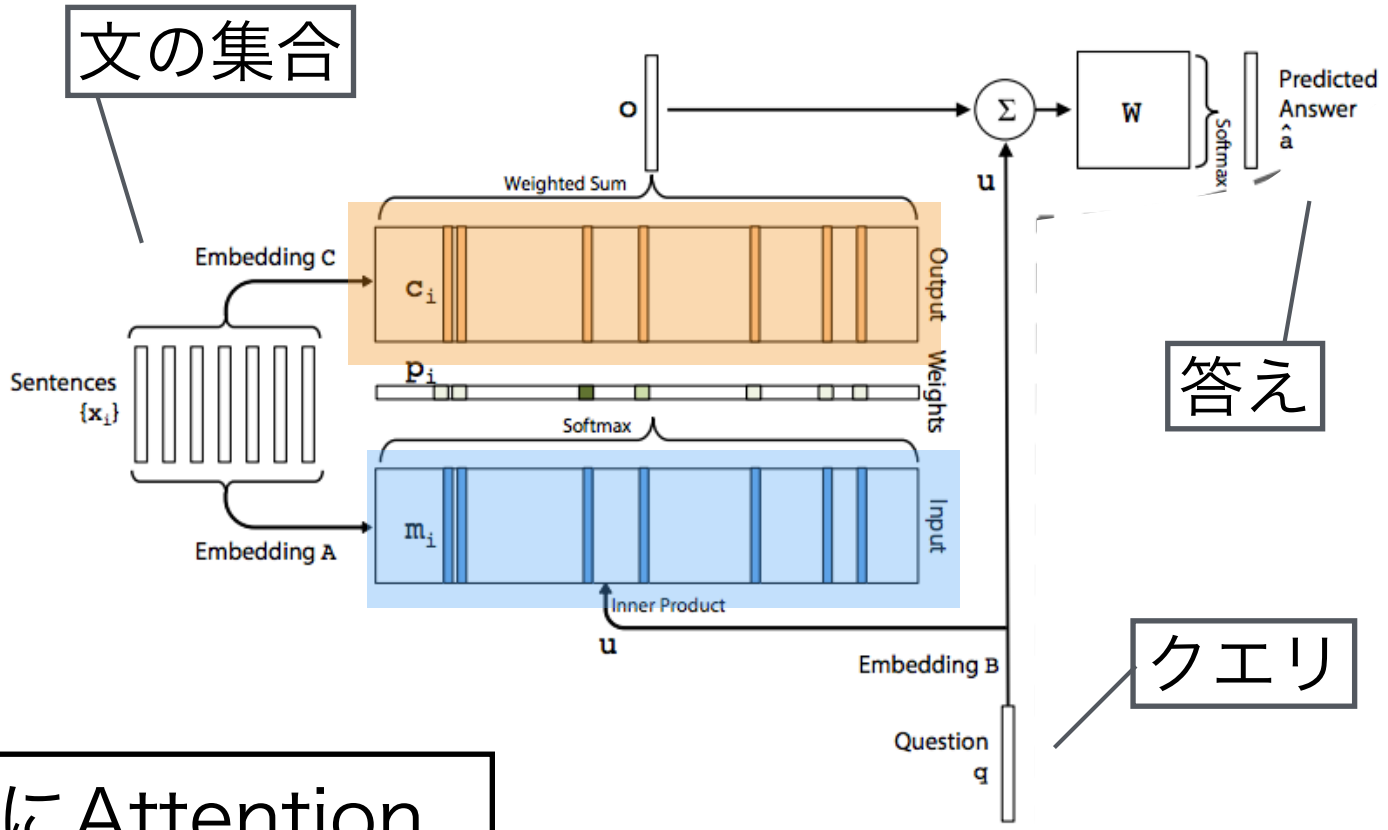
End-To-End Memory Networks[Sukhbaatar+2015]

タスクはQA

```
Sam walks into the kitchen.
Sam picks up an apple.
Sam walks into the bedroom.
Sam drops the apple.
Q: Where is the apple?
A. Bedroom
```

$$m_i = \sum_j A x_{ij}$$

文は単語ベクトルの和



実際はわりと普通にAttention

$$p_i = \mathrm{Softmax}(u^T m_i)$$

$$o = \sum_i p_i c_i$$

入力のattention荷重を決めるためのLookupTableと，実際に加重平均されるものを別に

Recurrent Memory Network for Language Modeling[Tran+2016]

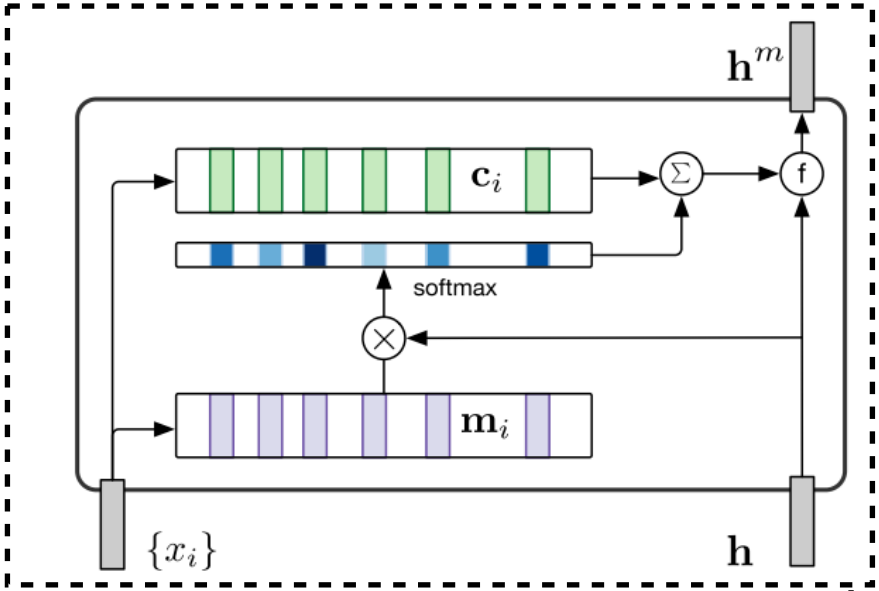
これも同様

過去の単語集合をAttentionする

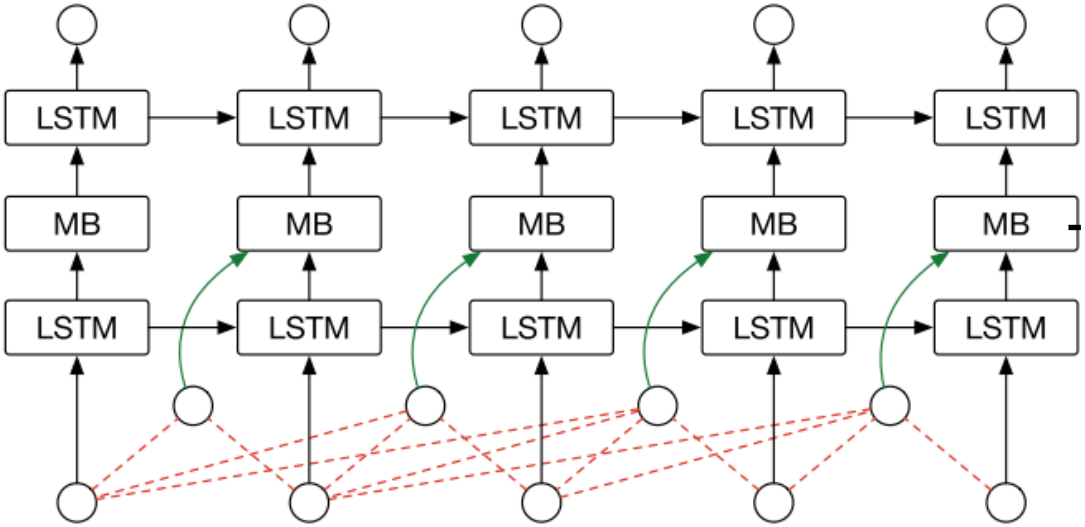
$$\mathbf{p}_t = \mathrm{softmax}\big((\mathbf{M}_i + \mathbf{T}_i)\mathbf{h}_t\big)$$

$$\mathbf{s}_t = \mathbf{C}_i^{\mathsf{T}}\mathbf{p}_t$$

$$g(\mathbf{s}_t, \mathbf{h}_t) = \mathbf{s}_t + \mathbf{h}_t$$

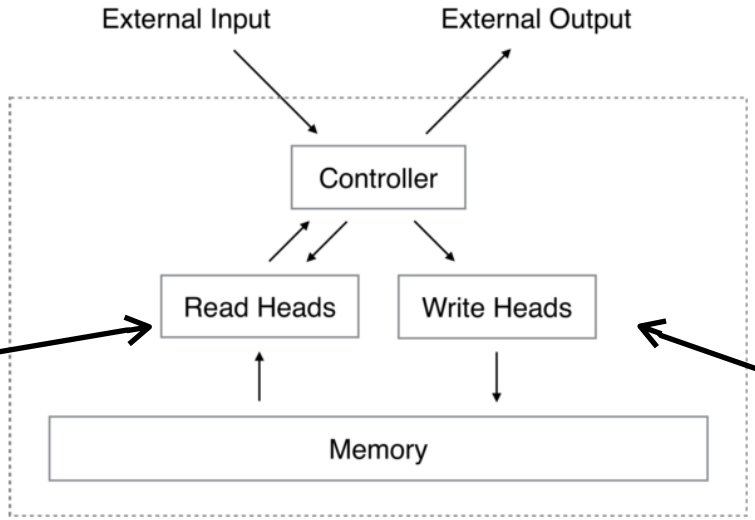


RNNLMを成す層の一部に



Neural Turing Machines [Graves+2014]

外部メモリー使います感が高い



Memoryのどの位置を
（重点的に）読むか

$$\mathbf{r}_t \longleftarrow \sum_i w_t(i)\mathbf{M}_t(i)$$

Attention($\sum_i w_t(i) = 1$)

Memoryのどの位置をどう書き換えるか

$$\tilde{\mathbf{M}}_t(i) \longleftarrow \mathbf{M}_{t-1}(i)\left[\mathbf{1} - w_t(i)\mathbf{e}_t\right]$$

元の記憶を弱めるgate

$$\mathbf{M}_t(i) \longleftarrow \tilde{\mathbf{M}}_t(i) + w_t(i)\,\mathbf{a}_t$$

今回の記憶を追記するgate

Agenda

- ~~Background1: Neural Network~~
- ~~Background2: Recurrent Neural Network~~
- Background3: Encoder-Decoder approach
  (sequence to sequence approach)
- Attention mechanism and its variants
  - Global attention
  - Local attention
  - Pointer networks
  - Attention for image (image caption generation)
- Attention techniques
- **NN with Memory**

おわり

Attentionはいろいろなところですでに使われまくり始めてる

Attentionの取り方(local attention, memory network)
Attentionをする対象（テキスト，画像，扱わなかったが音声や動画も）
Attentionの使いかた（pointer network, memory network）
よりよいAttentionのための小技
（doubly stocastic attention, gated attention , weak supervision）

Reference

[Graves2013] Alex Graves, “Generating Sequences With Recurrent Neural Networks”.

[Cho+2014] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”.

[Sutskever+2014] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, “Sequence to Sequence Learning with Neural Networks”.

[Luong+2015] Minh-Thang Luong, Hieu Pham, Christopher D. Manning, “Effective Approaches to Attention-based Neural Machine Translation”.

[Denil+2011] Misha Denil, Loris Bazzani, Hugo Larochelle, Nando de Freitas, “Learning where to Attend with Deep Architectures for Image Tracking”

[Cho+2015] Kyunghyun Cho, Aaron Courville, Yoshua Bengio, “Describing Multimedia Content using Attention-based Encoder–Decoder Networks”.

[Rush+2015] Alexander M. Rush, Sumit Chopra, Jason Weston, “A Neural Attention Model for Abstractive Sentence Summarization”

[Ling+2015] Wang Ling, Isabel Trancoso, Chris Dyer, Alan W Black, “Character-based Neural Machine Translation”.

[Vinyals+2014] Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, Geoffrey Hinton, “Grammar as a Foreign Language”

[Shang+2015] Lifeng Shang, Zhengdong Lu, Hang Li , “Neural Responding Machine for Short-Text Conversation”

[Hermann+15] Karl Moritz Hermann, Tomáš Kočiský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, Phil Blunsom, “Teaching Machines to Read and Comprehend”

[Vinyals+2015] Oriol Vinyals, Meire Fortunato, Navdeep Jaitly, “Pointer Networks”

[Xu+2015] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio, “Show, Attend and Tell: Neural Image Caption Generation with Visual Attention”**.**

[Vinyals+2015b] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan, “Show and Tell: A Neural Image Caption Generator”

[Mansimov+2016] Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba, Ruslan Salakhutdinov, “Generating Images from Captions with Attention”

[Meng+2016] Fandong Meng, Zhengdong Lu, Zhaopeng Tu, Hang Li, Qun Liu , “A Deep Memory-based Architecture for Sequence-to-Sequence Learning”.

[Sukhbaatar+2015] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus, “End-To-End Memory Networks”..

[Graves+2014] Alex Graves, Greg Wayne, Ivo Danihelka , “Neural Turing Machines”.

[Tran+2016] Ke Tran, Arianna Bisazza, Christof Monz, “Recurrent Memory Network for Language Modeling”.