

RNN 系列変換モデルを用いた 高階論理からの文生成

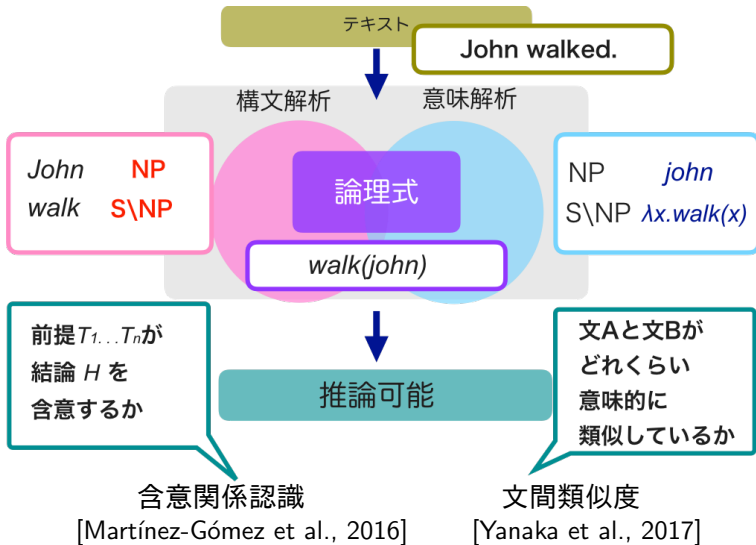
馬目 華奈

お茶の水女子大学 戸次研究室

言語処理学会第 24 回年次大会

March 13, 2018

はじめに

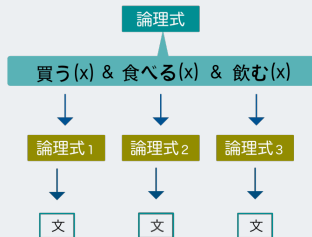


意味表示に論理式を採用する利点

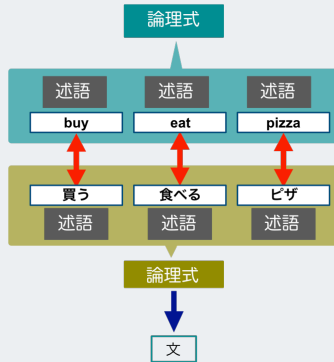
- 従来の係り受け解析や述語項構造解析を超える深い意味解析が可能
- 副詞や前置詞などは、イベントを項に持つ述語として扱うため、修飾表現を含む文も簡潔に扱える
- 量化や否定などの論理推論，モダリティや内包的述語など複雑な意味を扱える
- 述語論理の論理式に変換し，推論ができる

論理式を意味表示にした場合の応用例

テキスト平易化

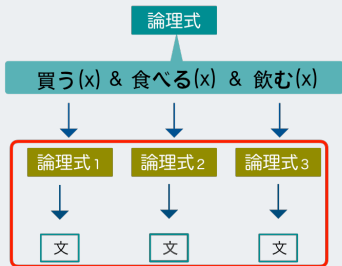


翻訳

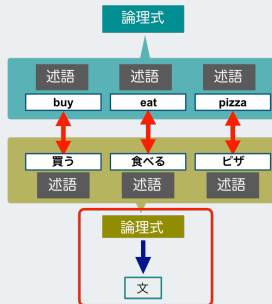


研究の目的

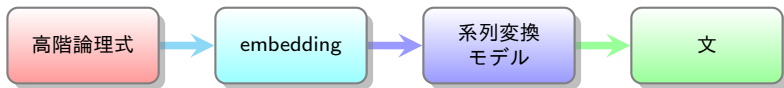
テキスト平易化



翻訳

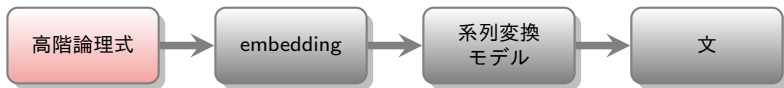


研究概要



- エンコーダ・デコーダを用いて高階論理式から文を生成する手法を提案.
- embedding の際, 4 種の手法を検討する.
(記号, トークン, 木構造, グラフ)

文から高階論理式への変換



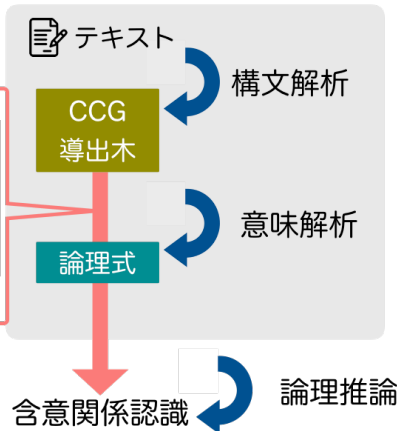
- エンコーダ・デコーダを用いて
高階論理式から文を生成する手法を提案.
- embedding の際, 4 種の手法を検討する.
(記号, トークン, 木構造, グラフ)

ccg2lambda を用いた論理式変換

Mineshima+ [EMNLP2015]

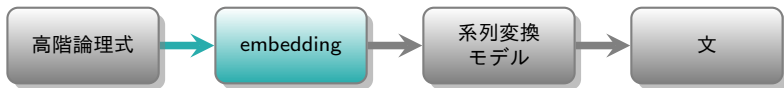
Martínez-Goméz+ [ACL2016]

	$\frac{\text{ate}}{(S \backslash NP) / NP}$	$\frac{\text{Pizza}}{NP}$	
Bob	$\lambda y \lambda x \exists e. \text{eat}(e, x, y)$	pizza	>
NP	$S \backslash NP$		
bob	$\lambda x. \exists e. \text{eat}(e, x, \text{pizza})$		<
	S		
	$\exists e. \text{eat}(e, \text{bob}, \text{pizza})$		



<https://github.com/mynlp/ccg2lambda>

embedding 手法の提案



- エンコーダ・デコーダを用いて高階論理式から文を生成する手法を提案.
- embedding の際, 4 種の手法 (記号, トークン, 木構造, グラフ) を検討する.

提案手法：記号区切りとトークン区切り

Bob walked の論理式：

exists x.(Bob (x) & exists e.(walk(e) & (Subj(e) = x)))

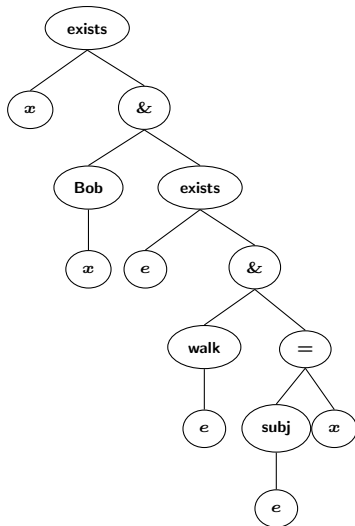
1：記号ごとに区切る

[e,x,i,s,t,s, , x, ., (, B,o,b, , (, x,), ...]

2：トークンごとに区切る

[exists, x, (, Bob, (, x,), &, exists, e, (, ...]

提案手法：木構造のベクトル化

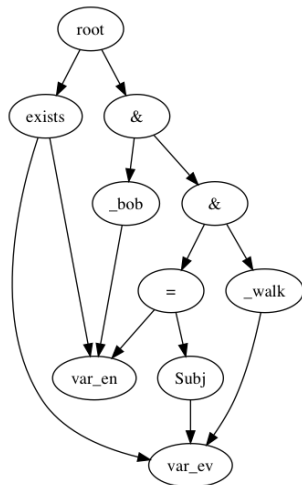


木構造を用いたベクトル

[exists, x, &, Bob, x, exists, e, &, ...]

- 論理式をポーランド記法に変換する
(論理演算子を前にもってくる)
- pre-order の深さ優先探索でたどる

提案手法：グラフ構造のベクトル化

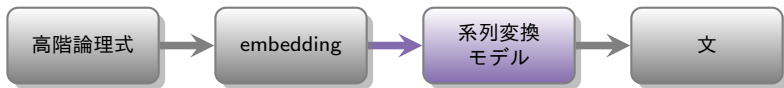


グラフ構造を用いたベクトル

[exists, e, x, &, **Bob**, &, =
, **Subj**, walk]

- 論理式をポーランド記法に変換する
(論理演算子を前にもってくる)
- 同じ変数をさす様に edge を変更
- pre-order の深さ優先探索でたどる

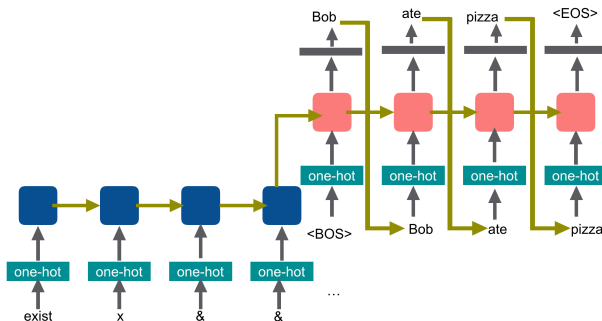
エンコーダ・デコーダ



- エンコーダ・デコーダを用いて高階論理式から文を生成する手法を提案.
- embedding の際, 4 種の手法を検討する.
(記号, トークン, 木構造, グラフ)

系列変換モデル [Sutskever et al., 2014]

- 入出力がシーケンスとなるニューラルネットのモデル
- エンコーダ：入力列を RNN により隠れ状態ベクトルに変換
- デコーダ：隠れ状態ベクトルを初期値とし、
隠れ状態と自身のこれまでの出力結果をもとに
次のトークンを生成



データセットの作成

- SNLI(The Stanford Natural Language Inference Corpus) [Bowman et al., 2015] を用い論理式と文のペアを作成
- 60 単語以内の文例を対象 train:9140/dev:2285/test:1500
- ccg2lambda を用いる

	<u>ate</u>	<u>Pizza</u>	
	$(S \backslash NP) / NP$	NP	
<u>Bob</u>	$\lambda y \lambda x \exists e. eat(e, x, y)$	<u>pizza</u>	
<u>NP</u>	$S \backslash NP$		>
<u>bob</u>	$\lambda x. \exists e. eat(e, x, pizza)$		
	<u>S</u>		<
	$\exists e. eat(e, bob, pizza)$		

テキスト(文)

高階論理式

実験設定

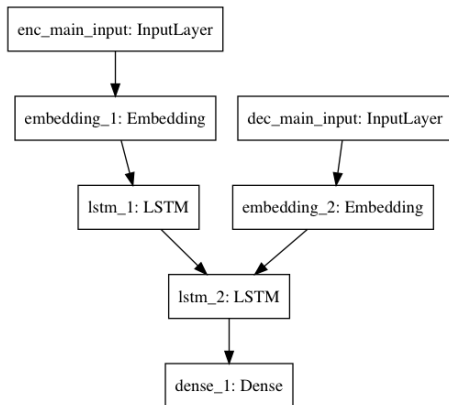
- 系列変換モデルによる文生成（入力：論理式，出力：文）
- トークン区切りの次元数を 256 次元に

	記号	トークン	木構造	グラフ
入力語数	70	5,118	5,107	4,991
出力語数	78	7,214	7,214	7,214
最長入力列	2,097	699	451	259
最長出力列	270	55	53	53

環境，ライブラリ

- tsubame サーバ（メモリ 240GiB, GPU × 4）
- python3 系ライブラリ（Keras, nltk）

モデルの詳細



学習モデル

embedding_1 → 256 次元
embedding_2 → 256 次元
lstm_1 → 256 次元
lstm_2 → 256 次元
dense_1(softmax) → 出力語数

評価方法

BLEU による評価

$$score = BP \exp \left(\sum_{i=1}^N \frac{1}{N} \log P_n \right)$$

$$BP = \begin{cases} 1 & (c \geq r) \\ \exp \left(1 - \frac{r}{c} \right) & (c < r) \end{cases}$$

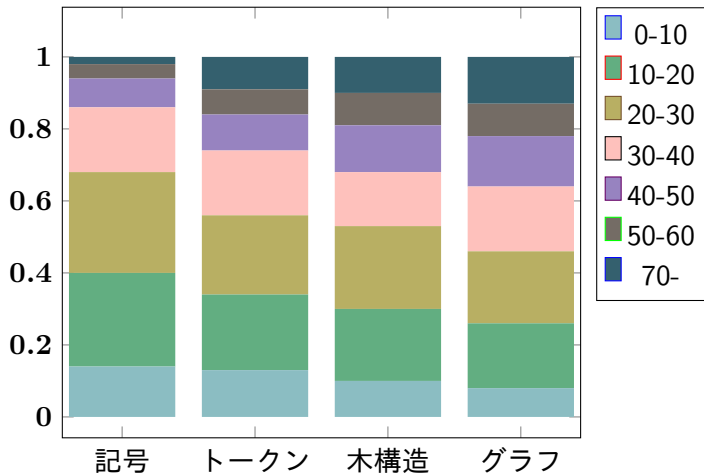
$$P_n = \frac{\sum_{i=0} \text{出力文 } i \text{ 中と解答文 } i \text{ 中で一致した } n\text{-gram 数}}{\sum_{i=0} \text{出力文 } i \text{ 中の全 } n\text{-gram 数}}$$

実験結果

BLEU 評価

指標	記号	トークン	木構造	グラフ
BLEU	34.9	39.7	41.8	44.7

BLEU スコア割合



出力結果 1

入力論理式

exists $x.(\text{_boy}(x) \ \& \ \text{exists } z.(\text{_football}(z) \ \& \ \text{exists } e.(\text{_play}(e) \ \& \ (\text{Subj}(e) = x) \ \& \ (\text{Acc}(e) = z))))$

正解文

boys play football .

出力結果

記号単位	The boy is wearing a blue shirt.
トークン単位	The boy is playing basketball .
木構造	The boys are playing sports .
グラフ	The boy is playing football .

出力結果 2

入力論理式

exists x.(_**woman**(x) & _**two**(x) & **exists** z.(_**package**(z) & **exists** e.(_**hold**(e) & (**Subj**(e) = x) & (**Acc**(e) = z))))

正解文

Two woman are holding packages .

出力結果

記号単位	Two women are	watching the paper.
トークン単位	Two women are holding	flowers .
木構造	Two women are holding	something .
グラフ	Two women are holding	a camera .

出力結果 3

入力論理式

exists x.(_man(x) & **exists** z01.(**exists** e.(_blue(e) & (Subj(e) = z01) & _standing(z01) & _shirt(z01)) & exists z02.(_front(z02) & exists z00. ...

正解文

A man in a blue shirt standing in front of a garage-like structure painted with geometric designs.

出力結果

記 A man in a blue shirt and a black shirt and blue jeans is walking ...
 ト A man in a blue shirt and a woman in a red dress holding up a small boy.
 木 man in a blue shirt with a beard is standing on a computer while a.
 グ A man in a blue shirt standing next to a cement mixer with one in a purple.

まとめ

- エンコーダ・デコーダを用いて高階論理式から文を生成する手法を提案した.
- 実験をする際, ccg2lambda を用いてデータセットを作成した.
- 提案手法の評価を行った結果, BLEU スコアは, トークン単位で区切り, 論理式の計算の順序を考慮することで高くなった.

今後の課題

- 他の意味表現を用いた文生成との比較を行う.
- 評価方法に, 文類似度を使用するなど評価方法を工夫する.
- 論理式の述語中に出てくる単語を出力するマスキング処理をかける
- 論理式の embedding を改良する

参考文献 I

- Lasha Abzianidze. A Tableau Prover for Natural Logic and Language. In *Proc. of EMNLP*, 2015.
- Daisuke Bekki. *A Formal Theory of Japanese Grammar: The Conjugation System, Syntactic Structures, and Semantic Composition*. Kuroshio, 2010. (In Japanese).
- Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. A large annotated corpus for learning natural language inference. In *Proc. of EMNLP*, 2015.
- Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. Neural AMR: Sequence-to-Sequence Models for Parsing and Generation. In *Proc. of ACL*, 2017.
- Pascual Martínez-Gómez, Koji Mineshima, Yusuke Miyao, and Daisuke Bekki. ccg2lambda: A Compositional Semantics System. In *Proc. of ACL System Demonstrations*, 2016.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of ACL*, 2002.
- Mark Steedman. Surface Structure and Interpretation. In *The MIT Press*, 1996.

参考文献 II

- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to Sequence Learning with Neural Networks. In *Proc. of NIPS*, 2014.
- Mingzhe Wang, Yihe Tang, Jian Wang, and Jia Deng. Premise selection for theorem proving by deep graph embedding. In *Proc. of NIPS*, 2017.
- Hitomi Yanaka, Koji Mineshima, Pascual Martínez-Gómez, and Daisuke Bekki. Determining Semantic Textual Similarity using Natural Deduction Proofs. In *Proc. of EMNLP*, 2017.