

Recurrent Neural Network Grammers

知能数理研究室 外山洋太，三輪誠，佐々木裕

導入

Recurrent Neural Network Grammars (RNNGs)

- ▶ 文の確率的生成モデル
 - ▶ 単語や句の入れ子的・階層的構造を陽に表現
- ▶ 問題：構文解析，文生成
- ▶ 動機：Sequential な Recurrent Neural Networks (RNNs) は自然言語の潜在的な入れ子構造を考慮できていない
- ▶ 構文解析（文 \rightarrow 構文木）と文生成のアルゴリズム

提案手法

RNNG の定義

$$RNNG := (N, \Sigma, \Theta)$$

$$N \cup \Sigma = \emptyset$$

$$\begin{cases} N : \text{非終端記号の有限集合} \\ \Sigma : \text{終端記号の有限集合} \\ \Theta : \text{NN のパラメータ} \end{cases}$$

提案手法

構文解析のアルゴリズム

$$f : X \rightarrow Y$$

$$\left\{ \begin{array}{l} x : \text{単語列 (入力)} \\ y : \text{構文木 (出力)} \\ S : \text{スタック} \\ B : \text{入力バッファ} \end{array} \right.$$

- ▶ 入力バッファの要素：まだ使われていない終端記号
- ▶ スタックの要素：終端記号, 未完成の非終端記号, 完成した非終端記号

提案手法

構文解析のアルゴリズム

▶ 遷移の制約

- ▶ n : スタック内の未完成の非終端記号の数

遷移	制約
NT(X)	$B \neq \emptyset \wedge n < 100$
SHIFT	$B \neq \emptyset \wedge n \geq 1$
REDUCE	スタック内の一番上の要素が 未完成の非終端記号でない $\wedge (n \geq 2 \vee B = \emptyset)$

提案手法

生成のアルゴリズム

▶ 遷移の制約

- ▶ n : スタック内の未完成の非終端記号の数

遷移	制約
GEN(X)	$n \geq 1$
REDUCE	スタック内の一番上の要素が 未完成の非終端記号でない $\wedge n \geq 1$