

第 1 章

Attention 付き Seq2Seq モデル

本モデルは2つのゲート付き回帰型ユニット (GRU) で構成されるニューラル機械翻訳のモデルである。一方の GRU は入力系列を一つの固定長ベクトルに符号化 (encode) し、もう一方の GRU により固定長ベクトル符号を出力の系列へと複合化 (decode) する。また、本モデルは出力系列を複合化する際に、出力系列と入力系列の参照を行う Attention 機構を備えており、これによってより長い分における精度向上を実現している。モデルは原文を目的文に翻訳する際の条件付き確率を最大化するためのパラメータを、原文-目的文の対訳コーパスから学習する。

$\text{Attention}c_t$ は Encoder 側の隠れ変数 \bar{h}_s と Decoder 側の隠れ変数 h_t で定義される。原文を x 、目的文を y とする時、 $\text{Attention}c_t$ を使って次に翻訳される単語の確率を以下の式で求めている。

$$p(y_t \mid y_1, y_2, \dots, y_{t-1}, x) = \text{softmax}(W_s \cdot \tilde{h}_s)$$

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

$$c_t = \sum_{s=1}^S a(s) \bar{h}_s$$

$$a_t(s) = \text{align}(h_t, \bar{h}_s) = \frac{\exp(\text{score}(h_t, \bar{h}_s))}{\sum_s' \exp(\text{score}(h_t, \bar{h}_s'))}$$

$$\text{score}(h_t, \bar{h}_s) = v_a^T \tanh(W_a(h_t; \bar{h}_s))$$

\bar{h}_s : *Encoder* 側の隠れ変数

h_t : *Decoder* 側の隠れ変数

$a_t(s)$: 両側の隠れ変数の荷重比率

W_s, W_c, W_a, v_a : 学習パラメータ

$(a; b)$: ベクトル a と b の連結