# 情報工学英語演習 「AttentionIsAllYouNeed」の和訳


## Abstruct

今までの主要な sequence transduction model は, エンコーダーとデコーダーを含む複雑な RNN や CNN に基づいていた. 最も良いパフォーマンスであったモデルも, エンコーダーとデコーダーをアテンションの仕組みを用いて接続しているモデルであった.

私達は,RNN や CNN を用いず, アテンションのみからなる Transformer と呼ばれる新しい簡潔なモデルを提案する.2 回の翻訳実験の結果,Transformer は今までのモデルより優れており, 更には, より並列化が可能で学習の時間も少ないことが分かった. Transformer は WMT2014 英独翻訳タスクにおいて,「プロの翻訳者の訳と近ければ近いほどその機械翻訳の精度は高い」という考え方に基づく機械翻訳の評価方法である BLEU スコアで 28.4BLEU を記録した. これは, 複数のモデルを融合させて 1 つの学習モデルを生成するアンサンブル学習を含めたこれまでの最高記録を 2BLEU 上回る結果であった. また、WMT2014 英仏翻訳タスクにおいては、8 個の GPU を用いた 3.5 日の学習というこれまでの最先端のモデルの学習よりも遥かに少ないコストで,41.0BLEU という単一モデルの最高記録を打ち立てた.


## 1 Introduction

RNN, 特に RNN において文章の長期的な依存関係を学習できるようにした LSTM や gated RNN は, 言語モデルや機械翻訳などの Sequence 問題への最適な手法として確固たる地位を築いていた. それ以来,Recurrent 言語モデルとエンコーダー-デコーダー構造の限界を押し上げる数々の努力がなされてきた. リカレントモデルでは, 通常, 入力と出力の時系列データの位置に沿って計算を行う.