

ニューラル機械翻訳の衝撃

鶴岡慶雅 (東京大学)

ニューラル機械翻訳

ニューラルネットワークを使って機械翻訳をするという一見突拍子もないアイデアは、コミュニケーションにおける「言葉の壁」の在り方を大きく変えることになるかもしれない。専門家の間では近年そのポテンシャルの高さからニューラル機械翻訳 (Neural Machine Translation, NMT) に大きな注目が集まっていたが、Google がニューラル機械翻訳に基づいたシステムを一般に公開したことにより、その優秀な翻訳性能が多くの人に知られることとなった。

最近公開された Google の機械翻訳システムに関する論文¹⁾ では、人手評価による翻訳精度が、従来手法であるフレーズベース機械翻訳と比べて、誤り率にして平均で 60% 減少したと報告されている。言語対によっては、人間の翻訳精度に迫るスコアを達成しており、機械翻訳システムは当分の間実用にならないという見方は完全に過去のものとなった。

翻訳性能以外にも、多言語翻訳の同時学習によって言語横断的なある種の共通意味表現 (interlingua) が自然に得られたり、学習コーパスには存在しない言語ペアの翻訳 (ゼロショット翻訳) が可能であることが示される²⁾ など、ニューラル機械翻訳ならではの興味深い研究成果が次々と報告されている。

誕生

ニューラル機械翻訳を実現するための重要な要素技術がリカレントニューラルネットワークと呼ば

れる機械学習モデルである。これは、通常が多層ニューラルネットワークと異なり、入力があるたびに変化する「状態」を保持しているため、同じ入力に対しても異なる出力を出すことが可能になっている。たとえば入力として、文章中の単語を順次受け取るようにすれば、「状態」によってそれまでの文脈を実数値ベクトルで表現し、次に出現するであろう単語を予測するようなモデル (言語モデル) を容易に作るができる。

現在のニューラル機械翻訳技術のベースとなったエンコーダ・デコーダモデル³⁾ では、2つのリカレントニューラルネットワークを使用する。1つはエンコーダと呼ばれ、翻訳元の文の単語を順次読み込み、文全体の内容を表す実数値ベクトルを生成する役割を担う。もう1つのリカレントニューラルネットワークはデコーダと呼ばれ、エンコーダが出力した実数値ベクトルを入力とし、上述の言語モデルとほぼ同様の仕組みで翻訳先の単語列を出力する。文全体の内容を1つの実数値ベクトルで表現してしまおうというのはいかにも乱暴に見えるが、このような単純な仕組みでそれなりに高い翻訳性能が得られたことに多くの研究者が驚いた。

発展

ニューラル機械翻訳のレベルを大きく引き上げたのは、アテンション (attention) と呼ばれる仕組みである⁴⁾。これは、デコーダが各単語を出力する際の情報として、エンコーダにおける各単語の隠れ状態の重み付き平均を入力として用いるという方法で、

翻訳元の文の文脈情報をより詳細に捉えることができるようになる。アテンションの機構が導入されたことによってニューラル機械翻訳の精度は大きく向上し、従来の統計的機械翻訳モデルに匹敵、あるいはその性能を追い越すこととなった。

当初ニューラル機械翻訳は低頻度語に弱いという弱点があったが、その弱点もさまざまな手法によって克服されつつある。単語レベルの翻訳だけでなく、文字ベースの翻訳や、サブワードと呼ばれる、単語よりも小さな単位を併用することで、現在のニューラル機械翻訳システムは計算コストを大きく増やすことなく低頻度語に対する頑健性を向上させている。

大きな注目を集めているニューラル機械翻訳モデルの性能を陰で支えているのは、モデルの学習に使われるパラレルコーパスである。パラレルコーパスとは、文同士の翻訳関係の対応がついている2言語のコーパスである。たとえば、機械翻訳モデルの学習・評価用データとしてよく用いられる WMT データセットでは、英仏では約 3,600 万文ペア、英独では約 500 万文ペアが提供されている。学習データの量が多ければ多いほど翻訳精度が向上するというのは従来の統計的機械翻訳と同様だが、ニューラル機械翻訳の場合は特にその傾向が強いと言われている。Google の翻訳システムでは、それらの標準的なコーパスの 10 倍から 100 倍以上の大きさのパラレルコーパスが学習に使われたとされている。パラレルコーパスの巨大さもさることながら、それだけの量のデータを使って実際にモデルを学習できるだけのハードウェアリソースにも驚かざるを得ない。

今後

もちろん最先端のニューラル機械翻訳システムとて完璧な翻訳ができるわけではない。文の意味を完全に取っ違えた翻訳文が生成されることもあるし、学習コーパスに含まれていないような種類のテキストでは翻訳精度が大きく落ちるといったほかの機械学習ベースのシステムと同様である。また、従来型の翻訳システムと比べて、翻訳誤りの原因を追究することが難しいという点も、システムの改良を難しくする要因になり得るかもしれない。

とはいえ、ニューラル機械翻訳は、それを実現するための仕組みが（少なくとも概念的には）非常にシンプルであるという大きなメリットがあり、これまで機械翻訳に関する研究の経験のなかった研究者や研究グループも機械翻訳の分野に続々と参入している。研究は急ピッチで進んでおり、当面ニューラル機械翻訳の進化は止まりそうにない。

参考文献

- 1) Wu, Y. et al. : Google's Neural Machine Translation System : Bridging the Gap between Human and Machine Translation, arXiv:1609.08144 (2016).
- 2) Johnson, M. et al. : Google's Multilingual Neural Machine Translation System : Enabling Zero-Shot Translation, arXiv:1611.04558 (2016).
- 3) Sutskever, I. et al. : Sequence to Sequence Learning with Neural Networks, In Advances in Neural Information Processing Systems, pp.3104-3112 (2014).
- 4) Bahdanau, D., Cho, K. and Bengio, Y. : Neural Machine Translation by Jointly Learning to Align and Translate, In International Conference on Learning Representations (2015). (2016年12月12日受付)

鶴岡慶雅（正会員） tsuruoka@logos.t.u-tokyo.ac.jp

2002年東京大学大学院博士課程修了。博士（工学）。マンチェスター大学研究員などを経て、2009年北陸先端科学技術大学院大学准教授。2011年より東京大学大学院工学系研究科 准教授。自然言語処理、ゲーム AI 等に関する研究に従事。