

自然言語処理における知識とは何か – 深層学習による機械翻訳

小町守

首都大学東京システムデザイン学部

本ワークショップの小町担当部分では、深層学習を用いた機械翻訳を例にとって、自然言語処理の哲学とは何であるかについての議論を深めることを目的とする。

まず自然言語処理の歴史を概観し、アノテーションを中心とした機械学習（教師あり学習）を用いた自然言語処理のパラダイムについて述べる。アノテーションとは言語テキストに何らかの注釈付けをすることであり、アノテーションされたコーパス（言語資源）から自動的に辞書やルールを抽出する、というアプローチが現在主流となっている。

次に、この 4-5 年急速に研究開発が進んでいる機械学習の一手法である深層学習がなぜ自然言語処理にとって重要であるかを説明する。深層学習以前と深層学習以後で教師あり学習のフレームワークはほとんど変わっていないが、自然言語処理においてはこの 10 年の間、表現学習および系列学習の 2 点において大きな展開が起きており、それ以前の自然言語処理と意味処理において大きな進展があったことを明らかにする。また、深層学習自身は突然出現した手法ではなく、いくつかの鍵となる手法が発見・成熟したことによって可能となった手法であり、それらが登場する背景について説明する。

そして機械翻訳を例に取り、対訳コーパスから深層学習によって翻訳モデルを学習するニューラル機械翻訳について、その射程と限界を確認する。機械翻訳自身がこの 20 年で統計的機械翻訳からニューラル機械翻訳へと大きく進化しているが、深層学習を用いることで何ができるようになり、また何が原理的に実現不可能であるかを確認した上で、自然言語処理におけるアノテーションがなぜ哲学の問題となるのかについて述べる。

最後に、自然言語処理の哲学がなぜ必要かについて論じる。自然言語処理は機械翻訳をはじめとしてウェブ検索システム、日本語入力など様々なアプリケーションで成功を収めているが、これらのタスクではなぜうまく行くのか、うまく行かないタスクは何であるかの認識を共有した上で、自然言語処理における哲学について議論する。