

AI 自習ドリル

第25回 ChatGPTで注目の自然言語処理に詳しくなる… 翻訳AIづくり

牧野 浩二

1 できること

AIで注目される技術として、翻訳サービスのDeepLのような翻訳や、ChatGPTのような会話があります。これらは自然言語処理という分野の発展形です。本来このような技術には、膨大なデータと学習時間が必要です。

TensorFlowのウェブ・ページにはチュートリアルが幾つかあります。今回は「アテンションを用いたニューラル機械翻訳」というチュートリアル⁽¹⁾のノー

トブック(nmt_with_attention.ipynb)を改変したものを使って、学習時間が30分程度で翻訳することのできるAIを作成します。自然言語処理に必要な技術を学ぶことができますでしょう。

今回扱う翻訳技術は、自然言語処理という技術分野に含まれます。自然言語処理とは人間が使う言葉を機械的に処理して、内容(意味)を解析する処理技術です。

2 イメージでつかむ

今回は翻訳を対象として自然言語処理を学びます。本章では、どのようなところが難しいのか、それをどのように解決しているのかを説明します。

自然言語処理の難しさ

● 言語によって語順が異なる

人間が行う翻訳を例に説明します。日本語を母国語とする人が英語を学ぶときの最大の難関は、辞書で単語の意味を調べることはできても、語順の並び替えができないといった問題です。

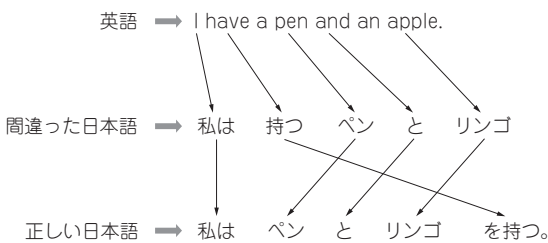


図2-1 言語による語順の違い

図2-1の例で考えてみます。単語単位でそのまま置き換えると間違った日本語になってしまいます。正しく翻訳するには順番を入れ替える必要があります。この例のように、文章全体を考えてしっかりした順番で処理しなければ正しい答えが得られません。

● 前後の単語との結びつき

人間は文章の意味を理解しながら単語の予測ができますが、AIではそれが難しいです。ここでは、以下の例文を考えます。

I eat an orange and an apple.

翻訳すると以下ようになります。

私はミカンとリンゴを食べる。

このように英語では、食べる(eat)に続く単語はほとんどが食べ物となります。以下のような文章にはなりません。

I eat a pen and an apple.

私はペンとリンゴを食べる。

つまり、単語には前後の結びつきというものが重要です。