Abstract geometric lines in the top left corner, consisting of several overlapping, irregular polygons and lines in a light beige color.

単語 → 文生成

2023/05/23

背景と目的

- なぜ、このテーマを選んだのか？

文章生成が面白い



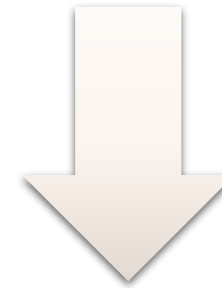
単語から文を生成できないか？

- 今回のプロジェクトの説明

単語を入力し、文を生成する

• 入力と出力

入力例：男 スキー 山



出力例：男が山でスキーをする

概要

- 使用した技術

Seq2seq+Attention

- 学習に使用したデータセット

HuggingfaceのCommon_Gen



概念から文章

[common_gen · Datasets at Hugging Face](#)

詳細

- 工夫した点

データセットの中身が英語だったので、全て翻訳した

詳細

- 学習に工夫した点

Amazon SaGeMaker Studio Lab



GPUのメモリを16Gまで使用可能
(ほぼ、Google Colabと同じ)

- 利点

ベクトルのサイズを大きくできる

結果

「単語 → 文章生成」プログラム

入力

～単語を入力して下さい(10単語まで)～

男 スキー 山

文章生成

出力

スキーで雪山を下る2人の男。

単語の特徴をとらえている文の生成に
成功

「単語 → 文章生成」プログラム

入力

～単語を入力して下さい(10単語まで)～

山 スキー 男

文章生成

出力

何人かの人々が山を滑り降りてサーフィンをしています。

森 山 川
入力: 森 山 川
出力↓
川の森にある谷と小さな青い朝雲がある森の空撮
山の間の峡谷にある深い緑の森に氷の小さな雲がある空中に撮影
底の秋の森に囲まれた小さな川のある川
川に森をゆっくりとする小さな谷の日の出のタイムラプス
川の森の美しい雲
遠くに山のあるある風の強い田園地帯のピンクの葉の接写
川が見える森の上の森の空撮
山の間の森に流れ込む小さな深い川のある町
川の近くの森の空撮
山のある山々に流れ込む小さな小川のタイムラプス

結果

単語の位置や別の単語を使用すると、文的に正しくない。単語をすべて含む文は稀にしか出ない

<u>ワグ</u> 尾 犬	犬は尻尾を振っています。
<u>ワグ</u> 尾 犬	犬が少年に尻尾を振る。

<u>s</u> テーション <u>トレー</u> ニング <u>プル</u>	駅まで電車が 入る。
駅 電車 プル	駅に 停車 する 列車

ファン <u>マッチ</u> ウォッチ	ファンと一緒に試合を観戦する	
ファン マッチ ウォッチ	試合を観戦するファン	
ファン マッチ ウォッチ	試合中に見るファン	

原因

データセットの翻訳の質が悪く、
データセットの質が悪い状態での学習

解決策

学習に使用するデータセットを翻訳せずに英語のまま学習させる

学習時に入力に使う単語をシャフルする

データセットの量を増やす

まとめ

- テーマ

単語から文を作る

- 結論

ある単語には強く、学習データが少ない単語には弱いモデルができた。