

LSTMを用いた作文の文章生成

T190461 浅羽 隆一郎

指導教員 三好 力 教授

1. はじめに

作文作成の一助にすべく自然言語処理に用いられるRNN(Recurrent Neural Network、回帰型ニューラルネットワーク)から勾配消失を改良したLSTM(Long Short-Term Memory)を用いて題名を入力することで自動で文章を生成するシステムを構築すること、さらにそれを例題として活用できるまでに質を高めることを目標としている。このシステムが実現すれば教育機関における国語学習の補助や、日本語の習得の手助けになるのではないかと考える。

2. 提案手法

文章を単語分けし、各単語にIDを振り、コーパスを作る。Encoderとして題名のコーパスを入力することで、Decoderとして本文のコーパス出力するよう学習させる。全体の概略図を図1に示す。 t が題名のコーパス、 w が本文のコーパス、 y が出力値である。

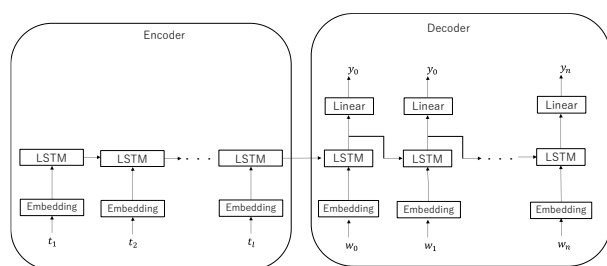


図1 全体の概略図

3. 実験結果

図1のような手法はSeq2Seqと呼ばれる。学習データとしてWeb上に掲載している作文を引用した。一般的なSeq2SeqとEncoderの出力値をDecoderのすべての入力値に連結したPeekyと呼ばれる手法とを比較し、実験を行った。その結果、一般的なSeq2Seqの方が単語や同じような文章が続くことがなく文法的に正

しい点が比較的多くみられることが分かった。このことから一般的なSeq2Seqの方がやや質が高い文章を生成することができるように思われる。一般的なSeq2Seqを使って学習させ生成した文章を下に記す。

明るい社会になる...最近、いっしょばんていた。あなたの後世の後世だめは、いっしょばんていた。あなたの後世辛かついまサラダかった。あなたの後世の後世だめは、いっしょばんていた。あなたの後世辛かつ可哀想どんどん弱い成績訴えは、いっしょばんいっしょ幸福ていた。あな たの後世の後世だめは、いっしょばんすわっていた。あなたの後世辛かつ栄養念願すうた。あなたの後世の後世だめは、いっしょばんていた。あなたの後世辛かついま、いっしょばんのみ弱い。あなたの後世の後世の後世だめは、いっしょばん幸福ていた。あなたの後世辛かつすうた。あな たの後世の後世の後世だめは、いっしょばんいっしょ幸福ていた。あなたの後世辛かついまサラダサラダかった。あなたの後世辛かつ栄養だめ親切な親切なのは、

4. おわりに

全体として文法的には正しいが、意味的には成り立たず、実用できるまでの段階には至らなかった。今後の課題として、自動で作文を評価するシステムを作り、強化学習に用いられるDQNを用いれば改善されるのではないかなと思われる。また、Encoderの重要なデータだけを自動で強調するAttentionという技術があるため、それを用いれば多少の改善が見られるのではないかと考えられる。

参考文献

- [1] ゼロから作るDeep Learning ② ―自然言語処理編 斎藤 康毅 2018/7/21