

画像

第4章

人間のアシスタントとして一大分野に
発展するかも

データ

音

データ分析向き RNN ②…

日本語文の自動生成

ご購入はこちら

牧野 浩二, 西崎 博光

リスト1 リカレント・ニューラル・ネットワークによる自動作文…それなりに意味の通る文章になっている？

```
$ python make_sentence.py -m model.10 -v vocab
```

```
単語を入力>>> 吾輩
自動作文結果: 吾輩 は また 主人 の 顔 を 見 て <unk> 主人 は 「 先生 」 と 云 う 。
```

(a) 自動作文の実行例

```
単語を入力>>> 我
自動作文結果: 我 は 猫 で ある 。
単語を入力>>> 猫
自動作文結果: 猫 の もの は なかなか ない 。
単語を入力>>> 僕
自動作文結果: 僕 は どう し て い る の か ？
単語を入力>>> 机
自動作文結果: 机 の 上 に は あ る が <unk> この 時 は 吾輩 の こ と く <unk> この 時 に は <unk> その 時 に は <unk> その 時 に は <unk>
単語を入力>>> 時計
自動作文結果: 時計 は 無 論 <unk> 主 人 は ま た 一 人 に な っ た 。
単語を入力>>> 主人
自動作文結果: 主 人 は ま た 主 人 の 顔 を し て い る 。
```

(b) さまざまな作文結果

● 今回やること&課題

ここでは、コンピュータに文章を生成させてみます。今回のサンプルは日本語の作文ができるようになっていきます。

第1部で示したリカレント・ニューラル・ネットワークの初期型では、過去の情報がどんどん薄まっていきます。それはそれで良いこともありますが、次のような場合はうまく答えることができなくなります。

「修学旅行で京都に行ったときに印象に残ったのは、そこに住む人たちの温かさ、歴史ある街並みで、中でも○○のすばらしさに圧倒されました。」

これを読むと、○○に入るのは「金閣寺」や「清水寺」と推測できます。しかし、京都という単語が離れすぎているため、初期型のリカレント・ニューラル・

表1 自動作文で使用するプログラム(ディレクトリ名: sakubun_RNN)

使用プログラム	ファイル名
学習用	train_lm.py
作文用	make_sentence.py
変換用	seikei.pl

ネットワークでは○○をうまく答えることができなくなります。

● 進化したRNN…LSTM

そこで、以下の2つを実現できる仕組みが必要です。

- 重要そうな単語はいつまでも覚えていること
- 重要でない単語はすぐに忘れること

これを実装したものをLSTM(Long Short Term Memory)と呼びます。そして、Chainerではこの2点をブラック・ボックスの処理に入れ込んで使いやすくなっています。

実験

● できること…単語を入力するだけで文章を出力

作文とはどのようなものかイメージがわからないかもしれません。そこで、筆者らが事前に学習させておいた「吾輩は猫である」のモデルを使って実行してみます。

実行結果をリスト1に示します。実行すると最初に「単語を入力」と表示されます。そこで、「吾輩」のように単語を入力すると、それで始まる文章が表示され