

計算機で川柳を書いてみた

芝浦工業大学：澤崎航輔

2017年11月3日

研究動機

前期の授業でテキストマイニングを行いそこから自然言語処理に興味を持ち始めた。その後、先輩と食事に行ったとき俳句の話聞き、書いてみたくなったのでやってみることにした。

1 概要

川柳のルールは簡単である。5・7・5のように読めば良い。例として

:数理研
:ふたを開ければ
:だらけてる

のような要領である。もちろん字数の制限をオーバーしたりする字余り・字足らずのようなものもある。また、川柳と俳句の違いは季語(季節を表現する言葉)が含まれているか・いないかであり、季語が含まれないものが川柳となる。今回はこの川柳を自動的に生成することを目標とする。以下の生成手順をプログラミング言語のPythonを用いて行う。

2 生成手順

2.1 キーワード入力

川柳を生成する場合のキーワードを1つ以上入力する。例えばコンビニのおにぎりについての川柳を書きたい場合は「コンビニ」、「おにぎり」とキーワードを2つ入力しておく。

2.2 単語の抽出

あらかじめ用意しておいた文章(コーパス)から語句を抽出する。その後、キーワードとの類似性を比較し、有用な語句を抽出する。また、抽出した語句をさらに名詞・動詞・形用紙・副詞などの品詞ごとに分けておく。

2.3 文法定義

次に語句の順番を決定するような文法を定義しておく。例えば「名詞」、「形容詞+名詞」などは正しいので使用する文法として記憶しておき、「名詞+助動詞」などの間違っている文法については使用しないので記憶しない。

2.4 部分的生成

語句の抽出と文法の生成が終わった後は5・7・5の各部分に該当する部分を生成する。この生成でもキーワードとの比較を実行しながら生成し、2つ以上の語句を使用する場合はそれら同士の類似性も比較する。例えば「美味しい+おにぎり」なら意味的にOKなので評価値は高い。逆に「美味しい電卓」となると電卓は基本的に食べれないので変な文章であるので評価値は低い。

2.5 結合

最後に生成された5・7・5の部分をすべて結合する。その中で全体の調和がとれていてキーワードとの類似性が高いものを実際に川柳として生成する。例えば「コンビニ」「おにぎり」「食う」という3つのキーワードを入力すると下のような川柳が生成される。

美味しい酢 / 美味しいお菓子 / いつも食う

おにぎりがどこかに言っしまい、酢を食べるといのはおかしいが美味しい+お菓子などが結合していたり、その美味しいおかしをちゃんと食べていることからおおよそ意味が通ってはいる。評価はこれまでに入力されたキーワード等との類似度を比較して行う。

今後の課題

接続詞等を効率的導入可能なモデルを考案する、意味をより正確に認識できるようにセマンティックスを導入する。

参考文献

- [1] Pythonによるスクレイピング&機械学習開発テクニック: BeautifulSoupScikit-learn、TensorFlowを使ってみよう / クジラ飛行機著。- ソシム (2016年)。
- [2] 自然言語処理の基礎 / 奥村学著。- コロナ社 (2010年)。
- [3] 社会科学情報のオントロジ: 社会科学の知識構造を探る / 斉藤孝編著。- 中央大学出版部 (2009年)。