

Markov 連鎖を用いた新しいひみつ道具の生成

理学部 1 回生 牧レオナ

はじめに

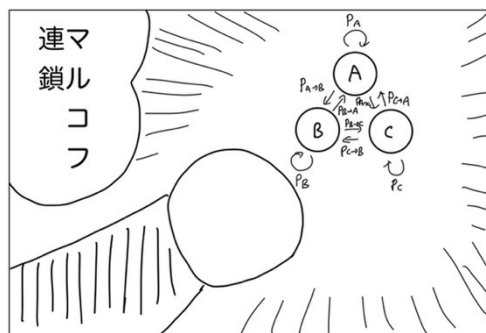
Markov 連鎖を用いた文章作成技術が登場し、既存の文章からその性格を捉えた新しい文章を作ることが可能になった。私は、この技術を利用することで、新しいひみつ道具を生成できるのではないかと考えた。

Markov 連鎖とは

Markov 連鎖 (Markov 過程) とは、次のようなものである (参考: 『機械学習のための確率と統計』, 杉山 将, 2015, 講談社, pp.69).

Markov 連鎖 (Markov 過程)

時間とともに変化する確率変数を確率過程 (stochastic process) とよび、ある時刻 t での標本 X_t が、1 つ前の標本 X_{t-1} にだけ依存する確率過程をマルコフ連鎖とよぶ。



Markov 連鎖

これを、文章作成に応用するとは、どのようなものであろうか。たとえば、次の 2 つの文章を用いて新しい文章を作ること考えよう。

「ドラえもんは、どら焼きが好き。」

「ドラえもんは、ネズミが嫌い。」

このとき、これらの文章をそれぞれ簡単のため次の 4 つの要素に分ける。

「ドラえもんは、 /どら焼き/が/好き。」

「ドラえもんは、 /ネズミ/が/嫌い。」

この状態で、「ドラえもんは、」から文章を始めるとしよう。ある言葉の次に繋がる言葉は、元のデータに示されたものの中から等しい確率で選ばれるものとする。この場合、「ドラえもんは、」の次に繋がる可能性のある単語は、「どら焼き」か「ネズミ」のどちらかである。ここでは、「どら焼き」が選ばれたとしよう。「どら焼き」の次に繋がるのは「が」しかないので、「どら焼き」につながる単語は「が」である。最後に、「が」に繋がる可能性のある単語は「好き。」か「嫌い。」のどちらかである。ここでは、「嫌い。」が選ばれたとしよう。すると、最終的に完成した文は次のものである。

「ドラえもんは、どら焼きが嫌い。」

これは、元のデータになかった新しい文章である。しかも、この文章は、文法として正しく、意味も理解可能である。このようにして、Markov 連鎖を使えば、元の文章になかった新しい文章を作り出すことができる。

もちろん、元の文章と同じものが生成されたり、意味の通らない文章が生成される可能性もあることは注意されたい。

これを既存のひみつ道具に適用して、新しいひみつ道具を作ってみよう。

方法

与えられたひみつ道具を、単語ごとに分解するために、形態素解析エンジンとして「MeCab」(<https://taku910.github.io/mecab/> 参照:2022/10/31)を使用した。プログラムは、Python を用いたコードが、『徒然なるままに技術/【Python 初心者向け】形態素解析とマルコフ連鎖を使ってひろゆきっぽい文章を生成してみた』, 2021/10/12(https://gakogako.com/python_hiroyuki/ 参照:2022/10/31)に掲載されていたため、これを使用した。このページには、Markov 連鎖についても分かりやすい解説が掲載されているため、ぜひ確認されたい。また、このコードは、Markov 連鎖を簡単に用いるために、「markovify」ライブラリを使用している。

ひみつ道具のデータは、『ドラえもん Wiki 道具一覧』(<https://doraemon.fandom.com/ja/wiki/道具一覧> 参照:2022/10/31)に掲載されていた 1858 個のひみつ道具を使用した。

私はプログラミングに精通していないため、コードの詳細を理解していないことを、声を大にしてここで申し上げます。

結果

数十回の試行を行った中で、新しいひみつ道具になる可能性があると考えられるものを選択し、次に示す。

- 十円なんでもドア
- いつでも地図
- いつでも窓
- 火事や交通事故を見るだけで動かせるメガネ
- なんでも誰でも窓
- 無敵ホコとタテ全自動のブリキ自動車セット
- SOS カプセルとコントローラー
- 毎日規則的になる虫眼鏡
- 服になる注射

- 即席海作り機と宝探しカウンター
- なんでも伸ばすことのできるホース
- 迷路遊びのできるホース
- 口を開けたら自動的になる薬
- 新聞日付変更チョーク

一方で、次のような、理解のしがたい道具も生成された。

- 塗ると魚の卵・フー子
- 歩かないでどこへでも卵にする道具
- ガスが固まってできた豆
- ゾウ印口紅とタコ印口紅とタコ印口紅と悪口
- 水ビル建築機水ビル建築機水ビル建築機水ビル建築機水ビル建築機水ビル建築機水ビル建築材
- 円ピツアルバイト料先払い円ピツアルバイト料先払い円ピツアルバイト料先払い円ピツ

まとめ

既存のひみつ道具の多くは、わずか数個の単語からなっており、組み合わせが限られたものであるから、それほど多くの新しい組み合わせは作り出せないかと思われたが、ひみつ道具に「ありそうな」新しい道具を生成することに成功した。

その理由は、「どこでも」や「なんでも」といった、汎用性の高い言葉が既存のひみつ道具に多く使われていることと、助詞を伴う道具が多数あったためであると考えられる。助詞を伴う道具とは、「春の足音を聞くことのできる機械」のような道具である。このような道具は、Markov 連鎖による文章作成と同様に、助詞によって多数の分岐を作るため、新しい道具が生成されやすかったと考えられる。

しかしながら、(i) 既存のひみつ道具の最後を変えただけのものや、(ii) ループを何周かしたことで、狂気的な語の連鎖になってしまったものも生成された。(i) としては、「機械」と「薬」が入れ替わったものが散見された。加えて、「○○と○○」という形の自由度が高く、脈絡のないものもあった。

(i) の例：新聞日付変更チョーク (新聞日付変更ポストというひみつ道具のポストがチョークに変わっただけ)

(ii) の例：円ピツアルバイト料先払い円ピツアルバイト料先払い円ピツアルバイト料先払い円ピツ

今後は、ドラえもんの名言データを利用して、新しい「ドラえもんっばい」名言を生成したい。