

話題語に注目したマルコフ連鎖型人工無能の改善

理工学部 数理情報学科
T110044 高山 一真
指導教員 佐野 彰

概要

人工無能は通称チャットボット、あるいは会話ボットとも呼ばれ、その名の通りユーザーとの会話を目的としたプログラムの総称である。しかし人間の様に文章の意味を理解して返答するのではなく、人工無能は会話文から特定のパターンを見出して、それを元に意味の通じるであろう文章を返しているだけである。本研究では、まずマルコフ連鎖を用いた人工無能を作成した。これは過去の事象に依存することなく、現在の状態にのみ依存して次の事象を決定づけるマルコフ性という性質を持つ確率過程を利用して返答文を作る手法である。入力された文に対して形態素解析を行い、単語の出現回数とともにマルコフ連鎖テーブルに加えることで入力文を学習し、連鎖のテーブルを用いて返答文を生成し、ユーザーとの会話を実現させる。しかしこの手法を用いた人工無能は製作者が意図しない文を生成することがあるため、会話が成り立たない可能性が高い。本研究はいい加減な返答しか出来ない人工無能をともに返答させるにはどのような改良を行えば効果的かを検討することが目的である。

本研究では基本的なマルコフ連鎖型人工無能のシステムに改善を加える度に会話実験を行い、実験結果からシステムの問題点を発見し、さらに変更を加えるという形を繰り返し、段階的にシステムを改良する。

まず最初に入力文に名詞を発見した場合、その名詞を会話の話題語として記憶し、返答文を生成する際にマルコフ連鎖で選択された名詞を、話題語となる名詞に置き換えるように変更をシステムに加えた。これにより、返答文を入力文と関連させることでの的外れな文を生成しないようにした。

次に話題語となった名詞を起点として、開始文字と終了文字に向かって文を生成するようにシステムを変更した。この動作を行うために、マルコフ連鎖で文を生成する方向とは逆の方向に向かって文の生成を行うアルゴリズムを加えた。話題語となる名詞が文章生成の起点となることで、開始文字から話題語までの文の生成に入力文との関連を持たせることが出来た。

上記のシステム変更の結果、返答文全体に入力文との関連を持たせることが出来たが、会話実験の結果は必ずしも改善されなかった。そこで、話題語となる名詞に対して適切な動詞を組み合わせるようにシステムを書き加えた。適切な名詞と動詞を組み合わせるために、入力文に名詞が発見され、引き続き動詞が発見されたときに、辞書型変数を使ってこの名詞と動詞の組み合わせを記憶させる。返答文生成の際に動詞が選択された場合、選択されている話題語に適切な動詞を作成した辞書型変数の中から決定し、動詞の置き換えを行う。これにより返答文では、話題語となる名詞と組み合わせる意味の通る動詞が選ばれるようにした。

適切な名詞と動詞を組み合わせるシステムで実験を行った際、終了文字が選択されないため文の生成が終わらずに極めて長い文になってしまう結果が複数出た。この問題を改善するために文の構成単語数に制限を設けた。文章を学習する機能に単語の個数を記憶する機能を追加し、返答文を生成する際に、マルコフ連鎖を用いて作成される返答文の単語数を決定する。そして決定した単語数ちょうどで終わる文章で返答文を生成する。しかし、この会話実験の結果は、システム変更前より悪くなった。よってシステムの変更を破棄し、単語数を制限する別の方法を考えた。

長い文が生成されるのは開始文字、もしくは終了文字が選択されにくいためである。したがって、マルコフ連鎖の文章生成で開始文字と終了文字が選択されやすくすれば極端に長い文章が生成されにくくなる。しかしこの確率を必要以上に高く設定すると逆に極端に短い文しか生成されない。そこでマルコフ連鎖で単語を選択される度に徐々に開始文字もしくは終了文字の選択確率を上げるようにシステムを変更した。

本研究で行った以上5つの改良と実験の結果、名詞と動詞の関連付けを行う機能を加えたときのシステムで、最も会話が成立した回数が多かった。この結果より名詞と動詞だけでなく様々な品詞も関連付けして学習させることが効果的であると考えられる。