

文脈を考慮したマルコフ連鎖型チャットボットの作成

理工学部 数理情報学科

T070107 八木 貴大

指導教員 佐野 彰

概要

チャットボット、日本では人工無脳とも呼ばれているものは、人工知能もしくは人工頭脳から派生した言葉である。チャットボットについて簡単に説明すると、人工知能とは別のアプローチで人間らしさをつくっていかうとする会話プログラムのことである。

本研究テーマにおいて究極の目標はマルコフ連鎖によって文生成を行うチャットボットを作成し、人間との自然な対話を実現させることである。しかし人間と同じような自然な対話をチャットボットにさせることは非常に難しい。そこで今回の研究ではそれを実現させるための1つの手段、可能性を示すことを目的としている。当然文字を扱うことになるのでプログラム言語はRubyを使用した。

チャットボットに人間と対話をさせるためには入力文の意図や文脈を読むことが必要になる。そこで本研究ではマルコフ連鎖で文生成する場合に重要な要素となる「開始文字の決定方法」についてと、マルコフ連鎖型の特徴でもある「サンプルデータが増えると精度が落ちる」という2つの課題に対するアプローチからそれらを実現しようと試みた。

「開始文字の決定方法」については入力文の名詞部分を抜き出しそれを開始文字にする方法を試した。これは文において名詞はその話題に直接に関連することが多いと推測したからである。「サンプルデータが増えると精度が落ちる」という問題については入力文の話題によって辞書を使い分ける方法を試した。こうすることでソースとなる辞書の量を減らし、かつ話題に沿った返答が期待できると考えたからである。本研究では「サッカー」と「政治」の2つの話題辞書を用意し試行した。

結果からこれらの方法を使えば、ある程度入力文に沿った返答が可能であることがわかった。今回は話題辞書数が少ないので対応できる話題の数も当然少ないが、より多くの数、種類の話題辞書を用意することが出来ればより多くの話題に対応できることが期待できる。しかしまだまだ自然な対話とは程遠いので改善が必要である。

これらの方法の問題点は、話題ごとに辞書をつくる作業を人力で行っているため辞書の充実が難しいことや、どの辞書を選ぶかを判断する処理が話題辞書の充実按比例して増えることである。加えて入力文に名詞がない「こんにちは」などの挨拶の文の場合には今回用いた開始文字の選択方法では意味がないという問題もある。これを解決するためには挨拶だけは辞書依存型のようにテンプレートを用意することや、挨拶専用辞書を作るなどの工夫が必要になるだろう。

今後の課題として、これらの問題点を解決するために、チャットボットの改良、文章をweb上から取り込み話題ごとに振り分けるプログラムの作成、話題辞書の選択方法の最適化などが挙げられる。