

会話の話題を判別する 人工無脳の研究

SATテクノロジー・ショーケース2015

■ はじめに

私は、人と会話のできる機械に興味を持っている。人と機械が会話することで、より機械を生活の場に活かせるようにしたいと考えたからである。

人工無脳とは、人と会話することを目的としたプログラムのことである。一般的に、ユーザである人が発言を入力し、人工無脳はその発言に対して、返答を作成・表示する、といった流れで処理が行われる。「ELIZA」1)や「ALAIN」2)といった人工無脳が有名である。

私はこれまでに、ユーザの発言に対応できるように様々な返答を用意した人工無脳を作成した。しかし、発言に対して返答が適切でない場合が多いことが課題だった。

この課題を解決するために、会話内容の話題の判別に取り組んだ。話題を特定することで、返答の選択肢をしぼり、より発言に合った返答の選択ができると考えたからである。発言に対して適切な返答を返すことで、人と機械との自然な会話を実現することを目標とした。

■ 研究内容

ユーザの発言から会話内容の話題を判別するプログラムの作成をした。話題は、「Yahoo!カテゴリ」というWebサイトを参考に、14に分類した。2段階の話題判別の手法を考え、2種類のプログラムを作成した。プログラム作成にあたり、Ruby言語(Ruby2.0.0)を用い、IDEとしてGeany(1.9.2.0)を使用した。

● 特定語話題判別プログラム(図1参照)

「Yahoo!JAPANが提供するテキスト解析WebAPI」³⁾を用いて、ユーザの発言の中からキーフレーズを抽出する。その後、事前に「Yahoo!カテゴリ」から抽出し、階層化した約6000語のカテゴリ名のデータベースの中に、キーフレーズと一致するカテゴリ名があるか検索する。ヒットした場合、最上位の階層のカテゴリ名を話題として選択する。

● 検索エンジンによる話題判別プログラム(図2参照)

「特定語話題判別プログラム」で、話題が判別できなかった場合に、このプログラムで話題判別を行う。このプログラムは、「検索エンジンを利用した未登録単語に関する単語間距離の測定」⁴⁾という論文を参考に作成した。検索エンジンのAPIを利用し、キーフレーズと最上位の階層のカテゴリ名をそれぞれAND検索し、ヒット件数取得する。その中で一番ヒット件数の多いものを話題として選択する。

この2段階の手法を用いた話題を判別するプログラムに、70通りの発言を入力し、正しい話題の判別ができるか実験を行い、話題判別の正答率を測定した。その結果、16通りの発言に対して、正しい話題を判別することができ、話題判別の正答率は23%となった。さらに正答率が高くなるように、話題判別プログラムの改善をしていきたい。

今後は、判別した話題を用いて関連した情報を収集し、返答を作成するプログラムをつくらうと考えている。その後、会話内容を記録できるようにし、より自然な会話をする人工無脳の作成に取り組みたい。

■ 主要参考文献

- 1) 「Eliza, computer therapist」
<http://www.manifestation.com/neurotoys/eliza.php3>
- 2) 「ALAIN A.I.」
<http://www.grokwork.com/alain/>
- 3) 「Yahoo!JAPANが提供するテキスト解析WebAPI」
<http://developer.yahoo.co.jp/webapi/jlp/>
- 4) 「検索エンジンを利用した未登録単語に関する単語間距離の測定」
http://www.anlp.jp/proceedings/annual_meeting/2006/pdf_dir/D2-1.pdf

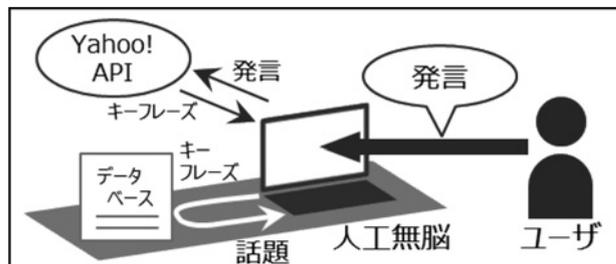


図1 特定語話題判別プログラムの流れ

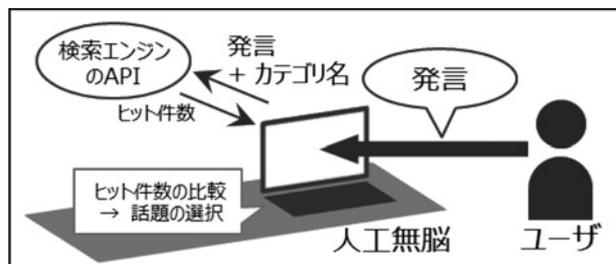


図2 検索エンジンによる話題判別プログラムの流れ

代表発表者 **岩本 和也 (いわもと かずや)**
 所属 **千葉県立柏の葉高等学校 情報理数科**
 問合せ先 **〒270-0882 千葉県柏市柏の葉 6-1**
TEL:04-7132-7521 FAX:04-7133-2435

■キーワード: (1) 人工無脳
 (2) 自然言語処理
 (3) 話題判別