

第二の自分：対話を通して自身の情報を蓄積・利用する対話システムの提案

三木 康太[†] 宮部 真衣[†]
[†]和歌山大学システム工学部

1 はじめに

近年、対話による天気や経路などの情報提供や、しりとりなどのタスクをこなすだけでなく、雑談も可能な対話システム^{*1*}が存在している。このような対話システムは、話し言葉の入力によりシステムを効率的に操作することや、会話により親近感を持たせるといったことを目的としたものである [1]。また、雑談を主目的とする非タスク指向の対話システムは、要求される話題の幅が広く、楽しさや会話の自然さなどの向上を目的とした研究がなされている [2][3][4]。しかし、自然な対話の困難さ、会話を盛り上げる難しさが課題として挙げられている。また、特定ユーザを対象としたシステムは少なく、幅広いユーザを対象としたものが多い。

本研究ではユーザの情報を蓄積・利用する対話システム“第二の自分”を提案する。提案システムは、特定の話者のみを利用対象とした Web 上で動作する対話システムである。このシステムは、対話で入力された会話文を解析することで、(1) 特定話者に関する知識を学習すると共に、辞書サイトや SNS サイトの解析により、(2) 一般的な知識を学習する。従来とは異なり、特定ユーザを利用対象とし、2 種類の知識を蓄積・利用することにより、“もう一人の自分”のように動作することを想定している。

本稿では、提案システムの概要と主要な機能について述べる。

2 関連研究

Twitter を用いた非タスク指向型対話システムの一つに、稲葉らの KELDIC[2] が挙げられる。この研究では Twitter のツイートとリプライのペアを使い、特定の話題語と応答文のペアを作成することにより、自然な会話を実現している。小林らのシステム [3] では、ユーザの人間関係や嗜好といった情報を記憶することで、特定の会話文の場合に、ユーザの情報を使用し応答している。柴田らの研究 [4] では、ソーシャルメディアに限

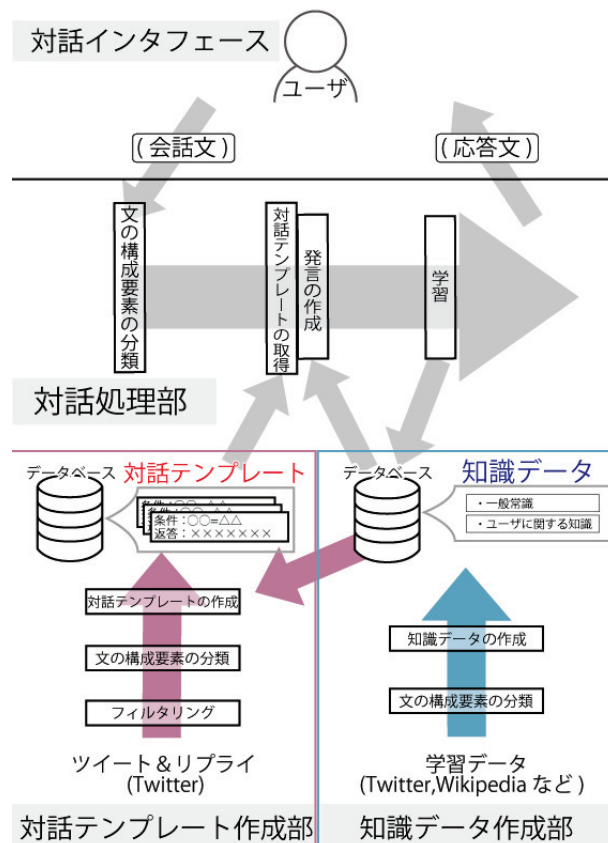


図 1: システム構成

らず、Web 上から検索を行い、応答の候補を抽出し利用している。

いずれのシステムも本稿で述べる対話システムと類似する点が多いが、本稿で提案するシステムでは、特定のユーザに関する知識を会話文から学習し、特定の話者に最適化したシステムとする。

3 第二の自分：対話を通して自身の情報を蓄積・利用する対話システム

本研究で提案する対話システム“第二の自分”は、従来のような一般的な知識を利用するだけでなく、対話を通して特定の話者に関する情報を蓄積することにより、ユーザにとっての“もう一人の自分”のような存在を目指したシステムである。

話者の情報を蓄積し、利用可能にすることで、話者の興味のある深い話題の提供や、過去の自分の言動を

Alter Ego: A Proposal of a Dialogue System to Store and Utilize the Speaker's Information through Conversation

Kota MIKI[†] Mai MIYABE[†]

[†]Faculty of Systems Engineering, Wakayama University

*1 シャベってコンシェル, <https://www.nttdocomo.co.jp/service/information/shabette.concier/>

*2 Siri, <http://www.apple.com/jp/ios/siri/>

回顧させることができると考えられる。また、特定話者以外の他者からの会話を可能にすることで、特定の話者の代わりに受け答えを行い、特定の話者と他者とのコミュニケーションを円滑化できるようなシステムを目指す。

本システムでは、ユーザによる入力 that 特定の条件を満たした場合に、予め用意した特定の応答をすることで、対話を実現する。この条件となる入力(以下、応答条件)と対応する応答(以下、応答フレーム)のペアを、本稿では対話テンプレートと呼ぶ。また、システム内に蓄積する話者に関する情報や一般的な知識等を知識データと呼ぶ。

図1に提案システムの構成図を示す。提案システムは対話処理部、対話テンプレート作成部および知識データ作成部から構成される。また、知識データ作成のために、Twitter¹, Wikipedia² と連携する。以降の節において、それぞれについて概説する。

3.1 知識データ作成部

知識データ作成部は、「会話者の名前=A」、「パンの原材料=小麦」といった同格関係を知識データとして構築し、知識データベースに蓄積するモジュールである。

客観性・主観性を問わない語句の知識データを得るために、Twitter から特定の語句で検索したツイート、および Wikipedia からテキストを取得し、構築する。

得られたテキストから主語や目的語といった文の構成要素を解析する。分解後に主語、目的語、形容動詞のみ、もしくは主語、形容詞のみ、もしくは主語、目的語のみの文章があった場合、「A=B」「A の B=C」のようなテキスト形式で知識データベースに格納する。

3.2 対話テンプレート作成部

対話テンプレート作成部は、Twitter から取得したツイートをもとに、応答条件と応答フレームのペアを作成するモジュールである。

対話テンプレートは、多数のデータが得やすく、先行研究 [2] でも用いられている Twitter のツイートとそのツイートへのリプライ(以下、ツイート対)から作成することとした。ツイート対を見ても会話の意味がわからない場合、対話に用いるのは適切で無いと考えられる。そこで本研究では、稲葉らのフィルタリング手法 [2] を参考に、以下の4つのルールでツイートをフィルタリングし、学習に用いる。

- A. URL を含まない
- B. 人名、代名詞が無い

- C. リプライに対応するツイートは、どこにもリプライしていない。
- D. 単語数が1~30である。

フィルタリングしたツイート対の語句間に関連性がみられる場合、関連性がある語句を変数に置換し、応答条件と応答フレームを作成する。

3.3 対話処理部

対話処理部は、ユーザの入力文を処理するモジュールである。

ユーザは対話インタフェースを通して任意の会話文を入力する。システムは会話文から主語や目的語といった文の構成要素の分類を行い、対話テンプレートの応答条件に合うかどうかを判定し、一致したテンプレートを候補とする。複数のテンプレートが候補として抽出された場合は、一致した条件の数が最も多いテンプレートを採用する。採用されたテンプレートの応答フレームの変数に値を代入し、システムからユーザへの応答とする。

また、対話処理部に入力された会話文をもとに、ユーザに関する知識を学習することで、特定ユーザへの最適化を行う。ユーザの会話文を3.1節で述べた知識データ作成部により分析し、ユーザに関する知識として知識データベースに格納する。蓄積後、話者の情報は対話テンプレートと組み合わせて利用可能になる。

4 おわりに

本稿では Web 上で動作し、対話を通して話者に関する知識と、一般的な知識を蓄積・利用する対話システム“第二の自分”を提案した。

今後、提案システムにより会話の楽しさ、対話者の情報を蓄積できたかという観点から有用性について検証を行う。

参考文献

- [1] ITmedia: 「ユーザーがやりたいこと」に答えてくれる——頼れるエージェント「しゃべってコンシェル」, <http://www.itmedia.co.jp/mobile/articles/1208/31/news002.html>
- [2] 稲葉通将 高橋健一: Twitter を用いた非タスク指向型対話システムのための発話候補文獲得, 人工知能学会論文誌, vol.29, No.1, pp.21-31(2014).
- [3] 小林峻也 萩原将文: ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム, 人工知能学会論文誌, vol.31(掲載予定).(2016).
- [4] 柴田雅博 富浦 洋一 西口 友美: 雑談自由対話を実現するためのWWW上の文書からの妥当な候補文選択手法, 人工知能学会論文誌, vol.24, No.6, pp.507-519(2009).

¹ Twitter, <https://twitter.com/>

² Wikipedia 日本語版, <https://ja.wikipedia.org/wiki/>