

## 対話と SNS を利用した情報を蓄積・利用する対話システム “ 第二の自分 ” の構築

三木 康太<sup>†</sup>宮部 真衣<sup>†</sup><sup>†</sup>和歌山大学システム工学部

### 1 はじめに

近年、対話によってスケジュール管理や Web 検索などのタスクを行う対話システムや、雑談を目的とした対話システムが増加している。雑談を目的とする非タスク指向の対話システムは、多数の話題に対応する必要があり、自然な対話を行うことが難しい [1]。そのため、Twitter から応答文を自動で獲得する研究 [2] や、対話からユーザの情報を獲得する研究 [3] がある。

現在、多くの対話システムは、対応できる話題の種類は多いものの、特定の話題に特化することができない。特定の話題で会話が続けられない場合、ユーザの関心が薄れ、会話に対するモチベーションが低下する可能性がある。それぞれのユーザに適した応答が可能になれば、この問題を解決できる可能性がある。

Twitter や Facebook などの SNS は、3000 万人以上の利用者が存在する<sup>1</sup>。SNS には個人の情報が投稿されており、ユーザに特化させるための情報を獲得できる可能性がある。そこで、従来のような一般的知識だけでなく、SNS と対話から特定ユーザの情報を蓄積・利用することにより、話者の興味に関連する話題の提供や、話者の言動の回顧に対応可能な対話システム“ 第二の自分 ” を提案する。特定ユーザの情報を一般的な知識と併せて用いることにより、特定ユーザの“ もう一人の自分 ” のように動作する対話システムを実現する。

本稿では、提案システムの概要を述べ、システムで用いる知識表現について検討する。

### 2 関連研究

SNS を用いた非タスク指向型対話システムの一つとして、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利用し、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し、特定の会話文の場合に、記憶したユーザ情報を用いて応答

する。平野らのシステム [4] では、対話文から対話行為の推定と述語項構造の解析を行い、述語項構造と、属性カテゴリやトピックなどのユーザに関する情報を抽出している。

これらの研究では、SNS を用いた応答文の作成や、チャットテキストからのユーザ情報の取得を行っているものの、特定ユーザに特化させるためのユーザ情報の蓄積は行っていない。本研究では、ユーザ情報の取得・蓄積により、ユーザ毎に最適化可能な対話システムを目指す。

### 3 対話と SNS を利用した情報を蓄積・利用する対話システム “ 第二の自分 ”

本章では、提案システムについて概説した後、本研究における知識について説明する。

#### 3.1 提案システムの概要

本研究で提案する対話システム“ 第二の自分 ” は、従来のように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。話者の情報を用いることで、話者の興味のある深い話題の提供や、過去の言動の回顧ができるようになると考えられる。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。

#### 3.2 知識表現のための格の分析

本システムで用いる話者に関する情報や一般的な知識は、Twitter などの SNS サイト、Wikipedia などの辞書サイト、新聞記事などをもとに作成する。

コンピュータで知識を扱う手法として、格フレーム、事態間知識などがある [5]。事態間知識は述語項構造と呼ばれる単位から構成されており、述語項構造は表層格および深層格の二種類が存在する。

本研究では知識を扱う手法として、事態間知識のうち、深層格を利用する。深層格は、表層格に比べ文章からの抽出が困難であるとされる [6]。一方、表層格は文章からの抽出が容易であるが、「一人で」「一緒に」の

Development of a Dialogue System “ Alter Ego ” to Store and Utilize the Speaker’s Information through Conversation and SNS

Kota MIKI<sup>†</sup> Mai MIYABE<sup>†</sup>

<sup>†</sup>Faculty of Systems Engineering, Wakayama University

<sup>1</sup>総務省 | SNS の利用率,

<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc242220.html>

表1: 述語項で出現した格

	分類	格	定義	例文	出現回数	出現割合
1	既存	S	主格	“コロッケは”美味しい	109/263	0.41
2	既存	O	目的格	“本を”買いに行く	70/263	0.27
3	既存	Place	場所格	“バヌアツで”地震が発生した	20/263	0.08
4	既存	Target	時間格	“筋肉痛に”なりそう	13/263	0.05
5	既存	Time	時間格	“昨日から”喉が痛い	13/263	0.05
6	既存	Tool	道具格	“アナログで”描く	2/263	0.01
7	既存	Source	源泉格	“牛から”見たら	1/263	0.01
8	A	Reason	理由	“横須賀に来たので”	11/263	0.04
9	A	Assumption	仮定	“夜間だったら”	8/263	0.03
10	A	Contrast	逆説	“賞をあてたけど”	4/263	0.02
11	A	Addition	継続の意味	“思い出であり”, 宝物です	3/263	0.01
12	B	Parent	親子関係	“ネコの”体は	40/263	0.15
13	B	Equal	主格(S)と一致	俺は“無口だ”	37/263	0.14
14	B	Parallel	主格(S)との並立関係	“東京と”大阪で	17/263	0.06
15	C	Degree	述語に対する程度	“とても”嬉しい	21/263	0.08
16	C	Represent	述語項の代表的な内容	参加した“こと”	13/263	0.05
17	C	As	様態・状態	“ソングライターとして”活躍する	10/263	0.04

分類の列が“A・B・C”の場合、新たに拡張した格であり、“既存”の場合、Fillmoreの格である。

ような表層格では異なる格が、意味的に同じである場合、誤った述語項が作成されるという問題点も指摘されている[4]。深層格は目的に合わせて拡張される場合があり、12格への拡張[3]や、より多数に拡張する場合[7]も存在する。深層格は元々8つの格が定義されているが、SNS上の多様な表現に対応できるわけではない。

提案システムはSNSから情報を取得するため、より様々な知識を表す必要がある可能性がある。そこで、既存の格で対応しきれない表現を、コーパスの分析により抽出する。コーパスから人手で作成した述語項構造に、既存の格に該当しない表現がある場合、新たな格として採用する。今回は、Twitterから取得したツイート34文、人間同士の雑談チャット19文、新聞記事などのウェブサイト34文、日本語学習サイトの例文53文、合計140文のテキストを分析用コーパスとした。平均文字数は15.5文字(標準偏差8.0)、最大文字数53文字、最小文字数4文字である。これらのテキストから263個の述語項構造を作成し、格を分析する。

### 3.3 分析結果

分析の結果出現した格と例文および出現頻度を表1に示す。今回用いたコーパスデータでは、Fillmoreによって提唱された8格のうち7格<sup>1)</sup>と、(A)述語項同士の関係(4種)、(B)項同士の関係(3種)、(C)その他(3種)が知識表現において必要となった。

Fillmoreの格では述語項と述語項の関係を表すことが難しいため、逆説や理由を表す4格(Reason, Assumption, Contrast, Addition)を加える。また、名詞同士の親子関係、並立関係、同格関係を述語項で表すために、3格(Parent, Equal, Parallel)を追加する。さらに、主に述

語を修飾する格(Degree)、1語で述語項を言い表す格(Represent)、様態や状態を表す格(As)の3格を設ける。

今後、今回の分析で定義した格を他のコーパスに適用し、十分な知識データを取得可能かどうか検証する。

## 4 おわりに

本稿では、SNSと対話を利用することにより、特定話者に最適化する対話システム“第二の自分”を提案し、SNSや対話のテキストを用いて、システムで必要となる述語項の格を分析した。分析の結果、Fillmoreが提唱した7格以外に、必要となりうる10格を知識表現のための格として定義した。

今後、Twitterやチャットテキストからのユーザ情報取得実験を行い、定義した格の妥当性を検証する。また、“もう一人の自分”のように動作する対話システムの構築を行う。

## 参考文献

- [1] 東中竜一郎, 船越孝太郎: 対話システムにおけるエラー分析, 情報処理学会誌, vol.57, No.1, pp.42-43(2016).
- [2] 稲葉通将 神園彩香 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, No.1, pp.1-10(2016).
- [4] 平野徹, 小林のぞみ, 東中竜一郎, 牧野俊朗, 松尾義博: パーソナライズ可能な対話システムのためのユーザ情報抽出, 人工知能学会論文誌, vol.31, No.1, pp.1-10(2016).
- [5] 柴田知秀: 知識獲得, 情報処理学会誌, vol.57, No.1, pp.22-23(2016).
- [6] 竹野峻輔, 松田真希子, 梶原智之, 山本和英: 機械学習を用いた二重深層格の自動付与の検討, 言語処理学会第20回年次大会, pp.1011-1014(2014).
- [7] 斎藤友伸, 松居辰則, 岡本敏雄: 格文法解析による対話文からの知識収集, 電子情報通信学会技術研究報告, 教育工学 100(682), pp.9-16(2001).

<sup>1)</sup>主格, 対象格, 源泉格, 目標格, 場所格, 時間格, 道具格