

未登録語データベースを活用した雑談型対話システム

Chat-oriented Dialogue System Using Unregistered Database

仲村 敦 土屋 誠司 渡部 広一
 Atsushi Nakamura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

近年、福祉の場などの様々な場面でロボットが人間のパートナーとして活躍することが期待されている。そのためには、ロボットと人間が円滑な意思疎通を行う必要がある。日常的な雑談対話を、ロボットが人間らしく行うことができれば、人間との意思疎通を円滑に行うことができると考えられる。

雑談を行う対話システムとして、会話履歴フレームを利用した雑談型対話システムが提案されている^[1]。そのシステムでは、入力文にデータベースに登録されていない言葉が含まれていた場合、未登録語データベースに登録して、今後の対話に活用する。そうすることで、ユーザに話を聞いていることが伝わり、自然で快適な対話となる。

しかし、既存システムの未登録語データベースでは、ユーザの返答を少ししか応答に反映できない。そこで本研究では、未登録語データベースに単語間の関係を登録することで、ユーザの返答をより反映した応答の生成を目指す。

2. 関連技術

2.1. 概念ベース

概念ベース^[2]は単語を概念とし、属性と重みの対の集合として定義している。属性は概念を特徴付ける単語であり、重みは概念に対する属性の重要度を表す。約 9 万語が登録されている。概念ベースの例を表 1 に示す。

表 1 概念ベースの例

概念	属性, 重み
医師	(患者, 0.63) (医師, 0.58) (医療, 0.50) ...
治す	(医療, 0.56) (為政者, 0.31) (治める, 0.17) ...

2.2. NTT シソーラス

NTT シソーラス^[3]とは一般名詞の意味属性についてその上位下位関係、全体部分関係を木構造で示したものである。約 30 万語が登録されている。

2.3. 未登録語

未登録語とは、概念ベース、NTT シソーラス、未登録語データベースに登録されていない語のことである。また、未登録語ではない語を登録語とする。

2.4. 未登録語データベース

未登録語データベースとは、過去の会話に登場した未登録語とそれに対応する登録語が登録されているデータベースである。表 2 に未登録語データベースの一部を示す。

表 2 未登録語データベース

未登録語	登録語
サイゼリア	レストラン

2.5. Cabocha

Cabocha とは、係り受け解析を行うシステムである。係り受け解析は文を文節に区切った後、文節の対応関係を出力するものである。

2.6. 意味理解システム

意味理解システム^[4]は、入力文を単語が持つ情報ごとに、7W1H と述語のフレームに分割するシステムである。7W1H とは英語の疑問詞に用いられる 6W1H に「誰と」を表す Who+フレームを追加したものである。例として、「私

は昨日、友達と動物園に行った」という文を入力した場合を以下の表 3 に示す。

表 3 意味理解システムの分割例

Who	Who+	When	Where	述語
私	友達	昨日	動物園	行った

2.7. 会話履歴フレーム

会話履歴フレームは、意味理解システムと直前のシステムの発話を用いて、会話履歴を 7W1H と述語に分割して格納したものである。図 1 にユーザとシステムの対話例とその際の会話履歴フレームの格納例を示す。

図 1 会話履歴フレームの格納例

1 U:私は昨日、水族館でイルカを観た					
Who	Who+	What	When	Where	述語
私		イルカ	昨日	水族館	観た
2 S:誰と見ましたか?					
3 U:友達とです					
Who	Who+	What	When	Where	述語
私	友達	イルカ	昨日	水族館	観た

システムの「誰と見ましたか?」という Who+について聞く 7W1H 応答に対して、ユーザが「友達とです」と返しているため Who+フレームに「友達」が格納される。

3. 会話履歴フレームを利用した雑談型対話システム

3.1 システム概要

このシステムによる雑談型対話は、話者が文章を入力すると、それに対して表 4 に記載している 10 つの応答からあらかじめ設定している設定確率を用いて応答を選択し、ランダムに返答を行う。雑談型対話システムによって展開される対話の流れを以下の図 2 に示す。

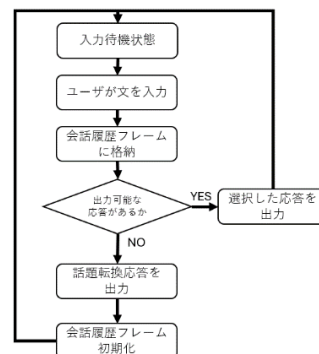


図 2 雑談型対話システムの流れ

システムを起動すると入力待機状態となり、そこで話者が文を入力する。入力文を、意味理解システムを用いて会話履歴フレームに格納する。会話履歴フレームの状態から、話題転換応答以外の各応答手法で応答可能かどうか判断する。応答可能な応答手法が 1 つもなければ話題転換応答が行われ、会話履歴フレームを初期化した後に、開始時の状態に戻る。1 つの場合はその応答手法を行う。複数あれば応答手法を重み付きランダムで選択し、その応答手法での処理が行われる。その後、入力待機状態に戻るという流れを繰り返す。

表 4 雑談型対話システムの応答手法一覧

応答手法	応答文の例
挨拶応答	おはようございます.
未登録語応答	インスタとは何ですか?
未登録語データベースを活用した応答	レストランといえば、サイゼリアですよ
感情判断応答	よかったですね.
7WH 応答	どこで食べましたか?
掘り下げ応答	友達の誰さんですか?
場所判断連想応答	海に釣りに行ったのですか?
自己開示応答	今日は寒いと思います.
話題連想応答	SMAP と思ったら 木村拓哉ですよ.
話題転換応答	最近何をしましたか?

3.2. 未登録語応答

ユーザの入力文に未登録語が含まれていたときの返答を未登録語応答という。図 3 に未登録語応答の一例を示す。

1 U: スプラトゥーンをしました
2 S: スプラトゥーンとは何ですか? (未登録語応答)
3 U: スプラトゥーンはインクで戦うゲームです

図 3 未登録語応答の例

3.3. 未登録語データベースの登録手法

出てきた未登録語と、未登録語応答に対するユーザの返答文の一番後ろにある名詞を登録する。図 3 の場合、未登録語「スプラトゥーン」と文の一番後ろにある名詞「ゲーム」が未登録語データベースに登録される(表 5)。

表 5 図 3 で未登録語データベースに登録された未登録語

未登録語	登録語
スプラトゥーン	ゲーム

3.4. 未登録語データベースを活用した応答

ユーザの入力文中に未登録語データベースに登録された登録語が出てきた際、未登録語データベースを活用した応答を返す。例を図 4 に示す。

1 U: 昨日、ゲームを買いました
2 S: ゲームといえば、スプラトゥーンですよ

図 4 未登録語データベースを活用した応答

4. 提案システム

4.1 提案する未登録語データベース

ユーザの返答文を係り受け解析した後、文節から助詞と記号を除いたものを MainWordTable に、係り受け関係、助詞、関係の種類を RelationTable に登録する。図 3 の対話の係り受け解析結果とデータベースの例を表 6, 7, 8 に示す。

表 6 係り受け解析結果

係り元	係り先
スプラトゥーンは	ゲームです
インクで	戦う
戦う	ゲームです

表 7 MainWordTable

ワード番号	メインワード	未登録語
0	スプラトゥーン	スプラトゥーン
1	インク	スプラトゥーン
2	戦う	スプラトゥーン
3	ゲーム	スプラトゥーン

表 8 RelationTable

ワード番号 1	助詞	関係の種類	ワード番号 2
0	は	上位	3
1	で	修飾	2
2		修飾	3

関係の種類は、「上位」、「感覚」、「略称」、「修飾」の 4 つがあり、2 つの単語間の助詞などを参考に、適切なものを選択する。

4.2 提案する未登録語データベースを活用した応答

未登録語データベースを活用した応答を図 5 に示す。

1 U: 昨日、スプラトゥーンを買いました
2 S: スプラトゥーンですか。
インクで戦うゲームですよ

図 5 提案する未登録語データベースを活用した応答

下線部の応答は、未登録語データベースから「スプラトゥーン」と「上位」の関係を持つ「ゲーム」を抽出し、「ゲーム」と関係を持つ単語を繋げて生成される。

5. 評価

提案した雑談型対話システムの快適さと自然さを評価するために既存システムとの比較実験を行った。実験協力者は男女 10 名である。実験協力者にはそれぞれのシステムと、システムの応答が 10 回出るか、未登録語データベースを活用した応答が 2 回出るまで対話を行ってもらった。その後、6 つの評価項目について、5 段階による評価を行った(表 9)。また、それぞれのシステムの未登録語データベースを活用した応答を含んだ対話例についても、自然さを 5 段階で評価してもらった(表 10)。

表 9 対話システムの評価の平均

	評価項目	提案	既存
快適さ	システムをまた使用したい	3	2.4
	システムを使用して楽しかった	3.6	3.1
	システムを使用して満足している	3.2	2.7
自然さ	システムが提供する話題は適切だった	3	2.2
	システムの応答は分かりやすかった	3.4	2.8
	システムの発話量は適切だった	3.7	3.9

「システムの発話量は適切だった」以外の 5 項目で提案システムの平均が既存システムの平均を上回った。

表 10 対話例の評価の平均

	提案	既存
対話例の自然さ	4.28	2.83

提案システムの評価の平均が既存システムの評価の平均を上回った。

6. 考察

評価から、未登録語データベースに関係を登録することで雑談型対話システムの性能が向上したことが分かる。これはユーザの返答をより応答に反映できたことが要因と考えられる。しかし、不自然な応答をすることもあった。これは複合名詞を未登録語として認識することや、単語間の関係を間違えて登録してしまう場合があることが原因と考えられる。

7. おわりに

本研究では、未登録語データベースに単語間の関係を登録し、ユーザの返答をより応答に活かすことで、雑談型対話システムの性能が向上した。しかし、考察で述べたような不自然な応答がみられたので、未登録語データベースの登録手法にはさらなる改良が必要である。

参考文献

- [1] 小山翔平, 土屋誠司, 渡部広一, “雑談型対話システムにおける自己開示発話の割合増加手法の提案”, 情報科学技術フォーラム, FIT2020, pp.127-128.
- [2] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5, pp.44-45, 2007.
- [3] NTT コミュニケーション科学研究所, “日本語語彙体系”, 岩波書店, 1997.
- [4] 小谷涼, 吉村枝里子, 土屋誠司, 渡部広一, “入れ子構造による文章の意味理解手法の提案”, 情報科学技術フォーラム FIT2016, pp.259-260.