

文の類似性を用いた未知語処理手法の提案と それに基づく円滑な対話応答システムの作成

福岡知隆† 税田竜一†† 久保村千明††† 服部峻† 亀田弘之†

†東京工科大学 ††日本工学院八王子専門学校 †††山野美容芸術短期大学

1 はじめに

近年, 人間の対話の相手が増えた. 人間に比べて膨大な情報の保持が可能なコンピュータである. チャットなどでの雑談相手, Web 上での商品の説明, 介護における話し相手など, 多種にわたって人間はコンピュータと対話を行うようになって来た.

しかし, 人間同士の対話と比較すると, その結果や過程が劣っている場合が多い. その原因の一つが円滑性 (発話者の意図に沿い, 対話が速やかに行われること) の欠如である. データベース内に情報が存在しない単語, 即ち未知語に遭遇した場合にその現象は著しい. 既存の処理では未知語に対して質問や話題転換が過剰に起こり, 対話の円滑さが損なわれる場合がある.

この問題を解決するため, 本研究では未知語処理を改善し人間とコンピュータ間の対話をより自然で円滑にする手法を提案する. また, それを用いた対話応答システムを作成したので, その報告を行う.

2 文の類似性を用いた未知語処理手法

ユーザの発言中に出現する未知語の品詞をシステムがデータベースの文例を用いて推定する. それにより, システムはユーザに未知語についての質問をすることなく, 応答文選択処理を行うことが出来る. 類似性は, 文の品詞並びと出現する単語の表層文字列の同一性を基に判断される. 以下, 未知語推定手順を詳細に記す.

2.1 類似文例の検索 入力文の品詞並びをキーとして, パターンデータベースから類似文例を検索する.

2.2 最類似文例の選択 入力文と呼び出された文例各々に対して, 出現する単語の表層文字列の一致数を重みとした類似度計算を行い, 最も類似度が大きい文例を1つ選択する. 重み w の計算には, X を入力文に出現する単語の集合, Y を候補文に出現する単語の集合とする以下のダイス係数を採用する. ただし, 一つの文中に同一の単語が複数回出現した場合は別の要素として扱う.

$$w = 2 \times \frac{|X \cap Y|}{|X| + |Y|} \quad (1)$$

以下に具体例を示す.

入力文の単語

$X = \text{音声/対話/という/こと/は/喋れ/ます/か/?}$

類似文例の単語

$Y_1 = \text{音声/対話/という/こと/が/重要/です/か/?}$

$Y_2 = \text{音声/対話/という/こと/は/話せ/ます/か/?}$

計算結果

$$w_1 = 2 \times \frac{6}{9+9} \doteq 0.67$$

$$w_2 = 2 \times \frac{8}{9+9} \doteq 0.89 \quad (\text{類似文例として } Y_2 \text{ を選択})$$

2.3 未知語の推定 入力文と選択された文例の品詞並びを比較し, 入力文の未知語の品詞を推定する.

Smooth Dialogue System based on Unknown Word Processing Method Using Similarity between Sentences

Tomotaka Fukuoka†, Ryuichi Saita††, Chiaki Kubomura†††, Shun Hattori†, and Hiroyuki Kameda†

†Tokyo University of Technology

1401-1 Katakuramachi, Hachioji, Tokyo, 192-0982, Japan

††Nihon Kogakuin College of Hachioji

1401-1 Katakuramachi, Hachioji, Tokyo, 192-0982, Japan

†††Yamano College of Aesthetics

530 Yarimizu, Hachioji, Tokyo, 192-0396, Japan

c01063316b@css.teu.ac.jp

ryus@pddin.net

ckubomura@yamano.ac.jp

{hattori,kameda}@cs.teu.ac.jp

	表層文字列情報	品詞並び情報
データベース の文例	音声 対話 ということは 話せますか？	名詞-名詞-助詞-名詞-助詞- 動詞-助動詞-助詞-記号
		+
ユーザの 発言	音声 対話 ということは 喋れますか？ 未知語	名詞-名詞-助詞-名詞-助詞- 未知語-助動詞-助詞-記号
		↓ 推定
	音声 対話 ということは 喋れますか？	名詞-名詞-助詞-名詞-助詞- 動詞-助動詞-助詞-記号

図 1: ユーザ発言中の未知語の品詞推定

3 作成した対話応答システム

図 2 に対話応答システムの流れを示す。

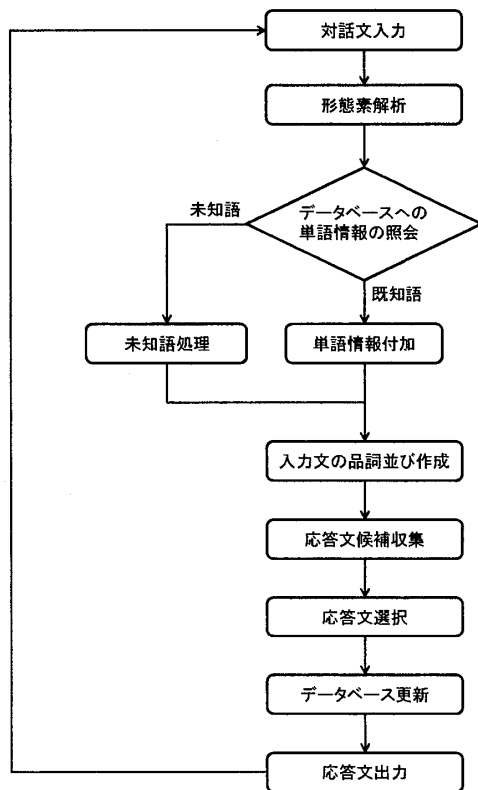


図 2: 対話応答システムの流れ

システムはユーザから対話文入力を受け取り、形態素解析を行い、文を単語単位に分解する。また、その結果得られた単語の表層文字列と品詞の組を検索のキーとし、データベースから入力文の出現単語の読み情報と頻度を得て、入力文の単語から品詞並びを作成する。

この際、読み情報と頻度が得られなければ、その単語を未知語とみなし、前述した未知語推定処理を行う。

次に応答文候補を選択する。応答文は過去の用例として、対話相手の直前の発言（以降前文とする）も属性情報として保持している。応答文候補は入力文の品詞並びと前文の品詞並びが同一の文例である。

入力文とそれぞれの候補文の前文から名詞、動詞、感動詞、副詞、形容詞のいずれかの単語をキー単語として抽出する。式 (1) において、 X を入力文のキー単語の集合、 Y を前文のキー単語の集合とし、前文の重みが最も大きい応答文候補を最終的な応答文とする。

続いてデータベースの更新を行い、入力文、応答文中に出現した単語とパターンの頻度を 1 増加する。また、未知語が出現していた場合は新規単語としてデータベースに登録する。最後に応答文を出力し、システムはユーザの次の対話文入力を待つ。

4 評価

本研究の未知語処理手法の妥当性とその成果について、東京工科大学の学生 27 人にアンケートを行った。

4.1 未知語処理手法の妥当性の確認

人間同士が対話を行うときの未知語への対応に関してアンケートを行った。その結果を表 1 に示す。

表 1: 未知の修飾語への対応

対応パターン	意味を尋ねる	話題を変える	相槌を打つ	その他
人数	10	1	13	3

上記の結果より、人間は未知語に出会った場合、必ずしもその言葉に関して質問するわけではなく、単語の意味を知らなくても対話を続ける傾向が強いと推察される。この点において、本研究の未知語の推定による対話の続行は妥当である。

4.2 対話応答の自然さの確認

対話応答システムを用いて、未知語処理手法を適用した対話応答例と、適用しなかった対話応答例のどちらが自然に感じるかのアンケートを行った。呈示した対話例を図 3 に、その結果を表 2 に示す。対話例は「喋れ」が未知語である。

ユーザ> PDDIN とは何ですか。
 COM> 音声対話型癒し系ロボットの事です。
 ユーザ> 音声対話ということは喋れますか。
 COM> a. はい、限定的ですが話せます。
 b. 重要なことです。

図 3: 呈示した対話の例

表 2: 対話応答選択結果

応答文 (未知語処理)	a (適用)	b (不適用)
人数	25	2

結果から、多くの人が未知語処理を利用した応答例をより自然だと感じており、本研究で提案した未知語処理の基本的有効性が確認された。

5 おわりに

人間とコンピュータの円滑な対話を目指し、「文の類似性を用いた未知語処理に基づいた円滑な対話応答システム」の概要とその評価について述べた。