

TAKE:情報提供エージェント

竹本祥柄 浦谷則好

東京工芸大学工学部コンピュータ応用学科

1. はじめに

対話とは「言語の発達によって共通の話題をやり取りするコミュニケーション形態」[1]とされている。社会的存在である人間にとって「対話」は、生活を享受するために欠くべからざるものである。

今日、ユーザとのコミュニケーションが可能な様々な自然言語処理プログラムが開発されている。しかし、未だその多くは、自然な対話を実現しているとは言い難く、人間同士が行う「対話」のような「ことばを交わす」ものではなく、内容を理解せず返答しているような不自然さを覚える。

本研究では、内容を理解し返答するような、より自然な「対話」を行い、専門的な発話に対しても適切な返答が可能なシステムの構築を目的とする。

2. 提案手法

本提案手法における返答生成のフローを図 1 に示す。

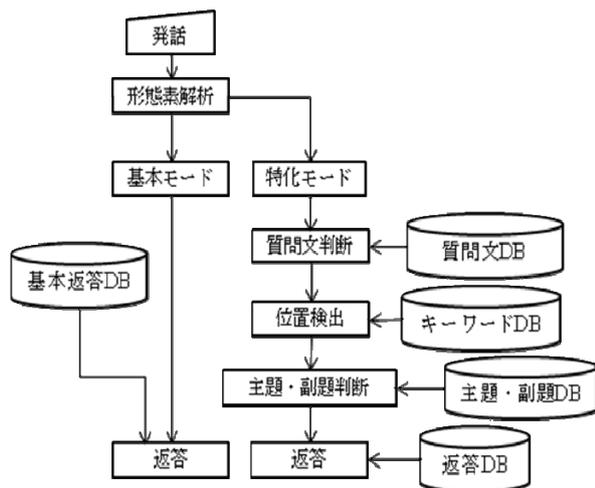


図 1 TAKE システム

まずユーザはシステムに対して発話する。システムは与えられた発話を、まず MeCab[2]による形態素解析を行い、単語の後の助詞の位置から主題を検出する。基本は与えられた発話に対して会話を繋ぐような返答をし、次の発話を待つが、与えられた発話に特定の単語が存在する場合、その単語に適したモード(特化モード)へ移行する。そして、質問文 DB とのパターンマッチを行い、発話がマッチした場合は、キーワード DB とパターンマッチを行う。質問文のパターンの必要に応じて、主題、副題、キーワードの位置を検出し、主題・副題 DB を用いて主題・副題を特定する。各条件を基に主題名で構成される返答 DB とのパターンマッチを行い、返答を生成する。質問文や各条件にマッチする返答が存在しない場合、基本動作に再度移行し、実行する。以下でそれぞれの処理と各ポイントについて記述する。

2.1. モードについて

本提案手法には「基本モード」と「特化モード」が存在する。

「基本モード」は、「こんにちは」という発話に対して「こんにちは！」と返すような、単純なパターンマッチ技法を用いており、これは Eliza の動作を参考に作成している。

「特化モード」は、より深い知識や推定を必要とする対話を実現するものであり、ユーザが特定の単語が含まれる発話をした場合に実行する。「主題」「副題」「キーワード」を記憶しながら、ユーザの発話に対し各 DB とパターンマッチを行うことで自然な対話となる文章を表示する。

また、本来であれば全ての話題に対して適切な対話を行うことが理想であるが、今回は時間的制限から「戦国時代」に特化した対話処理とする。

2.2. 主題・副題について

主題や副題になりうるものは「織田信長」のような固有名詞である。条件として形態素解析

の結果、品詞が名詞、品詞細分類 1 が「固有名詞」である単語が候補となる。条件に合う単語をあらかじめ主題・副題 DB に登録しておくことで主題・副題を認識する。

2.3. キーワードについて

「家臣」のような普通名詞を指す。条件として形態素解析の結果、品詞が名詞、品詞細分類 1 が「一般」である単語で、システムが単語として意味を理解できるものである。条件に合う単語をあらかじめキーワード DB に登録しておくことで認識している。

2.4. 質問文 DB について

質問文 DB は形態素解析での原型にあたる単語で構成されている DB である。省略可能である単語は()内に記入する。「Main」は主題、「Sub」は副題、「Keyword」はキーワードを示す。また、構成パターンは以下の 5 つである。一部を図 3 に示す。

1. 主題, 副題, キーワードが存在する
2. 主題, 副題が存在する
3. 主題, キーワードが存在する
4. 主題のみ
5. どれも存在しない

2.5. 返答 DB について

返答 DB は主題となりうる単語名が付けられた DB である。1 列目に質問番号, 2 列目に副題, 3 列目にキーワードが記入されており, 各条件とマッチした場合, 4 列目の返答文を表示する。一部を図 4 に示す。

3. 実行結果

実行結果の一部を記述する。
以降、>から改行までがユーザの発話である。

私に何か話しかけてください!

> こんにちは

こんにちは!

> 織田信長ってわかりますか?

わかりますよ! 有名な武将ですもの。

> 息子は誰がいますか

織田信忠、信雄、信孝、四男に羽柴秀忠を挟んで、勝長、信秀、信高、信吉、信貞、信好、長次、疑問視されていますが信正の 12 人です。

> 娘は誰がいますか

徳姫、冬姫、秀子、永姫、報恩院、三の丸殿、於振、鶴姫、源光院、月光院、正確性に欠けませんが、足利夫人と慈眼院がいます。

> 亡くなったのは何年でしたっけ

1582 年 6 月 21 日。本能寺の変のときですね。

> 信長は稲生の戦いをしたんですよね?

はい。織田信長と弟の信勝の家督争いから起きた戦いですね。稲生合戦、稲生原合戦とも呼ばれています。

> 場所はどこですか

尾張国稲生ですね。現在の名古屋市西部です。

> そうなんですか!

そうなんです!

4. まとめ

本稿では「対話」を通して、ユーザとのコミュニケーションを実現する 1 つの手法を提案した。課題として、現状では主題、副題が 3 つ以上存在する文には対応できないこと、DB の入力の手動であり、数を揃えるのに時間がかかる点であり、これを自動化することが今後の課題として考えられる。

参考文献等

[1] 言語処理学会 編: 言語処理学事典, 共立出版(2009)

[2] MeCab

<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

11 (Main	は)	Sub	を	する	End
8 (Main	は)	何	年	亡くなる	End
9 (Main	は)	何	年	生まれる	End
7 (Main	の)	Keyword	は	何	End

図 3 質問 DB(一部)

5 Null	側室	生駒吉乃、興雲院、原田直子、坂氏、養観院、土方氏、慈徳院、あここの子がいますね。
6 Null	生年月日	諸説ありますが、1534年6月23日ですね。
9 Null	Null	諸説ありますが、1534年6月23日ですね。
6 Null	没年	1582年6月21日。本能寺の変のときですね。

図 4 返答 DB(一部)