

244 遺伝的アルゴリズムを用いた帰納的学習による 雑談を対象とした音声対話システム

木村泰知† 荒木健治‡ 桃内佳雄† 枅内香次‡
北学園大工† 北大工‡

1 はじめに

最近、自然な対話を行うための音声対話システムの研究 [1] が多く行われている。しかし、対話には観光案内のように目的が明らかな状況と、雑談のように目的が不明確な状況が存在する。普段の生活でも何気ない雑談 [2] を行っていることは非常に多く、話しかけた目的を尋ねた方が、対話の流れが不自然になる場合もある。

雑談は、目的が達成されることよりは、自然に“対話すること”、“対話し続けること”自体に意味がある。精神科医のインタビュー代行システムとして ELIZA [3] がある。ELIZA が行うインタビューは雑談である。ELIZA はキーワードからスロット法を用いて文を生成する。キーワードで行うため、音声認識の誤りが存在する音声対話システムにおいても、頑健な応答ができる。ELIZA を音声対話システムに改良し、学習を導入した研究 [4] もある。しかし、テンプレートに代入する単語をユーザから獲得する学習では、予め用意したテンプレート以上のことはできず、対話を継続すると同じようなパターンの応答が繰り返し行われ、ユーザが不自然に感じてもそのまま対話が続けられるという問題がある。この問題点を解決するために、本稿では“遺伝的アルゴリズムを用いた帰納的学習手法 [5]”を雑談に応用した結果について述べる。

人間は対話を行いながら、多くのことを学ぶ。本手法でも、対話を行いながら学習を進めることが重要と考えている [6]。学習を行うためには、ユーザとシステムの対話を継続し、多くの対話例を収集することが必要になる。また、学習が進んでも、質問応答システムにおいてデータ及びルールの不足により応答ができない場合もある。本システムは、そのような場合には ELIZA 的応答システムが起動し、対話を継続させる。継続方法は、

1. キーワードの存在しない場合の応答
2. キーワードに対応した応答

を予め用意して、雑談に適した応答を行う。

ELIZA 的応答システムの役割は、次の 2 つにまとめることができる。

1. 対話の継続
2. 対話例の収集

システムは雑談をしながら対話例を収集し、対話例から遺伝的アルゴリズムを用いた帰納的学習により対話処理に必要な知識を獲得する。その際に、内容に関

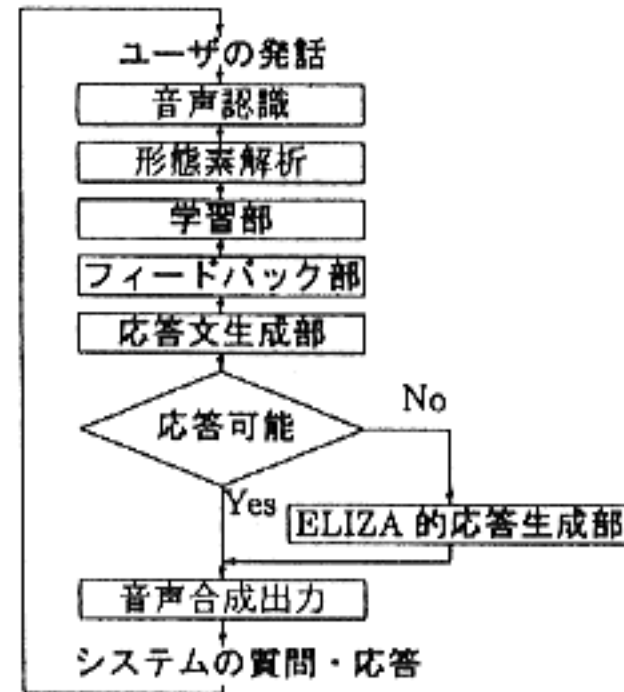


Fig. 1 処理過程

する知識も獲得するため、ELIZA より内容に立ち入った応答も行える。学習が進むにつれて、獲得したルールを用いて応答できるようになるので、ELIZA 的応答システムは、その起動頻度が減少する。

2 処理過程

図 1 に処理過程を示す。ユーザによる発話を市販の音声認識ソフトで認識する。ユーザの発話の音声認識結果を形態素解析ツール JUMAN [7] を用いて形態素解析し、その結果を利用して、自立語の抽出を行う。

学習部では、雑談に適した応答を行うために遺伝的アルゴリズムを用いた帰納的学習により応答文生成ルール及び、表層文生成ルールの獲得を行う。帰納的学習では、はじめに、ユーザとシステムの対話から自立語のみを取り出し、それらに対したルールを生成する。そして、それら 2 つの対に対して単語列及び、形態素解析結果の共通部分、差異部分を抽出し、この対の組の差異部分を変数としたルールと、この対の組より差異部分のみを抽出し対したルールの生成を再帰的に行う。表層文生成ルールは、右辺が表層文であり左辺が自立語の組及び、形態素列である。表層文生成ルールは、応答文生成ルールから助詞等を付加し、自然な応答を行うルールであり、全ての対話例から獲得する。

遺伝的アルゴリズムの適用においては、1 つのルールを染色体とし、各単語を遺伝子とする。

遺伝的アルゴリズム適用の諸条件を以下に示す。

1. 環境：ユーザとの対話

Table 1 雑談の例

1 ユーザ :何か楽しい事ないかな?
2 システム:最近楽しい事って何かあった?
3 ユーザ :海へ行った事だね
4 システム:混雑してましたか?
5 ユーザ :すごい人だったよ

Table 2 自立語の抽出と応答文生成ルールの例

1. 自立語の抽出例				
①	何	楽しい	事	ない
	名詞	形容詞	名詞	形容詞
②	海	行った	こと	
	名詞	動詞	名詞	
2. 応答文生成ルール				
①	海		混雑	
	名詞		名詞	
②	行った		してました	
	動詞		動詞	
③	@0	@1	事	@0 @1
	名詞	動詞	名詞	名詞 動詞
3. 応答文生成ルールの獲得				
@0	してました		すごい	@0
名詞	動詞		形容詞	動詞

2. 初期集団の発生:獲得した全ルール
3. 個体の適応度の計算:フィードバックを利用
4. 選択:品詞の並びが同一なルール
5. 交叉:同一の品詞の並びで一様交叉
6. 突然変異:突然変異率2%で突然変異を行う
遺伝的アルゴリズムのランダム性を用いて、雑談に適したルールを生成することができる。本手法では、品詞情報を用いて交叉を行い、信頼性の高いルールを獲得している。

フィードバック部では、予め誤応答の判定を行うために誤りキーワードを決めておき、ユーザ発話の音声認識結果に、誤りキーワードが存在した場合には、使用されたルールの誤応答度数に1を加え、存在しない場合には、自然な対話が行われていると考え、正応答度数に1を加える。

応答文生成部では、ユーザ発話の音声認識結果と応答文生成ルール的一致率や、フィードバックで得られた情報を利用して、最適な応答文生成ルールを選択し、表層文生成ルールを用いて助詞等を付加する。もし、ルールが不足していて応答文を生成できない場合には、ELIZA的応答生成部において応答文の生成を行う。生成された応答文は、市販の合成音声応答ソフトにより音声として出力する。

3 実行例

システムの動作を示すために表1の雑談の例を用いて説明する。システムは、ユーザ発話の音声認識結果を形態素解析し、表2の1のような自立語を抽出する。表1の2行目のシステムの応答は、適した応答文生成ルールが存在しなかったため、前のユーザ発話の音声認識結果を引用し、ELIZA的応答を行ったものである。4行目の応答は、表2の2の応答文生成ルールの3つの応答文生成ルールが存在したので、③のルールの変数"@0"及び"@1"に①②を代入し、表層文生成ルールにより助詞等を付加し生成されたものである。このように、学習を行いながら、対話を続けていくことにより、表2の3のように抽象的なルールの獲得も行える。

4 おわりに

本稿では、システムが実際にユーザと雑談を行い、その対話例から学習を行う手法について提案してきた。今後は、実際の人間同士の雑談と本手法による雑談を比較し、本手法の有効性を確認したい。

謝辞 なお、本研究の一部は科学研究費(No.09878070, No.10680367)及び北海学園大学ハイテク・リサーチ・センター研究費による補助のもとに行われた。

参考文献

- [1] 伊藤, 小暮, 中川: 協調的応答を備えた音声対話システムとその評価, 情報処理学会論文誌, Vol.39, No.5, pp1248-1257(1998).
- [2] 岡田: 何気ない行為を科学する, bit, Vol.30, No.12, 共立出版(1998)
- [3] WEIZENBAUM J. Computational Linguistics. ACM 9, No.1, (January, 1966), 36-45.
- [4] 加藤, 谷, 桐越, 菅田 "音声認識・音声合成を用いた Eliza 型 3DCG 会話ロボット" 茶飲み友達" に関する検討", 情処第 57 回全大, 6C-3(1999-10)
- [5] 越前谷, 荒木, 桃内, 棚内: 実例に基づく帰納的学習による機械翻訳手法における遺伝的アルゴリズムの適用とその有効性, 情報処理学会論文誌, Vol.37, No.8(1996).
- [6] 木村, 荒木, 桃内, 棚内: 遺伝的アルゴリズムを用いた多段階帰納的学習法による対話システムの評価, 情報処理北海道シンポジウム'99, pp31-34(1999)
- [7] 黒橋, 長尾: 日本語形態素解析システム JUMAN version3.61 1999