

17 遺伝的アルゴリズムを用いた多段階帰納的学習手法による 対話システムの評価

木村泰知^{*} 荒木健治[†] 桃内佳雄[†] 柄内香次[†]
(北学園大工)[†] (北大学工)[†]

1 はじめに

高度情報化社会の進展により情報の伝達は、重要な役割を果たしている。最近では、計算機が一つの独立した行為者(エージェント)として人と深い関係を持つようになり、人と計算機の意味疎通はさらに重要性を高めている。主に人間の通常の情報伝達是对話によってなされる事が多いため、知的なマンマシンインターフェイスを実現するためには自由でかつ個人に適應する対話システムの実現は重要な課題である。

自由な対話は最初から知識を与えておくだけでは表現・応答に制限が有り未知語の処理が難しくなる[1]。従来のシステムは未知語の処理としてすべての可能な表記を予め辞書に登録することにより処理するか、未知語が出現する度毎に辞書内にあるもので代用することが多かった[2]。しかし、既存の辞書だけでは限られたものなので、ユーザーの負担が大きくなる。それを解決していくためにも計算機自体が各ユーザの対話に適應していくことが必要である

解決策の一つとしてユーザに対して未知語の意味を尋ね、その意味を学習することによって、計算機側で「言い換え」[3]を行ったルールを作成し、応答できるようにする方法が考えられる。この処理を繰り返すことによって、ユーザに依存したシステムが作成でき、ユーザに適應した表現を多く生成することができると考えられる。

そこで、本稿ではこのような考えを実現する方法として「遺伝的アルゴリズムを用いた多段階帰納的学習手法による対話システム」を提案する。本手法はルール中に存在しない表現や応答は、システムが能動的に質問する[4]ことにより獲得する。本手法に基づくシステムは獲得したルールに対し、意味表現変換ルールによる言い換え、遺伝的アルゴリズムを用いた帰納的学習[5]により、さらにルールを生成し、多くの表現に対応できるようになる。また、協調的応答により類似した表現が出現した際にはその時点でシステムが持っている知識をユーザに回答することができる。

実験データは「Wizard of Oz方式」[6]を用いて集めた模擬対話を使用した。実験は「札幌観光案内」を題材とした対話を用いて行った。キーボード入力で行ない本手法の有効性を確認した。

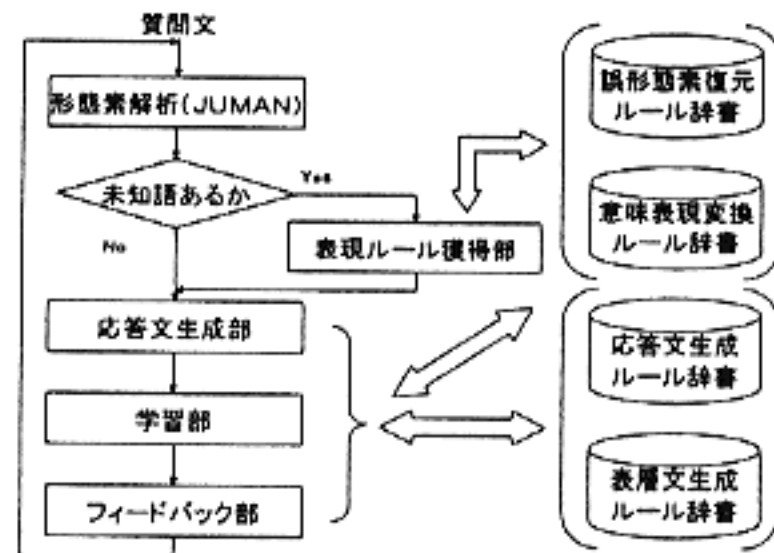


Fig. 1 処理過程

2 処理過程

2.1 処理過程の概要

本手法の処理過程を Fig. 1に示す。質問文を形態素解析ツール Juman [7]で解析し、自立語を抽出する。意味表現変換ルール辞書に登録されていない自立語を未知語とし、未知語が存在した場合には表現ルール獲得部により意味表現変換ルールを獲得する。次にその獲得されたルールを用いて学習を行う。応答文生成部で質問文に対する応答文を作成し出力する。その応答に対するユーザの応答により、学習部・フィードバック部の処理が行われる。辞書は意味表現変換ルール辞書、応答文生成ルール辞書、表層文生成ルール辞書、誤形態素還元ルール辞書からなる。本手法のルール格納方式を Table 1に示す。システムとユーザが質問と応答を繰り返しながら対話を進めていく。

最初に与えておく知識は

1. 意味表現変換ルールを獲得するための「～の他の表現を教えてください」
2. 応答文生成ルールを獲得するための「～の答え方を教えてください」
3. フィードバック部と表現ルール獲得部のための「わかりました」「違います」等の23単語のみである。

2.2 ルール獲得例

システムが対話よりルールを獲得する例を Fig. 2に示す。ここで、システムの処理として「スーパーはどこですか」の質問文を例に挙げる。質問文は形態素解析され自立語「スーパー」と「どこ」が抽出される。「スーパー」が未知語であった場合には表現ルール獲得部により他の

^{*}kimura@ai.eli.hokkai-s-u.ac.jp

[†]札幌市中央区南26条西11丁目北海学園大学工学部

[‡]札幌市北区北13条西8丁目北海道大学工学部

Table 1 ルールの格納方式

意味表現変換ルール			
見出し語	単語列	正	誤
応答文生成ルール			
質問文の単語列	応答文の単語列	正	誤
表層文生成ルール			
自立語の単語列	表層文の単語列	正	誤
誤形態素還元ルール			
形態素誤り単語列	正しい単語列	正	誤

正：正生成度数

誤：誤生成度数

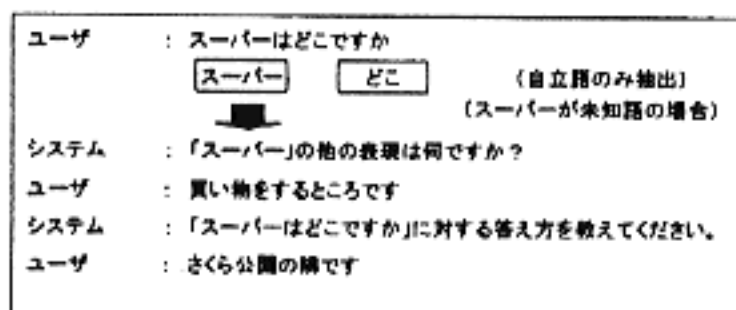


Fig. 2 対話からのルールの獲得例

Table 2 獲得されたルール例

意味表現変換ルール	
スーパー	買い物する ところ
応答文生成ルール	
スーパー どこ	さくら公園 隣
買い物する ところ どこ	さくら公園 隣
表層文生成ルール	
スーパー どこ	スーパー は どこ ですか
買い物する ところ	買い物する ところ です
さくら公園 隣	さくら公園 の 隣 です
買い物する ところ どこ	買い物する ところ は どこ ですか

正生成度数, 誤生成度数の初期値は0である.

表現を質問する. その結果 Table 2の意味表現変換ルールを獲得する. 応答文生成部で「スーパー」が使用されている応答文生成ルールを検索し, 存在する場合には, 置き換えて新しいルールを作成する. その後, 全ての応答文生成ルールと質問文を比較する. 適合するルールが存在する場合には応答し, 存在しない場合にはユーザ

に質問に対する応答の仕方を質問する. この対話例から Table 2に示すルールを獲得することができる.

2.3 各アルゴリズムの説明

2.3.1 表現ルール獲得部

ユーザからの質問文を形態素解析ツール Juman [7]で解析し, 自立語を抽出する. 自立語が意味表現変換ルール辞書に存在しない場合には, ユーザに質問する. 意味表現変換ルールとは「言い換え」を行なうための辞書で, 他の表現方法をユーザの応答から獲得する. 誤形態素還元ルールは意味表現ルール獲得の際に誤った形態素解析結果において, ユーザが応答出来ない際に, 形態素の区切り方を質問し, 獲得する. 例えば, 「えーと」などは「え」「ー」「と」に解析され, システムが他の表現を質問した場合, 応答が困難なためユーザに区切り方を修正してもらう.

2.3.2 応答文生成部

応答文生成部では意味表現変換ルールを既に獲得した応答文生成ルールや表層文生成ルールに対して置き換えを行い新たなルールを作成する. 質問文を自立語のみで考え, 応答文生成ルールとの一致率を計算する. 完全に一致したものは表層文生成ルールで処理され応答文として出力される. 一致率が60%以上であった場合はシステムが協調的応答を行う. 協調的応答は獲得したルールで最適な応答を試みるものである. 60%以下の場合ではシステムが質問文に対する応答の仕方をユーザに質問する. 質問文とユーザからの応答により, 応答文生成ルールが作成される. 表層文生成ルールは全てのユーザの入力文に対し行われ, 入力文から自立語のみ抽出した単語列と入力文全ての単語列の対応をルールとして獲得する. 応答は最終的に表層文生成ルールで行われる. 表層文生成ルールは応答文生成ルールとの一致率80%以上のルールを使用する. 応答文生成ルールと表層文生成ルールの選択の際に一致率が重複した際の最適なルールを選択する規準を以下に示す.

1. 正生成率が最大のルール
2. 正生成度数が最大のルール
3. 誤生成度数が最小のルール
4. 最も新しく獲得されたルール

2.3.3 学習部

遺伝的アルゴリズムを用いた帰納的学習 [5]を行なっている. 帰納的学習は2つのルールに対して単語の共通部分, 差異部分を抽出し, 差異部分が一つであるルールの組を検索する. このルールの組より差異部分を変数としたルールと, 差異部分のみのルールを作成する. 遺伝的アルゴリズムは1つのルールを染色体とし, 各単語を遺伝子として, 交叉を行なう. 交叉は2つのルールに対して単語の共通部分を検索し, 存在した場合には共通部

Table 3 実験結果

ルール名	実験者 1	実験者 2	実験者 3
意味表現変換ルール	61	36	62
表層文生成ルール	3,063	467	2,293
応答文生成ルール	110,812	823	3,480
誤形態素還元ルール	10	19	16
総対話数	240	139	206

分を交叉点として交叉を行う。突然変異は、乱数を用い、突然変異率を2%として単語に対するランダムな変更を行う。淘汰処理はフィードバック部で行なっている。

2.3.4 フィードバック部

フィードバック部ではシステムが質問に対して答えた結果にユーザが「わかりました」等の答えをした場合にはその際に用いられたルールの正生成度数を+1し、「違います」等の答えをした時にはその際に用いられたルールの誤生成度数を+1する。5回以上の使用で25%以下の場合のルールは淘汰する。正生成度数と誤生成度数は応答文生成部において応答するルールを決定する際にも用いられる。

3 実験

3.1 実験方法

人を機械になぞらえてVirtualなシステムとみなす「Wizard of Oz方式」[6]を用いて、理系の大学院生及び大学生の男性3人から「札幌観光案内システム」として集めた模擬対話を使用し実験を行なった。事前に丁寧な質問をするようにユーザに要求し、札幌観光について、知りたい情報を質問してもらった。そのユーザ役とシステム役の対話を書き下ろし、対話データとした。

3.2 実験結果

最終的に、獲得したルール数と総対話数をTable 3に示す。総対話数が多いほど応答文生成ルールと表層文生成ルールが多く作成されている。システム側からの有効な応答は実験者1のデータを使用した際の、一回であった。

4 考察

応答に最も影響を及ぼすルールは応答文生成ルールである。総対話数を基準に応答文生成ルールの獲得数をみた場合、実験者1と実験者3では、対話数の差は34対話であるのに対して、応答文生成ルールの獲得数の差は7000以上である。これは、同じ単語を含むルールが存在した事と、実験者1がひとつの観光地について詳しく質問したのに対して、実験者3は様々な観光地を質問したためと考えられる。実験者1の応答例でルールの獲

Table 4 対話例 1

U: 羊ヶ丘展望台はどこですか
S: 「羊ヶ丘展望台」の他の表現は何ですか?
U: 札幌を一望できるところ

Table 5 対話例 2

S: 「羊ヶ丘展望台には何がありますか」に対する答え方がわかりません。教えてください。
U: クラーク像のほか、展望台内のレストハウス、ウェディングパレスなどがあります

Table 6 対話例 3

U: 札幌の街を一望できるところはありますか
S: 「札幌一望できるところには何がありますか」が「クラーク像のほか、展望台内のレストハウス、ウェディングパレスなどがあります」と言う事は知っていますが「札幌の街を一望できるところはありますか」はわかりませんので、教えてください

得を考察する。その対話例をTable 4, Table 5, Table 6に示す。対話の流れは昇順に進んでいく。まず、対話例1により「羊ヶ丘展望台」の表現変換は「札幌を一望できるところ」という意味表現変換ルールを獲得した。その意味表現変換ルールを対話例2で獲得されたルールにより「羊ヶ丘展望台には何がありますか」の応答文は「クラーク像のほか、展望台内のレストハウス、ウェディングパレスなどがあります」である。その応答文生成ルールに対して意味表現変換ルールを置き換えることにより新しい応答文生成ルールを獲得する事ができる。その獲得したルールにより協調的応答をしたのが対話例3である。このように、能動的に質問し、獲得したルールを利用することによって、多様なルールを作成し、個人の表現に迅速に適応していく事ができる。

5 おわりに

本稿では、ユーザに適応した表現を生成する方法として「遺伝的アルゴリズムを用いた多段階帰納的学習手法による対話システム」を提案した。実際に実験対象者の質問は様々で、表現の違いも多い。そのためにも最初から知識を与えるのではなく、獲得方法のみを与える本手法は、ユーザの表現を学習する事により個人への適応が迅速に行うことができる。

今回の実験では、入力文が不足して、十分な学習が行われなかったため、システムからの応答は、ほとんどなかったが、応答の仕方及び学習が行える可能性が示され

た。しかし、他の表現をユーザに質問する際に、言い換えるのが困難な質問が存在した。今後は、本稿での問題点を修正し、音声対話に発展させ、同一ユーザによる様々な分野の対話を実験し、本手法のユーザへの適応の有効性を明らかにしていく予定である。

謝辞 なお、本研究の一部は文部省科学研究費 (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] 伊藤敏彦, 小暮悟, 中川聖一: 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] 亀田弘之, 桜井友子: ベタ書き日本語文からの未知語獲得システム, 電子情報通信学会技術研究報告, Vol.94 (TL94-11), pp17-24(1994).
- [3] 佐藤理史: 論文表題を言い換える, 情報処理学会研究報告, Vol.98 (98-NL-127), pp187-194(1998).
- [4] 安倍直樹, 中村篤祥: 能動的学習概要, 情報処理学会誌 Vol.38, No.7, pp559-561, July (1997).
- [5] 越前谷博, 荒木健治, 桃内佳雄, 栃内香次: 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] 堂本修司, 新美康永, 白井克彦, 田中穂積, 溝口理一郎: 音声による人間と機械の対話.
- [7] 松本祐治, 黒橋禎夫, 山地治, 妙木裕, 長尾真: 日本語形態素解析システム JUMAN 使用説明書 version3.4 1997