

コミュニケーション支援を目的とした帰納的学習による ペットを対象とした対話処理手法の性能評価

前田 宏幸[†] 荒木 健治[†] 栃内 香次[‡]

[†]北海道大学大学院工学研究科 〒060-8628 札幌市北区北13条西8丁目
北海道大学大学院工学研究科

電子情報工学専攻情報メディア工学講座信号処理工学分野

[‡]北海学園大学大学院経営学研究科 〒062-8605 札幌市豊平区旭町4-1-40

E-mail: [†]{maeda_h, araki}@media.eng.hokudai.ac.jp, [‡]tochinai@econ.hokkai-s-u.ac.jp

あらまし 現在ペットは多くの家庭で飼育されており、人間の言語を利用したペットとのコミュニケーションツールの実現が望まれている。しかし、動物の鳴き声のみを利用した翻訳等の方法で満足度の高いコミュニケーションを実現することは困難である。そこで、我々はこれまでペットとして飼育される動物を対象とした対話処理手法を提案している。本手法では、人間の発話と動物の行動の組からなる入力と動物側の返答である応答文としての出力の組から応答規則を帰納的学習により獲得し、飼い主とペットとの対話を用いた満足度の高いコミュニケーション手段を実現する。本稿では実験システムを作成して実験を行った結果から、提案した手法の有効性と動物の状態を入力に利用したことの影響について述べる。

キーワード 帰納的学習, 対話処理, ペット動物, コミュニケーション

Performance Evaluation of Dialogue Processing Method Using Inductive Learning Intended to Support a Communication with Pet Animals

Hiroyuki MAEDA[†] Kenji ARAKI[†] and Koji TOCHINAI[‡]

[†] Graduate School of Engineering, Hokkaido University

Kita 13 Nishi 8, Kita-ku, Sapporo-si, Hokkaido, 060-8628 Japan

[‡] Graduate School of Business Administration, Hokkai-Gakuen University

4-1-40 Asahi-machi, Toyohira-ku, Sapporo-shi, Hokkaido, 062-8605 Japan

E-mail: [†]{maeda_h, araki}@media.eng.hokudai.ac.jp, [‡]tochinai@econ.hokkai-s-u.ac.jp

Abstract It is wished to realize the communication tool with pet animals used human language. But, it is difficult to realize the satisfied communication in the method such as translation used only the roar of the animal. Therefore, we propose a dialog processing method using inductive learning intended for pet animals. This method acquires answer rules by inductive learning from the input which consists of human utterance and the animal's behavior, and the output as an answer of the animal side. As a result, it realizes the satisfied communication by the dialog with the owner and the pet animal. In this paper, we state the validity and the influence on use of the animal's behavior as the input from the experiment by the system.

Keyword Inductive learning, Dialog processing, Pet animals, Communication

1. はじめに

現在、ペットとして飼われている動物は人間社会の精神の健康を保つために重要な役割を果たしている。人間同士のコミュニケーションが疎遠になり、孤独な状況に陥りやすい現代社会において、人間の心理的な支えとしてペットの果たす役割は大きく、その重要性は今後ますます増加していくと考えられる。このことから、動物とのより良い信頼関係を築くためのコミュニケーション支援として、人間の言語を利用したペッ

トとのコミュニケーションツールの実現が望まれている。

人間の言語を利用したペットとのコミュニケーションを実現する方法として第一に考えられるのは、ペットの鳴き声を人間の言語に翻訳することである。翻訳によって満足なコミュニケーションを行うためには、ペットである動物が音声を発することによって多彩な意味を伝達している必要がある。人間の言語には文法が存在し、これによって有限の記号から無限の意味を

表現できると考えられているが[1], ペットである動物にも人間の文法に近いものが存在する知見として, ある種の小鳥(ジュウシマツ)は, 複雑な産出規則(文法)に基づく歌をうたうことが報告されている。しかし, この小鳥の歌に見られる(文法)は人間の言語の文法とは性質が異なり, 人間の言語では単語に相当する小鳥の歌の要素がそれ自体は特定の意味と結びついてはいない。小鳥の(文法)は様々なバリエーションの歌をうたうことを可能にするために存在し, 複雑な意図を伝達することは不可能である[2]。このように現時点では, 動物で人間の文法に匹敵するものを持つという事例は報告されていない[3]。

その他に, 複雑な意味を音声によって伝達することを可能にするためには, バリエーション豊かな音声を発することができる調音構造が備わっている必要がある。人間は二足歩行を行うことで調音に関わる骨格の構造に制限が無くなった結果, 自在に調音を行うことが可能となっており, 人間に近い知能を持つといわれるチンパンジーでも人間のようなバリエーションのある音声を発することは不可能である[4]。

以上のことから, ペットの鳴き声を直接日本語へ変換しても, その日本語は具体性を欠いた曖昧なものとなり, ペットの飼い主にとって不満を感じるものになってしまうと考えられる。結果として, ペットの鳴き声のみを用いた翻訳によって, 満足度の高いペットとのコミュニケーションツールを実現することは困難であると考えられる。

ペットと飼い主である人間との間におけるコミュニケーションの代表的なものとして, 言語によるペットとの対話が挙げられる。人間はペットである動物に話し掛け, そのときの動物の反応から, 動物の意図することを解釈している。このとき, 人間が解釈した意図は本当に動物が思っていることであるとは限らず, 実際には解釈した意図は飼い主の人間が動物の反応を人間側の立場から意味付けたものであると考えられる[5]。このことから, 人間の解釈した動物の意図は多種多様なものとなり, その解釈した意図をペットである動物の応答とする対話処理が, 現時点で実現可能な人間の言語を利用したペットとのコミュニケーションツールであると考えられる。

以上のような背景から, 我々は人間の言語を利用したペットとのコミュニケーションツールとして, 帰納的学習によるペットを対象とした対話処理手法を提案している。本手法では, 人間の発話と動物の状態の組からなる入力と飼い主である人間の解釈による動物側の返答となる応答文から, 応答規則を帰納的学習[6]により獲得する。この方法により, ペットである動物と飼い主である人間との間で対話処理を行うことによ

って, 従来では存在しなかった, 言語によるペットとの満足度の高いコミュニケーションの実現を目指す。本稿では作成したシステムによるシミュレーション実験において, 動物の状態を入力に利用することで, 動物の状態を利用しない場合と比較してより具体的な満足度の高い応答文が出力されることを確認し, また実験環境下においてもシミュレーション時と同様な有効性があることを確認するために, システムの構築と実験を行ったので, その結果について述べる。

2. 対話処理部の概要

本手法の対話処理部分の流れを図1に示す。帰納的学習を用いた対話処理手法[7]では, 過去の2つの対話例から字面上の共通部分と差異部分を見つけだし, その組み合わせを応答文生成ルールとして獲得し, ルール辞書に登録する。そして, このときにすでに登録されている辞書中のルールとも比較を行い, 共通部分と差異部分を見つけだすことによって再帰的にルール獲得を行う。これにより, 比較的少数の事例から対話応答を行うことが可能となり, 例外的な事例に対してもそれをルールとして保持することで対処することが可能となる。過去の対話例からのルール抽出例を図2に示す。入力文から日本語形態素解析ツール JUMAN[8]を用いて自立語列が生成され, 文中の自立語がルール抽出に利用される。また, 自立語列と入力文の組から応答文復元ルールが生成される。

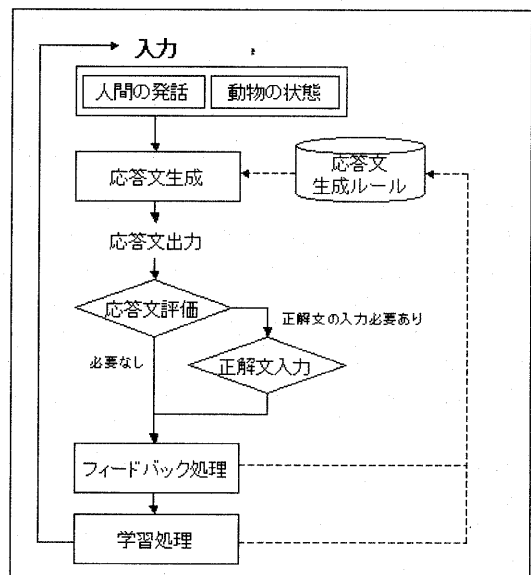


図1: 対話処理の流れ

	入力	出力
対話例1	今日のごはんはおいしいですか	今日の魚は大好きです
対話例2	今日のおやつはおいしいですか	今日のマタタビは大好きです



抽出されるルール1	今日 @ おいしい: 今日 @ 大好きです
抽出されるルール2	ごはん: 魚
抽出されるルール3	おやつ: マタタビ

図2: 対話例からのルール抽出例

対話例から抽出されたルールには、対話スコアと状況スコアが記録される。対話スコアは応答文が対話として成り立つ度合いを示す尺度で、状況スコアは応答文が現在の動物の状態と対応しているかを判断する尺度である。本手法で入力として利用している動物の状態は以下の通りで、それぞれの状態が複合したものを入力することが可能である。利用する動物の状態については藤崎[5]の研究を参考にした。

- (1) 距離 (近い, 遠い, 中間の3段階)
- (2) 鳴き声
- (3) しっぽを振る
- (4) 発話者を見る

状況スコアは動物の状態毎に値を個別に保持している。これらのスコアは応答文を生成する時のルール選択に利用される。

応答文生成時に使用可能なルールの組み合わせが複数存在する場合、以下の評価式による尤度計算を行い使用するルールを決定する。

$$\text{尤度} = \sum_{\text{使用するルール}} \left[(\alpha * DS) + \beta * \sum_{\text{動物の状態}} ST \right]$$

(DS:対話スコア, ST:状態スコア, $\alpha=0.3, \beta=1.0$)

使用可能なルールが複数存在する場合に、使用するルールを決定する例を図3に示す。この例では「こっちに来てちょうだい」という入力に対して、ルール1と3、ルール2と3という2つのルール選択候補が存在するが、入力時の動物との距離が近い場合には、尤度計算の際に近距離の状況スコアが利用された結果、ルール1と3の組み合わせが応答文生成に利用される。反対に、入力時の動物との距離が遠い場合には、遠距離の状況スコアが尤度計算に利用され、ルール2と3の組み合わせが応答文生成に利用される。なお、この例では説明の都合上状況スコアを近距離と遠距離のみとしたが、実際には本手法で入力として利用している

ルール	DS	ST:近距離	ST:遠距離
1 ①に来てちょうだい①に行きます	2	3	-1
2 ①に来てちょうだい①に行きたくありません	1	-1	3
3 こっちそっち	2	2	2

入力時に動物との距離が近い

入力時に動物との距離が遠い

使用ルール	評価式
1と3	$(2+2)*0.3+(3+2)*1.0=6.2$
2と3	$(1+2)*0.3+(-1+2)*1.0=1.9$

出力:そっちに行きます

使用ルール	評価式
1と3	$(2+2)*0.3+(-1+2)*1.0=2.2$
2と3	$(1+2)*0.3+(3+2)*1.0=5.9$

出力:そっちに行きたくありません

図3: 使用するルールの決定例

動物の状態全てに対して状況スコアが記録され、ルール選択の際に利用される。

フィードバック処理部では、応答文生成部によって出力された応答文に対してユーザが行った評価に基づいて、使用したルールに対してフィードバック処理を行い、対話スコアと状況スコアを増減させる。その結果、次回以降のルール選択においてより有効な応答文を出力する。詳しいフィードバックの仕様については実験の仕様を説明する際に述べる。

学習処理部では、ユーザが出力された応答文に対して校正を行った際に、その校正された応答文と応答文生成ルール辞書とを比較して、共通部分と差異部分に注目して帰納的学習によるルール抽出を行う。その際に、校正された応答文そのものを実例ルール、帰納的学習により抽出されたルールを抽出ルールと呼び、それぞれのルールの初期状態を以下のように設定する。

- (1) 実例ルールの初期状態:
DS = 1, 入力に対応する ST = 1
- (2) 抽出ルールの初期状態:
DS = 0, ST = 0

これにより、初期段階においてはルールの信頼度が高いと考えられる実例ルールの方が優先的に選択されやすくなる。

3. シミュレーション実験

2章で述べた手法をシステムに搭載し、システムの入力に動物の状態を導入したことによる影響を調査するため、オープンデータによるシミュレーション実験を行った。実験に際して、第一著者宅でペットとして飼育されている猫と第一著者以外の飼い主とのやり取りをビデオカメラに収録し、後日その映像を基に対話データとなる入力文、動物の状態、不正解時の正解文を手で作成した。総対話データ数は566発話となった。作成したデータからシステムが出力した応答文に

対して、第一著者が評価を行った。評価基準とそれぞれの情報のフィードバック方法は以下のとおりである。

- (1) 正応答：応答文が意味的に正しく、動物の状況と適応 (DS と入力に対応する全ての ST の値を 1 増加)
 - (2) 準応答：応答文が意味的に正しいが、状況と適応せず (DS を 1 増加, 全ての ST の値を 1 減少)
 - (3) 誤応答：応答文が意味的に誤り (DS と対応する全ての ST の値を 1 減少)
- ※助詞の重複、抜け等の文法の少々の間違いについては考慮しない。大きく誤っている場合には意味的に誤りとする。

ルール選択の評価式における動物の状態のパラメータである β の値を以下の 2 通りに設定して比較評価を行い、提案手法のシステムの入力に動物の状態を導入したことによる影響を調査した。学習の影響を確認するため、ルール辞書の初期状態は空の状態とした。

- ・ $\beta = 0.0$ (動物の状態を考慮せず)
- ・ $\beta = 1.0$ (動物の状態を考慮する)

シミュレーション実験の結果を表 1 に、動物の状態を考慮する場合としない場合とで評価が異なった発話の内訳を表 2、評価が異なった発話の例を表 3 に示す。結果を見ると、動物の状態を考慮しない場合と比較して正応答率が上昇しており、特に準応答から正応答に評価が改善された発話が多い。このことから、動物の行動を入力に用いない場合と比較して、より動物の状況に適応した具体的な応答が可能となっていることがわかり、提案した手法における入力に動物の状態を利用したことによる効果が現れたと考えられる。しかし、一方では誤応答率も上昇しており、特に準応答から誤応答に改悪された発話が多いことがわかる。これは、動物の状態を考慮しない場合と比較して対話スコアの影響が相対的に小さくなり、対話として成り立たない応答が増加したことが原因と考えられる。今後誤応答率の上昇を抑えるためにルール選択方法の改良などを行っていく必要がある。

また、学習が良好に行われていることの確認を行うため、全 566 発話中の後半 288 発話に対して、応答文生成ルール辞書を空の状態から評価したものと前半から継続しているものととの比較を行った。結果を表 4 に示す。この結果から正応答率が上昇し、未応答率が大きく減少していることがわかり、学習の効果が現れていると考えられる。しかし、一方では誤応答の上昇も大きく、これについてもルール選択方法などに改良を行っていく必要があると考えられる。

表 1：シミュレーション実験結果

	$\beta = 0$	$\beta = 1$
正応答	111(19.6%)	133(23.5%)
準応答	178(31.4%)	143(25.3%)
誤応答	90(15.9%)	101(17.8%)
未応答	187(33.0%)	189(33.4%)

評価変更	79(14.0%)
------	-----------

表 2：評価が異なった発話の内訳

改善数	43(54.4%)
-----	-----------

準応答⇒正応答	27(34.2%)
正応答⇒準応答	3(3.8%)

誤応答⇒準応答	11(13.9%)
準応答⇒誤応答	23(29.1%)

(評価変更総数 79 発話中)

表 3：評価が異なった発話の例

準応答⇒正応答の例

入力：おいでおいで ご飯

応答：やったご飯だ ⇒ ご飯はもういいです

(この時のペットとの距離は遠距離)

準応答⇒誤応答の例

入力：シンチャンはこっち来ないの

応答：ん？きたかな ⇒ ちょっと食べようきたかな

表 4：後半 288 発話の比較結果

	辞書空から	前半から継続
発話数	288	288

正応答	49(17.2%)	63(21.9%)
準応答	74(25.7%)	77(26.7%)
誤応答	34(11.8%)	55(19.0%)
未応答	132(45.5%)	93(32.3%)

4. 実環境下でのリアルタイム実験

次に、実環境下においてもシミュレーション実験時と同様の有効性が得られることを確認するために、実環境下で実験を行った。実験の際の状況設定は、ペットを写しているビデオカメラを固定し、その手前に飼い主がいる状況とする。これにより、背景差分でペットが写っている部分を判断することが可能となり、その画素数によってペットとの距離を区別することが出



図4：リアルタイム実験システムの稼動の様子

来る。以上のことから、システムの入力に使用する動物の状態を距離に絞ったリアルタイム実験システムの構築を行った。2章で述べた手法に、以下のような変更点を加えた。また、人間の発話の音声認識にはViaVoice[9]を使用した。

4.1. システムの追加部分の概要

システムはビデオカメラからの画像を100msec毎に320*240の解像度で取得する。その際に、ユーザは任意のタイミングでビデオカメラからの画像を背景画像として指定することができる。背景画像が設定されてからは、背景画像とビデオカメラからの画像を比較し、RGB(各256段階)の輝度値が全て50以上異なる画素を背景画像と異なる画素と判断し、ペットが写っている部分とする。そして、ペットが写っていると判断された画素数を $X(t)$ として、以下の条件で距離の認識を行い、それぞれに状況スコアの値を個別に保持させる。距離認識の条件については予備調査により設定を行った。リアルタイム実験システムの稼動の様子を図4に示す。

- (1) 近距離： $X(t) \leq 2000$
- (2) 中距離： $2001 \leq X(t) \leq 10000$
- (3) 遠距離： $10001 \leq X(t)$

そして出力された応答文に対して、満足度（以下に示す4段階の評価を表示）と応答文修正の必要があるかを尋ねる。修正の必要がある場合は満足度の評価に応じて、以下のようなフィードバック処理を行う。フィードバック処理の判断基準は、応答文に対して満足であるか不満であるかの2通りで判断し、その程度の差については考慮しない。なお、応答文修正の必要が無い場合には、対話スコアと状況スコアをともに1増加させるフィードバック処理を行う。

- (1) 非常に満足・・・DSを1増加，STを1減少。
- (2) やや満足・・・DSを1増加，STを1減少。
- (3) やや不満・・・DSを1減少，STの値を1減少。
- (4) 非常に不満・・・DSを1減少，STを1減少。

4.2. 発話内容限定による実験

実環境下においてもシミュレーション実験時と同様の有効性が得られることを確認するために、リアルタイム実験システムにおける学習の進み具合を調べることを目的とした、発話内容を限定した実験を行った。

第一著者宅でペットとして飼育されている猫に対して、飼い主である第一著者が「こっちにおいで」と呼びかけた後に出力される応答文に対して評価を行った。評価基準として、距離が近い場合には同意した内容の応答文、距離が遠い場合には拒否した内容の応答

表5：発話内容限定による実環境下での実験結果

	全体の結果	前半の結果	後半の結果
非常に満足	67.2%(86)	46.9%(30)	87.5%(56)
やや満足	7.0%(9)	12.5%(8)	1.6%(1)
やや不満	15.6%(20)	21.9%(14)	9.4%(6)
非常に不満	1.6%(2)	1.6%(1)	1.6%(1)
未応答	8.6%(11)	17.2%(11)	0.0%(0)

表6：距離毎に区別された代表的なルール

	DS	ST (近)	ST (中)	ST (遠)
入力：こっちにおいで 応答：いまはいません	18	-1	-2	18
入力：こっちにおいで 応答：どうしようかな	10	-1	11	0
入力：こっちにおいで 応答：そっちにいます	24	23	-1	0

文、その中間では迷っているような内容の応答文であれば満足度の高い評価を与えた。学習の影響を確認するため、ルール辞書の初期状態は空の状態とした。総発話数は128発話である。

128発話全体の評価結果と、前半と後半の64発話の評価結果を表5に、距離毎に区別された代表的なルールを表6に示す。前半よりも後半の方が評価が高くなっており、後半では未応答が無くなっていることから、実環境下においても学習が進んでいることが確認できる。また距離ごとに区別された代表的なルールを見ると、近距離・中距離・遠距離それぞれに状況スコアが高いルールが獲得されており、実環境下においても学習が進めば動物の状態(この場合は距離)に応じた応答ができるようになると思われる。

次に、音声認識誤りによる影響について考察する。「こっちにおいで」がそのまま入力となったのは59発話(46.1%)であり、単語認識率は68.9%であった。よく見られた音声認識誤りの例としては、「こっちにおいて」「国旗において」「こっちに多いで」等が挙げられる。音声認識誤りが存在した69発話の評価結果を表7に示す。前半35発話は「非常に満足」の評価が25.7%だったのに対して、後半34発話は73.5%となっており、発話に音声認識誤りが生じても学習によるルール獲得やフィードバックにより、距離に応じた満足度の高い出力がされるようになったことがわかる。

5. まとめ

本稿では、コミュニケーション支援を目的とした帰納的学習によるペットを対象とした対話処理手法を提案し、作成したシステムによるシミュレーション実験

表7：音声認識誤り69発話の評価結果

	全体の結果	前半の結果	後半の結果
非常に満足	49.3%(34)	25.7%(9)	73.5%(25)
やや満足	10.1%(7)	14.3%(5)	5.9%(2)
やや不満	23.2%(16)	28.6%(10)	17.6%(6)
非常に不満	2.9%(2)	2.9%(1)	2.9%(1)
未応答	14.5%(10)	28.6%(10)	0.0%(0)

を行った。シミュレーション実験の結果から、提案した手法によってペットの状態に基づいたより具体的な応答が可能になることを確認した。また、動物との距離をリアルタイムに認識する実験システムを作成し、実環境下における発話内容限定による実験を行った。その結果、実環境下においてもシミュレーション時と同様に学習が進み、動物の状態に応じた応答ができるようになることを確認した。今後の課題として、実環境下での実験においてより自由な発話での実験を行うこと、システムの使い勝手についての考察を行うことが挙げられる。また、対話処理部において誤応答を減少するための改善策として、ルール選択時の尤度評価式の改良や、動物への発話は全体的に短い文が多いことから、過去の発話と動物の状態を入力とルールに含ませることで発話のバリエーションを増加させ、より具体的な応答を行うことを目指すこと等が挙げられる。

謝辞 本研究の一部は文部科学省科学研究費補助金(第14658097号)によって行なわれている。

文献

- [1] 岡ノ谷一夫：“小鳥たちの歌の〈文法〉—アナロジーに基づく言語の起源の生物学—”日本音響学会誌 54巻6号 pp.457-465(2001-07)
- [2] 岡ノ谷一夫：言語の発生，渡辺 茂(編)，"心の比較認知科学"ミネルヴァ書房(2000)
- [3] 渡辺茂：“動物に文法はあるのか”，「言語」，大修館書店、2002年4月号
- [4] 正高信男：“言葉の誕生 行動学からみた言語起源論” 紀伊国屋書店(1991)
- [5] 藤崎亜由子：“人はペット動物の「心」をどう理解するか：イヌ・ネコへの言葉かけの分析から”発達心理学研究，2002，第13巻，第2号，109-121
- [6] K. Araki and K. Tochinal: "Effectiveness of Natural Language Processing Method Using Inductive Learning", Proceedings of the IASTED International Conference ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING, pp.295-300, May, 2001.
- [7] 木村他：“遺伝的アルゴリズムを用いた帰納的学習による音声対話処理手法”，電子情報通信学会論文誌，Vol.J84-D-II,No.9,pp.2079-2091(2001).
- [8] 黒橋禎夫，長尾真：“日本語形態素解析システムJUMAN”，version3.61,
- [9] ViaVoice <http://www.ibm.com/jp/voiceland/>