2D4-3

サービスロボットによるインタラクションを通した語彙の拡張

Increase the Vocabulary Using Interaction for Service Robots

渡邉祐太[*1] 田口亮[*1] 服部公央亮[*2] 保黒政大[*2] 梅崎太造[*1]
Yuta Watanabe Ryo Taguchi Koosuke Hattori Masahiro Hoguro Taizo Umezaki

[*1] 名古屋工業大学 Nagoya Institute of Technology
[*2] 中部大学 Chubu University

There's been an increase in research on service robots cooperating with humans using voice. Conventional robots don't receive the instruction and request repeated speaking when a user instructs in unknown wordings since recognizable wordings are designed previously. In this paper, we propose a method for a robot to increase an existing vocabulary using interaction to users by using the voice understanding section based on descriptive grammar and the word acquisition section using syllable recognition.

1. はじめに

近年，企業や大学など様々な研究機関でロボットが開発されていることや，音声認識・画像処理技術の発達に伴い,工業用ロボットだけでなく，福祉や介護，警備やアミューズメント，教育など幅広い分野でのロボットの活躍が期待されており，人に近い環境内で活動することが求められている．

特に家庭やオフィスで活動するロボットは，一般のユーザがインタラクションすることが想定される.初めてロボットに触れるユーザでも直感的にロボットとインタラクションすることが求められるため，人間が普段使用している言語を認識し応対する音声対話機能が必要となる．ロボットが人と対話するためには，言葉と実世界の事物・事象の対応関係をロボットが理解できなければならない．家庭やオフィスなどでは，未知の物や場所等に対応する必要があるため，それらを表す単語知識，すなわち語彙をユーザとのインタラクションを通して学習できることが望まれる．

ロボットによる語彙学習に関する先行研究では，ロボットに物や動作を見せながら対応する単語を発話することで，「箱」や「青い」と物を表す単語や，「乗せて」や「近づけて」といった動作を表す単語を学習させた[Roy 02]．

これらの研究では単語単位に区切られた発話や，決められた文法に沿った発話が学習に用いられてきた．しかし実運用を考慮すると，ユーザの自然な発話から学習できることが望ましい．

こうした背景から未知語のクラスが持つ音響的，文法的なモデルを学習・利用することで，発話に含まれる未踏力語を抽出する手法が提案されている[山本 04]．発話と指示対象の対応関係を，音響，文法，意味を統合したモデルで表現し，それを統計的モデル選択に基づいて最適化することで，単語の音素系列とその意味を学習する方法が提案されている．この研究を発展させ，実際にロボットが取得するセンサ情報のカテゴリ化を語彙学習と同時に行う手法も提案されている[田口 10]．

上記の手法では，初期知識ゼロからの学習を対象としており，トップダウンで与えられた対話知識を併用した学習モデルは提案されていない．そこで本研究では，開発者が設計した対話知識を持つロボットが，サービス遂行に係るユーザの発話から対話知識を拡張していく手法を提案する．

2. 提案手法

ロボットには事前に認識可能な単語と文法，およびそれに対する応答方法が与えられる．各単語にはその単語の意味を表す意味 ID が割り当てられるものとする．文法は発話時の意味 ID の順番が定義されている．例えば，“まえ”という単語には意味 ID“FORWARD”が与えられ，“にいって”という単語には意味 ID“MOVE”が与えられている．また，“まえにいって”と発話すると，“FORWARD”と“MOVE”という順番の文法に対応し，ロボットは前進することができる．

従来の音声対話システムでは，予め決められた単語や文法通りの発話でなければ認識・応答することができない．しかし，一般的な人間同士の対話では，意味は同じだが異なる言い表し方（本稿ではこれを「言い回し」と呼ぶ）が多く用いられ，それら全てを事前にシステムに与えることは困難である．そこで，本稿では登録されていない未知の音韻系列からなる単語と，その意味 ID を人とのインタラクションを通してロボットに学習させる．

提案手法の全体像を表す概略図を図1に示す．例えば，人がロボットに“まえにすすんで”と命令を行う．ロボットは命令を理解できなかったとき，再び命令するよう求める．人は“まえいって”とロボットが理解できる言い回しで命令を与える．ロボットは前に発話された命令が，今発話された命令と同じ意味 ID を示すとみなし，未知の発話に対し意味 ID を付与する．命令を複数回行った後，ロボットは聞き取った発話を辞書なし形態素解析により分節し，単語を学習する．提案手法全体の処理フローを図 1 に示す．形態素解析には，G. Neubig らの教師なし形態素解析を行う手法[Neubig 12]を用いる．

(1) 未知の命令　　(2) 意味の学習

図 1　提案手法の概略図

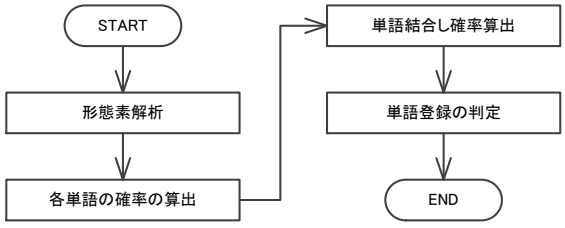

図 2　全体の処理フロー

連絡先：　渡邉祐太，名古屋工業大学工学研究科，
watanabe@ume.mta.nitech.ac.jp

2.1 単語の学習

複数回の発話データを蓄積，オフラインで学習を行う．形態素解析によって分けられた単語を，単語同士結合することで意味 ID に対応した一単語とする．そこで，本研究では DP マッチングを用いる．分割された単語wが現れたときに，意味 ID xである確率$P(x|w)$を求める．この確率はnを単語の出現頻度として，式(1)で求められる．

$$P(x|w) = \frac{n(x,w)}{n(w)} \tag{1}$$

求めた条件付き確率を用いて，文法とのマッチングを行う．発話文とそれに対応する文法をマッチングすることで，同じ意味 ID の単語同士を結合し，ロボットの単語辞書に登録する．

ここで，ローカルディスタンス$d(i,j)$は式(2)，グローバルディスタンス$D(i,j)$は初期条件を式(3)として式(4)で求められる．

$$d(i,j) = 1 - P(x_i|w_j) \tag{2}$$

$$\begin{aligned} &D(0,0) = d(0,0) \\ &D(i,0) = d(i,0)\ \ (i = 1,2,\cdots,I) \\ &D(0,j) = d(0,j)\ \ (j = 1,2,\cdots,J) \end{aligned} \tag{3}$$

$$D(i,j) = min\begin{cases} D(i-1,j)+1 \\ D(i-1,j-1)+d(i,j) \\ D(i,j-1)+1 \end{cases} \tag{4}$$

例えば，“まえにすすんで”という発話文は形態素解析により，“まえ/に/す/すんで”と分けられる．さらに，DP マッチングにより，“まえ/にすすんで”と単語が結合し，“まえ”には意味 ID “FORWARD”，“にすすんで”には “MOVE”が与えられる．

2.2 登録単語の決定

未知の言い回しでの命令の後，それと異なる意味の命令がされた場合がある．このとき，発話文には間違った意味 ID が与えられる．そこで，DP マッチングを行い結合された単語に対して，再び式(1)で表される確率$P(x|w)$を求め，確率が 0.7 以上のものを単語辞書に登録する．

3. 実験

今回の実験において，音声認識器には大語彙連続音声認識システム Julius[1] を利用した．Julius の単語辞書は，既存の大語彙が登録された辞書を用いず，表 1 の単語とそれに対応する意味 ID と文法，日本語音節を登録した単語辞書を使用し，1-best 認識によって結果を得る．マイクには，オーディオテクニカ社の AE6100 を使用した．形態素解析には，latticelm[2] を使用した．latiicelm のパラメータは初期設定のままとした．

表 1 単語辞書

単語	意味 ID	文法	行動
まえに	FORWARD	FORWARD MOVE	前進
うしろ	BACK	BACK MOVE	後進
みぎ	RIGHT	RIGHT ROTATE	右に 90° 回転
ひだり	LEFT	LEFT ROTATE	左に 90° 回転
いって	MOVE		
むいて	ROTATE		

[1]使用バージョン：dictation-kit-v4.0-win,
http://julius.sourceforge.jp/index.php
[2]使用バージョン：latticelm 0.4,
http://www.phontron.com/latticelm/index-ja.html

3.1 語彙学習

(1) 実験条件

本実験では，命令をテキストデータとして与えた場合と，話者1名が実際に発話した音声データを与えた場合の 2 条件で実験を行う．意味 ID が“FORWARD MOVE”，“BACK MOVE” “RIGHT ROTATE” “LEFT ROTATE”を示す命令をそれぞれに対し 9 個の言い回しを含む合計 36 回分の未知の発話をロボットに与えた．各発話文における言い回しを表 2 に示す．

表 2 発話文における言い回し

FORWARD MOVE	BACK MOVE
まえにすすんで	うしろにすすんで
まえすすんで	うしろすすんで
まえのほうにすすんで	うしろのほうにすすんで
まえにむかってすすんで	うしろにむかってすすんで
まえにいって	うしろにいって
まえへいって	うしろへいって
まえのほうにいって	うしろのほうにいって
まえむかって	うしろむかって
まえにむかって	うしろにむかって

RIGHT ROTATE	LEFT ROTATE
みぎみて	ひだりみて
みぎをみて	ひだりをみて
みぎのほうをみて	ひだりのほうをみて
みぎまわって	ひだりまわって
みぎのほうにまわって	ひだりのほうにまわって
みぎかいてんして	ひだりかいてんして
みぎにかいてんして	ひだりにかいてんして
みぎがわみて	ひだりがわみて
みぎがわをみて	ひだりがわをみて

(2) 実験結果

命令をテキストデータとして与えた場合に，登録された単語を表 3 に，音声データを与えた場合に登録された単語を表 4 に示す．テキストデータとして命令を与えた場合，すべての発話文が，方向を示す単語と行動を示す単語が分けられ，正しい意味 ID が与えられていることがわかる．音声データの場合でも，音声認識誤りは多く見られるが，方向を示す単語と行動を示す単語が分けられており，提案手法の有効性が確認できた．形態素解析でラティスを用いる等，音声認識誤りを吸収する手法を用いることで，より良い結果が得られるといえる．

表 3 登録された単語(テキストデータ)

FORWARD	BACK	RIGHT	LEFT
まえ	うしろ	みぎ	ひだり

MOVE	ROTATE
にすすんで	みて
すすんで	をみて
のほうにすすんで	のほうをみて
にむかってすすんで	まわって
にいって	にまわって
へいって	のほうにまわって
のほうにいって	かいてんして
むかって	にかいてんして
にむかって	がわをみて

表 4 登録された単語(音声データ)

FORWARD	BACK	RIGHT	LEFT
わ	うしろ	みぎ	ひだり
あまえ	むしろ	みぎの	ひがり
まえう		みぎいに	きだり
まいにめ		みぎがあ	にだり
ばえ			
のあんぐほお			

MOVE	ROTATE
えにすんで	みて
すすんでお	よみて
のほおにすん	おおみて
かってすすんれう	むいえ
にいっええ	のほおにむいっえ
えいって	かいでんして
にいっええ	にかいてんして
むかって	がわみけっ
にむかう	のほおおみて
いってお	よむいて
にめかってすんでえ	にむいってか
にいっぺ	のほおりむいっけ
のほおにいってお	かいてして
むかっれえ	がわおみて

3.2 意味 ID に誤りを含む発話を用いた語彙学習

(1) 実験条件

表 2 の命令のうち 4 個の命令の意味 ID に対して間違った意味 ID を与えたデータを使用する. 誤りを含む意味 ID を持つ命令とその意味 ID を表 5 に示す.

表 5 誤りを含む命令と意味 ID

命令	意味 ID
まえにむかって	LEFT ROTATE
うしろにいって	RIGHT ROTATE
みぎのほうをみて	FORWARD MOVE
ひだりかいてんして	BACK MOVE

(2) 実験結果

図 3 にはテキストデータを用いたときの各単語の意味 ID “MOVE”の確率を示す. また, 図 4, 図 5 には音声データを用いたときの各単語の意味 ID“RIGHT”の確率, 意味 ID“MOVE”の確率をそれぞれ示す.

図 3 から, “のほうをみて”という実際には意味 ID“ROTATE”を示す単語が登録されないことがわかる. また, “にいって”や“にむかって”の意味 ID が正しい単語も登録されていない. これらの単語を含む発話は 2 回のみであり, そのうち片方に割り当てられた ID が違うため, 登録されなかったと考えられる. 図 4, 図 5 からは音声データを用いたときにおいても, 同様のことが発生している. また, “りかいてして”という, 意味 ID“ROTATE”を示す単語が意味 ID“MOVE”として登録されている. “りかいてして”は“ひだりかいてんして”という発話の音声認識結果であるが, DP マッチングにより, “ひだり”と“かいてんして”と上手く分けられず, 学習データに含まれる“みぎかいてんして”の“かいてんして”と違うものとして処理されてしまったため, 確率が 1 となっている.

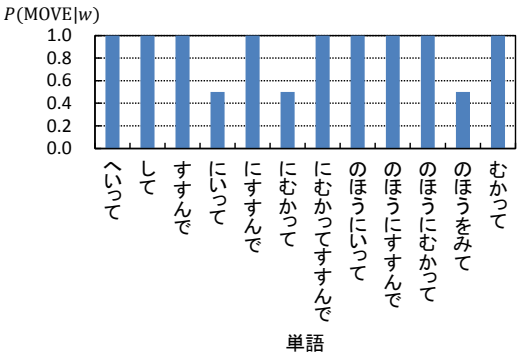

図 3 各単語の意味 ID“MOVE”の確率(テキストデータ)

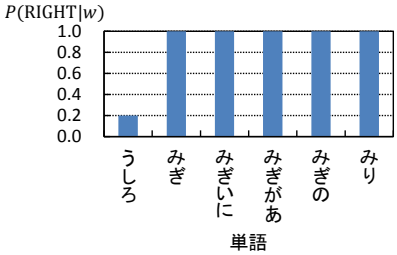

図 4 各単語の意味 ID“RIGHT”の確率(音声データ)

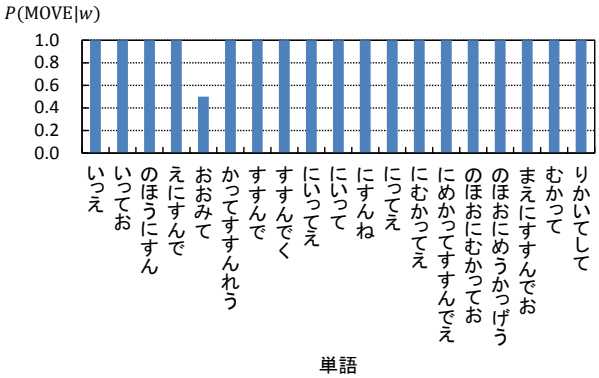

図 5 各単語の意味 ID“MOVE”の確率(音声データ)

4. おわりに

本稿では, 事前知識を持つロボットがユーザとのインタラクションを通して語彙を拡張していく手法について述べた. 提案手法は, 発話文から意味を持った 2 つの単語に分けることができ, 語彙を拡張できることを確認した. 今後は, より音声認識誤りを吸収できるラティス形式を用いた実験や, リアルタイムで学習することのできるアルゴリズムの開発, 実際のロボットを用いた実験も行っていきたい.

参考文献

[Roy 02] Roy, D. and Pentland, A.: Learning words from sights and sounds: a computational model, *Cognitive Science*, Vol. 26, No.1, pp. 113-146, 2002

[山本 04] 山本博史, 小窪浩明, 菊井玄一郎, 小川良彦, 匂坂芳典: 複数マルコフモデルを用いた階層化言語モデルによる未登録語認識, 電子情報通信学会論文誌 D-2, Vol. J87-D-2, No. 12, pp. 2104-2111, 2004.

[田口 10] 田口亮, 岩橋直人, 船越孝太郎, 中野幹生, 能勢隆, 新田恒雄: 統計的モデル選択に基づいた連続音声からの語彙学習, 人工知能学会論文誌, Vol. 25, No. 4, pp. 549-559, 2010.

[Neubig 12] Graham Neubig, Masato Mimura, and Tatsuya Kawahara: Bayesuan learning of a language model from continuous speech, *IEICE TRANSACTIONS on Information and Systems*, Vol. 95, No.2, pp. 614-625, 2012.