3B1-OS22b-4

ロボットの語意獲得のためのユーザの発話分類

Utterance Classification for Word Meaning Acquisition of Robots

中谷 仁 Hitoshi NAKATANI　植村 達也 Tatsuya UEMURA　荒木 修 Osamu ARAKI　西垣 貴央 Takahiro NISHIGAKI　尾関 基行 Motoyuki OZEKI
岡 夏樹 Natsuki OKA

京都工芸繊維大学
Kyoto Institute of Technology

It is desirable that robots who respond to spontaneous speech can learn the meaning of new words that are proper to a user. In order to learn the meaning of a new word, the robots must detect the things which the word refers to. In this paper, we propose a method of word meaning acquisition for robots who act in accordance with orders, and a method of utterance classification which is useful for the meaning acquisition. Utterances are classified into five types: command, correction, evaluation, description, and filler. Commands, evaluations, and fillers are further classified according to positiveness/negativeness. The classification is made based upon the prosody of utterance. Utterances were collected under a situation in which a user order a robot on a screen to move things, and the collected utterances were classified by the proposed method. The results showed that the utterance-type classification yielded F-measure of 0.622, and the positiveness/negativeness classification achieved F-measure of 0.846.

1. はじめに

家庭用のロボットはユーザの指示を，人が普段話すような自然な発話から理解することが求められている．そのためには，働く環境において与えられた仕事をこなしながら，言葉の意味を学習できる機能を持つ必要がある．本研究は，自由発話から行動を生成するロボットの語意獲得システムを提案する．

自由発話から語意を獲得する研究には Roy[Roy 00]，Yu ら [Yu 04]，田口ら [田口 10] や小野ら [小野 09] の研究がある．[田口 10] では，様々な言い回しの教示に対応できるよう，単語区切りのない連続音声から単語の音素系列を学習し，指示対象との関連付けを行っている．しかし，意味の学習に，発話と指示対象候補の共起のみを扱っているので，依頼などの，指示対象と共起しない発話については正しく学習できない．例えば，リビングにいるロボットに「玄関の靴をベランダに持っていって」と命令した場合，発話時に得られるセンサ情報からは「玄関」も「靴」も「ベランダ」も学習できない．そこで [小野 09] では，発話の直後のエージェントの行動を指示対象として，語を学習する事でこの問題を解決している．ただし，このような行動に着目した語意学習は，ただ語と行動の共起を用いるだけでは正しく学習できない．これは，語意を獲得していない状態では，発話意図通りに行動できることが稀であるので，語とまったく関係のない行動を学習してしまう可能性があるからである．この問題を解決するために，本研究では以下の２つの処理が可能である語意獲得手法を提案する．

1. 学習データとなる「語」と「概念」(物，動き，空間など) の対を選択できる
2. 学習データに対する評価を検出できる

まず，[小野 09] では前者について，前述した通り，発話の直後のエージェントの行動を指示対象として，語を学習する手法を採っている．だが，[小野 09] でも述べているが，必ずしも語と指示対象の関係はそうなっていない．インタラクションの流れによって，学習データ対は決定されなければならず，また，それは学習効率の増加をもたらす．本研究では，発話の種類によってインタラクションの流れを判断し，学習データ対を決定する手法を提案する．

そして，[小野 09] では後者について，発話にはエージェントへの命令以外に，エージェントの行動に対する評価を意味する発話があるとして，得られた評価を語意学習に用いている．しかし，人とロボットの自然なインタラクションにおいては，ロボットが行動した後に必ず評価を表す発話が得られるとは限らない，また，それを強制することは少なからずユーザに負担を与えてしまう，という短所がある．よって本研究では，発話には明示的な評価発話でない場合でも，直前の行動に対しての評価の情報が含まれていると仮定し，その評価を検出して語意獲得に利用する手法を提案する．

2. 想定する行動生成と語意獲得

本研究が提案する語意獲得手法は，行動に着目して語意を学習していくため，まず，2.1 で語意の更新方法について述べ，次に 2.2 で今回得られた発話データに対して行った発話分類について説明し，そして 2.3 で，発話分類をどのように語意獲得に利用するかを示す．

2.1 語意更新手法

本研究で想定している語意獲得は，語が指し示す概念（物，色，形，動きなど）を学習することである．「ユーザからの発話中の語」が指示する対象が「行動を構成する概念」[*1]であるとして学習する．どの語と概念が対になるかは，図 1 に示すように，発話分類を用いる事によって決められる発話と行動の対（以下発話行動対と呼ぶ）により決まる．発話行動対の中の語と概念の全ての組み合わせに対して，発話行動対に与えられた評価により語意を更新する．発話行動対に与えられる評価は，

*1 みかんを渡す，という行動の場合ならば，＜みかん＞や＜渡す＞という概念を指す

正と負の２通りであり，正の場合は語が概念を指し示している事を学習し，負の場合は語が概念を指し示さない事を学習するように働く．例として,「みかんとって」という発話と,「ぶどうをとる」という行動が対になった時，その発話行動対に負の評価が与えられたとする．この場合,「みかん」は＜ぶどう＞や＜とる＞という概念を指さない事を学習し，同様に「とって」は＜ぶどう＞や＜とる＞という概念を指さない事を学習する．

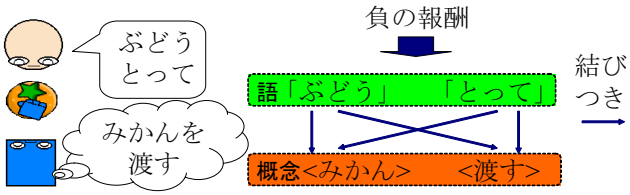

図 1: 発話行動対とその評価を用いた語意の更新方法

2.2 発話分類

本研究では，ロボットに音声命令を与える際のユーザの発話を，以下のように分類した．この分類を，機能での発話分類と呼ぶ．

- 命令発話：行動を命令する発話
- 訂正発話：間違っている行動を訂正させる発話
- 評価発話：行動に対しての評価をする発話
- 描写発話：行動を構成する概念を描写する発話
- フィラー

さらに命令発話と評価発話とフィラーについては，ロボットの行動への評価を基準として，以下のように分類する．この分類を，評価での発話分類と呼ぶ．

- 肯定的発話：直前の行動に対する肯定的な評価を含む発話
- 否定的発話：直前の行動に対する否定的な評価を含む発話

最初に，機能での発話分類について述べる．まず，命令発話とは，ロボットに新規の行動を命令する発話である．命令発話の例としては，みかんを食べたくなったので言った「みかんをとって」という発話がある．次に，訂正発話とは，ロボットのある行動に対して，訂正を促す発話である．例としては「醤油を目玉焼きにかけて」と言ったのにロボットが間違ってりんごにかけた時の「目玉焼きにかけて」という発話がある．次に，評価発話とは，ロボットのある行動に対しての，ユーザからの評価を示す発話である．例としては,「みかんとって」と言われたときに，ぶどうを取ってきた時の「違う」という発話がある．また，描写発話とは，行動や状況について描写している発話である．例としては,「みかんとって」と言われた時に，りんごを取ってきた時の「それはりんご」という発話がある．最後のフィラーとは,「殆ど実質的な意味を持っていない，会話に頻繁に現れる言語表現」[蔡 07] であり,「んーと」や「そのう」などが例として挙げられる．今回，言い淀みや独り言もこのカテゴリに含める．

次に，評価での発話分類について述べる．まず，肯定的発話とは，ユーザが与えた命令通りロボットが行動したために，パラ言語情報として肯定の意味が含まれている発話である．反対に否定的発話は，命令通りに動かないロボットに対しての，否定の意味が含まれている発話である．訂正発話と描写発話についてこの分類が行われないのは，これらの発話はすべて否定的発話だからである．なぜなら，訂正発話は，行動が間違っているから訂正されるという点から否定的な発話であるのは明らかであり，描写発話は，本研究のような行動を命令する場面では，行動を失敗した時のみ出現する発話であり，行動に対して否定的な発話となるからだ．

2.3 発話分類による語意獲得手法

本研究で提案する語意獲得手法は，次のアルゴリズムで動く．まず，発話が入力されたら，機能での発話分類を行い，その分類が訂正発話と描写発話以外なら，評価での発話分類を行う．次に，機能での発話分類により，発話行動対を更新する（発話行動対の更新と呼ぶ）．そして，発話分類によって得られた評価を，対応する発話行動対に与える（評価の付与と呼ぶ）．最後に，評価を与えられた発話行動対に対して語意の更新を行う．

次に，各々の発話で，どのように「発話行動対の更新」と，「発話行動対への評価の付与」を行うかを示す．それぞれの発話種類ごとの発話行動対の更新および評価の付与について，命令発話は図 2，訂正発話は図 3，訂正発話とフィラーは図 4，描写発話は図 5 に示す．

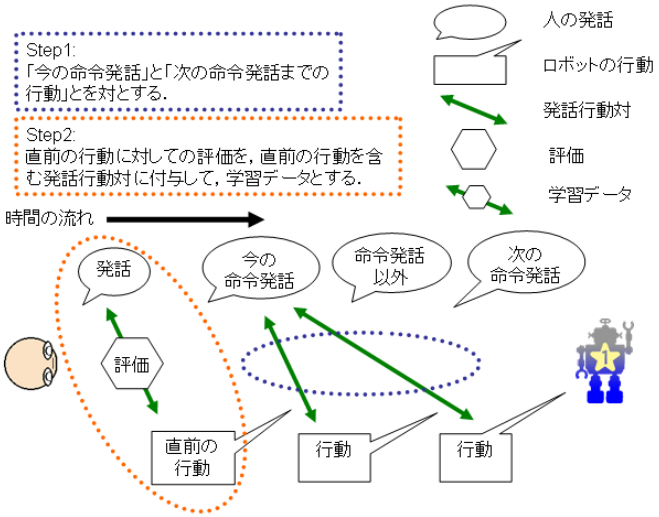

図 2: 命令発話での発話行動対の更新と評価の付与

まず，発話行動対の更新について示す．最初に，命令発話の場合を図 2 に示す．次の命令発話と今の命令発話の間に生成された行動と，今の命令発話とを対とする．こうしたのは，新しい命令が発話された時（この瞬間を命令の変わり目と呼ぶ），命令の内容が変わったために，古い命令と関連する行動はもう出現しないためである．命令の変わり目までに行った行動に対して，発話行動対を生成する利点は，従来研究よりも多くの学習データを得ることである．次に，訂正発話の発話行動対の更新を図 3 に示す．この場合は，前の命令発話から命令の変わり目までに生成した行動と，今の訂正発話とを発話行動対とする．これは訂正発話が，表現が違う可能性があるにしても，直前の命令発話と同じ行動を目的として発話されたからである．また，評価発話とフィラーの場合は，発話行動対の更新を行わない (図 4)．最後に, 描写発話の発話行動対の更新を図 5 に示す．この場合は，直前の行動と今の描写発話を発話行動対とす

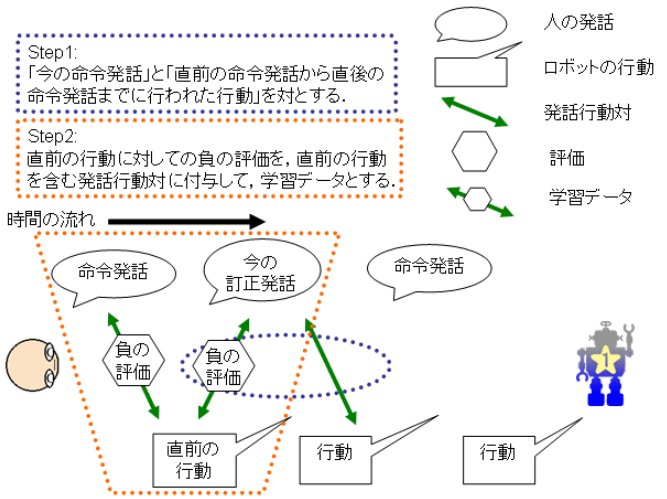

図 3: 訂正発話での発話行動対の更新と評価の付与

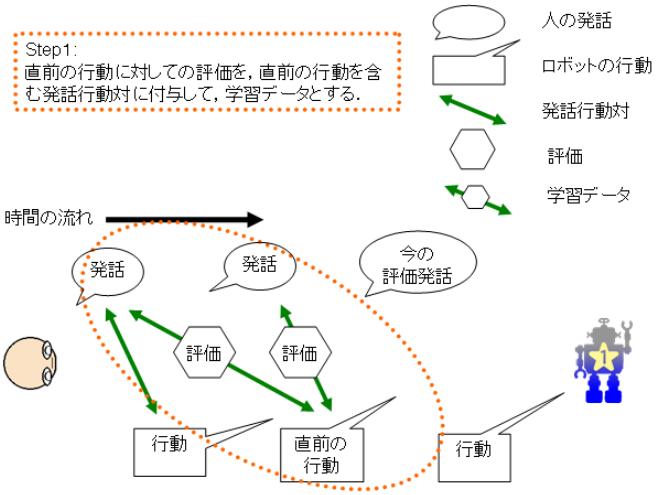

図 4: 評価発話とフィラーでの発話行動対の更新と評価の付与

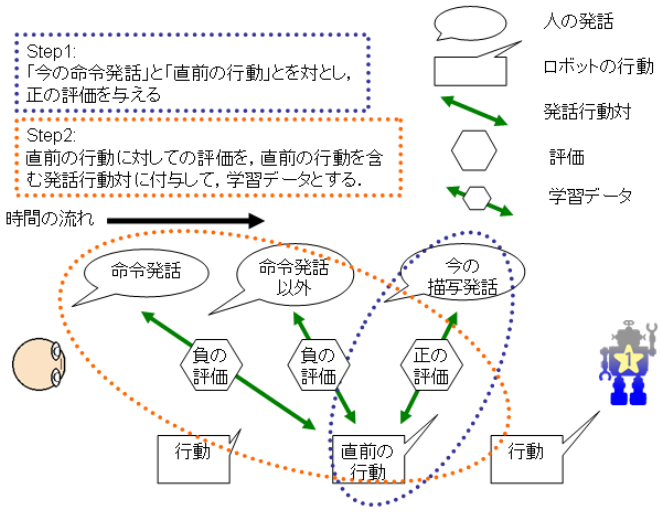

図 5: 描写発話での発話行動対の更新と評価の付与

る．これは，直前の行動以外について描写することはまずないからである．

評価の付与は，どの場合も直前の行動を含む発話行動対に対して行う．命令発話 (図 2 の Step2) と評価発話とフィラー (図 4 の Step2) の場合の付与する評価は，評価での発話分類が肯定的発話ならば正の評価，逆に否定的発話ならば負の評価である．訂正発話の場合，付与する評価は負の評価である (図 3 の Step2)．また，描写発話の場合，今の描写発話と直前の行動の発話行動対に対しては正の評価を与え，直前の行動の他の発話行動対に対しては負の評価を与える (図 5 の Step2)．このようにするのは，例えば，「みかんとって」と言われ，「りんごをとる」行動をした際に，「それはりんご」という描写発話を得たとすると，直前の行動「りんごをとる」を「それはりんご」という発話で描写しているので，この発話行動対には正の評価を与えるべきであり，また，直前の行動「りんごをとる」自体は，ユーザから見れば意図していない行動であるので，直前の行動を含む他の発話行動対に負の評価を与える．

3. 発話分類手法について

本研究では発話の分類に Support Vector Machine(SVM) を用いた．SVM に入力した発話の特徴量を以下に示す．一つの発話ごとに，これらの特徴量が計算され，ラベルと組み合わせて学習データとなる．ラベルは，機能での分類では，命令発話，訂正発話，評価発話，描写発話，フィラーの 5 種類であり，評価での発話では，肯定的発話，否定的発話の 2 種類である．

- パワーと基本周波数の平均，標準偏差，最大値，最小値，最大値と最小値の差，最大値の出現箇所，最小値の出現箇所
- パワーと基本周波数の線形近似の傾き（発話の全体，始端，終端での値）
- 直前の行動開始時から発話開始時までの時間
- 行動停止時における，行動停止開始から発話開始時までの時間
- 発話継続長
- 発話速度

4. 発話収集実験

4.1 実験概要

本研究の実験の目的は，何らかのタスクを遂行するために，教示者（ユーザ）がエージェントに対して誘導を行った際の，命令の変わり目を持つ発話データの収集を行い，提案手法での発話の分類精度を評価することである．

そのために，図 6 に示す食卓場面のシミュレータを用いて，音声のみで画面上のエージェントに話しかけ，エージェントの行動をうまく誘導することにより，目標を達成してもらうというタスクで発話収集実験を 6 名に対して行った．エージェントは，ユーザの音声に対して，音声認識が完了するとランダムに行動する．また，エージェントが取る行動は，「何かを掴みどこかに置く」という動作のみとした．そして，ユーザに達成してもらう目標として，「何かをある色の皿に置く」という状態をシミュレータの画面上部に示した．

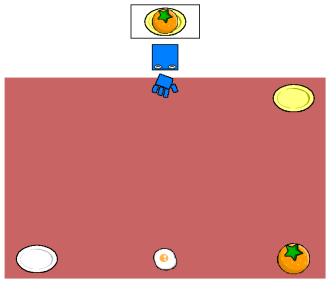

図 6: 食卓シミュレータ

5. 発話分類実験

5.1 実験方法

本研究での発話分類実験を説明する．まず，発話分類については，機能での分類である命令発話，訂正発話，評価発話の 3 種について，評価での分類である肯定的発話，否定的発話の 2 種について発話分類を行った．描写発話とフィラーが発話分類の対象となっていないのは，発話収集実験で得られた，描写発話とフィラーの出現頻度が，二つを合わせても全ての発話の頻度の 1 %に満たず，統計的な手法の SVM で扱うにはデータ数が足りないと判断したからである．

また，得られた発話のラベル付けは，実験風景を撮ったビデオデータと音声データを用いて著者の一人が行った．

発話の単位は，実験中に連続音声認識ソフトウエア Julius を用いて得られた一つの音声区間を，一つの発話とした．Julius では音声区間の検出に，零交差数とパワー（振幅レベル）のしきい値を用いている．

評価には，10 fold cross-validation を用いる．SVM のパラメータを変えて実行し，全ての発話の F 値を平均した値が高いものを選出する．

5.2 実験結果

機能での分類の結果を表 1 に，評価での分類の結果を表 2 に示す．機能での分類では，命令発話，訂正発話，評価発話の F 値の平均が 0.622，評価での分類では，肯定的発話，否定的発話の F 値の平均が 0.846 となった．

表 1: 発話の機能での分類

	発話数	精度	再現率	F 値
命令発話	273	0.724	0.778	0.750
訂正発話	220	0.429	0.441	0.435
評価発話	189	0.400	0.320	0.356
平均	682	0.626	0.619	0.622

表 2: 発話の評価での分類

	発話数	精度	再現率	F 値
肯定的発話	378	0.875	0.907	0.891
否定的発話	304	0.783	0.720	0.750
平均	682	0.846	0.848	0.846

5.3 比較実験

提案手法と従来手法の比較をシミュレーションにより行う．比較方法は，ユーザの発話とそれに対するエージェントの行動をランダムに 100 万組生成し，その系列から正しい学習データを何割検出できるかで行った．正しい学習データとは，提案手法で発話分類精度が 100% である時に検出される学習データである．比較する二つの手法は，提案手法を発話分類実験での分類精度で行った手法と，発話とその直後の行動を発話行動対とし，行動の直後の発話からその対の評価を，提案手法と同様に得る手法である．それぞれの分類精度は，前者が 11.38%，後者が 18.68%となった．よって，現在の分類精度の提案手法では，従来手法で学習できる語意の獲得には用いるべきではなく，発話行動対を構成する行動が発話の直近に来ない場合，発話行動対の評価が遅れる場合に用いるべきである．

6. まとめ

本研究では，自由発話を理解し行動するロボットの語意獲得手法を提案し，その手法の中核である，命令の変わり目を検出するための発話分類手法を構築，評価した．発話収集実験を行い得られた発話データについて，発話分類手法の分類精度は，機能での分類及び評価での分類の F 値がそれぞれ 0.622 と 0.846 という結果になった．

現状，提案手法では，訂正発話と評価発話をうまく認識できない．よって今後の課題は，発話を誘導できるタスクの設計や，発話分類に有用な特徴量の選出により，分類精度を上げる事である．

謝辞：本研究は科研費 (21500137) の助成を受けたものである．

参考文献

[Roy 00] D. Roy, “Integration of speech and vision using mutual information,” Acoustics, Speech, and Signal Processing, 2000. ICASSP ’00. Proceedings. 2000 IEEE International Conference on, vol.6, no.4, pp.2369–2372, Aug. 2000.

[Yu 04] C. Yu and D. Ballard, “On the integration of grounding language and learning objects,” in AAAI-04, ed. A.A. forArtificial Intelligence (AAAI), pp.488–494, The MIT Press, Cambridge, Massachusetts, 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，May 2010.

[小野 09] 小野広司，左 祥，伊丹英樹，尾関基行，岡 夏樹，”最終行動ヒューリスティクスを用いた状況推定による自由発話音声データからの語句意味学習”，HAI シンポジウム 2009，2B-2，東京，Dec. 2009.

[蔡 07] 蔡　嘉綾，“日本語学習者の会話におけるフィラーの研究–中国語母語話者を中心に，” 東北大学高等教育開発推進センター紀要，vol.2，no.311，pp.311–314，March 2007.