1J1-OS13-9

食卓場面での参照的意味と機能的意味の獲得

Acquisition of the referential meaning and the functional meaning in a dining table situation

高田 宏明 TAKADA Hiroaki　植村 竜也 UEMURA Tatsuya　中谷 仁 NAKATANI Hitoshi　尾関 基行 OZEKI Motoyuki　岡 夏樹 OKA Natsuki

京都工芸繊維大学
Kyoto Institute of Technology

Words have two kinds of meanings: referential meaning which refers to objects, properties, events, relations and actions in the physical world, and functional meaning which exerts an influence upon listeners. Robots in everyday situations such as a dining table situation need to understand both the meanings, and the paper proposes a method for acquiring the both meaning through interaction in a dining table simulator. A robot in a simulator learns the referential meaning of words based on the cooccurrence between the words and the objects or actions, and the functional meaning of words by reinforcement learning.

1. はじめに

ロボットが人との自然なやりとりの中で言葉の意味を獲得することができれば，たとえ未知の場面であっても，人の意図を正しく理解し，その場に応じた振る舞いができるようになる．これは予め言葉と行動の対応表をロボットに登録しておくことに比べて，言葉を覚えるまでに時間がかかる一方，人それぞれの言い回しや省略の仕方にも柔軟に対応することができる．また，人が明示的にロボットを訓練する場合に比べて，訓練しなければいけないという心理的な壁がなく，人が意識していなかった言葉の意味まで獲得できる可能性もある．

獲得すべき言葉の意味には，物や動作を指し示す参照的意味と，聞き手に影響を与える働きを持つ機能的意味の2つがある[1]．参照的意味の獲得については Roy[2]や Yu と Ballard[3]らの研究があり，機能的意味の獲得については左ら[4]の研究があるが，これらの研究では参照的意味と機能的意味のどちらか片方だけに注目している．しかし，参照的意味と機能的意味には相補的な性質があり[5]，同時に学習していくことで効率的に獲得することが期待できる．

そこで本研究では，参照的意味と機能的意味を同時に学習する方法を提案する．本手法では，発話と状況（または行動）の共起の数え上げと強化学習を組み合わせることで，参照的意味と機能的意味の学習を相補的に進める．

本稿では，人とロボットのやりとりの場面として，人とロボットが食卓に向かい合って座り，食べ物や食器などを指差したり渡してもらったりする状況を想定する．ただし，現時点では実際の食卓ではなく，食卓を模した食卓シミュレータを用い，ロボットを模したエージェントとGUIを通して対話する．

本稿の構成は次の通りである．まず2節で参照的意味と機能的意味，および，本研究で使用する食卓シミュレータについて説明し，3節で提案する参照的意味と機能的意味を同時に学習するアルゴリズム（提案手法）について説明する．4節でそれを検証するための実験について説明し，5節でまとめと今後の展望を述べる．

連絡先：岡 夏樹
京都工芸繊維大学 大学院工芸科学研究科 情報工学部門
〒606-8585 京都市左京区松ヶ崎御所海道町
e-mail: nat -at- kit.ac.jp

2. 食卓シミュレータにおける語意獲得

2.1 参照的意味と機能的意味

参照的意味とは，物やその属性，また，動作などを指し示す言葉の意味である．参照的意味は，言葉と共起する実空間中の事象とを対応付けることで獲得できる．例えば，「ごはんがあるね」や「ごはんちょうだい」などの発話がなされた時に「ごはん」とテーブル上の物体＜ごはん＞は対応付けの候補となる．本研究においては，このような言葉と事象の共起度数に基づいて参照的意味を獲得する．

機能的意味とは，聞き手に影響を与える働きである．例えば，「ごはんちょうだい」という発話は，聞き手に対して，ごはんを話し手に渡す行動をとるよう働きかける．そして，話し手は自身の発話の（機能的）意味が正しく受け取られたことに対して「ありがとう」などという．本研究ではこの行動－報酬のプロセスに着目し，話し手の発話とロボットの行動の対応付けを強化学習することで機能的意味を獲得する．

ある発話における機能的意味は，その発話中の各語句の参照的意味を組み合わせることで推測できることが多い．また，各語句の参照的意味は，それらの語句で構成される発話の機能的意味の要素になっていることが多い．そこで本研究では，獲得した参照的意味を用いて未知の機能的意味を推測し，また逆に，獲得した機能的意味を要素に分解することで未知の参照的意味の獲得を支援する．

2.2 食卓シミュレータ

図1に本研究で用いる食卓シミュレータの実行画面を示す．人は，キーボード上のキーを押すことで，ロボット型エージェントに対して発話できる．エージェントは受け取った発話内容とテーブルの状況から行動を決定する．人はエージェントの行動が自身の意図に沿ったものか否か，評価キーを押すことでエージェントに伝える．評価を受けたエージェントは機能的意味の学習を行い，ユーザから次の発話を受け取るまで待機する．

なお，本研究では以下の前提を置く．

- エージェントは発話中の語句を分節できる
- エージェントは意味付けの対象となる概念を獲得している
- エージェントは物体と属性と動作を分節できる
- 人の発話は状況を描写する発話と，エージェントに対して命令する発話の2種類とする

3. 提案手法

3.1 参照的意味の獲得

参照的意味は発話に含まれる各語句と状況／行動の要素との共起を数え上げることで獲得する. 状況と行動のどちらの要素との共起を数え上げるかは話者がどちらを対象として発話しているかによる. テーブル上の状況の要素である物体やその属性について描写する発話(以下, 描写発話)であれば, 物体や属性との共起を数え上げる必要がある. また, 聞き手の行動に働きかける命令や依頼といった発話(以下, 命令発話)であれば, 行動の要素との共起を数え上げる必要がある. ただし, 命令発話の場合, 必ずしも話者の意図通りに聞き手が行動するとは限らないため, 共起を数え上げる際は, 話者の意図通りに聞き手が行動したことを確認する必要がある.

なお, 語句 w との共起度数が最も高い意味 m が語句 w の参照的意味であるとするのは問題がある. 各物体や属性の生起頻度には偏りがあるため, その語句が持つ本来の意味によらず, 生起頻度が高い概念と結びついてしまうのである.

そこで, 各概念の生起頻度を考慮して語句の意味づけを行うために Fisher の直接法を用いる. Fisher の直接法を用いて, 語句 w と特異的に高い共起度数を持つ意味 m を抽出し, その意味 m を語句 w の参照的意味とする[6].

3.2 機能的意味の獲得

機能的意味は強化学習の手法の1つであるQ学習を用いて獲得する. Q学習ではある状態における各行動の価値の期待値をQ値として表現する. このQ値は正の報酬が与えられると上がり, 負の報酬が与えられると下がる. 各行動の選択確率はQ値に依存する.

状態はテーブル上の状況と人の発話からなるとする. ロボットが人の発話に対して意図通りに行動すると褒められる(正の報酬が与えられる)とする. 正の報酬が与えられると, その状態(ある状況である発話がなされた状態)でのその行動のQ値が上昇するため, その行動の選択確率が上昇し, ロボットは徐々にその状況におけるユーザの発話の意図通りに行動できるようになる. これが機能的意味の獲得になる.

3.3 参照的意味を用いた機能的意味の推測

「ごはん食べて」の発話の機能的意味は(状況にもよるが)＜ごはんを食べる＞ことである. 語句「ごはん」の参照的意味が＜ごはん＞, 「食べて」の参照的意味が＜食べる＞であるとすでに獲得している時(＜＞は概念を表す), これらの知識を用いることで「ごはん食べて」という発話を初めて聞いた場合でも, その発話の機能的意味である＜ごはんを食べる＞を推測できると考えられる. これが参照的意味を用いた機能的意味の推測になる. ただし, 機能的意味は同じ発話であっても状況によって変化するため, 語句の参照的意味からだけでは正確に推測することは難しい. そのため, この方法は機能的意味の獲得の支援に留まることに注意する必要がある.

3.4 機能的意味を用いた参照的意味の獲得の促進

機能的意味を, 状況における行動として表現し, かつ, 行動がいくつかの要素から構成される時, 獲得した機能的意味を用いて参照的意味の獲得を促進できる. 前述の通り, 話者の意図通りにエージェントが行動した時, 発話と行動の要素同士の共起を数え上げることができる. 機能的意味を獲得している場合, 毎回, 正しい行動を行うことができ, 共起の数え上げを毎回行えるため, 参照的意味の学習を促進することができる.

3.5 参照的意味と機能的意味の同時獲得

参照的意味を元にして機能的意味を推測することができると同時に, 機能的意味を元にして参照的意味の獲得を促進することができることを述べた. これら2つの方法を併用することで両方の意味を相互にブートストラップ的に獲得することができる.

3.6 アルゴリズム

本節では参照的意味および機能的意味を獲得するアルゴリズムを説明する.

参照的意味は発話と状況／行動との共起を数え上げ, それを Fisher の直接法を用いて獲得する. 機能的意味は強化学習を用いて獲得する. また, 獲得した参照的意味を用いて発話から機能的意味を推測することで強化学習の効率を上げる. 同時に, 獲得した機能的意味を利用して参照的意味の獲得効率を上げる.

アルゴリズムの流れを次に示す:

1) 現在の状態を取得する.
2) ユーザの発話を識別する.
3) 描写発話時の数え上げを行う.
4) 参照的意味の学習を行う.
5) 行動を選択し実行する.
6) 報酬を獲得する.
7) 機能的意味の学習を行う.
8) 命令発話時の数え上げを行う.
9) 1)に戻る.

(1) 状態の取得

物体とその属性からなるテーブル上の状況とユーザの発話を取得し, 状態を取得する. なお, 「ごはんどうぞ」と「どうぞごはん」のように語順が異なるが, 同じ語句を用いた発話は同じものとして扱う. また, 「ごはん A が近くにありごはん B が遠くにある」状況と「ごはん B が近くにありごはん A が遠くにある」状況も同じものとして扱う.

(2) 発話の識別

取得した発話が描写発話か命令発話かを識別する. 「ごはん食べて」や「どうぞどうぞ」のような命令／依頼の語句が1語以上含まれている発話を命令発話, 「みそ汁近くだね」や「あるねあるね」のような描写／叙述の語句が1語以上含まれている発話を描写発話として識別している.

現在は発話中の語句の知識を用いて識別を行っているが, 将来は, 発話の韻律情報や語尾の助詞などから識別することを計画している.

(3) 描写発話時の共起の数え上げ

ユーザの発話が描写発話である時, 描写発話に用いられている各語句とテーブル上の各物体および各属性との共起を数え上げる.

(4) 参照的意味の学習

ステップ(3)および(8)で数え上げた共起度数から Fisher の直接法を用いて語句の意味づけ(参照的意味の獲得)を行う.

まず, 数え上げた共起度数からある語句 w1 とある意味 m1 の共起度数の四分表を作成する. 次に, Fisher の直接法を用いて, 作成した四分表以上に偏った四分表が得られる確率を求める. そして, その確率の逆数を計算する(以下, こうして求めた値を確信度と呼ぶ). なお, Fisher の直接法を用いた結果, 共起

度数が特異的に低いとされた意味については，確信度を0とし，意味づけの対象から外す．この一連の処理を m1 以外の各意味 m2～mN に対しても同様に施す．

描写発話時の各語句に対する各意味の共起度数から前述の処理を用いて確信度を求める．同様にして，命令発話時の各語句に対する各意味の共起度数から確信度を求める．この2つの確信度を足し合わせた値を語句 w1 が各意味 m1～mN である確信度とする．

最後に，確信度の総和が1になるように正規化し，正規化後の各値を語句 w1 が各意味 m1～mN である確率とする．

(5) 行動選択

発話が描写発話の時，＜笑顔で会釈する＞行動を選択する．

発話が命令発話の時，学習した参照的意味(語句 w が意味 mである確率)から各行動の確からしさを求め，その総和が1になるように正規化する．次に，Q値からボルツマン選択を用いて各行動の確率を計算する．この2つの確率の平均を各行動の選択確率とし，その選択確率に応じて行動を1つ選択する．

(6) 報酬の獲得

ユーザに褒められた場合，正の報酬を獲得する．逆に，ユーザに叱られた場合，負の報酬を獲得する．また，報酬を獲得する前に次の発話がなされた場合，報酬は与えられない．

(7) 機能的意味の学習

得られた報酬と現在のQ値を元に，Q値を更新し，機能的意味の学習を行う．

(8) 命令発話時の共起の数え上げ

ユーザの発話が命令発話であり，かつ，正の報酬が与えられた時，命令発話に用いられている各語句と行動の要素との共起を数え上げる．

4. 実験

参照的意味の学習精度および機能的意味の学習性能を調べるための実験を計画している．

4.1 参照的意味の学習精度

参照的意味の学習精度を次のように定義する．

各語句 w1～wN の真の意味がそれぞれ m1～mN である時，w1 の意味が m1 である確率，w2 の意味が m2 である確率，…，wN の意味が mN である確率の平均を参照的意味の学習精度とする．

4.2 機能的意味の学習性能

実験開始から実験終了までに正の報酬が与えられた回数を，命令発話が発せられた回数で割った値を機能的意味の学習性能とする．

4.3 実験概要

学習した参照的意味を用いた機能的意味の推測の効果，獲得した機能的意味を用いた参照的意味の獲得促進の効果，および，両方の機能による相互作用の効果を評価するために次の4つの実験を行う．

実験A：両方の機能を使う．

実験B：参照的意味から機能的意味を推測する機能のみ使う．

実験C：機能的意味から参照的意味の獲得を促進する機能のみ使う．

実験D：どちらの機能も使わない．

4.4 実験内容

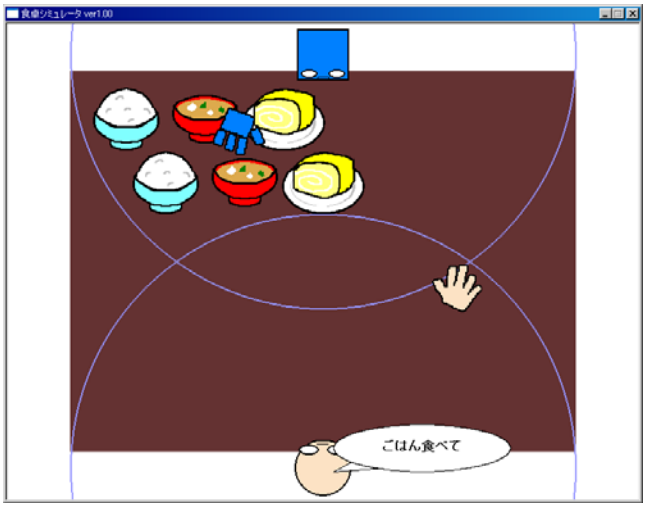

図1：実験で使用する食卓シミュレータ

食卓シミュレータを用いて実験を行う．実験で使用する食卓シミュレータの画面を図1に示す．

実験はごはん2つ，みそ汁2つおよび卵焼き2つがエージェントの近くに並べられている場面から始まる．ユーザとエージェントは食卓上の料理を渡したり貰ったりしながら食事をする．食卓上の料理が無くなったら試行を終了し，次の試行に移る．試行は1つの実験に対し3回行われる．

4.5 実験設定

(1) 実験の要素

実験の要素は次の通りである(「」は語句を表し，<>は意味付けの対象となる概念を表す)．

語句 = {「ごはん」「みそ汁」「卵焼き」「あるね」「近くだね」「遠くだね」「ちょうだい」「どうぞ」「食べて」}

料理 = {<ごはん> <みそ汁> <卵焼き>}

属性 = {<ある> <近い> <遠い>}

動作 = {<もらう> <あげる> <食べる>}

(2) 発話

発話は2語からなり，キーボード上の2つのキーを組み合わせて押すことで発話される．発話を構成する各語句は次の図2のようにキーボード上の各キーと対応している([ほめる]と[しかる]および[ごちそうさま]は語句ではない)．また，実験参加者に分かりやすいように，実験時には各キーに対応する語句を書いたシールを貼り付ける．

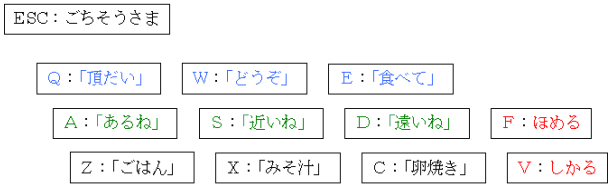

図2：各キーと各語句の対応

(3) 操作方法と手の届く範囲

食卓シミュレータ上の手はマウスを使って動かすことができる. 料理に手を重ねて左クリックすることで料理を掴み, 掴んだ状態で右クリックすることで料理を離すことが出来る. また, 料理を掴んだ状態で, 人またはエージェントの顔に手を重ねて, 右クリックすることで掴んだ料理を食べることができる.

なお, 人とエージェントの手の届く範囲に制限を設けている(図1中の半円形の領域). 手の届く範囲を設けることでインタラクションの必要性を作っている.

(4) エージェントの行動

エージェントの行動は対象と対象に施す動作の2つの要素からなる. また, 行動中に叱られた場合, 行動を中断する(負の報酬は与えられる). なお, エージェントは発話が与えられた時のみ行動を行う.

4.6 実験手順

食卓シミュレータの操作方法を書いた文を読んでもらい, 操作に慣れてもらうために, 練習を行う. 練習後, ユーザへの教示を含んだ実験の説明書きを読んでもらい, 実験A～Dをランダムな順番で実施する. 全ての実験が終わった後, アンケート欄に実験の感想を記入してもらう.

ユーザに与える教示は次の通りである.

- 褒める／叱ることでロボットは言葉の意味を学習する
- 描写発話時に言葉の意味を学習する
- 期待通りの行動を行った時に褒めて下さい
- あなたとロボットの両方がまんべんなく料理を食べるようにして下さい

5. 結論

言葉の意味には参照的意味と機能的意味の2つがあり, ロボットは両方の意味を理解できることが重要である. それぞれを共起の数え上げ(と Fisher の直接法), および, 強化学習を用いて獲得する方法について述べ, 両意味の間にある相補性に着目して, 互いにブートストラップ的に獲得する方法について述べた.

今後は, 計画した実験を実行し, 得られた結果を元にアルゴリズムの性能を評価する予定である.

参考文献

[1] Roy, D.. Semiotic Schemas: A Framework for Grounding Language in the Action and Perception. Artificial Intelligence, 167(1-2): 170-205 (2005).

[2] Roy, D.: “Learning from sights and sounds: a computational model,” Ph.D. Thesis, MIT Media Laboratory, Cambridge (1999).

[3] Yu, C., and Ballard, D. H.: “A multimodal learning interface for grounding spoken language in sensory perceptions,” ACM Trans. Applied Perception, 1(1): 57-80 (2004).

[4] 左祥, 北川憲, 林口円, 小野広司, 荒木雅弘, 岡夏樹, “時間的に切迫した状況におけるインタラクションデータからの意味学習, ” 第 23 回人工知能学会全国大会, 1F2-OS7-1 (2009).

[5] Leech, G. N., “ Principles of Pragmatics, Longman, ” London(1983); 池上, 河上(訳) : 語用論, 紀伊国屋書店 (1987).

[6] 岡夏樹, 増子雄哉, 林口円, 伊丹英樹, 川上茂雄, “Fisher の直接法を用いたインタラクションデータからの意味学習, ” 知能と情報(日本知能情報ファジィ学会誌), 20(4): 461-472 (2008).