

敵対的模倣学習により獲得した方策を用いた語彙学習の効率化

松井 壮哉 田口 亮
名古屋工業大学 大学院工学研究科

1. はじめに

日常生活の中で人を支援するロボットは、環境固有の単語とその意味を発話から学習することが望ましい。語彙学習の手法として人が未知の単語の意味を直接教示する方法が提案されている[1]。しかし、未知の発話が入力されるたびに操作が必要となり効率が悪い。そこで、本研究ではタスクの実行に必要な知識(タスク知識)を敵対的模倣学習により学習し、獲得されたタスク知識を用いて効率的に語彙を学習する手法を提案する。提案手法では、ロボットに未知の命令発話が入力された際、状況に対する動作の適切さをタスク知識に基づいて推定し、人の発話の意味を合理的に推定する。例えば図 1 のように小売店で働くロボットが未知の命令を聞いた時に、「在庫が少ない商品を補充する」というようなタスク知識があれば、在庫状況から発話の意図を推定することができる。このように学習したタスク知識を利用し、人による意味の教示を最小限にすることで語彙学習の効率化を図る。

2. タスク設定

本研究では小売店での品出しを想定して、簡易的なタスクを設定する。店内には複数の棚があり、各棚に陳列する商品が決まっている。棚の商品は時間とともに減っていく。在庫状況に合わせて適切に商品を補充することがロボットの目的となる。人からの指示がない場合、ロボットは自身の判断で品出しを行うが、指示がある場合には、その指示に従わなければならない。指示は「ポテトチップ」のように商品名を示す単語で行われる。命令を受けたロボットは、バックヤードで対応する商品を選択するが、その商品が間違っていた場合、人から誤りを指摘され、正しい商品を選ぶまでこれを繰り返す。その後、ロボットはその商品を棚に運ぶ。

当初、ロボットは効率的な品出し方法や、商品名の知識を持たない。そのため、人の品出しを観察することで方策を学習し、その後、バックヤードでの人とのインタラクションを通して、商品名を学習する。

3. 提案手法

提案手法の概要を図 2 に示す。提案手法では、まず Generative Adversarial Imitation Learning (GAIL)[2]を用いて、人の動作例からタスク知識を学習する。その後、獲得したタスク知識と単語分布から行動の事後

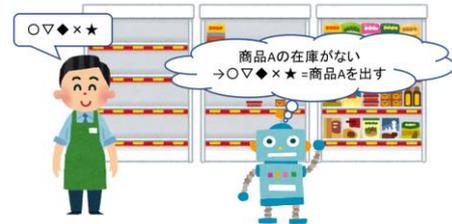


図 1 タスク知識を利用した発話意味推定

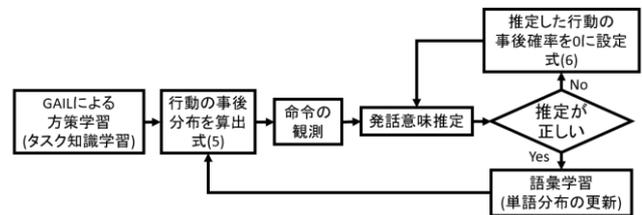


図 2 提案手法の概要

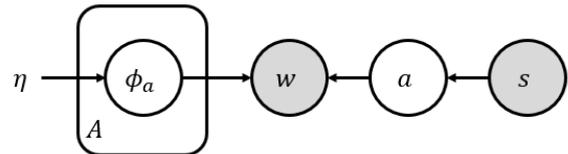


図 3 意味推定手法のグラフィカルモデル

分布を算出することで、在庫状況と単語の意味を考慮した行動決定を行う。

意味推定手法のグラフィカルモデルを図 3 に示す。灰色のノードは観測変数、白いノードは潜在変数を示す。商品を選択する行動の集合を A 、入力された単語を w 、実行する行動を a 、環境の状態を s と記す。行動 a は式(1)、 a の生成に用いるパラメータ ψ は式(2)、入力発話 w と単語分布 ϕ は式(3)、(4)で生成される。ただし、 $\text{Mult}(\cdot)$ 、 $\text{Dir}(\cdot)$ 、 $\pi(\cdot)$ はそれぞれ多項分布とディリクレ分布、獲得した方策であり、 η はディリクレ分布のパラメータを表す。

$$a \sim P(a|s) = \text{Mult}(\psi_s) \quad (1)$$

$$\psi_s = \{\pi(s, a) | a \in A\} \quad (2)$$

$$w \sim P(w|a) = \text{Mult}(\phi_a) \quad (3)$$

$$\phi_a \sim \text{Dir}(\eta) \quad (4)$$

単語と状態が入力された際の行動 a の事後確率は式(5)で表される。ここから行動の事後分布を算出する。この分布に従い行動を確率的に決定することで、単語と方策に基づいた行動決定を実現する。

$$P(a|w,s) \propto P(a|s)P(w|a) \quad (5)$$

命令に対してロボットが選択した行動 \hat{a} が誤っていた場合、ロボットは異なる行動を選択し直す必要がある。その際に、再び推定結果として \hat{a} を出力することを防ぐために、行動 \hat{a} の確率を0とし、それ以外の行動 a_j について式(6)のように事後確率を変更する。

$$p(a_j|w,s) = \frac{p(a_j|w,s)}{\sum_{k \in A'} p(a_k|w,s)} \quad (6)$$

ただし、 A' は誤りを指摘されていない行動の集合である。また、推定が正しかった場合には、その情報を元にディリクレ分布のパラメータ η を更新する。これにより、徐々に正しい単語分布が推定される。

4. 実験

4.1 実験条件

提案手法の動作を確認するため、シミュレーションにより実験を行う。表1に示す3種類の変数を用いて、状態を表現する。商品を置く棚は6つあり、各棚に対応した1種類の商品が置かれる。ロボットは1ステップで1回行動を実行できる。行動の詳細を表2に示す。ロボットは隣接する棚、およびバックヤードの7か所に移動できる。また、バックヤードで1種類の商品を選択し、対応する棚まで移動することで補充を実行する。商品は1ステップごとに0~3個購入されるものとし、その確率をそれぞれ0.85, 0.08, 0.05, 0.02とした。

まず、この環境で、エキスパートとなる人が品出しタスクを実行する。100ステップ分の操作データ(状態と実行した行動)を20個収集し、GAILにより方策を学習する。

方策の学習後、人の指示に基づいた品出しを行う。指示命令で用いる単語は「ポテトチップ」、「チョコレート」、「醤油」、「カレー」、「牛乳」、「水」の6つとする。ロボットがバックヤードにいるときに、人が状態を見て補充すべき商品を決め、単語をロボットに伝える。1回の指示に対し、ロボットは正しい商品が見つかるまで、繰り返し商品選択を行う。その回数(以下、確認回数)を評価指標とする。なお、評価データには、様々な在庫状況で行われた96個の指示(1単語あたり16個)を用いる。6回教示をするごとに、評価データに対する確認回数を出力する。この実験を10回行い、その平均により、性能を評価する。

方策のみ、および単語分布のみで意味推定をする手法と比較し、提案手法の有効性を検証する。なお、ディリクレ分布のハイパーパラメータは $\eta=(1.0, 1.0, 1.0, 1.0, 1.0, 1.0)$ と設定した。

4.2 実験結果と考察

教示回数による確認回数の変化を図4に示す。教示回数が0回の場合、ロボットは単語知識を持たないため、3.3回の確認が必要となるが、GAILで学習した方策を用いることで、確認回数を1.6回に抑えることがで

表1 状態の変数とその説明

ロボットの位置	棚1~6の周辺+バックヤード(7か所)
持っている商品	棚1~6に対応する商品+未所持
各棚の在庫状況	在庫数を0~20までで示す

表2 行動の種類とその説明

移動	隣接した棚及びバックヤードへの移動+その場で停止
補充する商品の選択	バックヤードで実行可能 商品を1種類だけ所持
補充	持っている商品と対応する棚で実行 棚の在庫を10個追加(20個を超えない)

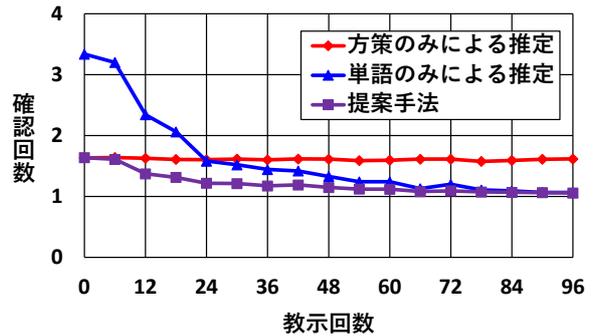


図4 教示回数による平均確認回数の推移

きる。一方で、教示回数を増やすことで、単語知識を用いる手法の確認回数が減少していき、方策のみを用いる手法よりも確認回数が少なくなる。特に提案手法ではより少ない教示回数で確認回数が減少した。

提案手法は、方策での意味推定をベースとして語彙学習をすることで、未学習時における推定精度の高さと、学習による推定精度の向上を両立している。この実験では商品と棚が一对一对応のため、単語知識が性能に大きく寄与する。しかし、1つの商品が複数の棚に対応するような場合、単語知識だけでは一意に棚を決定できないため、方策を併用できる提案手法の優位性が増すことが考えられる。

5. まとめ

本稿では、タスク知識の学習により語彙学習を効率化する手法を提案した。提案手法では状態と行動の関係性を事前に学習することで、語彙学習における人の教示を減らし、未知の単語や教示回数が少ない単語に対して高い精度で単語の意味推定をできることを確認した。今後は、より複雑な環境や自然発話に対応した実験を行いたい。

参考文献

- [1] 藤本ほか:キズ検査ロボットによる音声インタラクションを通じた語彙の拡張,人工知能学会全国大会,104-OS-22a-5, (2015).
- [2] J. Ho, et al.: Generative Adversarial Imitation Learning, Advance in Neural Information Processing Systems, vol.29, pp.4565-4573, (2016).