

友好関係構築を目的としたユーザ適応行動学習エージェント

三治祐太[†] 佐久間拓人[‡] 加藤昇平[‡]

[†]名古屋工業大学 工学部情報工学科

[‡]名古屋工業大学大学院工学研究科工学専攻

1 はじめに

近年、人間とロボット間でのコミュニケーションに関する研究が増加しており、その中には人間との友好関係や信頼関係の構築に着目した研究が存在する。しかし、人間との友好関係の構築においては、効果的な応答が人によって異なる点を考慮する必要がある。

福田ら [1] は親しみやすい対話エージェントの作成に斎藤 [2] の対人感情と対人欲求-行動の円環図式を用い、人間の行動を再現した。しかし、福田らのモデルの行動はどの状況下でどのような反応をするかを事前に設定された確率に基づいて決定されていた。

内田ら [3] は自律対話ロボットの対話戦略において、ユーザの価値観に対する肯定・否定の適切な割合を調査した。結果より対話意欲の喚起には肯定の割合を50%以上100%未満に設定することが対話戦略として有効と示唆された。よって、友好関係構築には毎回肯定的反応を返すことが必ずしも最良ではないと考えられる。

本研究では人間が他人と友好関係を構築する思考過程を再現する。エージェントは自身の経験を元に学習し、各状況における最適な行動を選択する。これにより、人間と共存するための友好関係構築を目指す。

2 提案手法

図1に提案手法の概観を示す。ユーザから受けた各行動に対し、エージェントはユーザから次にポジティブな反応を得られると思われる行動を選択して返す。エージェントの返答後、次に得られたユーザの返答に応じてエージェントの価値関数を更新する。エージェントは全ユーザに対する経験をまとめた「全体経験」と各ユーザに対する「個人経験」の両方について価値関数を更新し、状態に応じて行動を変化させる。

本研究では行動の中でも発言のみを扱い、図2に示した斎藤の対人欲求-行動の円環図式に基づきユーザの発言を8種類に分類し、これを状態と認識する。またエージェントの行動についても8種類から1つ選択するものとする。

2.1 学習過程

学習過程では以下の項目を考慮する。

1. 各ユーザに対して個別に学習
2. 離散的な行動かつ時系列を考慮した学習

User Adaptive Behavior Learning Agents for Friendship Building

Yuta SANJI[†] Takuto SAKUMA[‡] Shohei KATO[‡]

[†]Dept. of Computer Science, Nagoya Institute of Technology

[‡]Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology

^{†‡}Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

{sanji, sakuma, shohey}@katolab.nitech.ac.jp

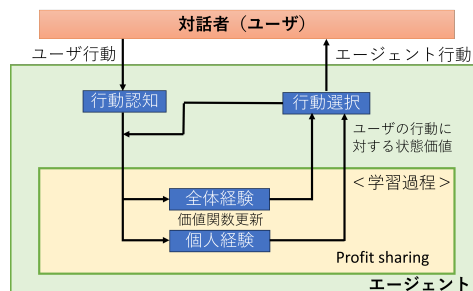
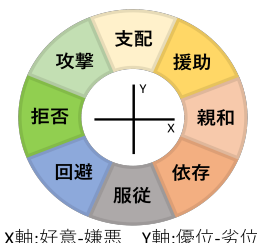


図1: 提案手法の概観



X軸:好意-嫌悪 Y軸:優位-劣位

図2: 対人欲求-行動の円環図式

項目1を満たすためには、連合学習の考え方をを用いる。エージェントは個人経験を用いて学習させた価値関数と、全体経験を用いて学習させた価値関数をそれぞれ更新しつつ行動を選択する。交流が増えるほど個人経験に比重を置いて行動選択することでユーザ適応を目指す。

本研究は8種類に分類された行動に対する学習であるため、項目2を満たす学習手法としてProfit Sharingを用いる。Profit Sharingではある一定数のステップが経過した時点で、それまでの各状態に対して一括で報酬を与える。会話において、最終的な相手の反応にはそれ以前の会話の流れも関わるため、報酬の与え方が適していると考えた。

エージェントは親和・援助・依存的行動のポジティブ反応が得られることを正の報酬とし、反対に攻撃・拒否・回避的行動のネガティブ反応が得られた場合は負の報酬として価値関数を更新する。また本実験においては支配・服従的行動をニュートラルとした。価値関数は以下の式に基づいて更新する。

$$R_T = \begin{cases} 1 & \text{ユーザの行動が援助・親和・依存的行動} \\ 0 & \text{ユーザの行動が支配・服従的行動} \\ -1 & \text{ユーザの行動が攻撃・拒否・回避的行動} \end{cases} \quad (1)$$

$$W(s_t, a_t) \leftarrow W(s_t, a_t) + f(t, R_T, T) \quad (2)$$

$$f(t, R_T, T) = \gamma^{T-t-1} R_T \quad (0 \leq \gamma \leq 1) \quad (3)$$

$W(s_t, a_t)$ は現在状態である s_t に対する行動 a_t の価値を返す関数であり、 $f()$ は強化関数である。 γ は割引率であり、 R_T は時刻 T でのユーザの行動に応じて得られる報酬を表す。また本研究では負の報酬を用いるため、鈴木ら [4] が提案した修正値を用いて計算する。

2.2 行動選択過程

行動選択過程では以下の2点を考慮する。

1. 交流経験が少ないユーザには個人経験が使いにくい
2. 探索と同時に一貫性を考慮する

項目1を満たすためには個人に対する経験が多くなるほど個人経験に基づき行動選択する確率を上げていけばよい。そのため、全体経験と個人経験のどちらを用いるか選択する場面では ε を以下のように定義し、 ε_n の確率で全体経験を元に行動を選択する。 n は各ユーザに対する応答回数である。

$$\varepsilon_n = k * \varepsilon_{n-1} \quad (\varepsilon_0 = 1, k = 0.90) \quad (4)$$

項目2を満たすためには、全体もしくは個人のどちらかを選択した上で ε -greedy 法と Softmax 法を組み合わせさせて行動を選択する。ここでは、固定された ε_{fix} ($\varepsilon_{fix} = 0.20$) を元に、 $1 - \varepsilon_{fix}$ の確率で価値最大の行動を選択し ε_{fix} の確率で Softmax 法により行動を選択する。これにより、探索と同時に一貫性を考慮しながら行動を選択する。一貫性を考慮するのは、親和的行動と拒否的行動のような真逆の行動が交互に選択されるといった情緒が不安定になることを極力回避するためである。

3 実験

シナリオを用いたユーザ・エージェント間の対話を実施した。本研究では行動決定の部分を重視しているため、シナリオの選択肢は予め斎藤のモデルに基づく8行動がラベル付けされているものを用いる。シナリオは自作であり、ラベル付けは3名によるアノテーション作業を実施した。

3.1 対話実験設定

20代の学生20名(男性16名, 女性4名)を対象にシナリオ対話実験を実施した。各実験協力者はエージェントと10往復の対話を3回実施する。初回はユーザの行動に対しエージェントはランダムに返答することで、各行動に対してポジティブな反応を得るために効果的な返答を学習する。2・3回目は別のシナリオにおいて、それまでの学習結果を踏まえた上で返答する。

実験協力者はエージェントが学習して返答する群とランダムで返答する群の2群に分け、エージェントの学習効果を検証する。2群の比較により、ユーザの返答の変化がエージェントの学習による効果であるか、エージェントに対する馴れによる影響であるか確認する。実験協力者に偏りが無いよう、TIPI-Jに基づく回答を元に外向性の値を用いて実験協力者を2群に振り分けた。

3.2 評価

ユーザから得られたポジティブ反応数の比較と、SD法を用いたエージェントに対する印象評価と記述回答によりユーザとの友好関係構築ができてきているのか、どの要素が友好関係構築に寄与しているか分析する。

4 結果

図3に各群でのユーザからのポジティブ反応数、図4に各回でのエージェントに対する平均好感度を示す。好感度は1から7(7が最良)の7段階評価で回答させた。

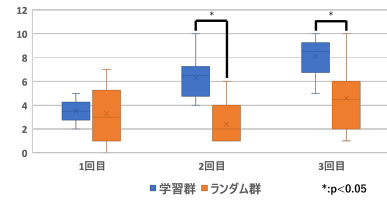


図3: 各群平均ポジティブ反応数 (n=20)

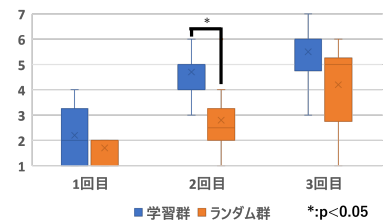


図4: 各群平均好感度 (n=20)

有意水準を5%としたt検定により、同条件の1回目については有意差が見られず、2・3回目のポジティブ反応数と2回目の好感度において有意差が確認された。

また全てのポジティブ反応数と好感度の相関を調べた結果、相関係数が0.66となり正の相関が見られた。

5 考察

実験結果より、20回程度の少ない応答回数でユーザからポジティブな反応を引き出すための学習効果が確認された。しかし、好感度に対する記述回答では「突然態度が変わった」との意見があった。これには行動選択過程での Softmax 法による探索が影響していると考えられるため、探索と現時点で最良と思われる行動を選択するバランスについては再度検討する必要がある。

6 今後の展望

今後の展望として、今回の実験によって得られた価値関数を用いて、エージェントが初対面のユーザに対してどのような反応を示すか実験する。また、本実験では行動の中でも発話のみに着目したが、身体的な動作まで含めた学習についても同様に検証していく。

謝辞

本研究は、一部、文科省科研費 (JP19H01137, および, JP20H04018), ならびに, NICT 委託研究の助成により行われた。

参考文献

- [1] 福田 和彰 他: 対人感情モデルを導入した親しみやすい感性対話エージェント, 情報処理学会第82回全国大会, 99-100(2020)
- [2] 斎藤 勇: 対人感情の心理学, 誠信書房 (1990)
- [3] 内田 貴久 他: 対話意欲を喚起する価値観肯定・否定割合に基づく自律対話ロボットの対話戦略, 人工知能学会全国大会 30, 115-1(2016)
- [4] 鈴木 敦司 他: 罰を考慮した Profit Sharing 強化学習法, 人工知能学会全国大会論文集, 241-241(2003)