

子ども向け選択式音声対話システムにおけるセリフ選択自動化の検討

A Study on Automatic Dialogue Selection in a Spoken Dialogue System for Children

花田 昂 岡 夏樹 田中一晶 高橋ともみ 久保祐喜

Hanada Subaru, Oka Natsuki, Tanaka Kazuaki, Takahashi Tomomi, and Kubo Yuki

京都工芸繊維大学
Kyoto Institute of Technology

要旨: ペチャットはぬいぐるみにつけるボタン型スピーカーである。養育者がアプリを操作することでぬいぐるみを通して子どもと音声対話することができる。我々は、養育者によるセリフ選択履歴と子どもの反応に基づいて、セリフ選択を自動化することを目指している。本論文では、1) セリフ選択履歴からの教師付き学習と、2) 子どもの音声応答の有無を報酬としたセリフ選択の強化学習を試みたので、その結果を自動化検討の第一報として報告し、教師付き学習と強化学習を併用することを提案する。

はじめに

ペチャット¹は、市販されているボタン型スピーカー（直径 4.5 cm×奥行 2.1 cm）である。ペチャットをぬいぐるみに装着しておくことで、養育者がスマートフォン上のアプリを操作することで、ぬいぐるみを通して子どもと音声対話することができる。ペチャットには、呼びかけ／質問／ごはん・おやつ／お片づけ／じゃんけんなどテーマごとに可愛い声でセリフが収録されており、養育者はその中からペチャットが発話するセリフを選択する。

我々の目標はこのセリフ選択を自動化することである。この課題の特徴は以下の通りである：

1. 養育者によるセリフ選択履歴が利用できる。
2. 対話相手が子どもである。
3. 子どもからの非定型な応答が多発することが予想される。
4. 入力デバイスは内蔵マイクだけである。
5. 応答候補が用意された非タスク指向対話である。

音声対話システムの研究は、

- タスク指向対話と
- 雑談等の非タスク指向対話を対象とするものに大別される。また、
- 自然言語理解、対話戦略の学習、自然言語生成などのモジュールで構成されたシステムを構築するアプローチ

- 応答候補から選択するアプローチ
- end-to-end で学習するアプローチ

がある。本研究では上記の 5 番目に記した特徴を有する課題を扱うため、対象は非タスク指向対話であり、応答候補から選択することにより対話するアプローチをとる。

また、特徴 1 から、養育者によるセリフ選択履歴からの教師付き学習が考えられるが、本研究ではこれに加えて、子どもの応答を報酬とした強化学習を試み、教師付き学習と強化学習を併用することを提案する。

セリフ選択履歴からの教師付き学習

学習データ

本論文で使用した学習データは、市川らの観察実験[1]により収録されたものである。データ収録実験の概要を以下に記す。なお、実験が行われた京都工芸繊維大学の倫理審査委員会から実験実施の承認を得ており、参加した母親全員から書面にて参加の同意を得ている。

3歳から6歳までの年少児から年長児の子ども（男児 8 名、女児 12 名）と母親、計 20 組が実験に参加した。全員、家庭ではペチャットは所有していなかった。実験は京都工芸繊維大学内の実験室で実施された。実験の様子は 2 台のビデオカメラにて撮影さ

¹ <https://pechat.jp/>

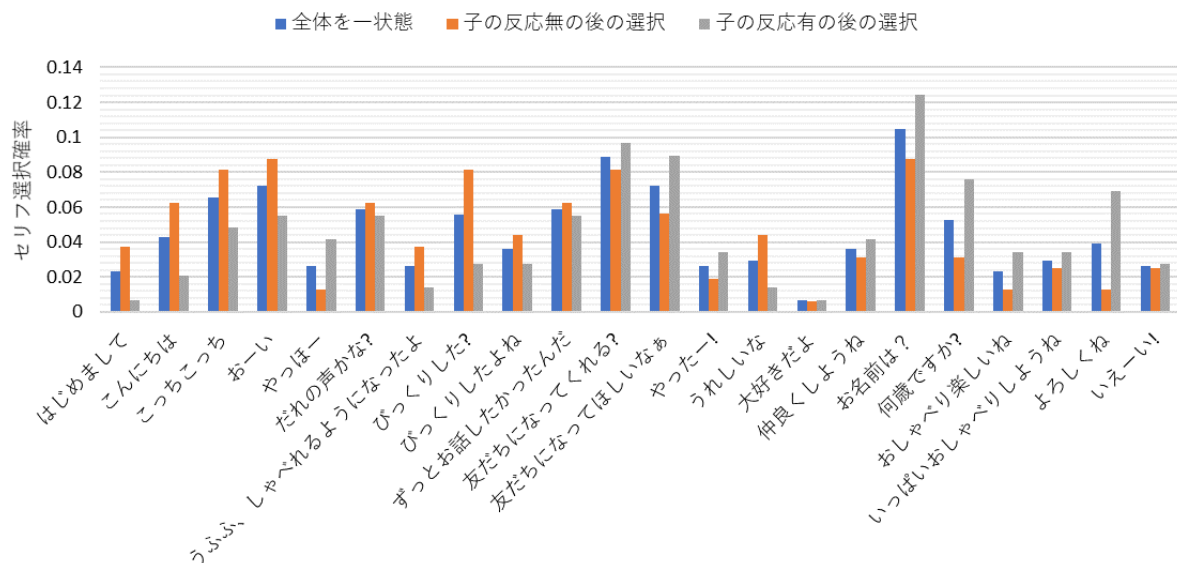


図 1: 母親によるセリフの選択確率

れ、アノテーションソフト ELAN²を用いて、子どもの発話を含むインタラクションの様子が記録された。

実験手続きのうち、本研究の目的と関連する部分の概要は次の通りであった。まず、実験室にきた母親は子どもに気付かれぬ状態でペチャットの操作練習を行い、その間、子どもは実験者と遊んでいた。その後、母子の自由遊びを経て、対話実験を実施した。対話実験では、実験者が指定したテーマ（(1) はじめまして (2) あいさつ (3) くしゃみ (4) 週末 (5) 頑張っていること）で、順に対話操作をしてもらった。原則、全ての指定したテーマで対話を行ってもらい、一区切りがついたと母親が判断するまで 10 分程度実施した。

テーマ (1) はじめまして の操作インタフェースは、22 種類あるセリフの一覧から 1 つ選択するもので、この固定された一覧からのセリフ選択を何回か繰り返すことになる。これに対して、テーマ (2) あいさつ (3) くしゃみ (4) 週末 (5) 頑張っていること の操作インタフェースは、提示される 3~5 個のセリフから 1 つ選択すると次の画面に遷移し、新たな選択肢が提示される一という流れを 6~7 回繰り返すものであった。

以上の実験により収録した対話データのうち、本論文では、テーマ (1) はじめまして におけるデータを学習・分析対象とした。これを対象とした理由は、このテーマの操作インタフェースが、固定した同一の選択画面からの選択を繰り返すものであったため、選択の傾向を学習・分析するために必要なデ

ータを比較的短い実験時間で得ることができたからである。

教師付き学習の方法

どのようなときに、どのセリフを選択するとよいかを、母親の選択をお手本に学習して自動化することがここでの目標である。後述の強化学習での用語と一致させると今後の議論に都合がよいので、「状態」「行動」という語句を用い、この目標を書き直すと、各状態でとるべき行動を教師付きで学習する一となる。

基本的には 22 種類のセリフの中から 1 つを選ぶことが「行動」であり、本論文ではこれを行動とした場合の学習・分析結果について報告する。ただし、ペチャットの操作インタフェースではセリフの選択以外に、別の選択画面に遷移する操作も可能であるため、これも行動の一つとして拡張することは今後の課題である。

一方、「状態」をどう定義するかについては、様々な可能性がある。最も大ぐくりすると、はじめまして のテーマ全体で状態は一つだけで状態遷移しないと見ることである。この場合の教師付き学習は単純であり、はじめまして のテーマ全体を通しての母親のセリフ選択確率と同じ確率でセリフを選択すればよい。この場合の母親のセリフ選択確率を図 1 に示した。横軸は 22 種類のセリフ、縦軸がセリフ選択確率である。3 本並んだ 3 色の棒グラフのうち、左側の紺色が全体を一状態とみたときの選択確率を示

² <https://archive.mpi.nl/tla/elan>

す。「お名前は?」「友だちになってくれる?」のように多く選択されたセリフと、「大好きだよ」のようにあまり選択されなかったセリフがあることが分かる。教師付き学習により、母親の選択頻度に応じた選択が可能となる。

状態を細分化して学習すると、より絞り込んだ選択ができるようになる。一例として、ペチャットの一つ前のセリフに対する子どもの音声応答の有無により、状態を2つに分けることを考えよう。図1には、この場合の母親のセリフ選択確率を併記してある。3本並んだ3色の棒グラフのうち、真ん中の赤色が直前の子どもの応答が無かった場合のセリフ選択確率を示し、右側の灰色が直前の子どもの応答があった場合の選択確率を示す。「はじめまして」「こんにちは」「びっくりした?」のような、子どもの反応が無かった場合に選択されやすいセリフと、「何歳ですか?」「よろしくね」のように、子どもの反応があった場合に選択されやすいセリフがあることが分かる。これらの選択傾向の違いは、教師付き学習により再現できると考えられる。

さらに、我々は、子どもの音声応答の有無に加えてその韻律情報により状態を細分化することで、お母さんの選択をよりきめ細かく再現できる可能性があると考えている。この検討結果については、[2]で報告する。さらに、応答内容による状態分割も考えられる。

以上の他にも、直前のペチャットのセリフが何であったかにより状態を分割することが考えられる。ただし、これを精度よく実現するためには、細分化の程度に応じた量の学習データが必要となるため、どのように分割するかを自動制御することも含めて検討中である。

子どもの応答からの強化学習

学習データ

上述の教師付き学習で用いたインタラクシオンデータをここでも使用する。ただし、強化学習では報酬を利用するため、報酬の値を求める必要がある。本論文では、行動(セリフの選択)に対する報酬を、そのセリフに対する子どもの音声応答の有無で定義した。具体的には、応答があった場合の報酬を+1とし、無かった場合を-1とした。

報酬の算出においても、子どもの応答の韻律情報を加味することで、よりきめ細かい報酬情報を利用できると考え、その効果を検討中である。この結果についても、[2]で報告する。子どもの応答は音声だけでなく、うなずきなどの仕草や表情などにも表れ

るため、これらを利用することでさらに精度良く報酬を求めることができるが、ペチャットの入力デバイスは内蔵マイクだけであるため、本研究では子どもの音声応答だけから報酬を算出する方針とした。また、実際の使用場面を考えると、対話相手の子どもの音声とそれ以外の音や音声(たとえば養育者の声)を区別する必要があるが、本論文では、これが区別できた場合の学習性能を評価することとし、人手により子どもの音声応答の有無をタグ付けしたデータを学習に用いた。

強化学習の方法

標準的な強化学習の手法であるQ学習[3]を採用した。Q値の更新式を次に示す:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left(r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right)$$

ここに、 $Q(s, a)$ は、状態 s で行動 a をとった場合の行動価値、 α は学習率、 r は報酬、 γ は割引率である。

教師付き学習と同じく、状態としては、はじめましてのテーマ全体で状態は一つだけとする場合と、ペチャットの一つ前のセリフに対する子どもの音声応答の有無により、状態を2つに分ける場合を検討した。また、行動についても、教師付き学習と同じく、22種類のセリフの中から1つを選ぶことを行動とした。

ペチャットのセリフ(行動)に対する子どもの音声応答の有無を報酬と定義したため、報酬は行動ごとに毎回即時報酬として与えられることになる。このため、本論文では割引率 $\gamma = 0$ として、即時報酬のみを考慮した。今後は、別の選択画面に遷移する操作を行動として含むように拡張する計画であるため、遅れのある報酬も考慮したQ値の更新式を使用する予定である。

行動は、次式の通り、softmaxで確率的に選択することとした:

$$\frac{e^{Q(s,a)/\tau}}{\sum_{b=1}^n e^{Q(s,b)/\tau}}$$

本論文では温度定数 $\tau = 1$ とした。

強化学習の結果

上述のインタラクシオンデータを用いてQ値を更新していくが、十分に学習を繰り返すとQ値は報酬の期待値に収束することが知られている。本データの場合は、 $Q(s, a)$ は、状態 s で行動 a をとった場合の報酬 r の期待値に収束するため、Q値の収束先は、インタラクシオンデータから求めることができる。この収束したQ値を用いてsoftmaxで行動を選択した

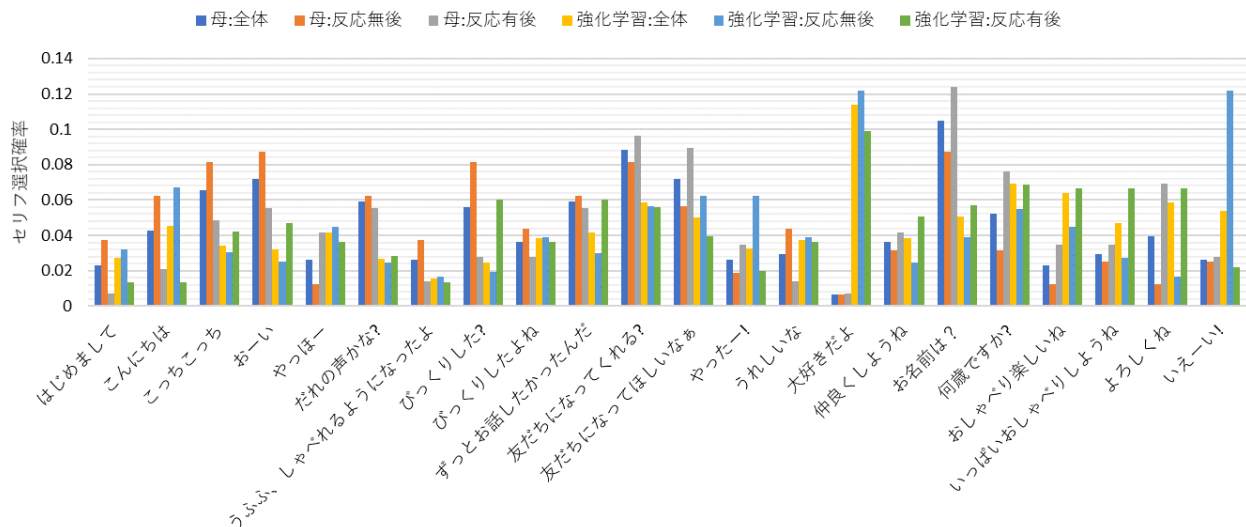


図2: 母親によるセリフ選択確率と強化学習によるセリフ選択確率の比較

場合の選択確率を図2に示す。図2は図1との比較を容易にするため、図1の棒グラフ(母親の選択確率3本)はそのまま再掲し、その右側に、強化学習による選択確率の棒グラフ3本を追加した。追加した3本は左から順に、黄色:強化学習(全体を一状態とみたもの)、深緑色:強化学習(子どもの応答が無かった場合の選択)、浅緑色:強化学習(子どもの応答があった場合の選択)一である。

図2から次のことが分かる。母親の選択確率と強化学習による選択確率(期待報酬に基づく選択確率)が似ているセリフもあるが、大きく異なるセリフもある。「はじめまして」「こんにちは」「びっくりしたよね」「よろしくね」などは、全体を一状態とみなした場合も2状態に分割した場合も、母親と強化学習による選択確率が似ている。これは、母親がほぼ期待報酬に沿った選択をしていたことを意味する。これに対して、「こっちこっち」「おーい」「だれの声かな?」「お名前は?」などのセリフは、母親の期待に反して子どもが応答しにくいものであったことが分かる。これは子どもが初めての実験室や雰囲気緊張していつもと異なっていたであろうことや、ぬいぐるみに話しかけられる状況での子どもの応答が母親にとっても初めてで予測しにくいものであったことが原因の可能性がある。逆に、「おしゃべり楽しいね」や「いえーい!」のように母親の予想より子どもが応答しやすいセリフもあったことが分かる。なお、「大好きだよ」の選択確率も大きく異なるが、これは、母親の選択回数が2回しかなく、たまたまその2回とも子どもが応答したことによると考えられるため、データが増えたときに、同様の結果が出るとは限らない。

教師付き学習と強化学習の併用

以上の結果と考察から、次のことが分かる。

- 母親の選択は常に最適(子どもの応答があるという意味)とは限らないため、母親の選択を教師信号とした教師付き学習により、子どもの応答の確率を最大化する選択ができるとは限らない。
- 逆に、次のことも言える。
- 子どもの音声応答の有無を報酬として学習することは、子どもの応答の確率を最大化する選択には必ずしもつながらない。これは、音声応答の有無の認識誤り、韻律情報の不使用、音声以外の身振りや表情などの情報の不使用に起因して、報酬の最大化が子どもの応答の良さの最大化には必ずしもならないからである。

以上のように、

1. 教師信号は最適ではない
2. 報酬も最適なものとは与えられない

この条件の下に、教師信号と報酬を利用した学習を行う場合は、教師付き学習と強化学習を併用する[4][5][6]ことが望ましいと考える。具体的には、教師付き学習による行動の選択確率と強化学習による行動の選択確率の平均をとった確率で行動するという解が考えられる。どちらがより信頼できそうかの情報が得られる場合は、重み付きの平均を使うことができるだろう。以上のような、教師信号も報酬も最適とは限らないという状況は、本論文で対象としたペチャットの発話選択に限らず、一般的に見られると考えられるため、そのような状況での学習手法を

開発することは、有意義であろう。

control with supervised and reinforcement learning,
arXiv:1702.03274, (2017)

まとめと今後の展望

本論文では、子ども向け選択式音声対話システムにおけるセリフ選択自動化を目指して、1) 養育者によるセリフ選択履歴からの教師付き学習と、2) 子どもの音声応答の有無を報酬としたセリフ選択の強化学習を試み、その結果を報告した。この課題は、養育者によるセリフ選択が最適であるとは限らず、また、報酬の算定も正確であるとは限らない、という性質を持つことを示し、そうした状況においては教師付き学習と強化学習の併用が必要であることを結論とした。

今後は、状態の自動分割、行動（発話）の選択肢を変更する行動の導入などにより、実用レベルの学習システムを構築し、さらに、使用履歴からのオフライン学習だけでなく、ユーザの使用場面でのオンライン学習を試みる計画である。

謝辞

本研究は JSPS 科研費 JP20H05564 の助成を受けたものである。また、博報堂 Pechat 開発チーム（堀紫・小山徹也・古矢真之介）の協力に感謝する。

参考文献

- [1] 市川 淳, 光國 和宏, 堀 紫, 池野 湧太, アレクサンドル ルブロン, 河本 徹和, 西崎 友規子, 岡 夏樹: 性格に着目した子どもと対話型スピーカーのインタラクションに関する検討, 日本感性工学会論文誌, Vol. 19, No. 2, pp. 173-179, (2020)
- [2] Kubo Y., Oka N., Hanada S., Tanaka K., and Takahashi T.: Automatic utterance selection based on prosodic features of children's vocalizations, the Annual Conference of Japanese Society of Artificial Intelligence (JSAI 2021), (to appear)
- [3] Watkins, C. J. C. H.: Learning from delayed rewards, PhD thesis, Cambridge University, Cambridge, England, (1989)
- [4] Henderson, J., Lemon, O., and Georgila, K.: Hybrid reinforcement/supervised learning of dialogue policies from fixed data sets, Computational Linguistics, Vol. 34, No. 4, pp. 487-511, (2008)
- [5] Xiujun Li, Lihong Li, Jianfeng Gao, Xiaodong He, Jianshu Chen, Li Deng, Ji He: Recurrent reinforcement learning: A hybrid approach, arXiv:1509.03044, (2015)
- [6] Williams, J. D., Kavosh Asadi, K., and Zweig, G.: Hybrid Code Networks: practical and efficient end-to-end dialog