

Sim2RealQA：ニューラル質問応答モデルの 仮想世界から現実世界への転移

宮西 大樹^{1,a)} 前川 卓也³ 川鍋 一晃^{1,2}

概要：日常生活で起きた出来事に対して質問応答できるようになれば、人間の記憶支援・忘れ物や落とし物の検索・人の監視や見守りといった実世界に根ざしたシステムが実現できるようになる。従来の実世界質問応答では、近年数多くの質問応答の課題で高い性能を示すニューラルネットワークで構成した質問応答モデルが用いられてきた。しかし、実世界のデータをラベリングする作業はプライバシーの問題を引き起こすため、実世界の質問応答データセットを作成することは困難である。その一方で、ニューラルネットワークを用いた質問応答モデルは、その能力を発揮するために大量の学習データが必要になる。本論文では、この制限を克服するため Simulation to Real QA (Sim2RealQA) という新たな枠組みを提案する。本手法では、プライバシーを侵害することなく十分な量の学習データを作るため、人の日常生活を模倣するシミュレータを使用する。そして、現実世界の質問応答の問題を高精度に解くため、仮想世界の日常生活行動のデータをもとに作成した大量の質問応答データセットを用いてニューラル質問応答モデルを訓練する。Sim2RealQA の枠組みを評価するため、我々は実際の家屋とライフシミュレーションゲーム内の日常生活行動のログデータをもとに仮想・現実双方の質問応答データセットを作成した。このデータセットを用いて、実世界の解答ラベルがない場合、仮想世界のデータが実世界の質問応答に役立つことを実証する。

1. はじめに

現実の状況を理解し、日常生活に関する様々な質問に答えことができる知能システムの構築は数多くの実用的なアプリケーションに繋がる。例えば、「昨晚私は何を食べたっけ?」「山田さんは今どの部屋にいるの?」「さっき失くしたスマートフォンは今どこにあるの?」「鈴木さんは食後にちゃんと薬を飲みましたか?」といった質問に解答することができれば、記憶補助・人や遺失物の探索・見守りシステムなどに応用することができる。このような知能システムの実現に向けて、実世界質問応答という課題が提案されている [1]。実世界質問応答の目的は、行動認識技術を用いて作成した人間の日常生活に関する行動のテキストデータを読み込み、実世界の状況を問う質問に対して正しい答えを出力することである (図 1 下)。実世界質問応答は、複雑な推論と多様な回答の出力が必要であるため、従来の方法では、外部記憶機構を持つニューラル質問応答モデル [2,3] が用いられてきた。ニューラルネットワーク質問応答モデルは高精度な質問応答が可能であるものの、性能を発揮

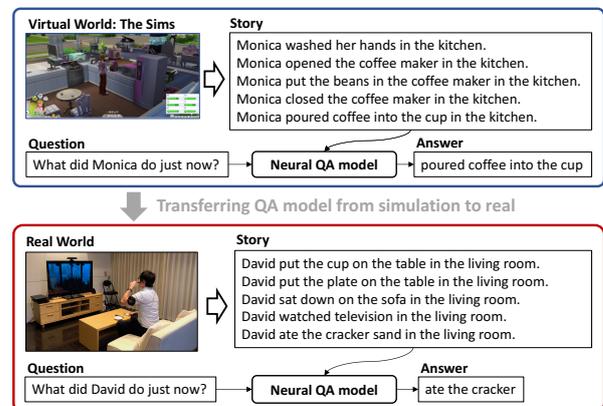


図 1 Sim2RealQA の概要

するために、学習データが大量になる。しかし、実世界のデータのラベリングは人手による確認が必要であるため、プライバシーを侵害する恐れがある。よって、実世界の解答ラベルの取得は困難である。本稿では、この問題を解決するため、Simulation to Real QA (Sim2RealQA) という新たな枠組みを提案する。

本手法では、プライバシーを侵害することなく十分な量の学習データを作るため、人の日常生活を模倣するシミュレータを使用する。そして、現実世界の質問応答の問題を高精度に解くため、仮想世界の日常生活行動のデータをも

¹ 国際電気通信基礎技術研究所

² 理研 AIP

³ 大阪大学

^{a)} miyanishi@atr.jp

ID	Question	Answer
1	Is [person] in [place]?	yes/no
2	Did [person] [activity] just now?	yes/no
3	Was [person] in [place]?	yes/no
4	Did [person] [activity] ?	yes/no
5	Where is [person]?	[place]
6	Where did [person] see [object]?	[place]
7	Where was [person] before [place]?	[place]
8	Where was [person] after [place]?	[place]
9	Where did [person] see [object] before [place]?	[place]
10	Where did [person] see [object] after [place]?	[place]
11	Who is in [place]?	[person]+
12	Who was in [place]?	[place]+
13	Who [activity]?	[place]+
14	Where did [person] [activity]?	[place]
15	What did [person] [activity]?	[object]
16	Which rooms did [person] go?	[place]+
17	What did [person] do just now?	[activity]
18	What did [person] do before [activity]?	[activity]
19	What did [person] do after [activity]?	[activity]
20	How many rooms did [person] go?	[number]
21	How many times did [person] [activity]?	[number]
22	How many times did [person] see [object]?	[number]

表 1 質問応答データセット作成するために使用した質問とその解答のテンプレート。[person][place][activity] は人物・場所・行動を表し、与えられたストーリーをもとに当てはめる。

とに作成した大量の質問応答データセットを用いてニューラル質問応答モデルを訓練する (図 1)。Sim2RealQA の枠組みを評価するため、我々は実際の家屋とライフシミュレーションゲーム内の日常生活行動のログデータをもとに仮想・現実双方の質問応答データセットを作成した。このデータセットを用いて、実世界の解答ラベルがない場合に、仮想世界のデータが実世界の質問応答に役立つかどうかを検証する。

2. Sim2RealQA データセット

2.1 日常生活ストーリーの収集

2.1.1 仮想世界の日常生活行動データの収集

仮想世界の日常生活行動データを取得するため、ライフシミュレーション The Sims 4 を用いる。The Sims では、屋内環境を自在に編集することができ、任意のレイアウトで家屋を設計、屋内に家電や家具を自由に配置することができる。また、この家屋にシムと呼ばれる人間を模したキャラクターを住まわせることができる。シムの内部には性格や感情や欲求などのパラメータが存在しており、このパラメータと周囲の環境に従って、シムは屋内を半自動的に行動することができる。例えば、空腹ゲージが下がる食べ物を食べる行動を取りやすくなり、膀胱ゲージが下がるとトイレに行く行動を取りやすくなる。本研究では、日常生活行動のシミュレーションを行うため、異なる 3 つの

	Virtual	Real
# Unique Questions Sentences	26,356	3,697
# Unique Answers Sentences	957	227
# Unique Story Sentences	7,216	692
Avg. Story Length	183	147
Max Story Length	351	160
Min Story Length	80	131

表 2 Sim2RealQA データセットの統計情報

住宅環境を用意した。各環境には、キッチン・ダイニング、リビング、バスルーム、ベッドルーム、エントランス、庭といった場所があり、人間が日常生活を送る上で必要な家具や家電を用意した。1人暮らしの環境では、男性1人と犬1匹、男性1人と猫1匹、女性1人と犬1匹、女性1人と猫1匹の単身世帯のシミュレーション。2人暮らしの環境では、男性2人、女性2人、男性1人と女性1人のルームシェアリングを想定したシミュレーション。3人暮らしの環境では、男性と女性の夫婦と子供1人の核家族世帯のシミュレーションをそれぞれ行った。1日の行動を1セッションとして、10セッション分のデータを取得し、各個人にどこで・だれが・何をしたかといった日常生活行動を表す文のラベリングを行った。

2.1.2 仮想世界の日常生活行動データの収集

現実世界の日常生活行動データを取得するため、屋内で行動認識のデータ収集でよく用いられる行動実験を行った。行動実験ではアイトラッキングカメラを被験者5人装着し、ワークシートに記述した場所ごと固有の19個の日常生活行動を連続して行った。例えば、寝室で本を読んだ後、テーブルの上をハタキで掃除し、その後、時計を確認したといった一連の日常生活行動を行う。このデータに対して、誰がいつどこで何をしたかといったテキストデータを人手付与し、実世界の日常生活行動のログデータとした。

2.2 質問応答データセットの作成

質問応答のデータセットを用意するため、表1の質問テンプレートを用意し、日常生活の内容や実世界の状況を問うような質問を自動的に作成した。さらに、実世界のデータの質問に対する解答を用意するため、文法構造をもとにしたルールベースの質問応答モデルを作成し、質問に対応する解答を自動的に生成した。仮想世界と現実世界の両質問応答データセットを総称して Sim2RealQA データセットと呼ぶ。表2に Sim2RealQA データセットの統計情報を掲載する。

3. 手法

仮想世界の質問応答データセットが実世界の質問応答を解くために役立つかどうかを以下の手法を比較することで検証する。

Method	Sim2Sim	Sim2Real
POP	0.280	0.272
RNN	0.639	0.357
RNN+AT	0.813	0.446
RNN+PG	0.882	0.647
DMN	0.709	0.394
DMN+PG	0.902	0.683

表 3 Sim2Sim と Sim2Real での RNN, RNN+AT, RNN+PG, DMN, DMN+PG の質問応答の精度

- **POP**: 訓練データの中で課題の種類 (表 1) ごとに最も頻出する解答を出力。
- **RNN**: sequence-to-sequence (Seq2Seq) [5] を用いた質問応答モデル。Seq2Seq を用いてストーリー文と質問文を読み込み、解答文を出力。
- **RNN+AT**: Seq2Seq に注意機構 [1], [3] を取り入れた質問応答モデル。
- **RNN+PG**: RNN+AT にコピー機構 [4] を導入した手法。解答文の単語の生成と入力文中の単語のコピーの両方を行うことが可能。コピー機構を用いることで未知語 (現実世界のデータには出現するが学習に使う仮想世界のデータには出現しない単語) に対処できる。
- **DMN**: 動的メモリネットワーク [2], [6] を用いた手法。文レベルの注意機構を用いて質問文とストーリー文の関係を表現。デコーダーについては上記 RNN と同じ。
- **DMN+PG**: 上記 DMN にコピー機構を導入した手法。POP とそれ以外を比較することでニューラルネットを用いた手法の有効性がわかる。RNN と RNN+AT を比較することで注意機構の有効性を検証でき、RNN+AT と RNN+PG を比較することでコピー機構の有効性を検証できる。また、RNN+PG と DMN+PG を比較することで、RNN と DMN のどちらのエンコーダーが Sim2RealQA に適しているかを検証することができる。

4. 実験

4.1 評価方法

Sim2RealQA データセットを用いて各手法を比較し、どのような要素が Sim2RealQA を解く上で重要であるかを示す。評価方法として以下の3つを考える。

- **Sim2Sim**. 仮想世界のデータセットで学習して、仮想世界のデータでテストする。
- **Sim2Real**. 仮想世界のデータセットで学習して、現実世界のデータでテストする。
- **Real2Real**. 現実世界のデータセットで学習して、現実世界のデータでテストする。

本稿の目的は Sim2RealQA の有効性を検証することであるため、Sim2Real が主な評価方法となる。さらに、Sim2Real と Sim2Sim の結果を比較することによって、Sim2Real の

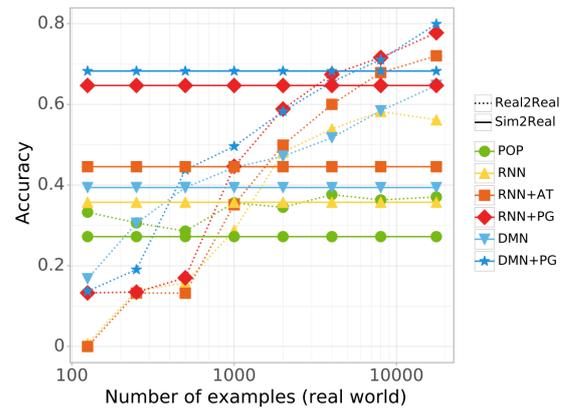


図 2 Sim2Real と Real2Real の条件下での質問応答の精度

課題としての難しさを知ることができ、どのような課題が難しいかを分析することができる。また、Real2Real と Sim2Real を比較することによって、Sim2RealQA の有効性とその限界を検証することができる。評価指標としては、真の解答文と予測した解答文が完全一致した場合の Accuracy を用いる。

4.2 実験結果

表 3 に Sim2RealQA データセットに対する各手法の Accuracy を掲載する。表から Sim2Sim の結果と比べて、Sim2Real の結果が低下していることがわかる。これは Sim2Real の場合、学習データに使用した仮想世界のデータと現実世界のデータの傾向が異なるためだと考えられる。次に、頻度ベースの手法 POP と学習ベースの手法 (POP 以外) を比較すると、学習ベースの手法が Sim2Sim, Sim2Real とともに上回っており、本データセットでの学習ベースの手法の有効性がわかる。また、Sim2Real の場合に関して、RNN+AT は RNN を上回り、RNN+PG は RNN+AT を上回っていることから、注意機構とコピー機構の有効性がわかる。さらに、DMN+PG と DMN を比較した場合でも、DMN+PG が DMN を大きく上回っていることから、Sim2RealQA によって、未知語に対処できるコピー機構が重要であることがわかる。また、DMN+PG が RNN+PG を上回っていることから、コピー機構を用いる場合、質問文とストーリー文の関係を表現できる動的メモリネットワークが Sim2RealQA に有効に機能することがわかる。

次に、Sim2Real と Real2Real を比較する。図 2 は横軸が実世界のデータを用いた学習に使う事例数で縦軸が Accuracy を示す。Sim2Real は実世界の正解ラベルが使えない設定なので Accuracy は一定になる。Real2Real の場合、予想されたとおり POP を除く学習ベースの手法は学習に使う事例数が増えるにつれて Accuracy が高くなる。そのため、十分に実世界の学習データがある場合、Real2Real は Sim2Real を上回っている。ただし、学習の事例数が少ない場合、学習ベースの Accuracy は著しく低下する。そ

のため、Sim2Realの学習ベースの手法がReal2Realの手法を上回ることがある。これは実世界の正解ラベルが取得できない場合、Sim2RealQAの枠組みが有効であることを示している。

5. おわりに

本稿では、仮想世界の質問応答データを使うことで、実世界の質問応答の課題を解くSim2RealQAの枠組みを提案した。Sim2RealQAの枠組みを評価するため、実際の家屋とライフシミュレーションゲーム内の日常生活行動データをもとに仮想・現実双方の質問応答のデータセットを作成した。このデータセットを用いてSim2RealQAの有効性を検証したところ、プライバシーの問題で実世界の解答ラベルが取得できない場合、Sim2RealQAが非常に有効な方法になりうることがわかった。

参考文献

- [1] Bahdanau, D., Cho, K. and Bengio, Y.: Neural Machine Translation by Jointly Learning to Align and Translate, *ICLR* (2015).
- [2] Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., Zhong, V., Paulus, R. and Socher, R.: Ask Me Anything: Dynamic Memory Networks for Natural Language Processing, *ICML*, pp. 1378–1387 (2016).
- [3] Luong, T., Pham, H. and Manning, C. D.: Effective Approaches to Attention-based Neural Machine Translation, *EMNLP*, pp. 1412–1421 (2015).
- [4] See, A., Liu, P. J. and Manning, C. D.: Get To The Point: Summarization with Pointer-Generator Networks, *ACL*, pp. 1073–1083 (2017).
- [5] Sutskever, I., Vinyals, O. and Le, Q. V.: Sequence to sequence learning with neural networks, *NeurIPS*, pp. 3104–3112 (2014).
- [6] Xiong, C., Merity, S. and Socher, R.: Dynamic memory networks for visual and textual question answering, *ICML*, pp. 2397–2406 (2016).