

# 視覚情報を用いた対話型物体特定タスクにおける 質問数削減手法の検討

金沢 壮真<sup>†</sup> 松森 匠哉<sup>‡</sup> 新行内 浩輔<sup>‡</sup> 福地 庸介<sup>‡</sup> 今井 倫太<sup>†</sup>

慶應義塾大学理工学部<sup>†</sup> 慶應義塾大学大学院理工学研究科<sup>‡</sup>

## 1. はじめに

エージェントが人とインタラクションしながら情報を獲得することは人工知能分野における最も重要な課題の一つである。とりわけ、実世界エージェントは、設計/学習時に存在しない未知の状況に遭遇することが多い。そのため、ユーザに対して適応的に質問をすることで情報獲得を行うことは有効である。

本稿では視覚的に提示された物体群から言語による質問を通じて相手の注目する物体を特定するタスク Goal-Oriented Visual Dialogue (GOVD) [1]を扱う。GOVD においては、質問生成と対象となる物体の特定を、深層学習を用いて end-to-end に学習する手法が提案されてきた[2]。しかし、これらの手法では予め決められた上限まで質問する方策を取ることが多く、冗長な質問が生じる課題が生じていた。これは人間とエージェントがインタラクションを行う際には大きなストレスとなり得る。

本研究ではこのような end-to-end なモデルにおいて強化学習の報酬関数を改良して学習を行うことで冗長な質問の削減に取り組み全体的な質問数の削減を図った。既に尋ねた質問と同じ質問を生成することに負の報酬を与える報酬関数や物体推定の確率分布から求まる相対的な情報量の推移から計算される報酬関数を提案し、これらの関数が質問数の削減に寄与していることが明らかになった。

## 2. Goal-Oriented Visual Dialogue

GOVD においては深層強化学習を利用して Guess What?! [2]タスクを解くモデルが多く提案されている。Guess What?! [2] タスクは、Oracle が画像中の物体を一つ選択し、エージェントが質疑応答を通して Oracle が選択した物体を特定するこ

Improving Goal-Oriented Visual Dialogue by Asking Fewer Questions

<sup>†</sup>Soma Kanazawa, Michita Imai

Faculty of Science and Technology, Keio University

<sup>‡</sup>Shoya Matsumori, Kosuke Shingyouchi, Yosuke Fukuchi

Graduate School of Science and Technology, Keio University

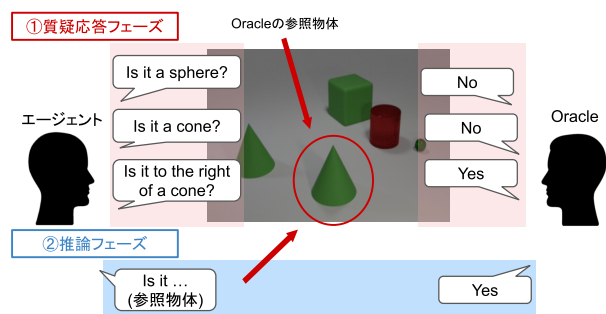


図 1: Goal-Oriented Visual Dialogue

とを目的としている。

本稿では図 1 のように視覚情報が与えられた対話において、エージェントの質問と Oracle の応答が繰り返されるタスクにおいて CLEVR[3]データセットを使用して学習し、検証を行った。Oracle とは実世界におけるユーザの役割をするものであり、エージェントが行った画像に関する質問に対して “Yes” または “No” の応答を行う。

画像中の物体を特定するためにエージェントは複数のターンに渡って質問生成を行う。しかし生成される質問の中には全く同じ内容の質問を繰り返して生成する場合や、既に行った質問によって、明らかに物体が特定できる状況においても予め決められた上限まで質問を続ける課題が生じていた。

本稿では既存のモデルにおいて三種類の報酬関数を組み合わせて学習を行うことにより、GOVD においてタスク成功までに要する質問数の削減を行う手法を提案する。

## 3. 提案

本稿では以下の三種の報酬関数を提案する。

Same Question Penalty は過去の質問と全く同じ内容の質問を生成した際に与えられる定数の負の報酬である。

表 1: 実験の定量的評価

手法	タスク成功率(%)	質問数	精度寄与度
提案報酬なし	83.88±0.59	4.00±0.00	0.21
提案報酬あり	83.00±0.33	3.35±0.35	0.24

Turn Penalty は質問生成が行われる度に負の定数の報酬が与えられる関数であり、一つのエピソードで与えられる報酬は式(1)のように質問数  $t$  が多いほど少なくなる。

$$R_{tp} = -C_p \times t \quad (1)$$

なお、本稿では  $C_p = 0.2$  として学習を行った。Kullback-Leibler (KL) divergence Reward は物体の推測確信度を表す確率分布において、質問生成を行う前の確率分布  $P$  と質問後の確率分布  $C$  の分布間擬距離を報酬  $R_{KL}$  として以下の式(2)で与えられる報酬である。

$$R_{KL} = \sum_{i=1}^n C(i) \log \frac{C(i)}{P(i)} \quad (2)$$

$i$  は画像中の物体の各インデックス、 $n$  は物体数を表している。質問生成によって変化する確率分布において、得られる情報量がより多くなるように学習を行うことを目的としている。また報酬  $R_{KL}$  においても、質問生成が行われる度に報酬が算出される即時報酬として実装を行った。

## 4. 実験

### 4.1 実験条件

Transformer アーキテクチャ [4] を用いたモデルにおいて、3 章で提案した報酬関数を用いて深層強化学習を行った。

実験は教師あり学習で事前学習を行った後、強化学習の方策勾配法の一つである REINFORCE を用い、学習率  $5e-4$ 、バッチサイズを 1024 に設定して行った。

表 1 に定量的評価としてタスク成功率、タスク成功までに要する質問数、質問一问あたりの精度寄与度を示す。質問一问あたりの精度寄与度とは、タスク成功率を平均質問数で割った値であり、生成された各質問がどの程度タスク成功に寄与したかを表している。また図 2 に、提案する報酬関数を使用しない場合と使用した場合で強化学習を行った際に実際に生成された質問の例を示す。

### 4.2 実験結果

実験結果の定量的評価を示した表 1 において、提案手法は既存モデルよりも質問数を少なくしてタスクを成功させていることが明らかになり、

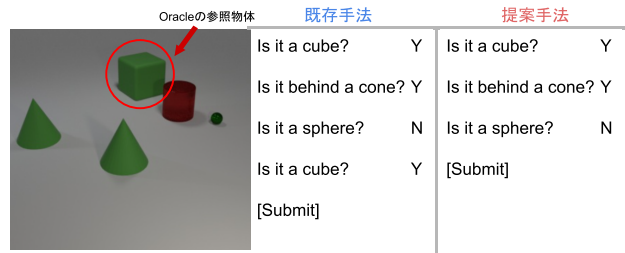


図 2: 生成された質問の例

結果として質問一问あたりの精度寄与度も高くなっていることがわかる。

図 2 の左側では本稿の提案手法を用いずに学習された既存モデルが生成した質問を示し、右側は提案する報酬関数によって学習されたモデルが生成した質問を示している。

既存のモデルでは一度行った質問と同じ質問を繰り返して生成している場合があるのに対し、提案する報酬関数によって学習されたモデルは同じ質問を繰り返していないことがわかる。

## 5. おわりに

本稿で挙げた三種類の報酬関数を使用して強化学習を行うことにより GOVD における質問数を削減し、生成される質問の一问あたりの精度を向上させることを示した。

今後の課題として質問生成を行うモデル自体の改良に取り組み、さらに精度を上げる必要があると考えられる。また、図 2 の質問例のように本来ならば一問目の質問で Oracle の参照物体を確定できる場合には、より早く質問を終えることができるような学習方法の改良に取り組むべきだと考えられる。

## 参考文献

- [1] Vries et al. (2017) “Guess What?! visual object discovery through multi-modal dialogue”. CVPR.
- [2] Strub et al. (2017) “End-to-end optimization of goal-driven and visually grounded dialogue systems”. IJCAI.
- [3] Johnson et al. (2017) “CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning”. CVPR
- [4] Vaswani et al. (2017) “Attention is all you need”. NIPS