合理的な論証構築ための情報探索対話戦略の最適化と評価

勝見 久央 [1] 吉野 幸一郎 [1,3] 平岡 拓也 [2] 本浦 庄太 [2] 山本 風人 [2] 定政 邦彦 [2] 中村 哲 [1]

[1] 奈良先端科学技術大学院大学
[2] 日本電気株式会社
[3] 科学技術振興機構 さきがけ

1 はじめに

論証のやり取りによって対話を行う論証対話システムは説得や交渉といった場面で有用であり，近年広く研究されている [1, 11, 12, 8]．論証は主張とその根拠から構成され，システムの主張を対話相手が受け入れるためには，システムは適切な根拠に基づいて合理的な論証を構築する必要がある．しかし，実際の対話においては，システムが事前にそのような根拠をすべて保持していることは稀である．そこで，システムが対話中に，その主張の根拠となる情報を必要に応じて対話相手から収集することができれば，不足している根拠を補って，合理的な論証を構築することができる．こうした対話は，例えば裁判において裁判官が，合理的な判決を下すために必要な情報を証人に問い合わせるような場面でみられる [7]．

対話中に必要な情報を収集する形式の対話は情報探索対話に分類される [13]．すなわち，論証対話システムにおいては，システムが質問者となり，情報探索対話を通して，合理的な論証構築に必要な根拠を収集するために回答者に対して質問を行う．この際，実際の対話においては，収集すべき根拠の候補はしばしば多岐にわたり，回答者の知識の欠落によって収集が失敗する場合や，限られた時間の中で根拠の収集を行う必要があるなどの制約を考慮する必要がある．このため，人手で合理的な論証構築のための情報探索対話の洗練された戦略を構築することは非常に困難である．

関連研究として，自律型エージェント間の情報探索対話における最適対話戦略に関する研究 [5, 6, 11] が存在するが，これらの研究では，上述の要因を考慮せず，対話戦略のルールを手作業で構築している．例えば，質問者の戦略は候補となりうるすべての事実に関して網羅的に質問を行うように構築されている．これに対して本研究では，情報探索対話における事実のやりとりをマルコフ決定過程によって定式化し，深層強化学習によって情報探索対話の対話戦略最適化を行う．

2 論証構築のための情報探索対話

本研究では，対話を通じた合理的な論証構築を目的とした情報探索対話における対話戦略を強化学習によって構築する．本章では，合理的な論証構築のための情報探索対話のマルコフ決定過程を用いた定式化と，その定式化において深層強化学習によって最適対話戦略を学習する方法について説明する．

2.1 合理的な論証構築のための情報探索対話

本研究では情報探索対話として，質問者であるシステム Que が合理的な論証構築のために回答者 Ans から必要な事実を収集する対話を取り扱う．

論証は主張 α とその根拠 Φ の 2 つ組 $\langle \Phi, \alpha \rangle$ によって表現される．Φ は事実と推論規則の集合によって構成され，α は単一の事実で構成される．なお，本研究では原子論理式 $q_1, \ldots, q_n, q$ を用いて $q_1 \wedge \ldots \wedge q_n \rightarrow q$ と表現できるもののうち $n = 0$ のもの，すなわち q を事実と呼び，それ以外のものを推論規則と呼ぶ．さらに，[4] らの論証の枠組みに従い，論証には収集した事実だけでなく，適宜補完された仮説も事実として根拠に含めることを許す．仮説された事実は記号 asm を用いて表現する．図 1 にある時刻 t での論証の例を示す．ここで，主張は，「少年は無罪である.」という事実であり，収集した事実は,「少年は若い.」と,「女性は少年が被害者をナイフで刺しているところを目撃していない.」という 2 つの事実である．そして，ここから 2 つの推論規則によって,「少年は危険な殺人犯ではない.」と,「老人は被害者の叫び声を聞いていない.」という 2 つの事実が仮説され，これらによって主張が支持されている．

主張 α と仮説推論モデル L が与えられたとき，論証 $\langle \Phi, \alpha \rangle$ は $\Phi = K_Q \cup H$ として構築される．ただし，K_Q は質問者が保持している事実と推論規則の集合からなる知識で，H は仮説された事実の集合である．また，今回は合理的な論証構築に必要な推論規則はすべて予め与えられているものとする．ここで H は次を満たすような仮説である．

$$\arg\min_{H \in \mathcal{H}} \mathrm{E}(\{\alpha\} \cup K_Q \cup H),$$
$$s.t., K_Q \cup H \models \alpha, \{\alpha\} \cup K_Q \cup H \not\models \bot.$$

なお，$\mathcal{H}$ は仮説候補である[1]．また E は推論モデルにおけるコスト関数を表す．例えば $\mathrm{E}(\{x, y, \ldots\})$ は，$x, y, \ldots$ が

[1] $\models$ は右項が左項から導出可能であること，$\cup$ は和集合をそれぞれ

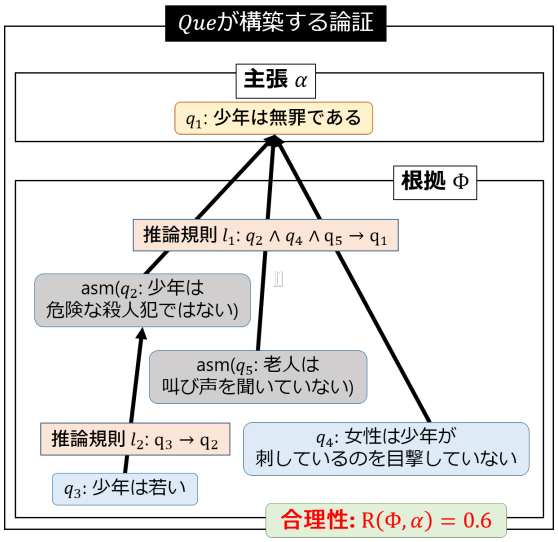

図 1: 質問者 Que が構築する論証の例

同時に真となる蓋然性を表す [3]．なお，本研究では [10] に従い，論証の合理性を次のように算出する．

$$\mathrm{R}(\Phi,\alpha) = (\min_{H'} \mathrm{E}(\{\alpha\} \cup H') + \min_{H''} \mathrm{E}(K_Q \cup H'')) - \mathrm{E}(\{\alpha\} \cup \Phi) \quad (1)$$

本研究では図 2 に示すように，システムは事前に論証の主張 α を与えられ，次の (1)〜(4) の手順に従い事実を収集する．(1) Que は Ans から事実を収集するために問い合わせを行う．(2) Ans は Que の問い合わせに応じた事実が自身の知識に含まれていればそれを返す．(3) Que は収集した事実 q を自身の知識 K_Q に加える．($K_Q \leftarrow K_Q \cup q$) (4) Que は事実が追加された知識に基づいて論証を構築し，その合理性が評価される．この (1)〜(4) のプロセスは，$\mathrm{R}(\Phi_t,\alpha)$ があらかじめ定められた閾値 Θ_R を上回るまで，もしくは，Que の問い合わせの回数があらかじめ定められた回数に達するまで繰り返される．なお，本研究では質問者であるシステムが収集の対象とするのは事実のみに限る．

2.2 合理的な論証構築のための情報探索対話の深層強化学習を用いた最適化

2.1 節で定義した情報探索対話をマルコフ決定過程を用いて定式化する．まず，時刻 t における $K_{Q,t}$ を Que の知識，Φ_t を $K_{Q,t}$ から L によって導出される論証の根拠，そして，q_t を Que の問い合わせによって Ans から収集された事実とする．

Que の行動 a_t は Que が Ans に問い合わせる質問に対応する．本研究では，Que は収集する事実に対応する問い合わせ候補のリストをあらかじめ保持しており，a_t はそのインデックスとして表現される．

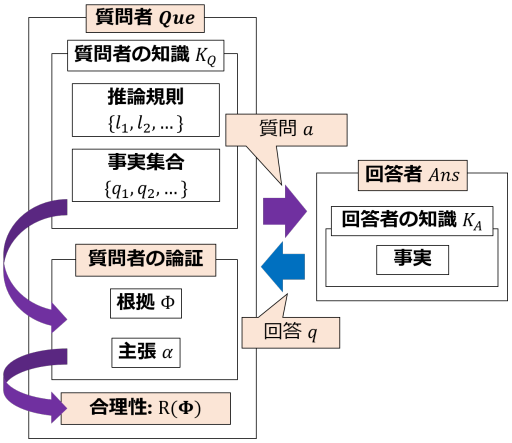

図 2: 合理的な論証構築ための情報探索対話の概要．

Que の内部状態 s_t は (1) 時刻 t までに Que がとった行動 (問い合わせ) の履歴，(2) 時刻 t までに Que が Ans から収集した事実からなる集合，(3) 時刻 t における Que が構築する論証の合理性から構成される．(1) の行動の履歴は $v_{h,t} \in \{0,1\}^{|A|}$ となるようなバイナリベクトルとして表現される．(2) は Que が収集した事実の集合 $\{q_1, q_2, ..., q_t\}$ をベクトル $v_{f,t}$ に変換して表現される．本研究では，$v_{f,t}$ の作成には，bag-of-facts 表現 (BoF) を利用した．BoF ベクトルの各次元は，bag-of-words と同様に，対応する事実が存在するかしないかのバイナリで表現され，$v_{f,t}$ はバイナリベクトルとなる．システムには，対話中に出現しうるすべての事実を登録した辞書があらかじめ与えられる．$v_{f,t}$ の次元数はこの辞書に登録された事実の数で，ベクトルの各要素は 0 で初期化される．事実が収集された場合，収集された事実の辞書のインデックスに応じたベクトルの要素を 1 とする．(3) は $[\mathrm{R}(\Phi_t,\alpha)]$ としてスカラー値を取る 1 次元ベクトルの形で表現される．すなわち，(1)〜(3) より，システムの内部状態は $s_t = [v_{h,t} \oplus v_{f,t} \oplus [\mathrm{R}(\Phi_t,\alpha)]]$ と表現される．

また，Que は下記のように報酬を得る．

$$r_t = \begin{cases} r_\mathrm{time} + r_\mathrm{goal} & (\Theta_\mathrm{R} \leq \mathrm{R}(\Phi_t,\alpha)) \\ r_\mathrm{time} & (otherwise) \end{cases},$$

3 評価実験

本研究では強化学習によって学習された最適対話戦略の評価を法廷審理ドメインとコンプライアンス違反検知ドメインの 2 つのドメインにおいて行った．

3.1 法廷審理ドメイン

本ドメインでは，最適対話戦略の学習と評価のために，およそ 20 種類の異なる事実から構成される K_A と 72 種

示す．また，K_Q に含まれる推論規則のうち α の導出に不要な推論規則は適宜 Φ から除外される．

類の推論規則から構成される K_Q を，Twelve Angry Men dataset [2] より構築し，最適対話戦略の学習と評価に使用した．なお，K_A については，学習用に 500 パターン，評価用に 50 パターンを用意した．

Twelve Angry Men dataset はテレビドラマ「12 人の怒れる男」 の作中で，12 人の陪審員が殺人事件の犯人として起訴された少年の無罪を結論付けるまでの議論を基に作成された論証データセットで，議論の過程で出現した発言のうち 80 ペアの発言について，支持もしくは反論の関係をアノテーションしたものである．本研究では，回答者 Que (システム) の主張である「少年は無実である」という事実も含めて，ここから 122 種類の事実と 72 種類の推論規則を抽出した [9]．

抽出された 72 種類の推論規則の集合は K_Q として，各対話の開始時に Que に与えられ，Que の質問候補は主張を除く 121 種類の事実となる．また，121 種類の事実からランダムに最大 20 種類の事実を選択して K_A としたものを 550 パターン作成し，そのうち 500 パターンを最適対話戦略の学習に，残りの 50 パターンを学習された最適対話戦略の評価に使用した．

3.2 コンプライアンス違反検知ドメイン

本ドメインにおける評価実験では，コンプライアンス違反検知データセット [14] から抽出された 20〜30 種類の事実で構成される 250 パターンの K_A と 106 種類の推論規則で構成される K_Q を学習と評価に使用した．

3.3 実験設定

評価実験においては，DDQN によって最適化された最適対話戦略をヒューリスティックによって作成した 3 つの比較手法と比較した．本節ではこれらのベースライン手法と実験設定について述べる．

ランダム戦略: Ans にランダムに質問を行う対話戦略．
深さ優先探索 (DFS) に基づく対話戦略: 推論規則の深さ優先探索結果に従って行動選択を行うヒューリスティックに基づく対話戦略 [5]．本対話戦略では，システムは事実グラフに対して深さ優先探索を行って行動選択を行う．ここで，事実グラフはシステムに K_Q として与えられた推論規則に出現する事実をノードとし，推論規則における導出関係をエッジとするような無向グラフである．例えば，K_Q 中に $p_1 \wedge p_2 \rightarrow q$ という推論規則が含まれるとき，事実グラフ上には p_1，p_2，q の 3 つのノードが含まれ，p_1 と q の間，p_2 と q の間にそれぞれエッジが張られる．図 3 に法廷審理ドメインにおける評価実験で使用した事実グラフの一部を示す．本研究では，システムの主張に当たる事実に該当するノードを探索の始点とした．なお，深さ優先探索中に各ノードにおいて次に探索対象とする隣接ノードはランダムに決定した．

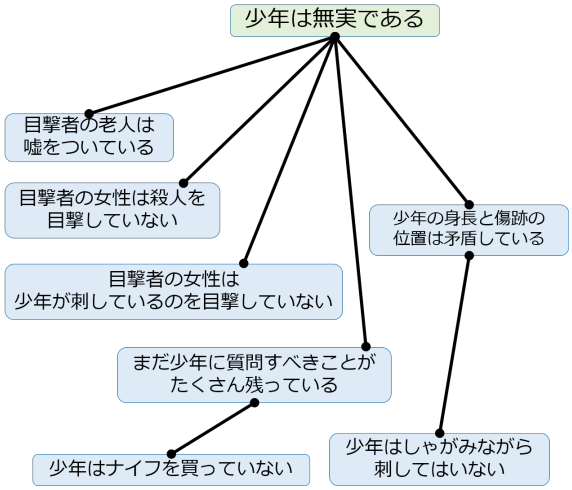

図 3: 法廷審理ドメインにおける評価実験で使用した事実グラフ (FG) の例の一部.

幅優先探索 (DFS) に基づく対話戦略: DFS に基づく対話戦略と同様に事実グラフの幅優先探索結果に従って質問を行うヒューリスティックに基づく対話戦略．なお，幅優先探索中に同一の深さのノードの展開順序はランダムに決定した．
DDQN: DDQN によって学習された最適対話戦略．Q-network は 2 層の非線形隠れ層と行動価値関数の値を出力する 1 層の線形出力層から構成され，$s_t (= [v_{h,t} \oplus v_{f,t} \oplus \mathrm{R}(\Phi_t, \alpha)])$ を入力として受け取る．2 層の隠れ層はそれぞれ hyperbolic tangent を活性化関数とする 50 のユニットから構成される．ε-greedy 探索は学習開始時は $\varepsilon = 0.1$ として開始し，2500 アクションを経て線形に $\varepsilon = 0.01$ まで減少させた．強化学習の期待報酬和の計算における忘却率 γ は 0.95 で固定し，学習は 5000 エピソード (対話) を通じて行った．

MDPs における報酬は，$r_{\mathrm{goal}} = 100$，$r_{\mathrm{time}} = -1$ に設定した．論証の合理性の閾値は，法廷審理ドメインにおいては $\Theta_{\mathrm{R}} = 0.5$，コンプライアンス違反検知ドメインにおいては $\Theta_{\mathrm{R}} = 0.7$ とし，学習時の各エピソードの質問の上限回数は $T_{\mathrm{limit}} = 10$ とした．

DDQN の学習はシステムを質問者 Que とし，シミュレータである回答者 Ans と対話させることによって行った．Ans は Que から事実に関する質問を受けたとき，該当する事実が対話開始時の初期化で与えられた K_A に含まれていればそれを返し，含まれていなければ「知らない」という答えを返す．なお，学習時の各対話の開始時にはそれぞれのドメインにおいて 500 パターンもしくは 200 パターンの K_A から一つを選んで Ans の知識 K_A を初期化し，評価時は，同様に各ドメインにおいて 50 パターンの K_A から一つを選んで K_A を初期化した．

4 結果

表 1 に DDQN によって学習された最適対話戦略とベースライン手法との比較結果を示す．評価は「テストスコア」,「成功数」,「平均質問回数」の 3 つの尺度で行った．「テストスコア」は各評価対話終了時の累積報酬の 50 回平均を取ったものであり，スコアが高いほど対話戦略の性能が良いことを示す.「成功数」は 50 回の評価対話のうち，質問回数の上限 (T_{limit}) に達するまでにシステムが構築する論証の合理性が閾値を上回った回数 (すなわち，合理的な論証の構築に成功した回数) で，回数が多いほど対話戦略の性能が良いことを示す.「平均質問回数」はシステムが対話終了までに質問を行った回数で，回数が少ないほど対話戦略の性能が良いことを示す．表の上段は法廷審理ドメインにおける結果を示し，下段はコンプライアンス違反検知ドメインにおける結果を示す．表 1 で示されるように，DDQN に基づく対話戦略がベースライン手法と比較して最も高い性能を持つことが両方のドメインにおいて確認された．

表 1: 表の各セルの数値は，それぞれの対話戦略に基づくシステムを 5 つずつ並列に構築，学習したときの，平均値であり，括弧内に 1 標準誤差を示している．

	テストスコア	成功数	平均質問回数
法廷審理ドメイン			
ランダム	-2.684	3.6	9.884
	(1.88)	(0.92)	(0.04)
深さ優先探索	7.176	8.4	9.624
	(8.46)	(4.14)	(0.18)
幅優先探索	0.66	5.2	9.74
	(3.30)	(1.61)	(0.08)
DDQN	**45.456**	**26.6**	**7.744**
	(5.74)	**(2.75)**	**(0.26)**
コンプライアンス違反検知ドメイン			
ランダム	-10	0	-10
	(0.00)	(0.00)	(0.00)
深さ優先探索	-10	0	-10
	(0.00)	(0.00)	(0.00)
幅優先探索	-10	0	-10
	(0.00)	(0.00)	(0.00)
DDQN	**65.304**	**35**	**4.696**
	(2.37)	**(1.13)**	**(0.12)**

5 まとめ

本研究では論証構築のための情報探索対話をマルコフ決定過程を用いて定式化し，深層強化学習の手法の一つである DDQN を用いて質問者 Que の最適対話戦略を学習した．また，DDQN によって学習された最適対話戦略を，ルールベースで作成した対話戦略，ランダムに行動選択を行う対話戦略と比較し，DDQN によって学習された最適対話戦略が最も高い性能を示すことを確認した．今後は，学習された最適対話戦略が，エピソードごとに変化する回答者の知識の変化に正しく対処できているかについて詳細な分析を行う必要がある．また，実際の対話において発生しうる回答者からの応答の音声認識誤りや，応答を論理式に変換する際の自然言語理解の誤りを考慮するため，情報探索対話の定式化を部分観測マルコフ決定過程に拡張する必要がある．さらに，情報探索対話のやりとりの対象を現在の原子論理式で表される事実だけでなく，推論規則にも拡張させることを検討している．

参考文献

[1] Elizabeth Black and Anthony Hunter. An inquiry dialogue system. *Autonomous Agents and Multi-Agent Systems*, Vol. 19, No. 2, pp. 173–209, 2009.

[2] Elena Cabrio and Serena Villata. Node: A benchmark of natural language arguments. In *Proc. COMMA*, pp. 449–450, 2014.

[3] Eugene Charniak and Solomon Eyal Shimony. *Probabilistic semantics for cost based abduction*. Brown University, Department of Computer Science, 1990.

[4] Phan Minh Dung, Robert A Kowalski, and Francesca Toni. Assumption-based argumentation., 2009.

[5] Xiuyi Fan and Francesca Toni. Agent strategies for aba-based information-seeking and inquiry dialogues. In *Proc. ECAI*, pp. 324–329, 2012.

[6] Xiuyi Fan and Francesca Toni. Mechanism design for argumentation-based information-seeking and inquiry. In *Proc. PRIMA*, pp. 519–527. Springer, 2015.

[7] Marga M Groothuis and Jörgen S Svensson. Expert system support and juridical quality. *Knowledge and Information Systems - KAIS*, 01 2000.

[8] Ryuichiro Higashinaka, Kazuki Sakai, Hiroaki Sugiyama, Hiromi Narimatsu, Tsunehiro Arimoto, Takaaki Fukutomi, Kiyoaki Matsui, Yusuke Ijima, Hiroaki Ito, Shoko Araki, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Yoshihiro Matsuo. Argumentative dialogue system based on argumentation structures. In *Proc. SEMDIAL*, pp. 146–147, 2017.

[9] Hisao Katsumi, Takuya Hiraoka, Koichiro Yoshino, Kazeto Yamamoto, Shota Motoura, Kunihiko Sadamasa, and Satoshi Nakamura. Optimization of information-seeking dialogue strategy for argumentation-based dialogue system. *CoRR*, Vol. abs/1811.10728, , 2018.

[10] Ekaterina Ovchinnikova, Niloofar Montazeri, Theodore Alexandrov, Jerry R Hobbs, Michael C McCord, and Rutu Mulkar-Mehta. Abductive reasoning with a large knowledge base for discourse processing. In *Computing Meaning*, pp. 107–127. Springer, 2014.

[11] Simon Parsons, Michael Wooldridge, and Leila Amgoud. An analysis of formal inter-agent dialogues. In *Proc. AAMAS*, pp. 394–401. ACM, 2002.

[12] Simon Parsons, Michael Wooldridge, and Leila Amgoud. On the outcomes of formal inter-agent dialogues. In *Proc. AAMAS*, pp. 616–623. ACM, 2003.

[13] Douglas Walton and Erik CW Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press, 1995.

[14] 勝見久央, 平岡拓也, 本浦庄太, 山本風人, 定政邦彦, 吉野幸一郎, 中村哲. 論証構築のための情報探索対話戦略の最適化. 言語処理学会年次大会発表論文集, pp. P639–642, 2018.