4S-05

複数回交渉のための多腕バンディットに基づくメタ戦略

川田 涼平 †　　藤田 桂英 ‡
† 東京農工大学大学院　工学府
‡ 東京農工大学大学院　工学研究院

1　はじめに

自動交渉は異なる効用をもつエージェント間の対立を解消し，お互いに協調するため手段として注目されている．特に，同じ条件で交渉を繰り返す複数回交渉は自動交渉の重要な研究領域の一つである．また，交渉エージェントのパフォーマンスは交渉相手の戦略や交渉ドメインなどに強く依存するため，状況に応じて適切に交渉戦略を変更することが有効である．

本論文では，複数回交渉において各交渉の開始時に適切な交渉戦略を選択するメタ戦略を提案する．提案手法は複数回交渉における交渉戦略の選択を多腕バンディット問題としてモデル化し，全交渉で獲得する個人効用の和が大きくなるように各交渉ごとに交渉戦略を選択する．また，実験により提案手法を使用するエージェントが他のエージェントよりも多くの効用を獲得することを示す．

2　複数回交渉と複数論点交渉

本論文では，複数論点交渉を繰り返す複数回交渉に注目する．複数回交渉では，同じ交渉相手と同じプロファイルで交渉を繰り返すため，エージェントは交渉中に得た情報を次回以降の交渉に活用することができる．複数論点交渉は $n > 1$ 個の論点で構成され，各論点 $i \in \{1, \cdots, n\}$ は $m_i > 1$ 個の選択肢をもつ．各論点 i について選択肢 $v_i \in \{1, \cdots, m_i\}$ をひとつずつ選んだ組を合意案候補という．また，論点や各論点の取りうる選択肢などの定義を交渉ドメインという．プロファイルはエージェントの選好を表し，各論点 i の重み $w_i \left(\sum_{i=1}^{n} w_i = 1.0\right)$ と選択肢 v_i の評価値 $eval_i(v_i)$ からなる．また，各プロファイルは時刻 $t \in [0, 1]$ に依存して効用を減少させる割引係数 δ と，合意失敗時に獲得する効用値 (留保価格)r をもつ．合意案候補 ω の効用値 $U(\omega)$ は，論点 i で選ばれた選択肢を $\omega_i \in \{1, \cdots, m_i\}$ とすると

$$U(\omega) = \delta^t \sum_{i=1}^{n} w_i \frac{eval_i(\omega_i)}{max_{v_i \in \{1, \cdots, m_i\}} eval_i(v_i)}$$

となる．

Meta-Strategy for Multi-Time Negotiation: A Multi-Armed Bandit Approach
†Faculty of Engineering, Tokyo University of Agriculture and Technology
‡Institute of Engineering, Tokyo University of Agriculture and Technology

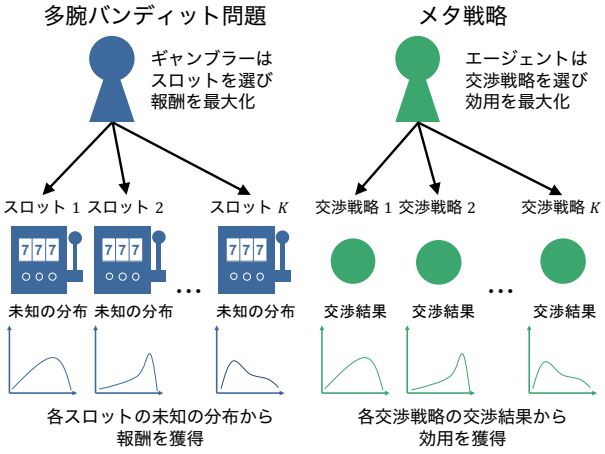

図 1: 多腕バンディットに基づくメタ戦略のモデル化

表 1: プロファイル生成のためのパラメータ

Name	$eval_i(v_i)$	w	δ	r
AAA	$Beta(2,4)$	$Dir(9,9,9,9,9)$	1.0	0.5
AAB	$Beta(2,4)$	$Dir(9,9,9,9,9)$	0.5	0.5
ABA	$Beta(2,4)$	$Dir(0.3,0.3,0.3,0.3,0.3)$	1.0	0.5
ABB	$Beta(2,4)$	$Dir(0.3,0.3,0.3,0.3,0.3)$	0.5	0.5
BAA	$Beta(4,2)$	$Dir(9,9,9,9,9)$	1.0	0.5
BAB	$Beta(4,2)$	$Dir(9,9,9,9,9)$	0.5	0.5
BBA	$Beta(4,2)$	$Dir(0.3,0.3,0.3,0.3,0.3)$	1.0	0.5
BBB	$Beta(4,2)$	$Dir(0.3,0.3,0.3,0.3,0.3)$	0.5	0.5

3　多腕バンディットアルゴリズム

多腕バンディットアルゴリズムは，未知の異なる期待値をもつ K 台のスロットマシンを T 回プレイする場合に報酬を最大化するためのアルゴリズムである．ϵ-greedy アルゴリズムは選択のたびに確率 ϵ で探索を行い，確率 $1 - \epsilon$ で活用を行う．探索ではランダムにスロットマシンを選び，活用ではそれまでの報酬の平均が最大のスロットマシンを選ぶ．UCB アルゴリズムは選択のたびに各スロットマシンの UCB スコアを計算し，最もスコアが高いものを選ぶ．スロットマシン s について，現時点での総試行回数を N，s の試行回数を N_s とすると，s の UCB スコア $UCB(s)$ は

$$UCB(s) = \hat{\mu}_s + c\sqrt{\frac{\ln N}{N_s}}$$

となる．ただし，$\hat{\mu}_s$ は s の過去の試行における報酬の平均で，c は探索の頻度を操作するパラメータである．

表 2: バンディットアルゴリズムごとの平均獲得個人効用

バンディット アルゴリズム	UCB $c = 0.01$	UCB $c = 0.05$	UCB $c = 0.1$	UCB $c = 0.5$	UCB $c = 1$	ϵ-greedy $\epsilon = 0$	ϵ-greedy $\epsilon = 0.1$	ϵ-greedy $\epsilon = 0.2$
Our Agent	**0.7788**	**0.7765**	**0.7725**	0.7382	0.7201	**0.7787**	**0.7706**	**0.7631**
Atlas3	0.7444	0.7443	0.7451	**0.7474**	**0.7477**	0.7441	0.7445	0.7451
CaduceusDC16	0.7138	0.7135	0.7134	0.7119	0.7104	0.7137	0.7130	0.7127
kawaii	0.7305	0.7304	0.7304	0.7274	0.7254	0.7312	0.7299	0.7293
ParsCat	0.6867	0.6872	0.6877	0.6875	0.6869	0.6869	0.6862	0.6862
Rubick	0.6658	0.6664	0.6652	0.6652	0.6648	0.6658	0.6654	0.6646
YXAgent	0.7132	0.7129	0.7121	0.7050	0.7006	0.7130	0.7110	0.7097

表 3: PRIANAC のエージェントとの交渉結果

エージェント	個人効用	社会的余剰
Our agent(UCB $c = 0.01$)	**0.7734**	1.4575
Agent33	0.6901	**1.4579**
AgentNP2018	0.7082	1.4362
Appaloosa	0.7067	1.3706
Ellen	0.6083	1.2223
TimeTraveler	0.7142	1.4573

4 メタ戦略

複数回交渉において各交渉の開始時に適切な交渉戦略を選択するメタ戦略を提案する．メタ戦略を，交渉戦略をスロットマシンとみなし，エージェントが交渉で獲得した個人効用を報酬とみなす多腕バンディット問題としてモデル化する (図 1)．提案手法は全交渉の総獲得個人効用が大きくなるように，各交渉で使用する交渉戦略を決定する．

提案したメタ戦略を自動交渉エージェントとして実装した．ϵ-greedy アルゴリズム ($\epsilon = 0, 0.1, 0.2$) と UCB アルゴリズム ($c = 0.01, 0.05, 0.1, 0.5, 1$) を交渉戦略の選択に使用した．交渉で使用する交渉戦略は，過去の自動交渉エージェント競技会 ANAC[1] の上位入賞エージェント (*Atlas3, CaduceusDC16, kawaii, ParsCat, Rubick, YXAgent*) のものを使用した．

5 評価実験

自動交渉プラットフォーム Genius[2] を使用して実験を行う．二者間交渉をエージェントとプロファイルの全組み合わせについて 100 回ずつ繰り返す．ドメインは 5 個の論点で構成され，各論点は 5 個の選択肢をもつ．プロファイルは評価値を *beta* 分布，論点の重みを *dirichlet* 分布で決定する．表 1 のパラメータごとに 2 個ずつ，計 16 プロファイルを使用する．交渉プロトコルは Alternating Offers Protocol[3] を使用し，交渉時間は 10 秒とする．

交渉戦略の選択アルゴリズムを変えて，提案エージェントとメタ戦略を使用しないエージェント (*Atlas3, CaduceusDC16, kawaii, ParsCat, Rubick, YXAgent*) 間で交渉を行った．各エージェントの獲得個人効用を表 2 に示す．UCB アルゴリズム ($c = 0.5, 1$) を使用する場合を除き，提案エージェントが他のエージェントよりも有意に多くの個人効用を獲得した．これはメタ戦略で状況に応じて適切な交渉戦略を選択することの有効性を示している．また，c や ϵ の値が大きくなるほど個人効用が減少していることから，複数回交渉の交渉戦略の選択で探索があまり重要でないことがわかる．これは同じ条件で交渉を繰り返した場合の効用値の分散が小さく，少ない探索で適切な交渉戦略を発見可能なためである．

UCB アルゴリズム ($c = 0.01$) を使用する提案エージェントと，自動交渉エージェント競技会 PRIANAC[4] に提出されたエージェントの交渉結果を表 3 に示す．実験結果から，提案手法が他の最新のエージェントとの交渉においても有効であることがわかる．

6 まとめ

本論文では，複数回交渉における交渉戦略の選択を多腕バンディット問題としてモデル化することで，状況に応じて効果的な交渉戦略を選択するメタ戦略を提案した．評価実験により，提案手法を使用するエージェントが他のエージェントよりも多くの個人効用を獲得可能なことを示した．

参考文献

[1] Reyhan Aydogan, Tim Baarslag, Katsuhide Fujita, Takayuki Ito, Dave de Jonge, Catholijn Jonker, and Johnathan Mell. The Eighth International Automated Negotiating Agents Competition(ANAC2017). http://web.tuat.ac.jp/~katfuji/ANAC2017/, 2017.

[2] Raz Lin, Sarit Kraus, Tim Baarslag, Dmytro Tykhonov, Koen Hindriks, and Catholijn M. Jonker. Genius: An integrated environment for supporting the design of generic automated negotiators. *Computational Intelligence*, 30(1):48–70, 2014.

[3] Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, pages 97–109, 1982.

[4] Takayuki Ito, Catholijn Jonker, Reyhan Aydogan, Tim Baarslag, Katsuhide Fujita, Satoshi Morinaga, Takashi Yoshida, and Yasser Mohammad. Pacific Rim International Automated Negotiation Agents Competition (PRIANAC). http://web.tuat.ac.jp/~katfuji/PRIANAC2018/, 2018.