

モンテカルロ木探索を用いた交渉の評価関数の提案

東京大学工学部電子情報学科 近山鶴岡研究室 4 年 伊藤 義章

2013 年 2 月 17 日

1 目的と背景

実世界では、競争的状況において自分の利益を最大化することを目指した多くの交渉が行なわれている。交渉に関する研究の多くは、交渉の定性的な研究であり、交渉の定量的な側面に着目した研究は未だ発展途上である。その理由のひとつは、交渉によって得られる可能性のある利益を数値化し、どのような交渉を提示するかを指標となる評価関数を用意することの難しさにある。本研究では、交渉において重要となる要素を定量的な観点から明らかにする事を目的とし、利益の数値化の方法として、行動の先読みおよび評価関数を用いる事を提案する。具体的には、先読みによる利益計算に、近年成功をおさめているモンテカルロ法を用いることを提案し、また、利益計算の指標となる様々な評価関数を提示する。

2 関連研究

2.1 UCT アルゴリズム (UCB applied to Trees)

状態の評価が難しい問題において有効とされる手法としてモンテカルロ法がある。ある局面から終局までランダムに行われる試行（プレイアウトと呼ばれる）に基づき、候補手を統計的に評価する。しかし、時間が有限の場合、明らかに有望でない手に対して、有望な手と同様の試行回数を行うべきではない。UCT アルゴリズムは、有望な手に UCB 値 (Upper Confidence Bound、以下 UCB)[2] を用いてバイアスをつけるモンテカルロ木探索である。UCT アルゴリズムは、「子ノードの選択」「子ノードの展開」「プレイアウト」「結果の更新」を終了条件（時間や探索回数など）となるまで繰り返し、平

均報酬の最も高い子ノードを選択する。

2.2 交渉のアルゴリズム

交渉決裂により、実際に得られるはずだった利益を得られず、機会損失を招いてしまう。交渉案を選択する評価関数の機会損失を図る指標として譲歩率がある [1]。ANAC にて優秀とされる結果を残した人工知能の IAMhaggler2011[1] は、一定の譲歩を行なっていることから、自分の利益のみを考える評価関数が優位な戦略でないと考えられる。

3 提案手法

3.1 利益の見積もり

UCT アルゴリズムの先読みによって得た勝率を、利益計算に用いる事にする。プレイアウトを行なう事で、各プレイヤーの勝率を計算することができる。交渉による損得は勝率に直結することから、UCT アルゴリズムの先読みによって予め得た勝率により利益計算を行なうことができると考えられる。

3.2 評価関数の作成

本研究では、前節で得られた利益計算の結果を用いて、交渉の成功率を考慮した評価関数を用いて各交渉の期待値を求める。ただし、交渉の成功率は「相手が得られる利益」に影響を受けるものだと仮定する。良い交渉案を選ぶ評価関数として、以下の 5 つの評価方法を提案する。まず、自分の利益のみしか考えない「自己中心的交渉」を行なうプレイヤーをベースラインとして扱う。次に、相手の利益を優先にした「利益優先交渉」と交渉の成功率を優先した「受諾優先交渉」を提示する。また、お互いの利益をバランスよく配分する手法として、「和交渉」「積交渉」を提示する。

4 評価

4.1 UCT の利益計算

ルールベースプレイヤーは全受諾プレイヤーに対して 42.1% とベースラインであるランダムプレイヤーに 41.25% と有意な実力差を示したことから、自分にとって有利となる利益計算を行い、交渉案を提示していることが分かる。一方で、UCT プレイヤーは受諾プレイヤーに対して 23.55%、ランダムプレイヤーに 25.0% と有意な実力差を得られなかった。ルールベースと交渉選択の一致率を検証した結果、要因として「プレイアウト回数が不十分であったこと」「相手の勝率が高くなった場合での精度悪化」が考えられる。

4.2 評価関数

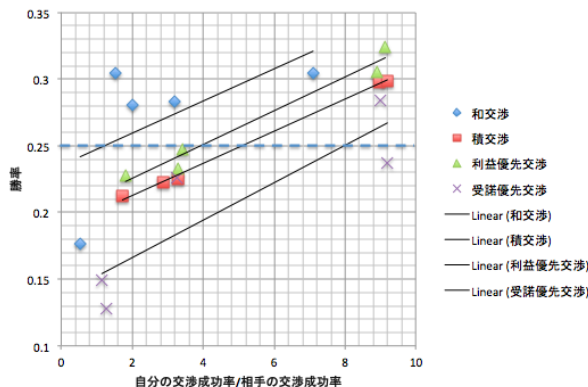


図 1. 評価関数の勝率と交渉成功率

図より、全ての評価関数において (自分の交渉成功率/相手の交渉率) の割合が高まれば高まるほど良い勝率を残すことができると考えられる。このことから、交渉における「提案成功率」と「受諾率」が関係していると考えられる。また、和交渉ではばらついたものの、利益優先交渉や、積交渉のように近似直線を描くと一定の直線上にのっていることがわかる。また、図よりこれら 4 つの直線の傾きは一定であるが、評価関数によってその近似直線が上下することがわかる。評価関数によって近似直線が上下する原因として、「自分にとっての利益」と「相

手に与える損益」のバランスによるものと考えられる。バランス型の交渉である和交渉では、相手に与える損益を加えたことが高い勝率に結びついたと考えられる。また、受諾優先交渉、利益優先交渉と積交渉では相手に与える損益がないため、自分の得られる利益が大きい順に高い結果を残すことができたと考えられる。利益優先や受諾優先ではなくバランスよく利益配分を行なう和交渉がもっともよい結果を残したことから、「自分にとっての利益」と「相手に与える損益」をバランスよく配分する評価関数が最もよい評価関数だと考えられる。

交渉の評価関数において「提案成功率」「受諾率」「自分の利益」「相手の損益」が重要な要素であると明らかにした。「相手の損益」を考慮するのは、自分では悪い交渉案だと思っている交渉案においても、相手との評価基準が自分よりもずれていることを利用して、悪い交渉案を混ぜることで勝率を高めることができると考えられるためである。

5 おわりに

5.1 まとめ

先読みによる利益計算に、近年成功をおさめているモンテカルロ法を用いることを提案し、また、利益計算の指標となる様々な評価関数を提示した。不確定不完全情報ゲームである「カタンの開拓者」をテストベッドとして実験を行なった結果、利益計算におけるモンテカルロ法の有効性を示す事はできなかった。しかし、交渉の評価関数を作成する上で、「提案成功率」「受諾率」「自分の利益」のみならず、「相手の損益」を考慮した評価関数を作成する必要があることが明らかになった。

5.2 今後の課題

利益見積もりを改善するため、UCT アルゴリズムのプレイアウト回数を向上することと、極端な局面でのバイアスを導入することがある。プレイアウト時間は線形に増加してしまうことから、一定の時間でシミュレーションを打ち切り、評価関数を用いて盤面を評価し、その時点で評価値が最も高いプレイヤーを暫定の勝利者とするといった工夫が必要となる。

参考文献

- [1] Tim Baarslag, Katsuhide Fujita, Enrico H. Gerding, Koen Hindriks, Takayuki Ito, Nicholas R. Jennings, Catholijn Jonker, Sarit Krause, Raz Lin, Valentin Robu, and Colin R. Williams. Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. *Artificial Intelligence*, 2012.
- [2] Levente Kocsis and Csaba Szepesvari. Bandit based Monte-Carlo planning. In *Machine Learning: ECML 2006*, pp. 282–293. Springer, 2006.