

2000年度

人工知能学会全国大会(第14回)論文集

Proceedings of the 14th Annual Conference of JSAI, 2000

発行日 2000年7月3日

発行所 社団法人 **人工知能学会**

東京都新宿区津久戸町4-7 〒162-0821

OSビル402 ☎(03)5261-3401

不確実性管理による競合戦略に基づくエージェント間競合方式

A Multi-Agent Competition Method Based on Uncertainty Management

園城 浩行 ベルーズ H. ファー
Hiroyuki ONJO Behrouz H. FAR

埼玉大学大学院 理工学研究科 情報システム工学専攻
Department of Information and Computer Science, Saitama University

Abstract: In this paper we propose a Multi-Agent competition method based on uncertainty management. There are many projects focusing on multi-agent system development and various technologies have been introduced. However, most of the works related to the software agents are based on the assumption that the other agent is trustworthy. In the growing internet environment, it may not be true. But the real method that allows agents to accomplish useful competitive conclusion is still missing. The proposed method based on the combination of game theory and belief network allows agents to decide the optimal strategy considering uncertainty about opponents' preferences and to reduce the uncertainty by predicting the movement of the competitor. We test this method via implementing electronic commerce agents.

1 まえがき

近年、マルチエージェントシステムに関する研究が活発に行われており、様々な技術が導入されている。従来の手法は全て他のエージェントが「信頼できる」という前提のもと、協力的にタスクを実行するものが多いが、現実世界においてはそのような環境ばかりではなく他のエージェントが信頼できない環境において、彼等と競合しなければならないケースが多数存在する。電子商取引 (EC) はその典型例である。しかし、エージェント間競合方式を提案している研究は数少ないのが現状である。

一般に、競合環境においてはエージェントが得られる効用の大きさは競合相手のエージェントの出力に左右されるため、自己の都合だけでなく相手のエージェントの動向を予測し、これらを統合的かつ戦略的に分析する必要があるが、競合相手は知識を隠してくるため情報は不確実になる。即ち、エージェントに不確実環境に対して自律的に対応させることが課題となる。

そこで、本論文では EC を実験対象に、不確実性管理に基づくエージェント間競合方式を提案する。本システムのエージェントは相手の動向に関する不確実性の度合いを自律的に判断し、そこから生じるリスクの大きさに応じた最適戦略を採択する。さらに、環境の変化をエージェント自身が長期的に予測することで競合戦略の信頼性を向上させる。

本論文の構成を示す。まず第 2 章にて本論文の応用例である EC 用マルチエージェントシステムについて述べる。第 3 章で実際の競合環境をゲーム理論の展開形でモデル化する。第 4 章にて不確実性を伴った環境での競合戦略決定方式を提案し、第 5 章では、エージェント自身が競合を経験するにつれて不確実性を緩和する手法につ

いて言及する。最後に第 6 章で本論文をまとめる。

2 電子商取引用マルチエージェントシステム

本論文では、図 1 に示す EC 用マルチエージェントシステム [1] を実験対象に研究を進める。本システムでは、*Cus-*

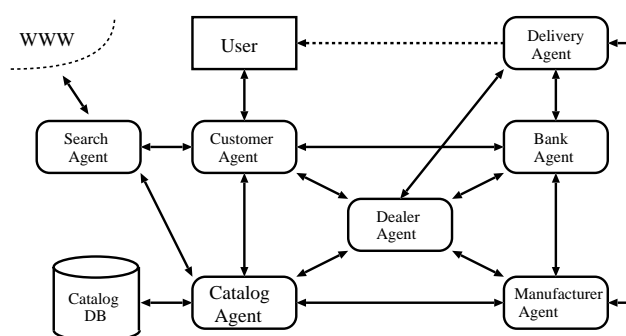


Figure 1: システムモデル

tomer, Search, Catalog, Dealer, Manufacturer, Bank, Delivery の 7 つのエージェントによって構成されており、それぞれが自律的に他のエージェントと協調や競合をしながら商取引の専門的なタスクを実行する。

本システムの流れでは、ユーザが商品を購入する際に複数の Dealer Agent と交渉をしたときに Dealer Agent 間で競合がおこる。この場合まずすべての Dealer Agent は等しく Customer Agent からユーザの選好に関する情報を得る。その後、各 Dealer Agent はそれぞれユーザ

連絡先: 園城 浩行 〒338-8570 埼玉県 浦和市 下大久保 255
埼玉大学 工学部 情報システム工学科 Far 研究室, Tel. 048-858-3489.
E-mail. onjo@cit.ics.saitama-u.ac.jp
WWW. <http://www.cit.ics.saitama-u.ac.jp/~onjo/>

の選好に応じることのできる代替案を戦略の候補としてピックアップする。最後に、ライバルエージェントの動向を予測し、戦略候補の中から効用の期待値が最高になる戦略を販売戦略として選ぶことになる (図 2)。

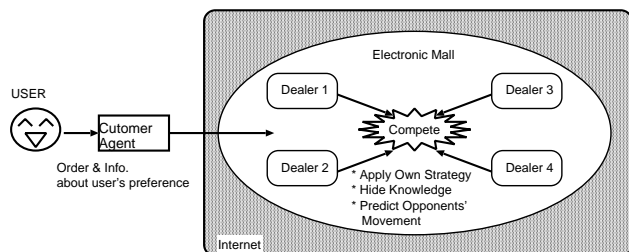


Figure 2: ECにおける小売業の競合

3 Dealer Agent 間競合環境のモデル化

相手 Dealer Agent がもつ効用関数が分からない場合は、相手のとる戦略を確実に予測できない。この場合、選好が不明なエージェントは「自然」と考え、その代替案を「自然の状態」と見なすが一般的である。本論文では Dealer Agent の集合を効用関数の分かっている Dealer Agent の集合とそうでない Dealer Agent の集合にわけ、前者に関してはゲーム理論的に、後者に関してはそれを一括して不確実な自然として取り扱う。本論文では、競合環境を情報不完備ゲームと考え、エージェント間競合環境 Γ を以下の 6 項組にて規定する。

$$\Gamma = (N, \Omega, X, \{A_i, \succeq_i, P_i\}_{i \in N}) \quad (1)$$

ここに、各要素は、 N : Dealer Agent の集合、 Ω : 自然の状態の集合、 X : 結果の集合 ($A_1 \times \dots \times A_n \times \Omega \rightarrow X$)、 A_i : Dealer Agent i の行動 (戦略) の集合、 \succeq_i : Dealer Agent i の X 上の選好関係 (これは基数推移性をもつ選好順序で基数効用関数 $u(x)$ が存在する)、 P_i : 集合 Ω の分割で Dealer Agent i の情報分割である。これをゲーム理論の展開形表現を用いて表現した例が図 3 である。この図では、Dealer Agent 間 1 対 1 の競合を想定して

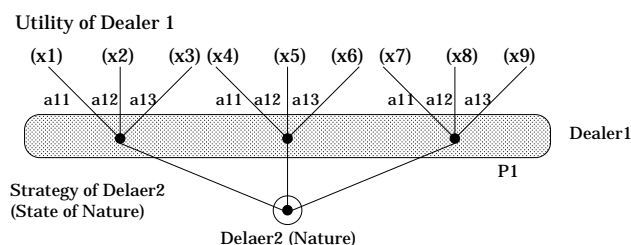


Figure 3: Dealer Agent 間競合モデルの例

おり、Dealer Agent 1 は Dealer Agent 2 の効用関数を知らないため、Dealer Agent 2 の戦略を予測できないことから、Dealer Agent 2 を自然と見なしている。 P_1 は Dealer Agent 1 の情報分割であり、相手の出方が分からない状態で Dealer Agent 1 は 3 つの戦略のうちどれかを決める。このとき、Dealer Agent 1 は自然の状態に関

して確率法則に反しない値を信念 (Belief) として当てはめ、その期待効用を最大化すべく戦略を決める。これがゲーム理論でいうベイズ完全均衡である。ここでの問題は Dealer Agent がどのように自律的に信念を見積もるかである。本論文では、不確実性に対する信念をエージェントが自律的に決定する方式を提案する。

4 不確実環境におけるリスク管理手法

4.1 情報の不確実性と意思決定

自然の状態の生起に関する信念の確実性の度合いが意思決定に影響を及ぼす。一般に、情報の不確実性は以下の 3 つにレベルに分けて考えられる [4]。

- Level-1: Decision Making under Certainty
自然の状態を 100% 特定できる場合。この場合、その代替案をとれば効用が最大化されるか特定できるので問題はない。
- Level-2: Decision Making under Risk
自然の状態を完全に特定はできないがその生起確率 p は分かっている場合。この場合はゲーム理論の期待効用最大化原理に従う。
- Level-3: Decision Making under Uncertainty
自然の状態を、その確率すら不確実な場合。この場合、人間は、楽観度の大きさに基づいて行動することが心理学者の研究により知られているため、本論文ではエージェントの楽観度を定量化し、確率の代わりに信念として割り当てる手法を提案する。

4.2 エージェントの楽観度の定量化

本論文では、OWA オペレータ [3] を応用して楽観度の定量化を行う。OWA オペレータは、自然の状態の集合を関数 F として平均化する n 次元オペレータである。これは確率法則に反しない n 次元重みベクトル、

$$\mathbf{W} = [w_1, w_2, \dots, w_n] \quad (2)$$

で、ある値 a_1, \dots, a_n に対して、以下の式を満たす。

$$F(a_1, \dots, a_n) = \sum_j w_j b_j \quad (3)$$

(ここに b_j は集合 a_1, \dots, a_n の中で j 番目に大きな要素)

OWA オペレータの重みベクトル \mathbf{W} は様々な解釈が可能だが、これを各自然の状態の疑似発生確率分布、つまり j 番目の最適解が起こる確率と解釈できることが数学的に分かっている [3] (この場合、疑似確率は、各 ω ではなく、結果の選好順序に割り当てられる)。ここで、重みベクトル \mathbf{W} をどのように決めるのかが問題となる。方法の 1 つは、Dealer Agent のユーザが適当に決める方法であるが、それではエージェントが自律的に意思決定を行っていることにならない。そこで本論文では Dealer Agent のユーザの楽観度を利用して \mathbf{W} を自動的に求める。Dealer Agent のユーザの楽観度を α とすると、

$$\alpha = \sum_j w_j (n - j) / (n - 1) \quad (4)$$

にて定義されることが分かっている[3]。そこで、エントロピー最大化原理を用いて以下の線形計画方程式に求めることで W が求められる。

$$\max - \sum_j w_j \log_2 w_j \quad (5)$$

Subject to:

$$\begin{aligned} \alpha &= \sum_j w_j (n - j) / (n - 1) \\ \sum_j w_j &= 1 \\ w_j &\geq 0 \text{ for } j = 1, \dots, n \end{aligned}$$

本手法で Dealer Agent のユーザが適当に楽観度を定めるだけで楽観度を定量化し、それを疑似発生確率として利用することで *Level-3* の問題を *Level-2* と同様に扱える。本論文では今後、これらを同じ問題として扱う。

4.3 リスク管理を可能にした効用関数

各 Dealer Agent の効用を決定するには効用関数が必要になる。Dealer Agent が商品売買によって得られる純利益の大きさ $u(x)$ は、「(商品の売り値-仕入れ値) × 販売量-戦略採択に必要な費用」となる。ここに自然の状態の生起確率 p に選好関係を導入し純利益の期待値、

$$\begin{aligned} p \succeq q &\iff u'(p) \geq u'(q) \\ u'(p) &= E(x, p) = \sum_{x \in X} u(x)p(x) \end{aligned} \quad (6)$$

にて一応、効用関数を得ることができる。しかし、この効用関数は、この理論に従えば Dealer Agent は期待値が同じ 2 つの戦略は無差別になる。しかし、*Level-2, 3* のような決定問題の場合は一般に戦略採択には費用がかかるため、期待値が等しくても分散が大きい時はそれだけリスクが大きくなり選好に影響を与える。例えば、競馬において期待値が同じでも重馬(ハイリスク・ハイリターン)を好む人、本命馬(ローリスク・ローリターン)を好む人に別れるのはこの典型例である。経済学の領域ではこのようなリスクの態度を様々な方法で表現しているが、人工のエージェントが自律的に考えるにはこれらの手法は複雑すぎるため、敷居が高すぎる。そこで、本論文ではリスクの大きさは分散の大きさによることに目をつけ、式(6)を以下のように修正する。いま、分散の大きさを $V(x, p)$ として、

$$\begin{aligned} p \succeq q &\iff u(p) \geq u(q) \\ u(p) &= E(x, p) - \eta V(x, p) \end{aligned} \quad (7)$$

ここに、 η は整数であり、リスク志向か回避かの重み係数である。リスク回避であれば η は正に、リスク志向であれば負に、リスク中立であれば 0 になる。さらに、戦略の集合から複数選択できる場合には、これらを混合することでリスク管理がはかれる。例えば $a, b, 2$ つの戦略を混合する場合、確率 p に従う a に対する確率変数を \tilde{x} 、確率 q に従う b に対しては \tilde{y} 、期待値をそれぞれ $E(\tilde{x})$ 、 $E(\tilde{y})$ 、分散を $V(\tilde{x})$ 、 $V(\tilde{y})$ とする。また、 \tilde{x} と \tilde{y} の共分散を $\text{Cov}(\tilde{x}, \tilde{y})$ とすれば、利用可能な資源のうち α を a に $(1 - \alpha)$ を b に配分した場合、それに対応する結果 \tilde{z} の期待値 $E(\tilde{z})$ 、分散 $V(\tilde{z})$ は以下のように計算される。

$$E(\tilde{z}) = \alpha E(\tilde{x}) + (1 - \alpha) E(\tilde{y}) \quad (8)$$

$$V(\tilde{z}) = \alpha^2 V(\tilde{x}) + (1 - \alpha)^2 V(\tilde{y}) + 2\alpha(1 - \alpha) \text{Cov}(\tilde{x}, \tilde{y}) \quad (9)$$

ここで、これらを式(7)に代入し、これを最大化する資源の配分 α^* を計算した結果、

$$\alpha^* = \frac{E(\tilde{x}) - E(\tilde{y}) + 2\eta(V(\tilde{y}) - V(\tilde{x}, \tilde{y}))}{2\eta(V(\tilde{x}) + V(\tilde{y}) - 2\text{Cov}(\tilde{x}, \tilde{y}))} \quad (10)$$

となり、戦略 a に資源を α^* 、戦略 b に資源を $(1 - \alpha^*)$ だけ割り当てたものを販売戦略として採択すれば、各 Dealer Agent のリスクに対する態度を反映した競合戦略を決定することができる。このような分散の異なる戦略を混合してリスク管理をする手法は経済学ではポートフォリオと呼ばれる手法である[5]が、本論文のようにカスタマイズすることで、ユーザがリスク回避度 η を適当に決めるだけでエージェントが自律的にリスク管理をすることができる。

5 不確実性の緩和

5.1 不確実環境における意思決定サイクル

前節の手法によって不確実環境においてもある程度満足のいく意思決定が行えるが、不確実環境ではどうしても戦略の信頼性が下がる。そのため、競合相手の動向を予測しながら、不確実性を *Level-3* → *Level-2* → *Level-1* と緩和することが必要である。ここで、本論文における「不確実性の緩和」とは Dealer Agent が環境の状況変化を知覚し、それを証拠として不確実性(自然の状態)に対する信念を変更することと等価である。

競合環境においては、競合相手は自己の戦略を悟られぬよう知識を隠してくるためエージェントの環境に対する知覚は不確実であり、かつ部分的にしかアクセス可能でない。よって、非常に限られた、競合相手が出た戦略に関する情報のみから競合環境の変化の推移を推論する必要がある。一般に、不確実環境かつ部分的アクセス可能な環境におけるエージェントの信念の更新サイクルは以下ようになる[2]。

- *Step-1*: 現在の知覚と一つ前の動作を含む利用可能なデータに基づいて現在の自然の状態の確率分布を計算する。
- *Step-2*: 動作記述と現在の自然の状態の確率分布とから各動作を行った時の確率分布を計算する。
- *Step-3*: 次状態の確率分布と効用情報とから各動作についての期待効用を式(7)にて計算し、それを最大化する戦略を選ぶ。

ここで興味の対象となるのは時刻 t までのすべての証拠を与えられた時の時刻 t における信念、即ち時刻 t までの知覚履歴を $E^1 \dots E^t$ 、戦略を $A^1 \dots A^t$ とすれば、

$$\text{Bel}(\Omega^t) = \text{P}(\Omega^t | E^1 \dots E^t, A^1 \dots A^t) \quad (11)$$

であり、この確率分布で不確実環境の推移をモデル化できる。そこで、本論文では、競合環境の推移に関する知識を動的信念ネットワーク(DBN)で表現し、その知識をゲーム理論の概念で自己分析することで、競合の特徴的分析・予測を行う(図4)。

5.2 DBN による信念の更新モデル

競合環境において、Dealer Agent が自然の状態に関する信念を更新する再には、

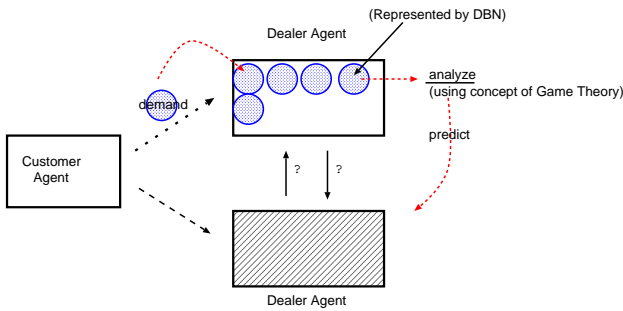


Figure 4: 自己分析と予測

1. 部分的にアクセス可能な環境からの利用可能な情報の知覚
2. 時間による環境の状態の変化の予測

を行う必要がある。ここで、本論文では、図 5 のような DBN を構築した。

Dealer Agent 間競争で直接知覚可能なのは点線で示した競争の結果と競争相手が出した戦略についてのシグナルである。ただし、競争相手がつねに完全なシグナルを出すとは限らないので (戦略の一部のみシグナルするなど)、まず、与えられた世界の状態での知覚の確率 $P(E^t|\Omega^t)$ を経験を基に計算する。そして、知覚した競争の結果と競争相手のシグナルを統合して自己分析を行う。つまり、競争の結果から相手の得た純利益を、シグナルから相手の戦略を予測し効用関数を予測することになる。もし、シグナルが観測できなかった場合は競争の結果のみを用いて信念を更新することにする。競争相手の効用関数を予測することで、次状態の自然の状態の確率分布 $P(\Omega^t|\Omega^{t-1}, A^{t-1})$ を計算し信念を更新する。

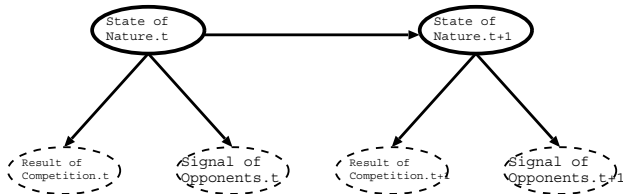


Figure 5: 信念の更新モデル

信念の更新サイクルは、図 6 に示されるように以下の手順で行われる。

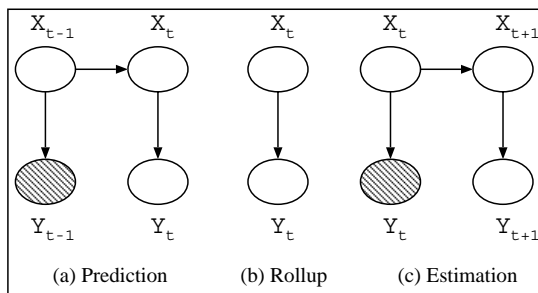


Figure 6: 信念の更新サイクル

1. 予測 (Prediction): まず前状態についての知識と動作が自然の状態に対してどう影響するかが与えられて、期待する状態の確率分布を予測する。これを \widehat{Bel} とし、時刻 $t-1$ でのある状態から t で望む自然の状態になる確率をすべての場合について合計することで計算され、次式を得る。

$$\widehat{Bel}(\Omega^t) = \sum_{\omega^{t-1}} P(\Omega^t|\Omega^{t-1})Bel(\Omega^{t-1}) \quad (12)$$

2. 畳込み (Roll-up): 分割 $t-1$ を削除し、時刻の状態変数に対する事前確率表を追加する。
3. 推定 (Estimation): 新しい知覚 E^t を使って、信念を更新する。状態変数と知覚はともに同一時刻のものを指しているため、ベイズの定理による更新をそのまま使える。よって、 $\widehat{Bel}(\Omega^t)$ を事前確率とすれば、次式を得る。ここに、 β は正規化のための定数である。

$$Bel(\Omega^t) = \beta P(E^t|\Omega^t)\widehat{Bel}(\Omega^t) \quad (13)$$

このような手順に従って信念を更新することにより、各 Dealer Agent は不確実性に対応しながら競争を進めることが可能である。つまり、Dealer Agent ごとの不確実環境における競争戦略を実現できる。

尚、上記の信念更新モデルは競争環境に対して一般的なものであり、具体的なシステムへの応用にあたっては実効レベルでのカスタマイズが必要である。具体的な EC への応用は今後の課題である。

6 むすび

本論文にて、不確実性管理に基づくエージェント間競争方式について論じた。本論文の手法により、短期的には不確実環境においてある程度満足のいく決定が行え、長期的にはエージェントが競争を経験することで不確実性を緩和できることが期待できる。今後は、本論文で提案した手法を EC に拡張し、模擬トレーディング実験を行う事で、競争の経験がどの程度不確実性の緩和に寄与するか分析する予定である。

References

- [1] B.H.Far et.al. : "An Integrated Reasoning and Learning Environment for WWW Based Software Agents for Electronic Commerce", *IEICE Trans. Info. & Sys.*, Vol. E81-E, No. 12, pp.1374-1386, (1998).
- [2] S.Russel, P. Norvig: "Artificial Intelligence: A Modern Approach, Prentice-Hall, (1995).
- [3] R.R.Yager: "On ordered weighted averaging aggregation operators in multi-criteria decision making", *IEEE Trans. Sys., Man, & Cybern.*, pp. 183-190, (1988).
- [4] 市川 淳信: 『意思決定論』, pp. 52-80, 共立出版, (1983).
- [5] 梶井 厚志, 松井 彰彦: 『ミクロ経済学 戦略的アプローチ』, pp. 200-212, 日本評論社, (2000).