確率モデルの学習によるエージェント間の合意形成

本村陽一 (Yoichi Motomura)

電子技術総合研究所 (Electrotechnical Laboratory)

Summary.
In dynamically changing and unpredictable environment, coming to an agreement for a multi-agent system without explicit communication is an important issue. We will discuss an advantage of probabilistic framework for coming to an agreement. Especially, complex agents that behave according to particular situations, can be modeled by Bayesian (belief) network. In this study, a process to achieve an agreement is formulated as learning for the probabilistic models. Learning algorithm can be derived from probabilistic hill-climb methods. Agents can evaluate a degree of consistency as frequency in learning process. Finally, learning agents find adequate parameters to go along with other agents, then they reach the agreement in the environment. We will show results of learning sigmoidal belief network in several cases. The convenient techniques based on probabilistic framework are also shown.

1 はじめに

動的に変化し, 予測が難しい環境における最適な意志決定の問題の例として, 異種の複数エージェントが構成する社会におけるエージェント間の合意形成がある. 未知の環境において柔軟な協調動作を実現するためにはアクション, 目的とするゴール, サブゴール, 動作対象など様々な選択についての一致, すなわち合意を適応的に形成する仕組みが不可欠である. 特に明示的な通信を用いることができない場合や, 未知のエージェントとの合意を形成する場合, 環境に対して不完全な観測しか得られない場合には状況や場の必然から自然に選択が一致する点, 焦点 (Focal point) [1] が複雑で動的な環境における知的システムを考える上でも本質的に重要な役割を果たし得ることが指摘されている [2, 3, 4]. これの具体的な応用としても未知の動的な要因が多数存在する環境や敵, 味方が混在する環境でのゴール達成, 人との対話的システムにおいてユーザーの意図を予測する内部モデルの構成などが考えられる. Focal point(焦点) は自律的に注視点を選択するための基準として, 問題の設定・その場の状況・文脈などにより顕著に他と区別されうる点と定義されている [1]. 特に人間にとっての Focal point は社会的な場の状況, 構図から規定されてくるものとして様々な社会的局面についての調査研究を行った例がある [1, 5]. マルチエージェントの枠組では, 対象を一つ選び出すタスクについて属性値の極端さや珍しさなどの直観的指標を用いて選択するアルゴリズムの提案 [7] や, 複数台の自律移動ロボットシミュレーターを用いた実験評価 [11] などがある.

しかし, 人間は現実社会における生活を通じて Focal point を決定する仕組みを既に獲得しているのに対して, 人工エージェントの場合には Focal point を決定する固定的なアルゴリズムを環境との関わりを抜きにしたまま事前に規定することは困難である. そのためエージェントが置かれている環境内での観測を通じて Focal point を決定する仕組みを自律的に獲得する枠組が重要である. ここではフォーカルポイントを特定の環境の中で探索していく過程を環境からのフィードバックによるエージェントの内部モデルの学習として獲得する方法を考える. 特に学習過程での現状態を評価するために選択の取り方について確率的な「幅」を持つ必要性について述べる. またより一般的な状況や文脈に依存した一致点を実現するためにエージェントの出力を条件付き確率分布として表現する. 条件付き確率を表現する確率モデルとしてはボルツマンマシンやベイジアンネットワーク (Belief net) があり, これの学習によって状況に依存して動作するエージェント間の合意形成を実現する。最後に確率モデルを用いることで適用可能となる確率・統計的手法について述べる。

2 合意形成

以下では Focal point を決定する仕組みを理論的に取扱うための定式化を行なう. 今, エージェント A とエージェント B がそれぞれアクション a^A,a^B を選択した時, この選択が一致する $(a^A = a^B)$ ことを合意の形成とする.

完全な情報が得られる場合, 各エージェントが全く同じ入力を受け, これに対し全く同じ決定的 (非確率的) アルゴリズムを用いて選択を行えば合意が成立することは明らかである. しかし完全な観測が期待できず各エージェントが得る入力が異なる場合や, そもそも各エージェントがヘテロジニアスであり全く同じ決定アルゴリズムを用いていない場合には未知の決定的なアルゴリズムを推定する必要がある. だが, 解空間を全探索できないような場合には接近法により推定を行なうための評価値を得ることが容易ではない.

一方, 確率的なアルゴリズムの場合にはエージェントのアクションの取り方に確率的な「幅」があり, 一定時間内の試行の間の選択頻度から選択が一致する期待値を得ることができ, 学習のために必要な現在の状態に関する評価が可能である. 未知のエージェントが存在する環境では他のエージェントが用いているアルゴリズムを推定することが必要であるがこれを近似する場合にも確率的な接近をはかる方法が有効である.

確率的に選択が行われる場合にエージェント k が各アクションを選択する確率を p_i^k とする. 全ての選択についての確率の集合

$$P^k(A) = \{p_1^k, p_2^k, \cdots, p_n^k\}$$

はアクションに関する選択を決定する全パラメータである. ここでエージェント $1, 2$ の選択が一致する確率は次式で表される.

$$P^1(A) \cdot P^2(A) = \sum_{i=1}^{n} p_i^1 \cdot p_i^2 \tag{1}$$

より一般的にエージェントの数が K の場合の全体の選択が一致する確率は次式により評価できる.

$$E \equiv \prod_{k=1}^{K} P^k(A) = \sum_{i=1}^{n} \prod_{k=1}^{K} p_i^k \tag{2}$$

3 学習による合意形成

以上の定式化のもとで合意形成のためには式 (2) を大きくするように各エージェントがアクションを選択するようにすればよい. 式 (2) を最大化するために最終的にはどれかのアクションを最大の確率値 $= 1.0$ で選択する必要があるが, 同様に学習する他のエージェントの集団の中でいかにして共通のアクションを選択するように仕向けるかということが重要なポイントである.

複数のエージェントが存在する環境において各エージェントが選択を行ない, 全てのエージェントの選択が一致した時に報酬が得られるものとする. 各エージェントが報酬を得ることによりそれぞれ各選択確率パラメータ $p_i^k(S)$ を修正していくと, エージェントはその環境での合意形成を実現しやすいような学習を行なうことになる. この時, どのパラメータをどれだけ修正すればよいかといった目安が確率的枠組によって与えられる. 例えば, エージェント α の選択 β についての各確率パラメータの勾配は

$$\partial E / \partial p_\beta^\alpha = \prod_{k=1, k\neq\alpha}^{K} p_\beta^k. \tag{3}$$

により求まるから, 学習アルゴリズムとしてこの値が最大となる確率パラメータを $\Delta p_\beta^\alpha = \partial E / \partial p_\beta^\alpha$ にしたがって変更するといったものが考えられる.

決定的動作をするエージェントが学習を行なう場合にもどの選択が良いかを評価する必要がある. 全エージェントが決定的な選択を行ない続けそれが一致していない場合にはエージェントは報酬を得る機会を失う. そこで一致性の評価のため様々な選択について探索を行なうと, これは確率的な選択を行なうことと同様の操作を行なっているとみなせる.

3.1 シミュレーション

確率パラメータ p によりアクション $(a1, a2)$ をそれぞれ $p, (1-p)$ で選択するエージェントを 10 用意し, 初期パラメータはランダムに設定する. 平均一致確率すなわち式 (2) をできるだけ大きくするようにそれぞれの確率パラメータの学習を行なう. 実験結果のプロットを図 1に示す.

初期値により収束に要する学習回数は異なるが, いずれの試行でも最終的にはエージェントの選択は全て一致した.

4 状況に依存した選択を行なうエージェントとの合意形成

次に問題設定をより実際的にするために, エージェントを環境からの入力を得て状況を判断し, それに応じて選択を行なっているものへと一般化する。

エージェントが決定的な動作に従っている場合にはこれはある状態 s におけるアクション a を確率 1 で選択するものとして表現することができ, 状況の文節 $S = \{s_1, \cdots, s_m\}$, アクション $A = \{a_1, \cdots, a_n\}$ とルール集合 $\{if\ s_i\ \ then\ a_j\}$ の組合せによって表すことができる. ここで S は環境を観測した結果のセンサ入力などから, 「物体を { 検出した, しない }」, 「物体の属性値が { 大きい, 小さい }」, 「検出した物体との距離が $\{l$ 以上, 未満 $\}$」などのように状況を文節するものである.

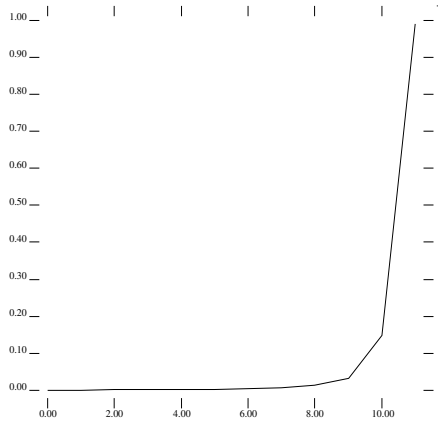

図 1: 学習による合意形成の過程 (縦: 平均一致確率, 横: 学習回数)

確率的動作を行なっている場合には, 各アクションを選択する確率を全て条件付き確率 $p(A|S)$ にすることでアクションの選択に状況依存性を持たせることができる. 状況依存性が複雑な場合には, グラフ構造により変数間の依存関係を表すボルツマンマシンやマルコフネット, ベイジアンネットを用いることでより複雑な状況依存関係を記述したり, 状況依存性を表すモデルのパラメータの入れ方を工夫することによって学習を容易にしたりすることができる。例えば Sigmoid 関数を用いた Sigmoidal belief net[10] や、ニューラルネットを用いたベイジアンネット [12] によりデーターからの学習を行なうものがある。

他のエージェントの内部構造がお互いに未知の場合に, 状況依存性を含めて学習することで合意を形成することを考える. この場合には事前に内部モデルを構成する際には真の S^*, つまり他のエージェントと選択が一致する最適な状況の文節については知ることができない. そこで適当な文節 S を用意してモデルを構成するしかなくここに不確実性が生じるため, もし他のエージェントが実際には決定的なルールにしたがって動作していたとしても, 観測結果は確率的な振舞を示す. そこで他のエージェントが決定的な場合でも予測のための内部モデルは条件付き確率分布 $P(A|S)$ により表現する方が適切であり, この場合にもやはり確率モデルを用いる方が都合が良い.

したがって式 (2) に状況を表す変数 S を導入した下式の最大化を考える.

$$\prod_{k=1}^{K} P^k(A|S) = \sum_{i=1}^{n} \prod_{k=1}^{K} p_i^k(S) \tag{4}$$

環境からの入力を x として次のような確率モデル (Sigmoidal Belief net) によりエージェントを動作させる。

$$P(a_1|x) \equiv p(x) = 1/(1 + exp(-(wx + b)),\ P(a_2|x) = 1 - p(x).$$

同一の環境内における様々な状況でそれぞれ選択を行ない, 全てのエージェントの選択が一致した時のみ報酬を与えることにより学習を行なうことにすると, 学習アルゴリズムは式 (4) を最大化するように次のようにパラメータ w, b を修正すれば良い。

$$\Delta w^\alpha = \partial \prod_{k=1}^{K} p^k(x)/\partial w^\alpha$$

$$\Delta b^\alpha = \partial \prod_{k=1}^{K} p^k(x)/\partial b^\alpha$$

4.1 シミュレーション

上で定義した確率モデルによるエージェントの学習能力を実験的に評価する. 各パラメータの初期値は [-1,1] の範囲でランダムに設定する。環境は様々な状況 x を発生し, 各エージェントはそれにしたがって動作する. 10 エージェントで共学習を行なった場合の選択の一致確率の遷移を図 2に示す。またある一つのエージェントの学習能力の評価として決定的動作, [if $x \geq 0$ then a_1 else a_2] を行なうエージェントに対してのの学習, 時々周期的に動作を変更する非協調的エージェントに対する学習結果をそれぞれ図 3, 図 4に示す. いずれも学習の結果, 十分高い選択一致確率が達成されている.

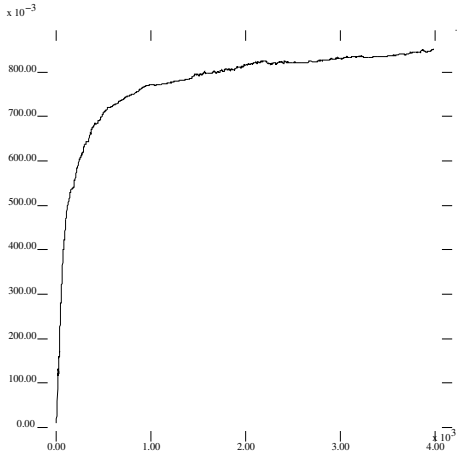

図 2: 10 エージェントの共学習による合意形成 (縦: 平均一致確率, 横: 学習回数)

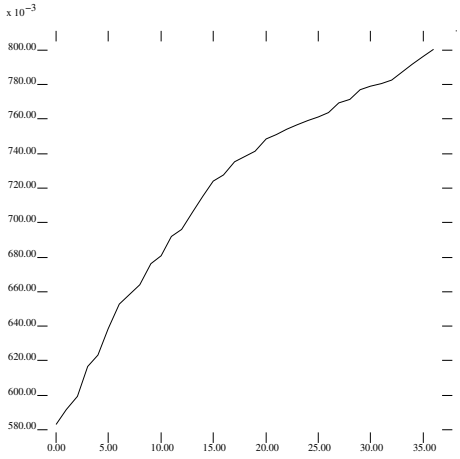

図 3: 決定的エージェントに対する学習 (縦: 平均一致確率, 横: 学習回数)

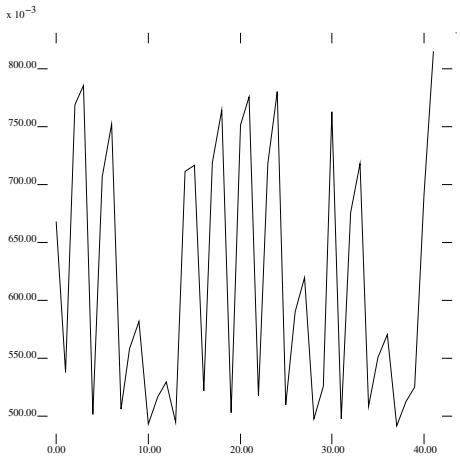

図 4: 非協調的エージェントに対する学習 (縦: 平均一致確率, 横: 学習回数)

5 議論

ここでは、確率モデルを用いることでエージェント間の合意形成について適用可能になる確率・統計的手法について述べる。

5.1 エントロピー

アクション選択の確率分布の条件付きエントロピー,

$$H(A|S) \equiv -P(A|S)logP(A|S) = -\sum_k \sum_i p_i^k(S) \log p_i^k(S)$$

が大きいとそのエージェントのアクションの選択はランダム性の高いものになる。学習の初期ではこれが大きい方が合意の探索に都合が良いが, 式 (4) の値をより大きくするためには最終的にはエントロピーは減少することになる。これを積極的にコントロールすることは焼きなましのプロセスであり、選択の空間が広大な場合によりよい一致点を探索する際に有効であると思われる。

また特に S によるエントロピーの減少分 $H(A) - H(A|S)$ は状況の文節 S の持つ相互情報量に対応する。これによって状況の文節の適切さを評価することができる。ただし、 S だけの学習を行なう場合には他の確率パラメータによる影響を取り除く必要があるため, その間他のパラメータを固定する必要があるだろう。

5.2 観測頻度が少ない場合の確率値の推定

一般に R 回の試行中, r 回行なわれたアクションが選択される確率パラメータの値に関する最尤推定量 (点確率) は, $p = r/R$ として求めることができる.

例えば 3 回の観測のうち a_1 の観測が 2 回であった場合, $p = \frac{2}{3}$ が確率パラメータの最尤推定量である. 観測数 R が大きい場合は最尤推定量は良い近似を与えるが, R が小さい場合には推定結果の分散が大きく不確実性が生じることがあるため, この確率パラメータ自身をさらに確率変数とみなすことで予測精度を高める手法がベイズ推定として知られている. ここで他のエージェントの $P(A|S)$ の予測に関してベイズ推定を行ない, これに対して自分の選択確率を学習することが考えられる.

各 p_i^k に関する確率密度関数を $f(p_1^k, \cdots, p_n^k)$ として, マルチエージェントのアクション選択の場合についての性質を調べてみる. アクション a_i が確率 p_i で独立に選択されている場合を考える. 全部で R 回の観測のうち各アクション a_i が選択された回数が r_i だとすると, 各アクションの選択回数の出方についての同時確率は以下の式で表される多項分布に従うことになる.

$$Pr(r_1, \cdots, r_n) = \frac{R!}{r_1! \cdots r_n!} p_1^{k_1} \cdots p_n^{k_n} \tag{5}$$

(ただし, $r_n = l - \sum_{i=1}^{n-1} r_i$, $p_n = 1 - \sum_{i=1}^{n-1} p_i$.)

ここで p_i としてある特定の値が点確率として存在していると考えるのではなく, ベイズ的に p_i をも確率変数とみて, 各アクションがある出方をした場合の各 p_i の値が x_i となっている $(p_i = x_i(i = 1, \cdots, n))$ 同時分布確率密度を考えると, これは Dirichlet 分布

$$f(P(A)) = \frac{\Gamma(R+1)}{\prod_i^n \Gamma(r_i+1)} \prod x_i^{r_i} \tag{6}$$

に従うことが知られている [8]. 特に観測数 R が比較的少ない場合には p_i の不確実性は比較的大きく, 分布の裾も広くなる. 例えば, 先の例で $R = 3, r_1 = 2$ の場合の p_1 の分布は図 (5) のようになっている.

そこで観測数が比較的少ない場合には, 他のエージェントを近似するための内部モデルとしては式 (4) の代わりにこれに式 (6) で重みづけた平均

$$\int_P \prod_{k=1}^{K} P^k(A|S) f(P^k(A|S)) dP \tag{7}$$

を考えた方が良いかも知れない. この分布はいろんな確率パラメータの分布をそれぞれ混合した形になっているので混合分布 (Mixture) とも呼ばれる.

6 まとめ

本研究では複数のエージェントが構成する環境における合意形成を確率モデルの学習により実現する方法について述べた。実験評価により, 山登り法的な学習により安定した合意形成が達成できた.

また確率モデルの学習による合意形成には以下の特長がある.

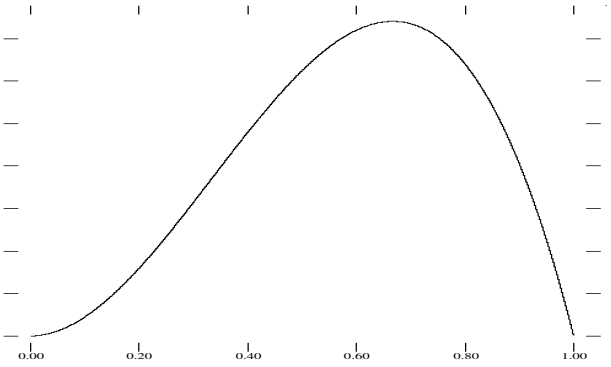

図 5: 3 回中 2 回観測された場合の確率パラメータの Dirichlet(Beta) 分布

- 学習の初期に確率的な選択を行なうことにより, 報酬を通じて, その選択の妥当性に対する評価を一致頻度として他のエージェントに伝えている.
- 未知のモデルが内部で用いている効用関数や状況の文節などを考慮せずに, 単に確率モデルのパラメータとして任意のモデルに対するフィッティングが行なえる.
- ネットワーク構造を持つ確率モデルを用いることで, より複雑な状況依存性を持つエージェントについても適用できる.
- 確率モデルのパラメタライズを工夫することで学習アルゴリズムの導出が容易になる.

今後の課題としては, 具体的な問題における合意形成への適用, より複雑な状況依存性を持つ場合の学習能力の評価などがあげられる.

参考文献

[1] Schelling, T.C.:"The Strategy of Conflict", NewYork, Oxford University Press,(1963).

[2] 大沢:" 焦点に基づくエージェント間の整合, 第 4 回インテリジェントシステムシンポジウム, pp.33-36(1994).

[3] 國吉: "実世界エージェントにおける注意と視点", 人工知能学会誌, vol.10No.4, pp.507-514,(1995).

[4] 生天目, 塚本: "複合エージェントによるエージェント社会の構築", マルチエージェントと協調計算 III, MACC'93, (1993).

[5] 松原:" 新版 意志決定の基礎", 朝倉書店 (1985).

[6] Sudgen,R.:"A Theory of Focal Points",The Economics Journal, 105(May), pp.533-550, (1995).

[7] Fenster, M. Kraus, S., J.S.Rosenschein:"Coordination without Communication: Experimental Validation of Focal Point Techniques", ICMAS-95, (1995).

[8] DeGroot,M.: "Optimal Statistical Decisions",McGraw-Hill, New York (1970).

[9] Geiger,D. and Heckerman, D.: " A Characterization of the Dirichlet Distribution with Application to Learning Bayesian Network",Maximum Entropy and Bayesian Methods,pp.61–68, (1995).

[10] Neal, R.: "Connectionist Learning of Belief Networks", Artificial Intelligence, 56, pp.71–113, (1992).

[11] 本村, 麻生, 國吉, 原, 赤穂, 松原, 開 :" 複数自律ロボット環境における Focal Point の形成" , MACC'95 (1995).

[12] Motomura,Y.: "Bayesian Network that Learns Conditional Probabilities by Neural Networks", Proc. of Int. Conference on Neural Information Processing '97, pp.584–587, (1997).