解説

確率的な知識の獲得と利用[†]

麻生 英樹*

1. はじめに

情報処理システムが、非限定的な開かれた実環境において、環境と密に相互作用しながら、固定的なゴールを持たないオープンなタスクを実行する場合、システムが置かれるであろう状況をあらかじめ完全に特定しておくことが困難であるため、なんらかの形で「情報の不完全さ」への対処が重要となる[17]。

人間が「もしかすると～かなあ。」「たぶん～だと思ふよ。」というような話法を使うように、「情報の不完全さ」への対処のために、「確率」あるいはそれに類した「確からしさ」の使用を考えることは自然である。

このような確率的推論を用いたシステムに関する初期の研究としては、古典的なエキスパートシステム MYCIN[35]に用いられた「確信度(Certainty Factor)」つきのルールがまずあげられる。MYCIN は血液感染症の診断と抗生物質処方についてのエキスパートシステムであるが、if-then ルールに「確信度」を付与することで、知識提供者であるエキスパートの持つ「確からしさ」の情報を取り入れるとともに、ルールの選択などの推論制御を柔軟化した。

また鉦脈探査エキスパートシステム PROSPECTOR では、Bayes 分析(Bayesian Analysis)

を使った帰納推論が行なわれた[11][12]。

これ以降も、この種の確率的推論の枠組としては、信念理論(belief theory)[34]証拠推論(evidential reasoning)[14][16]、必然性と可能性の理論(necessity and possibility theory)[46]、データ解析手法にもとづくもの[3]などさまざまなものが提案されている。また、述語論理に確率を付与する試みも、確率論理(Probabilistic Logic)[29]などがある。

確率的推論の研究の方向を次の二つに分けてみるのは有効かもしれない。

- ・合理的な確率的判断の実現によるシステムの効率化をめざすもの
- ・人間の確率的判断の説明、再現、利用をめざすもの

上の例でいえば、MYCIN は後者の色合いを持ち、PROSPECTOR は前者の色合いが強いと言えるだろう。また、後者は主に心理学、認知科学において研究されている。

確率的推論を古典的な確率をベースとして展開することについては、依然として賛否がわかれるところであるが、本稿では、確率的な知識の獲得および利用として、Bayes 決定理論をベースに、通常の確率にできるだけ準拠する立場から、できるだけ合理的な確率的判断の実現をめざす場合の問題について考察し、最近の研究の展開を概観する。

[†] Probabilistic Knowledge Acquisition and Probabilistic Inference

* Hideki ASOH
電子技術総合研究所
Electrotechnical Laboratory

2. Bayes 決定理論

確率的推論の問題は、特定の条件の下で、仮説的命題の確からしさを評価し、そこから適切な行動制御戦略を導くことである。行動選択に関係する各命題の真偽が完全に決定されていれば、通常の最適化問題であるが、そうではないところに、確率の必要性が生まれる。

この種の問題は、古くから、統計学の下で研究されてきた。その一つの到達点は Bayes 決定理論 (Bayesian Decision Theory) である [6]。その理論によれば、対象となっている世界の記述、つまり Universe of Discourse、その上の同時確率分布、効用関数がすべて既知であれば、期待効用を最大化する Bayes 最適 (Bayes Optimal) な行動選択が行なえる。

そこでは、知識獲得は、過去の多くのデータからの同時確率分布の推定であり、知識の利用は、獲得された同時確率分布と、現在の状況についての情報を用いた条件つき確率値の計算、そして、その値にもとづいた、期待効用最大化である。

このように、Bayes 決定理論は概念的には明快な枠組みであるが、それを実際に適用しようとすると、数多くの問題が生じる。確率的推論の研究の主要部分は、そうした問題への対応策であるとも言えるだろう。以下、それらの問題を以下のような四つの一般的原因に分けて論じる：

- ・ Universe of Discourse が不確定
- ・ 効用関数が不確定
- ・ 確率分布が不確定 (必要なデータ量の爆発)
- ・ 計算量／記憶量の爆発

3. Universe of Discourse の不確定

ある状況で、ある選択に関与する条件およびその記述のしかたがあらかじめわかっていることは少ない。よく使われる病気診断の例で説明すると、熱がある、というような症状から原因となる病気を診断する場合に、熱の高さのカテゴリ化の変

更や新たな病気カテゴリの導入の可能性にどう対処するか、という問題である。自己組織データベースのように、概念の階層構造を変更するというような場合もここに含まれる。

最終的には、確率推論システム自らに、そうした新たなカテゴリや仮説命題を自己組織的に導入する仕組みを持たせることが望ましいし、そのためにクラスタリング技法を使う試みなどが行なわれている。しかし、Universe of Discourse の変化は、確率的知識の大幅な変更を伴うことが多いため、一般的には、Universe of Discourse に変更があった場合には、またははじめから考え直す、という第 0 次近似解がとられ、この不確実性が正面から取り上げられることはあまり多くない。今後の重要な課題であることは確かだが、本稿でも、これ以上は論じない。

4. 効用関数の不確定

多くの実世界的な意思決定においては、異なる選択を比較するための価値尺度がよくわからないことも普通に起こる。Bayes 決定理論はこのような場合に無力であり、これも実際的には非常に重要な問題である。しかし、著者は、これに対する一般的な対処方針を知らない。強いてあげれば、とりあえずの評価を用いて意思決定を行ない、結果が望ましくなければ、評価自身も修正する、というような探索的な技法であろうか。

これに対し、たとえば、一連の行動の最終結果に対して評価が与えられるが、個別の要素行動についての評価は与えられないような場合に、要素個別行動の評価を推定するという問題は、強化学習 (Reinforcement Learning) として研究されている。効用が時間とともに一律に通減することを仮定したマルコフ連鎖モデルに基づく推定問題に帰着させるのであるが、代表的なアルゴリズムとしては Temporal Difference (TD- λ) [37] や Q-Learning [43] などがよく知られ、確率的要素の強いゲームである Backgammon の戦術学習 [38] やロボットの行動計画の学習などに使われている。

5. 確率分布の不確定(データ量の爆発)

確率分布の推定は、一般には、不良設定問題であり、安定な推定結果を得るには多数の偏りのないデータが必要である。しかし、実際にはそうしたデータが得られないことはごく普通である。そこで、確率分布自体の不確実性が問題になる。

この問題はまた、確率推論における「無知」と「等確率」の扱いにも関係している。「確率の値がわからない」ということと「等確率」ということは明らかに異なるが、しばしば、わからない場合は等確率とする、という0次近似解が用いられるためである。

この問題への対処法は大きく二つに分けることができる。一つは、Dempster-Shafer 理論[8][9][22][23][34]に代表される、確率分布の不確実性をそのまま扱えるように確率論を拡張する、というもので、理論的にも美しく、応用例もいろいろあるが、確率値の不確実性を扱うために、通常の確率計算よりもさらに計算量が増えるという問題がある。従って、対象の規模拡大への対処が一般には難しい。

これに対し、もう一つの方法は、分布推定問題を良設定にすることである。具体的には、正則化条件や確率分布モデルによって記述される構造的制約を明示的に導入することである。

古典的な統計においては、正規分布を中心とした分布モデル族が扱われてきたが、主に計算機の発達と有効な算法の発明によって、現在では、使用可能なモデルの種類は大きく広がってきている。中でも、Bayesian Network は、最近研究が盛んである[39][40]。Bayesian Network は、命題間の確率的依存関係を有効グラフのネットワークによって表現し、確率計算をネットワークを通じた局所的な情報伝播によって効率的に実行するものである。その基本的なアイデアは、事象間の独立性の利用である。たとえば、4つの事象の間に図1のようなネットワークで表される確率依存関係があるということは、同時確率分布 $P(A; B;$

$C; D)$ が $P(D|B; C)P(C|A)P(B|A)P(A)$ という構造を持つという意味である。これを、一般的に成立する $P(D|A; B; C)P(C|A; B)P(B|A)P(A)$ という分解と比べると、同時確率分布が単純化されていることがわかると思う。そして、Bayesian Network の有効性は、単に確率分布の構造をグラフによって表現する表現法を与えただけではなく、そのグラフ表現を用いることで、確率計算が効率的に実行できるというところにある。

図1 最も簡単なBayesian Networkの例

この他にも、確率オートマトン、確率文法、Hidden Markov Model(HMM)Random Markov Field(RMF)[7]、Boltzmann Machine[19]に代表される確率 Neural Networks、Bayesian Neural Network[27][42]など階層的な Neural Network の確率解釈、Deformable Template[45]、複数の分布を確率的に足し合わせた混合分布(mixture)[24]、など対象および用途に応じていろいろなモデルが提案されている。また、Nils-son の Probabilistic Logic と同じような、より大きな汎用性をめざす方向としては、述語論理の記述力を生かそうとする確率制約論理や、確率論理プログラムなどがある。

対象と用途にあったモデルおよびそのモデルを用いた知識獲得と利用のアルゴリズムを発明/発見することは、確率推論における最も重要な問題の一つである。

6. 計算量・記憶量の爆発

最も単純な構造無しの離散分布の場合、同時分布を記憶するために必要な確率値の数は、仮説命

題の数 n について指数的に増加する。従って、たとえデータが例外的に豊富に与えられても分布推定の計算量や結果を記憶するための記憶量、条件つき確率を計算する操作に必要な計算量などが現実性を失うことがしばしば起こる。

前節で述べたように、確率分布に構造を導入することは、推定の安定に必要なデータ量の削減だけではなく、計算量および記憶量を減らす効果もあり、対象とモデルがうまくフィットした場合には大きな成功をもたらす。最近の例では、音声認識や遺伝子情報処理における Hidden Markov Model や確率文法の利用[28]画像処理、画像理解における Random Markov Field、Deformable Template の利用[2][15][45]などがあげられる。

次節の内容ともかかわるが、Bayesian Network の例でもわかるように、モデルによる構造表現のみならず、それを用いた場合の分布推定アルゴリズムと確率計算(あるいはたとえば、事後確率最大解探索)アルゴリズムに実用的なものがあるかどうかも重要である。

たとえば HMM が応用されるにあたっては、高速な学習のための Baum-Welch アルゴリズムや事後確率最大解探索のための Viterbi アルゴリズムの存在が大きかったと思われる。Baum-Welch アルゴリズムや Boltzmann Machine の学習アルゴリズムは、直接観測不能ないわゆる隠れ変数を持つ確率モデルを導入した場合の分布推定アルゴリズムとして提案された Dempster の EM (Expectation Maximization) アルゴリズム[10]の一種であり、混合分布モデルの推定などにも使われる。

一方、確率的知識の利用で重要な条件つき確率、条件つき期待値の計算に関しては、近似計算法として、モンテカルロアルゴリズム[21]、あるいは中間場近似 (Mean Field Approximation) [33]が使われる。これらの計算技法は、統計力学の分野で物性量の計算のために発達してきたものであるが、最近では、より広い範囲で応用されはじめている。

7. 計算論的学習理論

機械学習の分野では、学習アルゴリズムの進歩や Valiant による PAC (Probably Approximately Correct) 学習概念の導入等[41][25]をきっかけとして、理論的研究が盛んに行なわれてきた。この分野は計算論的学習理論 (computational learning theory) と呼ばれている[5]。

確率的な知識の獲得も学習の一種であり、これについても、前節までに述べたような、新しい種類の確率モデルの増加もあいまって、研究が盛んである[1][44]。理論的研究の主な目的は、たとえば、ある確率分布モデルを一定以上の精度で推定するために、必要なデータ量や計算量を評価すること、である。必要なデータ量の評価の問題はまた、できるだけ少ないデータで安定した推定結果を得ることにともつながるため、「汎化 (generalization)」の問題とも呼ばれている[4]。

結果として、多くの複雑な構造を持つモデルに関しては、安定な推定に必要なデータ量はそれほど増えないが、推定に必要な計算量は、かなり増えてしまうことが知られている。つまり、多くの分布推定問題が NP 困難であることが判明している。そこで、このような場合については、現実的に高速に近似解を求める手段として、構造部分とパラメータ部分の推定を分離し、構造決定部分については、遺伝的アルゴリズムや遺伝的プログラミングの技法を用いる試みなどもある[26][20]。

8. 応用

情報処理システムの扱う情報の種類やシステムが処理を行なう状況が多様化し、実世界とリアルタイムに相互作用するシステムが求められるようになりつつある現在、こうした技法の重要性はますます大きくなっていると思われる。大津らの提唱する「柔らかな情報処理」を実現するシステムの基盤となるべき「柔らかな論理」においても、確率的推論は、重要な要素の一つとして取り上げ

られている[31]。

情報処理における確率推論の応用事例として最も多いものは、パターン認識[13]と時系列予測であろう。いずれも、多様性の高いパターンの情報を扱うために、確率的要素の導入が不可欠であった。

そこで用いられている推論は、ほとんどの場合、入力信号からその属するクラス、あるいは予測値を推定する、という一段の浅いものであった。従って、これらを、いわゆる「確率推論」とは呼ばないことも多い。しかし、最近になって、パターン認識の研究が、文字認識や音素認識といったレベルから音声理解、談話理解、画像理解、のような、より高次の情報構造の認識理解を目的とするようになるにつれ、より複雑な確率モデルを導入して確率的な推論を行なうことが行なわれるようになっていく。すでに述べたように、音声理解における HMM の成功は象徴的であるし、画像理解においても、確率分布による認識対象のモデル化が、より高次の対象(画像内の「車」、「顔」などのような)に対しても行なわれるようになっていく。遺伝子情報解析においても確率文法などの技法が活用されているし、自然言語理解においても確率的手法が再評価されつつある。

これに対し、より記号処理的な、深い、数段に渡る推論については、依然として、故障診断、医療診断が代表的であるようだが、最近では、自律ロボットの行動計画において、確率的な推論を利用することで、計画を外乱に対してロバストなものにする、といった研究も見られる[30]

複雑なモデルを応用する場合、簡単に使えるソフトウェアツールがあると便利である。確率・統計・データ解析のソフトウェアは SAS や S をはじめいろいろあるが、これらの、主に古典的なモデルのためのツールに加えて、ニューラルネットワークモデルのためのツール PDP++[32]、Bayesian Network のためのツール Barron(KC Associates)、Hugin、Hidden Markov Model Toolkit(HTK: Entropic Research Laboratory Inc.)、など最近のモデルのためのツールも充実し

つつある。

9. おわりに

確率論にできるだけ則した立場から、確率的推論の問題を考察し、いくつかの技法および応用を概観した。情報処理技術の対象の広がりに伴って、確率的な技法の導入／再評価が進んでいる。

確率的推論における最大の問題は、確率分布の推定や確率値計算に伴うコストである。このために、さまざまな技法が提案されているが、古典的な確率に準拠することで、しばしば ad hoc になりがちな heuristics をできるだけ明示的に定式化できるようにする。また、複雑な確率モデルを設定することで、より柔軟な推論が可能になる。

しかしながら、見方を少し変えると、実世界的なシステムに確率推論を導入することの第一のメリットは、状態が不定な場合でも、確率値計算にもとづいてなんらかの処理を行なうことができる、とも言えるだろう。その上で、もしも計算された確率値が適切なものであれば、それはシステムの処理効率を上げるのに役立つ。

この意味では、確率推論の応用においては、確率値の正確な推定や計算にあまりこだわるのはよくない、とも言えるのであり、ファジィ推論が成功した理由もこのあたりにあるように思われる。従って、確率推論の応用に関して実際の観点から、適切な構造モデルの構成と並んで重要なことは、システムに、自らの状態を常に評価し、誤った判断から復元できるような力を持たせておくことであろう。これはまた、確率的推論という不安定なツールを全体的なシステム設計の中で、どのように使いこなすか、という問題でもある。

論じられていない話題としては、カオスとの関係がある。情報の不完全性の裏には多くの場合、カオス的なシステムの複雑さがあるのであり、そこに確率を持ち込むのは既に一種の近似であるとも言える。カオス理論が、今後、確率「的」推論の中でどのように使われてゆくのかは興味深い。

謝 辞

本稿の執筆にあたっては、Real-World Computing Program の確率的知識の獲得と利用ワークショップでの講演、議論を大変参考にさせていただきました。ワークショップ主査の上坂吉則東京理科大学教授、ワークショップメンバーのみなさま、特に、統計数理研究所の伊庭幸人さん、外部講演者の方々に深く感謝いたします。

参 考 文 献

- [1] Abe, N. and Warmuth, M.: On the computational complexity of approximating distributions by probabilistic automata, *Machine Learning*, **9**, 205-260(1990).
- [2] Amit, Y., Grenander, U., and Piccioni, M.: Structural image restration through deformable templates, *Journal of the American Statistical Association*, **86**, 376-387(1991).
- [3] 麻生英樹：確率的知識と推論について，電子通信学会技術研究報告 AI-86-22(1986).
- [4] 麻生英樹：汎化能力の高い学習アルゴリズムについて，電気学会光応用・視覚研究会資料 LAV-93-3(1993).
- [5] 有川節夫他：特集：「計算論的学習理論とその応用」，情報処理，**32**，217-281(1991).
- [6] Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis* (2nd. Ed.), Springer-Verlag (1985).
- [7] Chellappa, R. and Jain, A. (eds.): *Markov Random Fields, Theory and Application*, Academic Press(1993).
- [8] Dempster, A.P.: Upper and lower probabilities induced by a multivaluated mapping, *Annals of Mathematical Statistics*, **38**, 325-339(1967).
- [9] Dempster, A.P.: A generalization of Bayesian inference, *The Journal of the Royal Statistical Society B*, **30**, 205-247(1968).
- [10] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm, *The Journal of the Royal Statistical Society B*, **39**, 1-38(1977).
- [11] Duda, R.O., Hart, P.E., Nilsson, N.J.: Subjective Bayesian mehotds for rule-based inference systems, *Proc. 1976 National Computer Conference* **45**, 1075-1082(1976).
- [12] Duda, R.O. et al.: Development of the PROSPECTOR Consultant System for Mineral Exploration, Final Report for SRI Projects 5821 and 6915, Artificial Intelligence Center, SRI International(1978).
- [13] Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Academic Press(1972).
- [14] Garvey, T.D., Lowrance, J.D., and Fischler, M.A.: An inference technique for integration knowledge from disparate sources, *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 319-325(1981).
- [15] Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, **6**, 721-741(1984).
- [16] Hajek, P.: Combining function for certainty degrees in consulting systems, *International Journal of Man-Machine Studies*, **22**, 59-76(1985).
- [17] 橋田浩一：人工知能とは何でないか—情報の部分性について，*bit*, Vol.20, No.8, pp.896-908(1988).
- [18] Heckerman, D.: A tutorial on learning Bayesian Networks, Technical Report MSR-TR-95-06, Microsoft Corporation, Microsoft Reserach Advanced Technology Division (1995).
- [19] Hinton, G.E., Sejnowski, T.J., and Ackley, D. H.: Boltzmann Machines: constraint satisfaction networks that learn, Technical Report, CMU-CS-84-119, Carnegie-Mellon University, Department of Computer Science(1984).
- [20] Iba, H., Kurita, T., de Garis, H., Sato, T.: System identification using structured genetic algorithms, *Proceedings of 1993 International Conference on Genetic Algorithms*, 279-286(1993).

- [21]伊庭幸人：マルコフ連鎖モンテカルロ法とその統計学への応用, Research Memorandum No. 570, 統計数理研究所(1995).
- [22]Ishizuka,M.: An extention of Dempster & Shafer's theory to fuzzy set for constructing expert systems, 生産研究, **34**, 312-315(1982).
- [23]石塚満：Dempster & Shafer の確率理論, 電子通信学会誌, **66**, 900-903(1983).
- [24]Jacobs,R.A., Jordan,M.I., Nowlan,S.J., and Hinton,G.E.: Adaptive mixtures of local experts, *Neural Computation*, **3**, 79-87(1991).
- [25]Keans,M.J.: *The Computational Complexity of Machine Learning*, The MIT Press(1990).
- [26]北野宏明(編)：遺伝的アルゴリズム, 産業図書(1993).
- [27]Mackay,D.J.C.: A practical Bayesian framework for backpropagation networks, *Neural Computation*, **4**, 448-472(1992).
- [28]中川聖一：確率モデルによる音声認識, 電子情報通信学会(1988).
- [29]Nilsson,N.J.: Probabilistic logic, *Artificial Intelligence*, **28**, 71-87(1986).
- [30]Nourbakhsh,I., Powers,R., and Birchfield,S.: DERVISH an office-navigating robot, *AI Magazine*, Summer, 53-60(1995).
- [31]大津展之：リアルワールドコンピューティング研究計画, 情報処理, **34**, 1423-1428(1993).
- [32]<http://www.cs.cmu.edu/Web/Groups/CNBC/PDP++/PDP++.html>
- [33]Peterson,C. and Anderson,J.R.: A mean field theory learning algorithm for neural networks, *Complex Systems*, **1**, 995-1019(1987).
- [34]Shafer,G.: *A Mathematical Theory of Evidence*, Princeton Univ. Press(1976).
- [35]Shortliffe,E.H. and Buchanan,B.G.: A model of inexact reasoning in medicine, *Mathematical Biosciences*, **23**, 351-379(1975).
- [36]Shortliffe,E.H.: *Computer-Based Medical Consultation : MYCIN*, Elsevier(1976).
- [37]Sutton,R.S.: Learning to predict by the methods of temporal difference, *Machine Learning*, **3**, 9-44(1988).
- [38]Tesauro,G.: Temporal difference learning of backgammon strategy, *Proceedings of the Tenth International Conference on Machine Learning(ML-92)*, 451-456(1992).
- [39]Pearl,J.: Fusin, propagation and structuring in belief network, *Artificial Intelligence*, **29**, 241-288(1986).
- [40]Pearl,J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann(1988).
- [41]Valiant,L.G.: A theory of the learnable, *Communications of the ACM*, **27**, 1134-1142(1984).
- [42]Watanabe,S. and Fukumizu,K.: The unified neural network theory and its application to new models, *Proceedings of 1992 International Joint Conference on Neural Networks (Beijing)*, II-381-386(1992).
- [43]Watkins,C.J.C.H.: Q-Learning, *Machine Learning*, **8**, 279-292(1992).
- [44]Yamanishi,K.: A learning criterion for stochastic rules, *Machine Learning*, **9**, 165-203(1992).
- [45]Yuille,A.L.: Generalized deformable models, statistical physics, and matching problems, *Neural Computation*, **2**, 1-24(1990).
- [46]Zadeh,L.A.: Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, **1**, 3-28(1978).

(1994年4月4日 受付)
(1994年11月24日 再受付)

[問い合わせ先]

〒305 つくば市梅園1-1-4
電子技術総合研究所
情報科学部情報数理研究室
麻生 英樹
☎：0298-58-5839
☎：0298-58-5841
e-mail：asoh@etl.go.jp

著 者 紹 介

麻生 英樹 (あそう ひでし)

電子技術総合研究所情報科学部情報
数理研究室

1981年 東京大学工学部計数工学科卒業。1983年 東京大学大学院工学系研究科情報工学専攻修士課程修了。同年 電子技術総合研究所に入所。現在情報科学部情報数理研究室主任研究官。知的システムの実現に関する研究、特に、確率的推論、学習システム、遺伝的アルゴリズム、自律知能ロボットなどの研究に従事。行動計量学会、日本神経回路学会、人工知能学会、電子情報通信学会各会員。