不確実性をもつ論理式の帰納推論に関する一考察†

松嶋敏泰†† 稲積宏誠††† 平澤茂一††††

不確実性をもつ論理式で構成される仮説の帰納推論について考察する．著者らは有限個の事実系列からの帰納推論の問題と不確実性を含んだ演繹推論の問題に対して一階述語論理と情報理論を用いた統一的基礎体系より一連の研究を行ってきた．本論文ではこの２つの研究の流れを統合して不確実性をもつ論理式（ルール）で構成される仮説の帰納推論問題について，帰納と演繹を個別で論じるのではなく一連の情報処理過程として情報理論と決定理論の立場から考察を行う．有限個の事実を得た時点で，1）真の仮説を選択する問題，2）新たな論理式の予測を行う問題，3）事実系列を最小の記述長で表現する仮説を選択する問題の３つの評価基準から考察を行い，それぞれの基準で最適な推論法を示し，その相互の関係についても言及する．

1. はじめに

不確実性をもつ論理式や確信度つきルールを用いた演繹推論の問題は初期のエキスパートシステムであるMYCIN[1]等の研究に始まり現在まで様々な研究[2),3]がなされているが，不確実性をもつ論理式で構成される仮説を帰納推論する問題の研究は最近まであまり見受けられなかった．その重要な理由の１つとして，従来の帰納推論の基本的アルゴリズムである与えられた事実に対して無矛盾な仮説を出力する推論戦略のみでは，この種の帰納推論を行うことが不可能であることが挙げられる．

従来の不確実性を含まない仮説の帰納推論や計算論的学習理論において，極限同定や PAC (Probably Approximate Correct)[4]等の成功基準のもとで学習可能性や各種効率を評価する場合，基本的アルゴリズムは任意の無矛盾な仮説を出力すれば十分であった．例えば，PAC 学習可能性の一般的証明技法である Vapnik-Chervonenkis 次元を用いた手法[5]においても，与えられた事実に対して無矛盾な任意の仮説であれば証明の展開に支障はない．ところが，この無矛盾な仮説を出力するアルゴリズムでは不確実性を含んだ規則や事例にノイズが入る場合には，いつまで例を与えても明らかに矛盾しているため切り捨てられる仮説が特定できず，仮説の候補が絞り込めないため，真の仮説あるいはそれに近い仮説を推論することは不可能となる．

本研究ではこの不確実性をもつ論理式の帰納推論の問題について，情報理論による帰納・演繹推論の統一的基礎体系に基づく知識情報処理に関する著者らの一連[6)~8]の研究の一環として考察を行う．この一連の研究では古典的演繹推論では取り扱えない知識情報処理における様々な推論の問題を個別のモデル化から考察するのではなく，第一階述語論理と情報理論による推論の基礎体系を提案し，推論を情報の視点から整理することにより統一的な観点から考察を行っている．

この一連の研究では現在までに，主に２つのタイプの推論を考察している．その１つは，帰納推論の問題であり，先に述べたように任意の無矛盾な仮説を出力する従来のアルゴリズムは極限での収束性を論じる評価では十分であるが，実際問題としては有限の事例を与えられた時点で無矛盾な仮説は複数存在しており，その中でどの仮説を選択するかはあまり論議されていなかった．そこで帰納推論を行う目的に立ち返り統一的基礎体系上でいくつかの目的関数を導入し，そのもとで最適な推論法を提案している[6),8]．

この著者らの帰納推論の研究は計算論的学習論で行われている研究と研究の立場に違いがあり，そのため研究のアプローチも異なっている．計算論的学習理論の場合，どのレベルの言語で表現される仮説がどの程度の効率で学習されるかに興味があるため，まず学習されたという基準，つまり学習成功（PAC 等）の基準を与えておき，その基準がどの程度の効率（事例の数等）で達成されるかを評価するアプローチで研究が行われる．成功の基準は制約条件であり，それを達成するための効率を様々な観点から論じている．著者らの研究では学習成功の基準を損失関数でとらえ，その損失関数自体を目的関数として最小化する推論法を

† Inductive Inference for Uncertain Formulas from the Viewpoint of Information Theory by TOSHIYASU MATSUSHIMA (Department of Management Information, Yokohama College of Commerce), HIROSHIGE INAZUMI (Department of Information Engineering, Shounan Institute of Technology) and SHIGEICHI HIRASAWA (Department of Industrial Engineering and Management, School of Science and Engineering, Waseda University).
†† 横浜商科大学経営情報学科
††† 湘南工科大学情報工学科
†††† 早稲田大学理工学部工業経営学科

考察している．この点で計算論的学習理論と研究の立場に違いがあると考えられる．

著者らのもう一方の研究は不確実性を含む演繹推論の問題であり，古典的演繹推論の健全性，完全性の基準では評価ができないこの種の推論に対して，新たに誤り確率の概念から最適な推論規則や論理式の不確実性の表現法を提案している[7]．

本研究では上記の著者らの帰納推論の研究と不確実性を含む演繹推論の研究とを統合することにより，不確実性をもつ仮説の帰納推論の問題を考察する．統合とは，単に不確実性を含む演繹推論の研究で提案した不確実性を含む論理式の表現法を用いて帰納推論をモデル化するというだけではない．２つの研究は統一的な基礎体系のもとで考察されているため，不確実性を含む仮説に対しても帰納と演繹を一連の推論過程として総合的に取り扱えることが本研究の特徴となっている．例えば本研究で提案する帰納推論法で帰納された不確実性を含む仮説を用いて先に提案した演繹推論法により演繹することも可能であり，さらに帰納と演繹を複合した推論として予測の問題についても整合性良く考察が可能である．また，論理式の不確実性の表現や推論の有効性についても帰納・演繹推論を通じ一貫した表現法や評価基準を用いている．研究の基本的アプローチとしては前稿と同様に推論の目的を明確にし，それを最適化する推論アルゴリズムを提案していく．

2. 準　備

2.1 確率世界論理

情報理論における情報量[9]はメッセージを確率変数とみなすことにより定義されるため，対象とする体系のどの部分が情報を担っている確率変数であると仮定するかによって情報量の定式化は異なってくる．著者らの一連の研究では推論の知識表現として主に古典的第一階述語論理を用いた場合を考えている☆．ある形式言語が与えられているとして，論理上で最も重要な情報はセマンティクスについての情報であるという観点から，解釈のパターンを確率変数とみなし論理に確率構造を導入することにより，推論の構造を情報理論からも考察可能な基礎体系の構築を試みている[6]．

定義 2.1 (確率世界論理)[6]　可算な第一階述語論理言語Lに対する古典論理的解釈のパターン☆☆を要素とした集合である解釈空間Ωを標本空間とし，その部分集合のσ集合体とその上の確率測度Pを定義することにより確率空間を構成する．このような確率空間をもつ述語論理を確率世界論理と定義する．　□

簡単な例として，２つの定数記号 a_1，a_2 と１つの一変数述語 Q で構成される第一階述語言語 L を考える．この言語上で以下の４つの解釈のパターンが解釈空間Ωの要素となる．

m_1：{$Q(a_1)$ は真 and $Q(a_2)$ は真}
m_2：{$Q(a_1)$ は真 and $Q(a_2)$ は偽}
m_3：{$Q(a_1)$ は偽 and $Q(a_2)$ は真}
m_4：{$Q(a_1)$ は偽 and $Q(a_2)$ は偽}

一般に，有限な第一階述語言語Lの解釈のパターン数を考えてみる．L の定数記号の個数を n_c，i-変数述語記号の個数を $n_{p(i)}$ $(i \leq K)$ とすると，サンプル空間となる解釈空間 Ω の要素数 $|\Omega|$ は以下のようになる．

$$|\Omega| = n_{p(i)} \sum_{i=1}^{K} 2^{n_c{}^i}. \qquad (1)$$

議論を簡単にして本質的部分を分かりやすくするため，以下では対象とする言語は有限な第一階述語論理言語Lとする．

古典論理はある１つの解釈（世界）を対象にするが，確率世界論理は上述のようにすべての解釈のパターンを同時に考えることになる．このように多くの世界を同時に対象とする論理としては，様相論理[10]があり，不確実性を含む演繹推論のために確率構造を導入した論理としては Nilsson の Probabilistic logic[2]☆☆☆がある．また確率構造を導入した帰納推論のモデルとしては，PAC 学習モデル[4]を拡張した Haussler のモデル[11]がある．PAC 学習モデルの特徴は事実の空間に確率測度を導入したことにより学習の確率的評価を可能にしたことにある．著者らの一連の研究でも解釈空間のみでなく定数記号の定義域上に確率を導入している．Haussler のモデルにおける事例の確率分布は，同じ性質をもった，つまり属性を表現する述語すべてに対して真理値が等しい定数記号の確率の和として求めることが可能である．さらに Haussler によって拡張された属性 X と概念 Y の条件つき確率による関

☆ 様々なパラダイムの帰納推論や学習問題が研究されているがほとんどの問題は述語論理を用いて表現可能と考えられる．

☆☆ 第一階述語言語Lのすべてのグランドアトムに対するある真理値の付置を１つの解釈のパターンとする．

☆☆☆ 確率世界論理が解釈空間上に確率構造を定義したのに対して，Probabilistic logic は解釈上に確率を仮定しているものの，与えられた論理式を１つの命題とみなし，その命題に対する真偽の解釈しか考慮していないため，帰納推論はまったく扱えない体系となっている．また，限定子を含んだ論理式に対してもそれを１つの命題とみなし真偽を割り振っているため，実質的には命題論理しか扱えない体系となっている．この違いについては付録Aを参照されたい．

係 $P(Y|X)$ や山西によって定義された確率的ルール[12]は，前稿[7]で定義し，本研究でも用いる条件つき信頼率と同等の概念と考えられる．よって，確率世界論理上では Haussler のモデルと同等のモデルを述語論理上に表現可能である．さらに，確率世界論理は解釈空間から確率構造を導入してるため決定理論的定式化が容易に行え，第1章で述べた最適なアルゴリズムを求める立場での展開が可能となる．また，条件つき信頼率以外の信頼率の表現も可能であるため，様々な曖昧性を含んだ知識の推論の基礎体系としても有用であり，前稿で詳しく考察を行っている[7]．

この確率世界論理上では論理式や帰納，演繹推論を情報量やエントロピーを用いて情報理論から考察可能である．その一部を簡単に以下に記す．ある閉論理式の真偽の情報，以下セマンティク情報と呼ぶ，はある解釈のパターン m が決まれば一意に決定される情報である．

定義 2.2[6] 解釈空間 Ω 上で解釈のパターン m の確率を $P(m)$ とすれば，論理式 W のセマンティク情報の自己情報量 $i(T(W|m)=1)$ は次のように定義される．

$$i(T(W|m)=1)=-\log\left(\sum_m T(W|m)P(m)\right), \quad (2)$$

ここで，

$$T(W|m)=\begin{cases}1\,;\ \text{解釈 } m \text{ において論理式 } W \text{ は真}\\ 0\,;\ \text{解釈 } m \text{ において論理式 } W \text{ は偽.}\end{cases}$$

□

本稿では特にことわらない限り対数の底は2とする．

例えば，任意の定数記号 a に関して論理式 $W: Q(a) \vee \neg Q(a)$ はトートロジーであり，どのような古典的解釈においても真であるため，この論理式の情報量 $i(T(W|m)=1)$ は 0 bit となる．この情報量は表現形式つまりシンタックスとは独立であり，セマンティクスのみに関する情報量である．つまり知識表現としてホーン節はもちろんのこと意味ネット，フレーム，決定木等を用いた場合でもセマンティク情報量は等しくなる．このセマンティク情報を用いると，例えば演繹可能性についても以下のように情報理論より表現することが可能となる．

補題 2.1[7] 仮説 A から論理式 W が導出される場合，A を与えたもとでの W の条件つきエントロピーは以下の等式をみたす．

$$H(T(W|M)|T(A|m)=1)=0. \quad (3)$$

ここで M は確率変数としての解釈を表している．□

以降，混乱を招く恐れのないときは $T(A|m)=1$ を A と略記する．このように情報理論からの演繹推論を考察することにより，前稿[7]では古典的な演繹推論を拡張し，不確実性を含む演繹推論を提案している．この拡張された演繹推論は本研究でも用いられるため付録Aに概要をまとめておく．

2.2 帰納・演繹推論の情報処理問題との対応

セマンティク情報を用いると，情報理論より帰納推論と演繹推論を総合的にある種の情報処理問題との対応によりモデル化可能である．ここでは著者らの一連の研究におけるモデル化の基本的考え方について簡単にまとめる．帰納推論は個々の事実から全体を説明する知識（規則，仮説）を推論するものであり，個々の情報，基礎情報と呼ぶ，から階層的に一段上のメタの情報である仮説を求めていると考えられる．これと逆の情報変換は演繹推論によって行われ，メタの情報である仮説を用いて基礎情報であるグランド論理式が導出される．第一階述語論理の場合，メタの情報は限定子つきの論理式で表現されていると考えられる[6]．従来，これらの推論は個別に議論されていたが，一連の情報処理過程として情報理論から考察することにより，その目的，機構，評価基準等が明確となる．

帰納推論の目的は帰納推論のみを単独で行う場合を考えると当然，

1) 真の仮説を推測することとなる．

しかし，帰納された仮説を演繹推論で利用するまでを考慮し，図1のように帰納・演繹の一連の推論過程ととらえると以下の2つのタイプの目的が考えられる．

2.1) 帰納された仮説から新たな論理式（まだ観測されていない事実）を演繹（予測）する．

2.2) 帰納された仮説からその帰納推論で用いられたすべての事例を演繹（復号化）する．

これら3種類の推論過程は情報階層の視点から考察すると，情報源符号化や予測の情報処理問題との共通点が浮かび上がってくる．

2.2)はデータを歪みなく圧縮する情報源符号化の問題に対応して考えることができる．情報源符号化の場合，情報源系列の符号化は情報源の確率構造を利用して行われる．この確率構造は，一般的にはパラメタライズされた分布族 $P(X^n|\theta)$ を仮定し，その確率パラメータ θ により表現されており，このパラメータ θ の情報を利用して情報源系列 X^n を理想符号長 $-\log P(X^n|\theta)$ の符号に圧縮していると考えることができる．ここで注意すべき点は，情報源系列 X^n によって

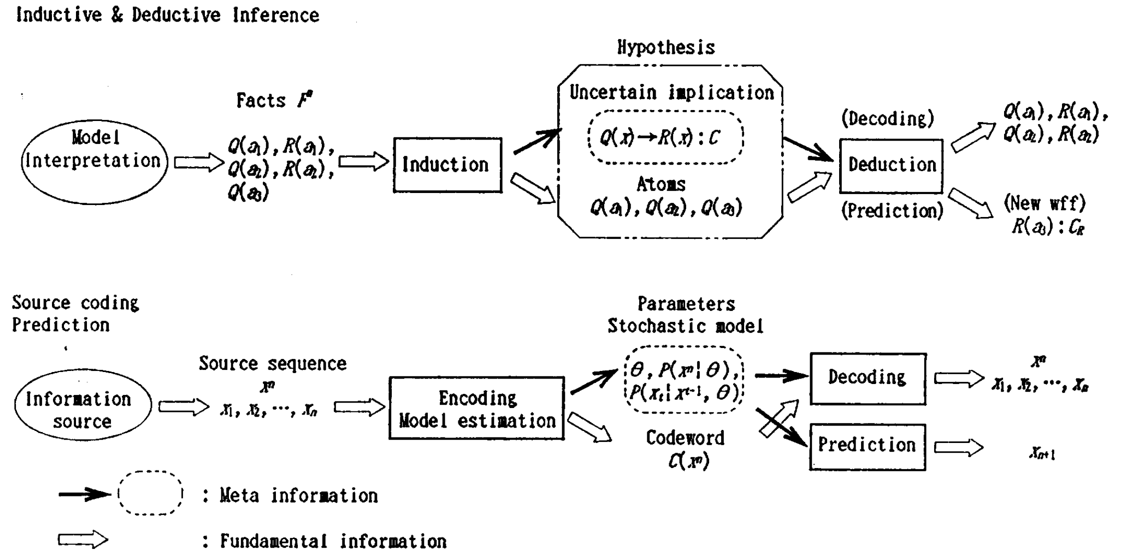


図 1 帰納・演繹推論と情報源符号化，予測の情報変換構造の対応
Fig. 1 Relationship between inductive・deductive inference and source coding・prediction from view point of information structure.

表現され情報と確率パラメータ θ により表現される情報の階層の違いであり，この確率パラメータが表現する情報は先の基礎情報に比べ情報階層が相対的に上位なメタ情報となっている．帰納された仮説が帰納推論の過程で与えられた事実のみを復元する一種のデータベースのような目的で用いられる場合，事実系列がメタの情報である限定子つき論理式により仮説に圧縮して記憶されていると考えることができる．情報源符号化の一般的評価規準は平均符号長や冗長度であり，第5章ではこの規準から推論法の評価を行う．

2. 1)の目的で，帰納した仮説で新たな論理式を演繹することは予測問題と関連して考えることができる．この予測問題においてもメタ情報が重要な役割を演ずる．自己回帰モデル等の時系列データの予測法は，データ系列を発生させる確率過程モデル $P(X_t|X^{t-1}, \theta)$ を仮定し，その確率モデルを用いて $t-1$ 時点までのデータ X^{t-1} から t 時点のデータ X_t を予測することになる．この場合の確率モデルのパラメータ θ 等がメタ情報に対応している．ただし，予測問題の場合はメタ情報として1つの確率モデルだけを用いて次のデータを予測する必要はなく，いくつかの確率モデルからの総合判断として予測を行うことも可能である．これと対応する帰納推論のパラダイムとして，Littlestone の学習モデル[13]がある．これは帰納推論と演繹推論の推論過程を分離せず，一挙に予測が行われているとも考えられる．この推論について予測誤り率を評価規準として第4章で考察を行う．

1)の視点では，データからそのデータが発生する構造を表現したメタ情報であるモデルを決定する問題と対応して考察可能である．真の仮説を仮説空間から選択する決定問題として，決定誤り率の評価規準から第3章で考察する．

このような帰納・演繹の一連の推論と情報源圧縮や予測などの情報処理との関連は定性的に論じられるの

表 1 帰納・演繹推論と情報源符号化，予測の関連
Table 1 Relationship between inductive・deductive inference and source coding・prediction.

	Fundamental information	Meta-information	
	Random variable	Parameters/Hypothesis	Distribution
Source coding (Prediction)	Sequence X^n Symbol X_t	Parameters θ^k	$P(X^n|\theta^k)$ $P(X_t|X^{t-1}, \theta^k)$
Probabilistic world logic	Interpretation m Ground atom $T(Q(a)|m)$	Hypothesis A: $Q(x) \rightarrow R(x) : C$	$P(m|A)$ $P(Q(a)|A)$

みではなく，確率世界理論を導入することによって表1に示すように確率モデル上でも密接に関係づけられる．本論文でもこの関係を用いて上記の3つの目的から不確実性をもつ仮説の帰納，演繹推論問題を考察していく．

3. 真の仮説の選択問題としての定式化

3.1 確率限定子つき論理式の帰納推論

本稿では，従来の論理では帰納される仮説を表現するメタ情報として限定子つき論理式を用いたのに対して付録Aで定義した確率限定子つき論理式[7]を用いることにより，不確実性を含んだ知識も仮説として帰納できる，より一般化された帰納推論について考察する．これは情報源符号化や統計的決定理論の立場からは，情報源あるいは確率モデルとして仮定する分布のクラスが拡張され，メタ情報の記述能力が高まり，より細かい記述が可能となった場合と考えることができる．

問題の本質を変えず論議を明快にするため仮説言語のクラスを確率限定子つき論理式全体ではなく，以下で定義する確率含意ルールのクラスに限定する．これは古典論理の含意を用いたルールに付録Aの式(51)で定義した条件つき信頼率[7]を付加することにより不確実な含意知識を表現できるように拡張したものである．この確率含意ルールのクラスは知識情報処理の応用分野で最も多く用いられる言語のクラスと考えられる．

定義 3.1 次の形式の論理式を確率含意ルールと呼ぶ．

$$(q_1(x)\wedge q_2(x)\wedge\cdots\wedge q_j(x))\rightarrow R(x):\ C. \qquad (4)$$

ここで，q_i は Q_i のリテラルで C は条件つき信頼率である．□

上式は以降では簡略化して以下のように表現する．

$$q_1, q_2, \cdots, q_j\rightarrow R:\ C. \qquad (5)$$

さらに，仮説の中に含まれる論理式は互いに以下に定義する包摂関係にないと仮定する．

定義 3.2 論理式 $B'\rightarrow A'$ は，$A'\sigma\subset A$ かつ $B'\sigma\subset B$ となる代入 σ が存在すれば，論理式 $B\rightarrow A$ を包摂する (*subsume*) と呼ぶ．□

与えられる事実系列としては前稿[6),8)]と同様にグランドアトムがランダムに与えられる場合を考え，n 個までの事実系列を F^n と記す．与えられる事実系列自体が不確実性をもつ場合や他の確率様相[7]を仮説の中に含む帰納推論も考えられるが，本稿では最も基本的なグランドアトムによる事実系列より確率限定子つき論理式を含む仮説を帰納する問題のみを扱うこととする．

3.2 決定問題としての定式化と最適推論法

第2.2節の1)の目的より帰納推論を決定理論的に定式化するため，まず決定関数を考える．前稿で考察した不確実性をもたない仮説の帰納推論問題[6)]では，仮説空間から論理式の集合である仮説を選択する多重決定問題[14)]とみなすことができた．本章の問題においても論理式あるいは含意ルールの集合を選択するまでは前稿とまったく同様の多重決定問題と考えられる．ところが，今回の問題はさらに選択された含意ルールの条件つき信頼率も求めなければならない．よって，個体定数の数が多い場合は多重決定問題と考えることに無理があり，条件つき信頼率の推測に関しては統計理論のパラメータ推定問題と同様に考えるべきであろう．つまり確率限定子つき論理式を含む仮説の帰納推論問題は広義の検定である多重決定問題と複数のパラメータ推定を同時に行う複合的な決定問題と考えられる．正確には含意ルールの集合 A^k とその各ルールの条件つき信頼率 C^k を推測する複合決定問題 (A^k, C^k) と定義される．ここで k は推測した仮説に含まれる含意ルールの数を示す．

不確実性をもたない仮説の帰納推論では多重決定問題として定式化可能であったため，損失関数としては真の仮説と一致している場合は0，不一致の場合は1とする損失関数 V_1[6)] を用いて最適な推論法を考察した．しかし，本研究のような複合決定問題の場合，一意的な損失関数を決めることは難しく，それぞれの損失関数の合成を考えることになる．複数のパラメータに対する同時推定問題の場合，それぞれのパラメータの2乗損失関数の線形結合を用いることが一般的である[15)]．そこでこの問題に対してもこの考え方を拡張して，次のような損失関数を定義する．

定義 3.3 真の仮説 $A^*:(A^{k*}, C^{k*})$ と仮説 $A:(A^k, C^k)$ との損失関数 V_4 は次式のように定義される．

$$V_4(A^*, A)=\begin{cases}(1/k)\displaystyle\sum_{i=1}^{k}(C_i^*-C_i)^2\,; & A^{k*}=A^k\\ 1\,; & A^{k*}\neq A^k.\end{cases} \qquad (6)$$

ここで，C_i は仮説の i 番目の含意ルールの条件つき信頼率である．□

この損失関数 V_4 の意味は，仮説として推論された含意ルールの集合は真の仮説の含意ルールと一致していることが望ましいので，一致していない場合は損失

を1としている．さらに，ルールの形がすべて一致していてもその条件つき信頼率の値が真の値と違っていては困るため，両者の2乗誤差を正規化した値を損失としている．例えば，真の仮説である含意ルール集合 A^*：$\{Q_1, \neg Q_2 \rightarrow R: 0.8, Q_2 \rightarrow R: 0.6\}$ に対して帰納推論した含意ルール集合が A：$\{Q_1, \neg Q_2 \rightarrow R: 0.7, Q_2 \rightarrow R: 0.8\}$ の場合には，損失 $V_4(A^*, A)=0.025$ となる．この損失関数は，前稿の不確実性を含まない仮説の帰納推論で用いた損失関数 V_1 の一般化にもなっている．

次に，損失関数 V_4 の決定問題に対するベイズ解を求める．

定理 3.1 損失関数 V_4 に対する最適な含意ルールの集合 A_O^k は次式のように求められる．

$$A_O^k=\{A^k \mid \max_{A^k}\int\cdots\int P(F^n \mid A^k, C^k) \cdot f(C^k \mid A^k)P(A^k)dC^k\}, \tag{7}$$

ここで，$P(A^k)$ は A^k の事前確率，$f(C^k|A^k)$ は含意ルール A^k が成り立つもとでの条件つき信頼率ベクトル C^k の事前確率密度関数である．

それぞれの含意ルールの条件つき信頼率 C_{iO} は次のように求められる．

$$C_{iO}=\int C_i f(C_i \mid A_O^k, F^n)dC_i, \tag{8}$$

ここで，

$$f(C_i \mid A_O^k, F^n)=\frac{P(F^n \mid A_O^k, C_i)f(C_i \mid A_O^k)}{\int P(F^n \mid A_O^k, C_i)f(C_i \mid A_O^k)dC_i}. \tag{9}$$

（証明） ベイズ多重決定と2乗誤差を損失関数としたベイズ推定の結果を組み合わせることにより求められる[15),16)]． □

含意ルール集合の決定は，ある k 個の含意ルールと事実集合 F^n の結合確率が最大となる，つまり事後確率が最大となる集合が選択される．さらに，それぞれの含意ルールの条件つき信頼率はその含意ルールが選ばれたもとでの事後確率による期待値が用いられる．

系 3.1 定理3.1の帰納推論法により含意ルールを推測する平均誤り確率 $Er(A_O^k)$ は次式で示される．

$$Er(A_O^k)=1-[\{\max_{A^k}\int\cdots\int P(F^n \mid A^k, C^k)f(C^k \mid A^k) \cdot P(A^k)dC^k\}/\{\sum_{A^k}\int\cdots\int P(F^n \mid A^k, C^k) \cdot f(C^k \mid A^k)P(A^k)dC^k\}]. \tag{10}$$

また，真の含意ルールが A_O^k であるもとで，定理3.1の帰納推論法により推測された条件つき信頼率 C_{iO} の真の値との平均2乗誤差 $Er^2(C_{iO})$ は次式となる．

$$Er^2(C_{iO})=\int(C_i-C_{iO})^2 f(C_i \mid A_O^k, F^n)dC_i. \tag{11}$$

□

例 3.1 定理3.1の帰納推論法の一部を簡単な例で示す．ある含意ルール A_i の条件つき信頼率 C_i の事前確率密度関数 $f(C_i|A_i)$ として次のようなベータ分布を仮定する．

$$f(C_i \mid A_i)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}C_i^{a-1}(1-C_i)^{b-1}, \quad 0<C_i<1. \tag{12}$$

定数記号の総数がある程度以上の大きさで，与えられた事実集合 F^n において，含意ルール A_i の前件（ボディ部）を充たす事実の数が $n(q)$，前件と後件（ヘッド部）の両方を充たす事実の数は $n(r, q)$ であった場合，含意ルール A_i の事後確率 $P(A_i|F^n)$ は以下となる．

$$P(A_i \mid F^n)=[\{\Gamma(a+b)\Gamma(a+n(r,q))\Gamma(b+n(q)-n(r,q))\}/\{\Gamma(a)\Gamma(b)\Gamma(a+b+n(q))\}]P(A_i). \tag{13}$$

ここで $P(A_i)$ は A_i の事前確率．

この含意ルールの条件つき信頼率 C_{iO} は定理3.1より次式のように帰納される．

$$C_{iO}=\int C_i f(C_i \mid A_i, F^n)dC_i=\frac{n(r,q)+a}{n(q)+a+b}, \tag{14}$$

次にこの推論による平均2乗誤差 $Er^2(C_{iO})$ は次式となる．

$$Er^2(C_{iO})=\frac{u(1-a/u)}{(u+n(q))(u+1)}, \tag{15}$$

ここで，$u=a+b$． □

このベータ分布は条件つき信頼率の事前分布として非常によい性質をもっていることが示せる．まず，$a=b=1$ とすることにより事前分布は一様分布となり，各信頼率の出現を均等と考えている場合の推論となる．さらに，$a=b=1/2$ とすることにより，第5章で考察する損失関数 V_2 をミニマックス基準で最適化する事前分布にもなる．その他，詳しくは文献17)を参照されたい．

つぎに，まったく別の視点から損失関数を考えてみる．本章では帰納推論を真の仮説を推論する目的から考察しているため，損失の測度としては上述のように

推論した仮説が真の仮説と一致しているかを論理式の違いから直接的に評価することを考えた．別の視点として，真の仮説が表現している解釈（世界）のパターンと推論された仮説が表現している解釈のパターンとの類似性で推論を評価することが考えられる．仮説が表現している解釈のパターンの類似性とは，言い換えればその仮説から導出☆できる論理式の集合の類似性であり，これを仮説間の類似度あるいは距離として損失関数を定義してみる．以上の考え方から2つの仮説間の距離としては，それぞれの仮説のもとで各解釈の生起確率の Kullback 情報量を用い，以下のような損失関数を定義する．

定義 3.4　真の仮説 $A^*:(A^{k*}, C^{k*})$ と仮説 A との損失関数 V_5 を次式のように定義する．

$$V_5(A^*, A)=\sum_m P(m|A^{k*}, C^{k*})\frac{P(m|A^{k*}, C^{k*})}{P(m|A)} \tag{16}$$

□

この Kullback 情報量は階層的なパラメータ構造をもちパラメータ数が異なる確率モデル間の距離を測る尺度として有効であり，本章の帰納推論問題のように含意ルールとその条件つき信頼率によって構成される仮説間の距離としても有用な測度と考えられる．

定理 3.2　損失関数 V_5 に対するベイズ解は次式で与えられる．

$$P(m|A_O)=\sum_{A^k}\int\cdots\int P(m|F^n, A^k, C^k)\cdot f(C^k|A^k, F^n)P(A^k|F^n)dC^k. \tag{17}$$

□

しかし，このベイズ解で得られる仮説 A_O は仮定した仮説言語である確率含意ルールでは一般的には表現できない．これは仮説空間の構造の問題であり，詳しくは次章で考察を行う．また，この事後確率で重み付け平均された仮説は，次章で考察する予測のための最適な推論法とも密接な関係がある．

4. 予測に用いる仮説の選択問題としての定式化

この章では第2.2節で述べた2.1)の目的より，新たな論理式を予測する推論問題を取り扱う．前章と類似のアプローチ法をとり，統計的決定問題として新たな論理式の真理値あるいは信頼率を予測する問題として定式化を試みる．前稿ではこの予測問題を次の2つの問題設定から考察した．

1) 直接的な予測：事例から直接新たな論理式の真理値を予測する．

2) 間接的な予測：事例から帰納された1つの仮説を用いて予測する．

第2.2節で予測は帰納推論と演繹推論を組み合わせた一連の情報処理過程とみなせることを述べたが，1)は帰納と演繹を結合して1つの情報処理として行っているとも考えられる．従来研究の Littlestone によって提案された学習モデル[13]はこの問題設定の中に含まれる．この学習モデルでは，帰納によって1つの仮説を選択するのではなく，多数の仮説を用いて予測を行う問題設定となっている．

他方，2)の問題設定は帰納推論と演繹推論を完全に分離して行うものである．まず，帰納推論により事実から1つの仮説を求め，その次にその仮説を用いて演繹推論を行うことによって，まだ事実として得られていない新たな論理式を推論する方法である．従来研究では Quinlan と Rivest の予測のための決定木の決定問題[18]がこの問題設定の1つの例と考えられる．

本章では，この2つの問題設定から新たな論理式の導出を考察する．

4.1 直接的予測問題

まず，直接的予測の問題設定で考察を行う．前稿[8]の不確実性をもたない論理式を仮説として用いた予測では，論理式の導出をその論理式の真理値 $\{0,1\}$ を決定する問題としてとらえ，損失関数としては後で述べる V_3 の絶対誤差を用いて考察した．本研究の場合，新たな論理式の予測結果としては，論理式の真偽のみでなくその信頼率も推定する．この推論された信頼率に対する損失を考察することも重要であるので，以下では真理値の決定問題とした場合と信頼率の決定を中心に考えた場合の2つの視点から考察を行っていく．

4.1.1 論理式の真理値を推定する問題として

まず前稿に準じて，事実集合 F^n からグランドアトム G の真理値を推定する直接的な予測問題として定式化を行う．決定関数 $D_{T(G)}(F^n)$ は事実集合 F^n から G の真理値 $T(G|m)\in 0,1$ への写像であり，損失関数 $V_3(T(G|m), D_{T(G)}(F^n))$ は前稿と同様に次式で与えられる．

$$V_3(T(G|m), D_{T(G)}(F^n))=|T(G|m)-D_{T(G)}(F^n)|. \tag{18}$$

さらに，損失関数 $V_3'(T(\cdot|m), D_T(F^n))$ も前稿と同様に次式で定義される．

☆ ここでの導出は前稿[7]で提案し，付録Aでもふれた拡張された演繹推論による．

$$V_3'(T(\cdot\,|m), D_T(F^n)) = \frac{1}{|S_G|}\sum_{G\in S_G}|T(G|m)-D_{T(G)}(F^n)|. \quad (19)$$

ここで，S_G はすべてのグランドアトムの集合.

この危険関数 $Rk_3(T(\cdot\,|A^k, C^k), D_T)$ は次式で与えられ，予測に関する平均誤り確率として意味づけられる.

$$Rk_3(T(\cdot\,|A^k, C^k), D_T) = \sum_m V_3'(T(\cdot\,|m), D_T(F^n))P(m|A^k, C^k). \quad (20)$$

定理 4.1 この直接的な予測問題として平均予測誤り確率最小の推論 $S_{T(G)O}(F^n)$ はベイズ解として次式で与えられる.

$$D_{T(G)O}(F^n)=\begin{cases}0;\ \sum_m T(G|m)P(m|F^n)<1/2\\ 1;\ \sum_m T(G|m)P(m|F^n)>1/2,\end{cases} \quad (21)$$

ここで，

$$P(m|F^n)=\sum_{A^k}\int\cdots\int P(m|F^n, A^k, C^k)\cdot f(C^k|A^k, F^n)P(A^k|F^n)dC^k. \quad (22)$$

$\sum_m T(G|m)P(m|F^n)=1/2$ の場合は確率 1/2 のランダム決定を行うこととする. □

定理 4.1 は，最適な予測は含意ルールとその条件つき信頼率の事後確率によるグランドアトム G の真理値の期待値により決定を行うことを意味している. これは各仮説の予測結果を重み付けして推論する Littlestone の weighted majority アルゴリズム[13]と定性的には類似のアルゴリズムとなっている☆.

系 4.1 定理 4.1 で推論された論理式 G の信頼率 $C(G)$ は次式となる.

$$C(G)=\sum_m\sum_{A^k}\int\cdots\int T(G|m)P(m|F^n, A^k, C^k)\cdot f(C^k|A^k, F^n)P(A^k|F^n)dC^k. \quad (23)$$

□

直接的予測の最適推論法で推論された論理式の信頼率 $C(G)$ はこの予測が正しい確率と一致している. これについては間接的予測の節でさらに詳しく述べることとする.

系 4.2 定理 4.1 の推論は次式の推論と同等となる.

$$D_{T(G)O}(F^n)=\begin{cases}0;\ \sum_m\sum_{A^k} T(G|m)\cdot P(m|F^n, A^k, C_O^k)P(A^k|F^n)<1/2\\ 1;\ \sum_m\sum_{A^k} T(G|m)\cdot P(m|F^n, A^k, C_O^k)P(A^k|F^n)>1/2,\end{cases} \quad (24)$$

ここで，C_O^k のそれぞれの条件つき信頼率 C_{iO} は次のように求められる.

$$C_{iO}=\int C_i f(C_i|A^k, F^n)dC_i, \quad (25)$$

条件式の値が 1/2 の場合は確率 1/2 のランダム決定を行うこととする.

(証明) 付録 B. □

これは式(25)の事後確率による平均の信頼率を各含意ルールの信頼率とした仮説を用い，拡張された演繹推論を行った結果をそれぞれの仮説の事後確率で平均した総合値で予測を行うことを表している.

4.1.2 論理式の信頼率を推定する問題として

ここまでは，新たな論理式の予測結果として，論理式の真偽に注目して損失関数を考えてきたが，以下からは推論された信頼率に対する損失について考察する. 事実集合 F^n から目的とするグランドアトム G の信頼率 $C(G)$ を予測する決定関数 $D_{C(G)}(F^n)$ と定義する. 信頼率のような実数に対する損失関数としては統計理論のパラメータ推定問題で用いられる 2 乗誤差が挙げられる.

定義 4.1 推論した信頼率の損失関数として V_6 を定義する.

$$V_6(C^*(G), D_{C(G)}(F^n))=(C^*(G)-D_{C(G)}(F^n))^2. \quad (26)$$

ここで，$C^*(G)$ は G の真の信頼率で真の仮説 (A^k, C^k) より以下のように定義される.

$$C^*(G)=\sum_m T(G|m)P(m|A^k, C^k). \quad (27)$$

□

この損失関数 V_6 を用い上記の V_3 の場合と同様な展開で危険関数を定義し最適解を求める.

定理 4.2 損失関数 V_6 に対する最適な信頼率の予測であるベイズ解 $D_{C(G)O}(F^n)$ は次式となる.

$$D_{C(G)O}(F^n)=\sum_m\sum_{A^k}\int\cdots\int T(G|m)P(m|F^n, A^k, C^k)\cdot f(C^k|A^k, F^n)P(A^k|F^n)dC^k. \quad (28)$$

□

定理 4.2 に示された最適な信頼率の予測は，系 4.1 で示した定理 4.1 の最適な真理値の予測を行った場合

☆ この weighted majority アルゴリズムは学習モデルとしては不確実性を含まない仮説に対する推論であり，mistake bounds と呼ばれる累積誤り回数で評価を行っている. このような累積損失からの定理 4.1 のアルゴリズムの評価については第 4.1.2 項で詳しく考察を行う.

の信頼率と一致しており，結局，2つの損失関数 V_3, V_6 のどちらに対しても最適な推論は定理 4.1 となり，実際の計算には系 4.2 が用いられることとなる．これは著者らの一連の研究において，論理式の不確実性を表す測度として信頼率を帰納と演繹の両推論において統一的に用いていることにも関連しており，論理式の真理値を推定する問題として，また，信頼率を推定する問題として帰納・演繹の一連の推論を考えた場合に最適な推論法が一致するという好ましい結果が得られた．

さらに，次のような損失関数 V_7 を考えてみても，最適な推論法は定理 4.1，4.2 と同様になる．

$$V_7(T(G|m), D_{C(G)}(F^n))$$
$$=\begin{cases}-\log D_{C(G)}(F^n)\,;\ T(G|m)=1\\ -\log D_{C(\neg G)}(F^n)\,;\ T(G|m)=0.\end{cases} \tag{29}$$

この損失関数は予測対象のグランドアトムが真理値が真の時，真であると予測した場合の信頼率の負の対数値を損失と考えている*．この V_7 の危険関数 Rk_7 を考えると，さらにこの損失関数の意味が明らかとなる．

$$Rk_7(T(G|A^k, C^k), D_{C(G)}(F^n))$$
$$=\sum_m V_7(T(G|m), D_{C(G)}(F^n))P(m|A^k, C^k)$$
$$=-\sum_m \sum_{g=G,\neg G} T(g|m)P(m|A^k, C^k)$$
$$\cdot\log D_{C(g)}(F^n). \tag{30}$$

式(30)はグランドアトム G の真の信頼率と推論により得られた信頼率との違いを Kullback 情報量で測った場合の分子の部分に対応している．つまり，真の信頼率と予測信頼率の距離を2乗距離で測った場合が損失関数 V_6 であり，Kullback 情報量で測った場合が損失関数 V_7 となる．

定理 4.3 損失関数 V_7 のベイズ解，つまり平均的対数予測誤差を最小化する推論は定理 4.1 の推論と等しくなる． □

損失関数 V_7 は確率的規則の学習の研究[12]において対数予測誤差として定義されているものと同等で，この研究では Littlestone の学習モデルと同様に損失の累積[13]で推論法を評価している．ここではこれらの研究で用いられている概念学習の予測モデルの枠組みにとらわれず，さらに一般化した次のような推論過程を考え，累積損失の評価を行う**．推論過程を逐次的に予測のサイクルを繰り返す過程と定義する．t 時点の予測サイクルは，まず今までに与えられた事実集合 F^{t-1} を用い，ランダムに決められた次の事実* F_t に対する真理値とその信頼率の予測 $D_{C(F_t)}(F^{t-1})$ を行い，つぎに正解の真理値が reinforcement として与えられる．このような推論過程の n 時点までの累積損失 V_7' は次式で定義される．

$$V_7'(T(\cdot|m), D_{C(\cdot)}(F^n))$$
$$=\sum_{t=1}^{n} V_7(T(F_t|m), D_{C(F_t)}(F^{t-1})) \tag{31}$$

このような累積的な損失 V_7' の考え方においても，平均損失に対して最適な定理 4.1 の推論法は有効である．

補題 4.1[17] 損失関数 V_7 を用いた累積損失 V_7' に対しても定理 4.1 の推論法はベイズ解として最適となる． □

この累積損失は次章の記述長からみた推論の問題と密接な関連があり，詳しくは次章で論じることとする．

このように V_3, V_6, V_7 の3つの損失関数から考察を行い，すべての場合に定理 4.1 の推論法が最適であることが示された．このように様々な視点からみて最適な推論法は，基本的で重要な推論法と考えられる．

4.2 間接的予測問題

仮説空間が凸集合の場合は確率的に重みづけられた仮説も仮説空間上に存在し，仮定した仮説言語で表現することが可能であり，定理 4.1 の結果が間接的予測にも利用できる**．つまり，式(17)と式(22)が等しいことより，定理 3.2 の仮説を用いて拡張された演繹推論を行う間接的予測と定理 4.1 の最適な直接的予測が同等になり***，定理 3.2 の仮説が間接的予測のために最適となる．

しかし，先に述べたように一般的には仮説空間は凸集合ではなく，本研究の確率限定子つき論理式を仮説言語とした場合にも仮説空間は凸集合とはならず，定理 4.1 の結果は直接的予測（仮説を複数個用いてもよい問題設定等）で有効となり，間接的予測にはそのまま利用できない．そこで，この定理 4.1 の最適解を間接的予測の問題設定上で見直すことにする．また，この節では真理値を推論する問題としての観点を中心に考察を行う．

この対象とする予測問題は帰納された確率限定子つ

---

* 偽の場合も同様に定義される．
** Littlestone と同様の概念学習の問題としての問題設定での評価については第6章で考察を行っている．

* 真理値は与えられていない．
** 計算論的学習理論の様々なパラダイムの中には凸集合の仮説空間を仮定することが可能なパラダイムもあるので，その場合には定理 4.1 の結果が間接的予測にもそのまま適用できる．
*** Kullback 情報量で仮説間の距離（損失）を測る1つの意味として，予測誤り確率を評価しているのと同等であると考えることができる．

き論理式を含む仮説 (A^k, C^k) を用いて任意の新たな論理式 G を間接的に予測する問題と考えられ，実際の予測の決定関数 $D'_{T(G)}(F^n)$ は事実集合 F^n から帰納された仮説 $A(F^n):(A^k(F^n), C^k(F^n))$ を用いて付録Aで述べた拡張された演繹推論を次式のように行うこととなる．

$$D'_{T(G)}(F^n, A(F^n)) = \begin{cases} 0;\ \sum_m T(G|m)P(m|F^n, A(F^n)) < 1/2 \\ 1;\ \sum_m T(G|m)P(m|F^n, A(F^n)) > 1/2, \end{cases} \quad (32)$$

ここで，$P(m|F^n, A(F^n)) = P(m|A^k(F^n), C^k(F^n), F^n)$，$A^k(F^n)$ と $C^k(F^n)$ は F^n から帰納された含意ルール集合とその条件つき信頼率である．

$\sum_m T(G|m)P(m|F^n, A(F^n)) = 1/2$ の場合は確率 1/2 のランダム決定を行うこととする．

最適な定理 4.1 の直接的予測は事後確率で平均した含意ルール集合で演繹推論を行うのと同等であったので，間接的予測では事後確率最大の次式の含意ルール集合を近似解として用いることが考えられる．

$$A_O^k = \{A^k | \max_{A^k} \int \cdots \int P(F^n|A^k, C^k) \cdot f(C^k|A^k)P(A^k)dC^k\}. \quad (33)$$

定理 4.4 含意ルールの集合 A_O^k が決まったもとで，新たな論理式の予測に最適なそれぞれの条件つき信頼率 C_{iO} は次式で与えられる．

$$C_{iO} = \int C_i f(C_i | A_O^k, F^n)dC_i. \quad (34)$$

（証明） 系 4.2 の証明と同様な展開で証明される．

□

系 4.3 真の仮説選択のための損失関数 V_4 に対して最適な含意ルール集合とその条件つき信頼率によって構成される仮説は，新たな論理式の予測に対して用いられた場合にも最適な予測の近似解となっている．

□

ある論理式 W の真理値の予測は上記の確率含意ルール集合 (A_O^k, C_O^k) を用い拡張された演繹推論により行われるが，推論結果として得られる論理式 W の信頼率 $C(W)$ は論理式 W が真となる確率，以後信頼係数と呼ぶ，を表しているわけではない☆．これは付録Aで述べた信頼率の定義からも明らかであり，仮説である確率含意ルールの集合が真の仮説であるときのみ信頼率と信頼係数は一致する☆☆．より正確には，事実 F^n から帰納された確率含意ルール集合 (A_O^k, C_O^k) により演繹された論理式 W の信頼率 $C(W)$ と W の信頼係数 CR は次式の関係となる．

$$CR = C(W)P(A_O^k|F^n) \quad (35)$$

また，この確率含意ルール集合 (A_O^k, C_O^k) のみを用いた演繹では真か偽か不明な確率は $1-P(A_O^k|F^n)$ となり，以後この確率を無知係数 IR と呼ぶ．

系 4.4 式(33)，式(34)によって得られた仮説を用い，拡張された演繹推論により論理式 W を真と予測した場合の信頼係数 CR は次式で与えられる．

$$CR = \sum_m \int \cdots \int T(W|m)P(m|F^n, A_O^k, C^k) \cdot f(C^k|A_O^k, F^n)P(A_O^k|F^n)dC^k \quad (36)$$

□

5. 記述長最小の仮説の選択問題としての定式化

この章では第 2.2 節で述べた2.2)の目的のもとで帰納推論を情報源符号化の視点から事実系列 F^n のセマンティク情報を仮説に圧縮して記憶する問題としてとらえ，F^n の仮説への符号化の問題として考察を行う．符号化を F^n に対する Kraft の不等式[9]を充たす符号長の与え方として決定関数 $D_L(F^n)$ を定義し，前稿[8]と同様の損失関数 V_2 を用い考察する．決定関数 D_L に対する危険関数 $Rk_2((A^k, C^k), D_L)$，つまり平均符号長は次式のように定義される．

$$Rk_2((A^k, C^k), D_L) = \sum_{F^n} P(F^n|A^k, C^k)D_L(F^n). \quad (37)$$

定理 5.1[17] 損失関数 V_2 のベイズ解である最適な符号化法 D_{LO} は次式のように与えられる．

$$D_{LO}(F^n) = -\log \sum_{A^k} \int \cdots \int P(F^n|A^k, C^k) \cdot f(C^k|A^k)P(A^k)dC^k. \quad (38)$$

□

この定理は事実系列を圧縮して記憶するためには，各含意ルールとその条件つき信頼率からの事実系列の生起確率を事前確率で平均化した確率を用いて符号化することが最適となることを示している．これと類似の解としては損失関数 V_5 に対する定理 3.2 と前章で述べた直接的な予測問題としての定理 4.1 の推論法があり，事後確率で平均化したものが最適として求められる．

しかし，前述したように一般的には帰納推論問題においてこれらの解はそのままでは利用することができ

☆ 前述したように直接的予測の場合は信頼率と信頼係数は一致する．

☆☆ 帰納された不確実性を含まない仮説を用いた演繹の信頼係数については文献 8) で詳しい考察を行っている．

ない．情報源符号化の問題では圧縮が目的であり，メタ情報を表現する空間として凸集合である確率分布空間を用いることが可能なため，符号化のためのメタ情報として事前確率で平均化した分布を用いることができる．ところが，帰納推論の目的は与えられた事実を単に圧縮して記憶していることにあるのではなく，仮説言語を用いて表現される仮説の形で知識（メタ情報）を得ることにある．例えば，帰納した知識を Prolog で書かれたプロダクションシステムで利用したい場合には，ホーン節で表現された仮説が帰納されることが望ましい．そのため，事前確率で重みづけられた含意ルールの集合の様に仮定した仮説言語では表現不可能なメタ情報を用いることは，帰納推論の目的から意味がない場合もあり得る．

帰納推論において単に記述長を短くして与えられた事実集合を記憶しておくという目的はあまり重要でなく，そのような目的での最適推論法も仮説空間が凸集合でない場合には直接的に用いることはできない．しかし，記述長という測度がいままで述べてきたような各種の目的に対する最適推論法を非常にうまく特徴づけていることは前稿[6),8)]で詳しく述べた．本稿でも前章の直接的予測問題としての定式化で損失関数 V_7 の累積損失 V_7' を考えた場合に，上記の記述長と非常に密接な関係がある．

系 5.1 定理 4.1，4.2 の最適な推論を行った場合 n 時点までの累積損失 V_7'，つまり最小の累積対数損失は，定理 5.1 の推論法により事実系列 F^n を最も圧縮して記憶した場合の記述長と等しい．

$$V_7'(T(\cdot\mid m), D_{C(\cdot)O}(F^n))=D_{LO}(F^n) \tag{39}$$

□

このように前章で定義した対数損失 V_7' の情報源符号化からみた意味は，逐次符号化を行った場合の符号長を表していることがわかる．この逐次符号化においてベイズ的に最短の符号長と事実系列 F^n を一まとめで符号化した場合の最短の符号長は等しいことが証明されている[17)]．

本研究で扱っている確率限定子つき論理式の帰納推論問題の情報源符号化の視点からの特徴は，情報を圧縮するために重要な役割をもつメタ情報が階層的構造をもっていることにある．情報源符号化の分野では，これと類似した問題の例として k 次マルコフ過程を情報源と仮定した場合等がある．この確率限定子つき論理式によって表現された仮説のセマンティク情報に関する記述長は古典論理の限定子つき論理式で記述された仮説の記述長と比べ短くなっている．これは情報理論において情報源として独立なベルヌーイ系列を仮定するより k 次マルコフ系列を仮定した方が一般的に圧縮率は向上することと対応している．古典論理の限定子つき論理式のみでは仮説として圧縮して表現できない事実集合も確率限定子つき論理式を用いることで仮説として表現できることになる．また，古典論理では予測できなかった論理式も確率限定子つき論理式を用いた仮説では精度良く予測が可能になる．

6. 確率含意ルールを用いた概念学習への応用

概念学習は帰納推論の重要な一分野であり，様々なパラダイムが存在する．概念学習とは，第一階述語論理で仮説（概念）を記述する場合，属性を表現するリテラルを論理記号で結合した論理式を前件として，概念を表現するリテラルを後件として含意するような形の論理式を仮説言語とする帰納推論と定義される．

本章で議論する概念学習と従来の一般の概念学習との違いは仮説言語が不確実性も表現できる定義 3.1 の確率含意ルールを用いることである．本章ではこの不確実性をもつ概念（仮説）の学習について，与えられた事実の集合から，不確実性も表現できる含意ルール集合 A^k とその条件つき信頼率 C^k を帰納する問題として，また新しい事例を予測する問題として考察を行う．

この章で考察する概念学習問題 $PIR(K, J)$ を定義する．確率含意ルールは，ある概念を表す述語 R のリテラルによる後件（ヘッド部）と J 種類の属性を表す述語から任意の個数のリテラルを用いた前件（ボディ部）から成る．仮説空間は，この確率含意ルールを高々 K 個用いた仮説により構成される．事例 E_i としてはある定数記号 a_i に関して学習対象の概念 R のリテラル r と，説明に使われる J 種類の属性 $Q_1, Q_2, \cdots, Q_J$ のリテラル $q_1, q_2, \cdots, q_J$ が以下の $J+1$ 重対で与えられる．

$$E_i=\{r(a_i), q_1(a_i), q_2(a_i), \cdots, q_J(a_i)\}. \tag{40}$$

よって，概念学習とは n 個の事例の集合 E^n：$\{E_1, E_2, \cdots, E_n\}$ から上記の確率含意ルールの集合を帰納推論する問題と定義される．

真の仮説を選択する問題として損失関数 V_4 のもとで最適な帰納推論法は定理 3.1 より次式の含意ルール集合 A_O^k とそれぞれの信頼率 C_{iO} を推論することとなる．

$$A_O^k=\{A^k|\max_{A^k}\int\cdots\int P(E^n|A^k,C^k)\cdot f(C^k|A^k)P(A^k)dC^k\},\qquad(41)$$

$$C_{iO}=\int C_if(C_i|A_O^k,E^n)dC_i,\qquad(42)$$

ここで，

$$f(C_i|A_O^k,E^n)=\frac{P(E^n|A_O^k,C_i)f(C_i|A_O^k)}{\int P(E^n|A_O^k,C_i)f(C_i|A_O^k)dC_i}\qquad(43)$$

次に，予測の問題を考える．まず，ある定数記号 b に関する概念 R の真理値 $T(R(b)|m)\in 0,1$ を事例集合 E^n と b の属性 $Q^J(b)$: $\{q_1(b),q_2(b),\cdots,q_J(b)\}$ から予測する直接的な予測問題から考察する☆．この決定関数を $D_{T(R(b))}$ とすると，損失関数 V_3 と V_3' は第4章と同様に次式となる．

$$V_3(T(R(b)|m),D_{T(R(b))}(Q^J(b),E^n))=|T(R(b)|m)-D_{T(R(b))}(Q^J(b),E^n)|.\qquad(44)$$

$$V_3'(T(R(\cdot)|m,D_T(Q^J(\cdot),E^n))=\frac{1}{|S_b|}\sum_{b\in S_b}|T(R(b)|m)-D_{T(R(b))}(Q^J(b),E^n)|.\qquad(45)$$

直接的予測の最適推論法は定理4.1より次式となる．

$$D_{T(R(b))}(Q^J(b),E^n)=\begin{cases}0;\ \sum_m T(R(b)|m)P(m|Q^J(b),E^n)<1/2\\1;\ \sum_m T(R(b)|m)P(m|Q^J(b),E^n)>1/2.\end{cases}\qquad(46)$$

ここで，$P(m|Q^J(b),E^n)=\sum_{A^k}\int\cdots\int P(m|Q^J(b),E^n,A^k,C^k)f(C^k|A^k,E^n)P(A^k|E^n)dC^k$，$\sum_m T(R(b)|m)P(m|Q^J(b),E^n)=1/2$ の場合は，確率1/2のランダム決定を行う．

この推論法が損失関数 V_6,V_7 に関しても最適で，Littlestone の mistake bound に準じた評価基準である累積損失 V_7' に対しても最適であることは定理4.2，補題4.1からも明らかであろう．

次に間接的な予測を考える．決定関数 $D'_{T(R(b))}(Q^J(b),E^n)$ は帰納された確率含意ルールの集合 $A^k(E^n)$，$C^k(E^n)$ を用いた拡張された演繹推論となり，以下のような方法を取ることになる．

$$D'_{T(R(b))}(Q^J(b),E^n)=\begin{cases}0;\ \sum_m T(R(b)|m)P(m|Q^J(b),E^n,A^k,C^k)<1/2\\1;\ \sum_m T(R(b)|m)P(m|Q^J(b),E^n,A^k,C^k)>1/2,\end{cases}\qquad(47)$$

☆ これは Littlestone の概念学習モデルを不確実性を含んだ問題へ拡張したモデルとも考えられる．

表2 含意ルールの事前確率と事後確率
Table 2 Prior probability and posterior probability of implication rules.

Hypotheses A^k	Prior probability $P(A^k)$	Posterior probability $P(A^k\|E^{20})$
$Q_1\to R:C$ $\neg Q_1\to R:C$	2.38×10^{-1}	2.18×10^{-1}
$Q_1\to R:C$ $\neg Q_1,Q_2\to R:C$ $\neg Q_1,\neg Q_2\to R:C$	1.16×10^{-1}	7.25×10^{-2}
$\neg Q_1\to R:C$ $Q_1,Q_2\to R:C$ $Q_1,\neg Q_2\to R:C$	1.16×10^{-1}	2.23×10^{-1}
$Q_2\to R:C$ $\neg Q_2\to R:C$	2.38×10^{-1}	8.32×10^{-2}
$Q_2\to R:C$ $Q_1,\neg Q_2\to R:C$ $\neg Q_1,\neg Q_2\to R:C$	1.16×10^{-1}	2.08×10^{-2}
$\neg Q_2\to R:C$ $Q_1,Q_2\to R:C$ $\neg Q_1,Q_2\to R:C$	1.16×10^{-1}	3.00×10^{-1}
$Q_1,Q_2\to R:C$ $Q_1,\neg Q_2\to R:C$ $\neg Q_1,Q_2\to R:C$ $\neg Q_1,\neg Q_2\to R:C$	5.66×10^{-2}	8.50×10^{-2}

表3 事例 E^{20}
Table 3 Set of facts E^{20}.

Type of examples	$R(a_i)$	$\neg R(a_i)$
$Q_1(a_i),Q_2(a_i)$	0	4
$Q_1(a_i),\neg Q_2(a_i)$	2	7
$\neg Q_1(a_i),Q_2(a_i)$	4	5
$\neg Q_1(a_i),\neg Q_2(a_i)$	1	4

$\sum_m T(R(b)|m)P(m|Q^J(b),E^n,E^n,A^k,C^k)=1/2$ の場合は，確率1/2のランダム決定を行う．

このモデルのもとで，近似的に予測誤差を最小化する仮説としては系4.3より，式(41)，(42)の含意ルールの集合 A_O^k と条件つき信頼率 C_O^k が推論される．

例 6.1 簡単な例として $PIR(4,2)$ を考えてみる．仮説空間に含まれる含意ルール集合 A^k のタイプは表2で示した7種類で，それぞれの事前確率を表2のように仮定した．また，それぞれの含意ルール A_i の条件つき信頼率 C_i の事前確率密度 $f(C_i|A_i)$ は例3.1で述べた一様分布と仮定した．

表3に示すような20個の事例 E^{20} が与えられた場合，各含意ルール集合 A^k の事後確率 $P(A^k|E^{20})$ は

表2のように求まる．まず，真の仮説を推測する問題から帰納推論を考え，損失関数 V_4 に対して最適な仮説は定理3.1と表2より，以下の確率含意ルール集合となる．

$$\begin{aligned} \neg Q_2 \rightarrow R &: 3.08\times10^{-1} \\ Q_2, Q_1 \rightarrow R &: 2.00\times10^{-1} \\ Q_2, \neg Q_1 \rightarrow R &: 8.33\times10^{-1} \end{aligned} \tag{48}$$

次に，予測の問題を考える．b の属性 $Q^2(b)$：$\{\neg Q_1(b), Q_2(b)\}$ が与えられたもとで $R(b)$：C を推論する直接的な予測問題では，損失関数 V_3'，V_6 等のもとで最適な推論は $R(b)$：6.41×10^{-1} となる．間接的予測では式(48)の確率含意ルールを用いて拡張された演繹推論を行うこととなる．推論結果は $R(b)$：7.14×10^{-1} で信頼係数 $CR=2.14\times10^{-1}$，無知係数 $IR=7.00\times10^{-1}$ となる．□

7. おわりに

従来の無矛盾性の基準によるアルゴリズムでは取り扱えない確率限定子つき論理式をメタ情報として用いた仮説の帰納推論について考察した．帰納推論の3つの目的に対していくつかの損失関数を定義し，それぞれについて最適推論法を提案した．

情報理論の視点からは，これらの最適な推論法は事実系列に含まれるセマンティク情報をそれぞれの目的に対して有効に利用した推論法となっている．例えば，真の仮説を選択する目的での最適推論法は，事実系列に含まれる仮説に関する情報のみを効率的に引き出していると意味づけられる．また，新たな論理式の予測を目的とした推論は，その新たな論理式の真理値に関する情報を事実系列から最大限抽出する推論となっている．

このように本稿ではいくつかの視点から最適な推論法を示したが，効率の良いアルゴリズムについては論じられなかった．一般的には全数探索を行い最適な仮説を選択することになるが，個別の問題では仮説のクラスの特別な関係や事前確率の条件により効率的な探索アルゴリズムが構成できる場合がある[6]．この点については今後の重要な研究課題と考えられる．

著者らの一連の研究では知識情報処理で用いられる様々な推論に対する統一的基礎体系をまず提案し，その基礎体系上で帰納推論への拡張と不確実性を含む演繹推論法への拡張を行っているため，本稿で扱った両方向への総合的拡張も容易に整合性良く行うことができた．その他，質問を許す学習等の様々な学習のパラダイムに対してもこの基礎体系は有効と考えられる．

謝辞 本研究の一部は文部省科学研究費（No. 04855075）の補助による．

参 考 文 献

1) Shortliffe, E. H. and Buchanan, B. G.: A Model of Inexact Reasoning in Medicine, *Mathematical Bioscience*, Vol. 23, pp. 351-379 (1975).
2) Nilsson, N. J.: Probabilistic Logic, *Artif. Intell.*, Vol. 28, pp. 71-87 (1986).
3) Cheeseman, P. A.: A Method Computing Generalization Bayesian Probability Value for Expert Systems, *Proc. 8th Int. Joint Conf. of Artificial Intelligence*, pp. 198-202 (1983).
4) Valiant, L. G.: A Theory of the Learnable, *Comm. ACM*, Vol. 27, No. 11, pp. 1134-1142 (1984).
5) Blumer, A., Ehrenfeucht, A., Haussler D. and Warmuth, M.: Classify Learnable Geometric Concept with the Vapnik-Chervonenkis Dimension, *Proc. 18th Annual ACM Symp. on Theory of Computing*, pp. 273-282 (1986).
6) Matsushima, T., Suzuki, J., Inazumi, H. and Hirasawa, S.: Inductive Inference Scheme at a Finite Stage of Process from a View Point of Source Coding, 電子情報通信学会論文誌 E, Vol. 73 E, No. 5, pp. 644-652 (1990).
7) 松嶋敏泰，鈴木 譲，稲積宏誠，平澤茂一，石渡徳彌：情報理論による不確実な知識の表現法と推論に関する一考察，早稲田大学理工学研究所報告 131，早稲田大学理工学研究所 (May 1991).
8) Matsushima, T., Inazumi, H. and Hirasawa, S.: An Inductive Inference Scheme Subject to Prediction, *IEEE Trans. Syst. Man Cybern.* (to appear).
9) Gallager, R. G.: *Information Theory and Reliable Communication*, Wiley, New York (1968).
10) Huge, G. E. and Cresswell, M. J.: *An Introduction to Modal Logic*, Methuen, London (1968).
11) Haussler, D.: Decision Theoretic Generalization of the Pac Model for Neural Net and Other Learning Applications, *Information and Computation*, Vol. 100, pp. 78-150 (1990).
12) Yamanishi, K.: A Loss Bound Model for Online Stochastic Prediction Strategies, *Proceedings of the Fourth Workshop on Computational Learning Theory*, pp. 290-302 (1991).
13) Littlestone, N.: Mistake Bounds and Logarithmic Linear-Threshold Learning Algorithms, Ph.D. thesis, University of California, Santa Cruz (1989).

14) Lehmann, E.L.: *Testing Statistical Hypotheses*, Wiley, New York (1959).
15) Hoel, P.G., Port, S.C. and Stone, C.J.: *Introduction to Statistical Theory*, Houghton Mifflin Company, Boston (1971).
16) Ferguson, T.S.: *Mathematical Statistics*, Academic Press, New York (1967).
17) Matsushima, T., Inazumi, H. and Hirasawa, S.: A Class of Distortionless Codes Designed by Bayes Decision Theory, *IEEE Trans. Inf. Theory*, Vol. 37, No. 5, pp. 1288-1293 (1991).
18) Quinlan, J.R. and Rivest, R.L.: Inferring Decision Tree Using the Minimum Description Length Principle, *Information and Computation*, Vol. 80, No. 1, pp. 227-248 (1989).

付 録

A. 不確実性を含んだ演繹推論と知識の表現法

A.1 不確実性を含んだ演繹推論と信頼率

前稿[7]で不確実な知識を取り扱うために確率世界論理上で演繹推論と論理式の表現法の拡張を行ったが，本研究においても，帰納推論と演繹推論を一連の推論として統一的に取り扱い，同様のモデル化のもとで議論を行っているため，前稿の結果について簡単にまとめておく．不確実な知識とは真理値が一意に決定できない論理式と考えることができる．古典論理においては，第2章で述べたように仮説から補題2.1をみたす完全に真偽が決まった論理式のみが演繹可能であるが，ある程度の誤り☆を許容すればセマンティク情報が不十分な論理式も推論することが可能であり，これが不確実性をもつ演繹と考えられる．

補題 A.1[7] 論理式の集合 A から論理式 W について誤りを許容しても導出を行う場合，次式が平均誤り確率最小の推論となる．

$$\arg\max_{W,\neg W}\Big(\sum_{m} T(W|m)P(m|A),\ \sum_{m} T(\neg W|A)P(m|A)\Big). \tag{49}$$

この推論を用いた場合の誤り確率を1から差し引いた信頼率 $C(W)$ は次のように与えられる．

$$C(W)=\sum_{m} T(W|m)P(m|A). \tag{50}$$

□

この推論は $C(W)=1$ の場合古典的な演繹推論と同等となり，古典的推論を拡張した推論となっていることは明らかであろう．この誤り確率最小の基準は，古典論理の演繹推論に対する評価基準である健全性を拡張し一般化した基準と考えることもでき，不確実性をもつ演繹推論の新たな評価基準と考えられる．また，この信頼率 $C(W)$ は，従来研究で明確な意味が説明されず用いられた確信度等の論理式の不確実性を表す尺度と異なり，真偽に対する誤り確率という明確な意味を持っている．

☆ 真の論理式を偽としてしまう，あるいはその逆の誤り．

A.2 全称限定子の拡張と推論規則

式(50)で不確実な論理式の信頼性の表現法を示したが，メタの情報，つまり限定子つき論理式についても信頼性が表現可能なように古典論理の限定子を拡張した確率限定子[7]について述べる．従来の確信度を用いる演繹推論法[3]では，含意の確信度を条件つき確率的に表現する場合が多い．前稿では確率限定子つき含意の信頼率について，いくつかの表現法を定義[7]しているが，その1つとして条件つき確率的な表現も定義しており，以下ではそれについて簡単に述べる．

定義 A.1 含意の論理式 $\forall x(Q(x)\rightarrow R(x))$ の条件つき信頼率 $C(R|Q)$ を次式のように定義する．

$$C(R|Q)=\frac{\sum_{m}\sum_{x} T(Q(x)|m)T(R(x)|m)P(x)A(m|A_T)}{\sum_{m}\sum_{x} T(Q(x)|m)P(x)P(m|A_T)}, \tag{51}$$

ここで，Q と R は述語記号，$P(x)$ は定数記号 x の定義域上での確率とする． □

この信頼率の直感的意味は，前件をみたす個体に対し後件もみたす割合を示していると考えられる☆☆．

このように条件つき信頼率で表現された含意を用いて，以下の仮定のもとで誤り率最小の推論法を示す．

仮定 A.1 解釈空間上の解釈のパターンの事前確率 $P(m)$ は一様分布とする．

この仮定は，まだ何も正しい論理式が知らされていないとき，考え得るすべての解釈のパターンは同等に考慮されるべきであろうという要請を確率世界論理上で表現したものである．

補題 A.2 式(52)の推論において，導出された $R(a)$ の信頼率 $C(R(a))$ は式(53)で与えられる．

☆☆ Nilsson の Probabilistic logic ではこのような限定子つき論理式について，その論理式自体の真偽の確率については定義されているが，このようにメタの情報に関して不確実性を表現することについては考慮されていない．そのため，Probabilistic logic は述語論理上で定義されているにもかかわらず，実質的には命題論理の演繹推論のみしか扱えない論理となっている．

$$\frac{\forall x(Q(x)\to R(x)):\ C(R|Q);\ C(R|\neg Q),Q(a):\ C(Q(a))}{R(a):\ C(R(a))}, \tag{52}$$

$$C(R(a))=C(R|Q)C(Q(a)) + C(R|\neg Q)(1-C(Q(a))). \tag{53}$$

(証明) 補題 A.1 より明らか. □

この推論では2つの条件つき信頼率 $C(R|Q)$, $C(R|\neg Q)$ を用いたが，一方の信頼率，例えば $C(R|Q)$ のみの情報しか与えられていない場合でも $C(R|\neg Q)=1/2$ とすることによって，与えられた論理式のもとで誤り確率最小の推論は可能である.

式(53)の計算法は仮説を充足する解釈のエントロピー最大の分布から求めた信頼率とも一致し，従来の確率的計算を利用する幾つかの研究において明確な理由なく用いられていた各アトムの確信度の同時確率のエントロピーを最大化する手法[3]の理論的根拠も与えている. なお，確率世界論理上のその他の導出規則やファジー論理，多値論理との関連については前稿[7]で詳しく考察を行っている.

B. 系 4.2 の証明

定理 4.1 の最適な予測 $D_{T(G)O}(F^n)$ は，次式の事後確率で重み付けされた G の信頼率 $C(G)$ が 1/2 以上かを比較し推論を行っていると考えることができる.

$$\sum_m \sum_{A^k} \int\cdots\int T(G|m)P(m|F^n, A^k, C^k) \cdot f(C^k|A^k, F^n)P(A^k|F^n)dC^k. \tag{54}$$

この式(54)の $\sum_m T(G|m)P(m|F^n, A^k, C^k)$ の部分は，G の信頼率 $C(G)$ を A^k の各含意ルールとその条件つき信頼率 C^k より拡張した演繹推論によって求めていることと同値なので，補題A.2の推論法より次式のようにそれぞれの含意ルールの条件つき信頼率 C_i の積を線形結合した式で置き換えられる.

$$\sum_m T(G|m)P(m|F^n, A^k, C^k) =\sum_i g_i C_i+\sum_{ij} g_{ij}C_iC_j+\sum_{ijk} g_{ijk}C_iC_jC_k+\cdots, \tag{55}$$

ここで，g はある定数とする. 各含意ルールの信頼率は仮定より独立なので，式(55)を事後確率 $f(C^k|A^k, F^n)$ で期待値をとる場合と，事後確率で以下のように期待値をとった C_{iO} を用いて式(55)を計算することは等しくなる.

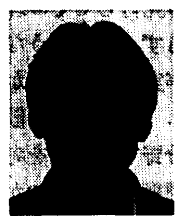
$$C_{iO}=\int C_i f(C_i|A^k, F^n)dC_i, \tag{56}$$

よって，定理 4.1 の最適な推論法と系 4.2 が同値であることが証明される.

(平成3年 1 月21日受付)
(平成4年10月 8 日採録)

松嶋 敏泰

昭和 53 年早稲田大学理工学部工業経営卒業. 昭和 55 年同大学院理工学研究科博士前期課程修了. 同年日本電気(株)入社. 昭和 61 年早稲田大学大学院理工学研究科博士後期課程入学. 平成元年横浜商科大学講師. 平成 4 年同大学助教授，現在に至る. 知識情報処理および情報理論とその応用に関する研究に従事. 工学博士. 電子情報通信学会，人工知能学会，情報理論とその応用学会等会員.



稲積 宏誠

昭和 57 年早稲田大学理工学部工業経営卒業. 昭和 59 年同大大学院理工学研究科修士課程修了. 同年同大大学院博士後期課程入学. 同年同大情報科学研究教育センター助手. 昭和 62 年湘南工科大学工学部情報工学科講師，現在に至る. 情報検索システムの評価および情報理論の応用に関する研究に従事. 工学博士.

平澤 茂一（正会員）

昭和 36 年早稲田大学理工学部数学卒業. 昭和 38 年電気通信卒業. 同年三菱電機(株)入社. 昭和 56 年早稲田大学理工学部工業経営教授，現在に至る. 工学博士. データ伝送方式，計算器応用システムの開発，ならびに情報理論とその応用の研究に従事. 昭和54年 UCLA 計算機科学科客員研究員. 昭和 60 年ハンガリー科学アカデミー，伊トリエステ大学客員教授. 電子情報通信学会，情報理論とその応用学会，人工知能学会などの会員.