

マーケティング・リサーチの自動化と帰納論理プログラミング —リサーチの形式化と人工知能の利用に向けて

Marketing research automation and Inductive Logic Programming
-toward formalizing the process of marketing research and the application of artificial intelligence

金城 敬太
Keita Kinjo

【Abstract】

In this paper, we conducted marketing research data analysis, especially quantitative research data by inductive logic programming (ILP) which are usually conducted by statistical analysis. As a result, we succeeded in verifying the effectiveness of the methodology as well as in getting many suggestions (toward the studies of marketing research) from the similarity between inductive reasoning and the process of marketing research. From the analysis method view point, our study showed that the ILP method can utilize the background knowledge actively, automatically search and test the hypotheses by setting up the bias, and detect the interaction. In addition, as the suggestions to marketing research theory, this paper proposes and discusses the respective formal definition of the hypothesis, the working hypothesis and the relation between them. This study can be seen as not only the introduction of the new analytical method, but also a first step toward formalizing the process of marketing research.

Keywords. Marketing research, Formalize, Inductive reasoning, Inductive logic Programming

【目次】

1. はじめに
2. 帰納推論の定義と帰納論理プログラミング
3. 帰納論理プログラミングを利用したリサーチ
 - 3.1 全体の概要
 - 3.2. 質問紙調査の設計
 - 3.3. マーケティング・リサーチへの ILP の適用方法の提案
 - 3.4. 分析結果
4. 考察および結語

1. はじめに

今日多くのマーケティング・リサーチが行われている。そのプロセスや分析について、すでに多くの経験的な蓄積もあり、その方法についての一般化や共有化も進んできている。またマーケティング・リサーチの土台となっている、社会調査のプロセスに関して基礎理論を構築する動きもある(新, 2005)。これらは調査において適切に仮説を設定し、検証することの基礎となるため重要な取り組みである。ただし、社会調査に関して数理論理学などを利用した形式的な議論はあまりなされていないことが指摘されている(稲, 1980)。こうした中、理論・思想的な背景としてマーケティング・リサーチと論理実証主義、さらには反証主義、批判的合理主義との関連性も指摘されている(栗木, 2007)。本研究ではこの考えを参考にしながら、人工知能のひとつである論理を用いた帰納的な推論システム、帰納論理プログラミング(Inductive Logic Programming; ILP)をマーケティング・リサーチに適用する方法を提案する。これにより、マーケティング・リサーチへのILPの有効性を検証するとともに、形式的な議論にむけたひとつのアプローチを提示したい。

導入するにあたり、マーケティング・リサーチの全体像やその流れについて述べ、議論する範囲を明らかにする。マーケティング・リサーチとは、企業が直面する特定の市場状況に関するデータと調査の体系的なデザイン、収集、分析、報告である、とされている(Kotler, P., Keller, K., Brady, M. et al., 2009)。これはマーケティング計画全体において機会をみつけたり、問題を規定したり、業績を評価するなどの一部を担っている(Birks, D. F., Malhotra, N. K., 2007)。マーケティング・リサーチの対象は、国、市場環境、競合他社の動きなどについて調べるマクロな調査と、ある社会における消費者の意識や行動などの捉えるためのミクロな調査に分類できる。後者の消費者を捉える調査は、大きく分けて量的調査と質的調査がある。量的調査のなかには質問紙調査、質的調査のなかにはインタビュー調査、参与観察などがある。様々な調査には共通点や差異、そして利点と欠点が存在しており、どのように理論を構築するかという目的も含めて様々な議論がなされている(Glaser, B. G. Strauss, A. L., 1967)。ひとつひとつの調査における推論は、複雑な構造を持っているためすべてを扱うことはできない。特に質的調査に関しては問題そのものが不明確な場合にも利用されるなど、そのプロセスは未解明な点が多い(Punch, K.F., 2005)。一方、真鍋が指摘するように量的調査のなかの特に質問紙調査については、「何を」「どのように」という観察の意図がすでに明確であるときに利用される(真鍋, 2003)。本研究では、形式的な議論が行いやすい問題の設定と仮説の設定が明確である、消費者の行動を捉えるための質問紙調査の議論を中心に扱う。

具体的に、調査やリサーチに利用されている考えやプロセスを確認する。調査の全体を広く捉えた場合は、「帰納推論」という推論形式のひとつであるという主張がある(安田・原, 1982)。帰納推論とは、個々の事例から一般的な結論を導き出す推論形式のことを指す。Wallaceも社会理論構築に関し数理論理学は用いていないが簡略的に推論として定式化を行っており、そのなかに調査を帰納推論として位置付けている(Wallace, W.L., 1971)。厳密には調査の一部である仮説構築の際に、演繹推論や発想推論も利用されているが(佐藤・山田, 2009)調査を行うことはデータから一般的なルールを導くことであるので広く捉えた場合、帰納推論と捉える事ができる¹⁾。ただし、

このような記述はあるものの数理論理学を用いた詳細な形式化はあまり進んでおらず、共通して議論する土台はあまりできていないと考えられる。そのため、論理的な議論が不足していると指摘する研究もある(石川, 1980; 稲上, 1980)。

具体的に「帰納推論」の考えに基づいた調査の作業プロセスが多く提案されており、これらは似た構造を持っている。はじめに一般的なモデルの例として、Punch らの出した仮説がある場合の調査のモデルを下記に述べる(Punch, K.F., 2005)。一般的な社会調査のプロセスは、ある調査領域から調査課題を導きだし、そのなかから課題に必要な特定の調査問題を導き出す。さらにその問題に対する答えとして理論などから導き出される仮説を設定して、その仮説を検証するために具体的な調査をデザインして、データを収集し、最後にデータ分析を通じて仮説の検証を行うという流れになっている。

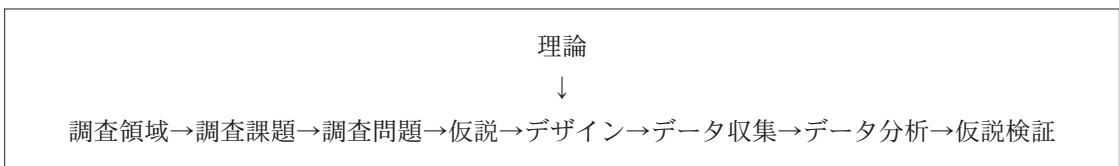


図1. 調査のモデル (Punch, K.F. 2005)

他にも新によって、同様に社会調査の理論化および、各作業におけるプロセスに必要な仮説やその下位概念などの詳細な定義が行われている(新, 2005)。またこのようなプロセスのモデルについてはいくつかの分類がある。Punch は、仮説がある場合と仮説がない場合の二つに分類している。それ以外の分類としては仮説の検証までを線形モデル、そこからさらに理論の修正、仮説、調査を行うプロセスを組み込んだ円環モデルとして分類しているものもある(指田, 2008)。前者については図1のような一方向でおわる調査のモデルがあり、後者の例では、見田による調査の循環プロセスがある(見田, 1965)。それ以外に Harvey ら(Harvey, L., MacDonald, M., 1993)も同様の円環モデルを提案している。こうしたサイクルは批判的合理主義における反証の考え方とも整合的である(栗木, 2007)。同様に KJ 法で有名な川喜田次郎は W 型問題解決を提案しているがこれも一種の円環モデルと捉えることができるであろう(川喜田, 1967)。これら円環モデルはより一般的なものであり、仮説構築から仮説の検証に至る一部のプロセスを切り出して単純に表現したものが線形モデルであると考えられる。このように調査の作業プロセスは、過去の研究をみても、似た構造を持っている(佐藤・山田, 2009)。

具体的なプロセスの内部構造について述べると、作業プロセスの共通な要素として「問題」、「仮説」(仮説生成)、「仮説検証」の三つが存在している。本研究は ILP の社会調査、特にマーケティング・リサーチへの適用に向けてこの三つに焦点をあてて説明する。なお、多くの調査のプロセスの研究における理論や用語は、定義の不統一性や多義性もあり統一的な議論を行うことは困難であるため主に Punch の議論を参考に定義をする。

プロセスのはじめにある「問題」をどのように捉えるかは議論が分かれるところである。最も単

純には「私たちは、なにを知ろうとしているのか」という問いに対する答えを求めることとして問題（調査問題）が導かれる（Punch, K.F., 2005）。それ以外にも調査のための共通の目的意識という意味や、理論やシステムにおける矛盾点を問題と表現することもある（佐藤・山田, 2009）。本研究では広義には過去の研究および帰納推論の理論的な定義に基づいて、ある経験的な事象に対して説明を求めること、とする。狭義には下記で採用する「仮説」の定義と対応させる関係上、ある経験的な事象を説明もしくは予測のための問い、と定義する。

「仮説」についても多くの定義が存在している。しかしもっとも単純で共通要素としては「（調査）問題への予想される答え」（Punch, K.F., 2005）というものがある。本研究では上記で述べた「問題」に対する答えであり、すなわち経験的な事象を説明もしくは予測するために未検証な命題（新, 2005, 真鍋, 2003）と定義する²⁾。また、仮説を調査可能なように抽象度を下げてより具体化し測定可能なものを「作業仮説」と呼ぶ（新, 2005, 佐藤・山田, 2009）ことがある。これは理論変数を複数の経験変数（測定可能な次元）によって捉えるものであり、概念とデータを接続したものである。この概念も複数の解釈がなされているが本研究では、単純に検証したい仮説の抽象度が高い場合に測定可能な変数に変換するものと捉える。なお、調査モデルの多くはこの「仮説」をどう作るかすなわち、「仮説生成」も含んでいるが、一般的にこのプロセスでは質的調査を利用し、それ以外の検証の部分に量的調査が利用されている。本研究では主に量的調査の検証プロセスの部分に焦点をあてるため、仮説生成については詳細には言及しない。なお、仮説形成についても一部論理による形式化が可能であるがこれについては考察で述べる。

最後の「仮説検証」は仮説が科学的に意味のあるものか否かを、サンプリングなどで収集したデータに基づいて分析・検証することである。多くの研究では、客観的に実証できる統計的検定や、複数の変数間の関連（相関）をみていく必要性から、重回帰分析や、パス解析や因子分析も含めた共分散構造分析などの統計的な因果推論が用いられてきた。この際、調査した仮説以外の変数を説明に用いることで新しい結論が導かれることも多く、発見的な利用も行われている。

以上のように、調査のプロセスについて述べてきたが、調査やそのプロセスの構造についての論理的な形式化や、帰納推論との対応関係があまり議論されていないほか、例えば仮説と作業仮説の関係など個別の技術についてのあいまいな部分も多い。

一方、人工知能の機械学習の分野で、このような問題の設定、仮説の設定、仮説の検証という帰納推論のプロセスを、形式論理学をもとにした論理プログラミング上でモデル化している帰納論理プログラミング（ILP）がある。論理実証主義に端を発する Plotkin の発見の論理（Plotkin, G.D., 1970）に起源をもち、その後批判的合理主義の影響下にある Shapiro の発見の論理および、論理に基づいた帰納推論のモデル化が行われてきた（Shapiro, E.Y., 1982）。その後、1990年代より論理プログラミングの拡張の過程で、ILP システム Progol などの開発などが行われ発展を遂げてきた（Muggleton, S., 1995）。ILP の詳細は2章、調査との対応関係は3章で述べるが、これは帰納推論という推論をベースにして構築されてきた議論であり、問題の設定や、仮説の設定や検証など、同じく帰納推論をもとにした調査、特にマーケティング・リサーチと多くの共通構造を持っている。

実際、マーケティング・リサーチと論理実証主義、さらには批判的合理主義との関連も指摘されている(栗木, 2007)。

以上をふまえ、本論文では、批判的合理主義と関連のあるマーケティング・リサーチや社会調査のプロセスのもとになっている「帰納推論」を厳密に形式化して応用した ILP の応用可能性を探る。さらにリサーチのデータに対して ILP を適用する一つの方法を提案および実証し、示唆を得る。これは単に分析方法として活用できるという以上の意味があり、マーケティング・リサーチとしてのメリットとして、ILP を対応させて実行ができるのであれば、一部形式化が可能になるということが示される。形式化がなされれば、概念が共通化、もしくは明確になるため、議論を行いやすくなる。例えば、ある概念をどう定義しているのか、概念間の関係はどのようになっているのかを自然言語ではなく、数学的な論理式として定義することができるようになるため専門家間でのコミュニケーションがより容易になる。仮に論理的な設定による形式化がうまくいってなくても、実際の調査と形式的な推論との差異を比べることができる。さらに ILP の手法は既存の回帰分析などの統計的手法を包含した方法になっていることもあり、より広範囲の問題の分析が可能であるとも考えられる。

以下、2章では ILP および、それ以外の分析方法との違いについて述べる。3章では ILP による社会調査データの分析を試みる。4章では分析結果をもとに議論を行う。

2. 帰納推論の定義と帰納論理プログラミング

1章で述べた ILP の詳細を述べ、また既存の調査の分析で用いられている方法との違いを述べる。具体的なマーケティング・リサーチへの適用については2章をふまえたうえで3章にて述べる。

はじめに帰納推論の論理的な設定を述べたうえで、次に論理プログラミングを説明し、最後に ILP を説明する。前節で述べたように Plotkin による定義もあるが、ここでは完結に表現するため以下の(1)から(4)の帰納推論の定義を用いる(井上, 2010)。

(1) $B \cup H = O$

(2) $B \cup H$ が無矛盾。

(3) H は L から構成される。

ここで B は背景知識、すなわち事前知っている情報や知識、規則(ルール)の集合である。 H は仮説、 O は観測データの集合であり、 L は仮説を構築するための制限(バイアスと呼ばれる)である。 \cup は和集合を表す記号である。また $=$ は論理的な帰結を示す。すなわち、「背景知識 B と仮説 H を組み合わせて演繹推論することで右側にある帰結 O を得ることができる」ということである。この定義は帰納推論および発想推論に共通のものであるが、特にある事象に当てはまる正事例とある事象に当てはまらない負事例を用いる帰納推論では O が正事例を表現し、また N を負事例として以下の(4)が加わる。

(4) $B \cup H \neq N$

上記の定義を端的に言えば、「背景知識 B は仮説 H を伴って正の事例の観察事象 O を説明することができる。一方、負事例 N は説明することができない」ということを表現している。このような仮説 H を求めることが帰納推論の問題となる。

つぎに ILP の前提となる、論理プログラミングについての簡単な説明を行う。論理プログラミングとは、一階の述語論理に基づいたプログラミングである。述語論理のうちホーン節と呼ばれる論理に限定し、融合原理 (Robinson, J.A., 1965) という証明方法を利用して実行するプログラミングである³⁾。このような論理プログラミングの枠組みは一般的に「演繹推論」を行うプログラムである。一方、この考えを利用しながら「帰納推論」を実現させたのが ILP である。ILP には複数のシステムがあるが、具体的な ILP のシステムとして多く利用されてきた Progol について説明を行う (Muggleton, S., 1995, 古川・尾崎・植野, 2001)。

具体的な処理について説明すると、目標概念 (target concept) を被覆するような仮説を背景知識とバイアスに基づいて構築するシステムである。まずこれらの用語をマーケティング・リサーチにおける事例と関連付けながら説明する。目標概念とは、ILP によって説明を行いたい概念のことを指す。これらはさきほどの O と N に該当し、正事例と負事例にわけて入力される。以下、例を示す。厳密な述語論理による形式ではなく、Prolog の記述形式をとっている。() 内は解説で [] は解釈を示す。

正事例 buy(p1).buy(p2).buy(p3).

負事例 buy(p4).buy(p5).buy(p6).

(buy は「商品を購入する」を示す述語。p1 ~ p6 は回答者の index とする。)

被覆とは、仮説によって事例をどれくらいの割合で説明できているかを指す。背景知識とは、事例に関する関連知識のことを指す。これらは事実としても与えられるし、ルール の形でも記述される。背景知識の例も述べよう。

背景知識 has_composure(X):-has_money(X),has_time(X).

has_money(p1).has_money(p2).has_money(p3).

has_time(p1).has_time(p2).has_time(p3).

people(p1).people(p2).people(p3).people(p4).people(p5).people(p6).

(has_composure は「X が余裕がある人である」。has_money,has_time はそれぞれ「お金がある」「時間がある」を示す。また people は p1 ~ p6 が人であることを示す。)

仮説とは目標概念を被覆するような規則 (ルール) であり、ひとつの命題の形をとっている。上記の目標概念に対して構築される仮説の例を述べると次のようなものが挙げられる。

仮説 buy(X):- has_composure (X).

(「X が余裕のある人ならば X は商品を購入する」というルール。「:-」は論理における含意「←」を表す。)

このような目標概念と背景知識を与えたうえで仮説を自動的に構築して絞り込みを行う。仮説は目標概念を説明する様々な述語の組み合わせで成り立つ。この仮説同士の組み合わせによって構築される空間のことを「仮説空間」と呼ぶ。さて ILP ではこの仮説空間のなかから実際に事例を被覆する最適な仮説を求めるのであるが仮説が膨大であるために仮説空間の絞り込みや探索のための工夫が存在する⁴⁾。具体的には、仮説の絞り込みのためにバイアスの設定が利用される。バイアスとは、仮説のためにあらかじめ与える情報のことでモード宣言と呼ばれるものがある。これは仮説となる述語をあらかじめ有る程度絞り込むために候補となる述語を与えるものである。これらを利用し、効率的に探索を行う。

また、仮説の候補から事例をどれくらい被覆するかの評価のみならず仮説の長さを罰金項として設けて、その基準がもっともよいものを仮説として採用する。これは統計学における情報量基準の考えかたに類似のものである。実際 Progol において情報量基準の一種であるベイズ情報量基準 (BIC) が利用されている。

他の手法との関連について述べる。まず、重要な点として述語論理の表現能力は非常に広範囲であるため、調査の分析で用いられる一般的な重回帰分析なども下位概念となっている。具体的には回帰で表現できるような変数間の関係を一つの述語として定義して ILP に組み込めばよい (Karalić, A. Bratko, I., 1997)。例えば、重回帰分析は下記のような形で表現することが可能である。

$$f(Y, X1, X2, \dots) := Y \text{ is } a1 * X1 + a2 * X2, \dots$$

上記の f は述語、Y は被説明変数、X1 と X2 は説明変数で、a1 と a2 は回帰係数である。is というのは等号を意味している。つまり、回帰における変数間の関係式は一つの述語として表現ができる。このように ILP は既存の統計的な手法をも包括した方法である。

似た分析法として、質的データ分析や事例研究に利用される、質的比較分析 Qualitative Comparative Analysis (略して QCA) がある (鹿又・長谷川・野宮, 2001)。これは正事例と負事例が与えられた場合に正事例に対して無矛盾なルールの構築を自動的に行うシステムである。ただし与えられるのがブール代数と等価なルールであり、命題論理を獲得するシステムであるといえる。また同様の分析方法として決定木がある (Breiman, L., Friedman, J.H., Olshen, J.H., et al., 1984)。これには数多くのアルゴリズムが開発されているが、これも正事例と負事例をわけるルールを平均情報量などの基準を利用して構築するシステムである。決定木はある程度の数のデータを用いることで必ずしもすべての事例を説明はせず、一般には統計的に妥当なルールが出力される。ILP との違いは決定木が QCA と同様に命題論理レベルの学習であるのに対して述語論理レベルの関係に関するルールの獲得ができることが挙げられる (古川・尾崎・植野, 2001)。

ILP の応用としては、初期のものに Srinivasan らが突然変異誘発性物質の構造推定問題への応用として ILP を利用したものがある (Srinivasan, A., Muggleton, S., King, R. et al., 1994)。ま

た、社会ネットワークの分析に応用した事例として、金城の研究などがある（金城・尾崎・相澤，2010）。ここではグラフ構造を持つ関係データの分析に ILP を用いてはいるが本研究のように社会調査データのプロセスとの対応させたうえで導入していない。

3. 帰納論理プログラミングを利用したリサーチ

3.1 全体の概要

マーケティング・リサーチのなかでも、特に質問紙調査による消費者の分析に対して ILP を利用し結果を考察する。調査における仮説の設計に対応した方法で記述を行い、実装の1つの方法を提案する。ただし、調査データの記述方法は他の方法も考えられるため必ずしもひとつではないことに注意されたい。重要なのは、問題の設定や仮説の半自動生成やその検証が形式的に行えるというところである。また本研究では ILP のシステムとして Progol を改良した aleph を利用する（Srinivasan, A., 1999）。以下、質問紙調査の設計と提案する適用方法を述べる。

3.2. 質問紙調査の設計

本研究で用いた調査は、2011年3月11日に日本で起きた東日本大震災の影響でその後に起きた「買いだめ行動」に焦点を当てたものである。この行動は特定の個人に物品が集中し、多数の人に物品がいきわたらないなどの社会的な問題を引き起こした。このような消費行動を起こした人々においてどのような共通の属性があるのだろうか。ここで、買いだめする商品として、具体的に「水」の購入を想定し、以下の問題を設定する。

問題「水の買いだめ行動を行ったひとはどのようなひとか。」

次にこの問いに対する答えとして、仮説を設定する。水の買いだめには地震が発生した際に水道水が止まり飲み水がなくなることへの備えと、原子力事故に伴う放射能の水道への汚染の不安が原因となっている可能性がある。しかし、後者に関しては妊婦や小さい子供がいる家庭などの一部での不安が大きく、世間一般で考えた場合は前者の可能性が高いと考えられる。そこでより有力な仮説として次の仮説を設定する。

仮説「地震への不安を持つひとが、水の買いだめ行動を行った。」

ただし、「地震への不安」という概念については抽象度が高く、直接測定することは難しい。そこでより具体的に測定可能な仮説に変換するために概念の抽象度を下げ、作業仮説⁷⁾とするために以下の仮説を設定する⁸⁾。

作業仮説「余震に対して不安を持つひとが、水の買いだめ行動を行った。」

作業仮説「今後の地震の増加に不安を持つひとが、水の買いだめ行動を行った。」

上記の作業仮説をもとにして、個人に対して、地震直後に水の買いだめをしたか否か、および余震や今後の地震に対する不安があるか否かの質問紙調査を行った。

また、「原子力発電に対する不安」や「モノ不足の不安」、「情報不足」、「漠然とした不安」なども水の買いだめ行動に関連する可能性があるため、これらについての仮説も検証が可能なように各概念を具体化し質問を設計している。これらの概念と具体的な調査票の質問項目との対応については補遺に記載している（補遺 1）。

調査対象者は、20歳以上で北海道、埼玉県、千葉県、東京都、神奈川県、静岡県、愛知県、大阪府、広島県、福岡県に住む（各地域 200名）ひととした。調査期間は、2011年3月25日～31日で、調査方法としてはインターネット調査を利用した。最終的に得られた有効回答数は、2,031となった。

3.3. マーケティング・リサーチへの ILP の適用方法の提案

調査設計と調査データに対して ILP を利用して、仮説の検証を行う。具体的にそれぞれの概念を ILP 上で表現して設定する方法を下記で提案する。

まず問題をどのように表現して ILP に設定するかを述べる。問題は、定義では経験的な事象に対して説明を求めることとしたが、これは ILP の文脈では目標概念として経験的な事象を設定することに相当すると考えられる⁸⁾。今回、説明を求めるとは問題の文からも明らかのように「水の買いだめ行動を行った」であったので、これをひとつの概念として捉え、それを $W_overbuy$ という述語で表現する。（aleph ではバイアス（モード宣言）の設定は「 $modeh(1, W_overbuy(people+))$ 」と記載する。）

目標概念 $W_overbuy(X)$.

（「X は水の買いだめ行動を行った」という述語。）

次に仮説についてどのように ILP で表現できるかを述べる。ILP を用いた場合、仮説（理論仮説）と作業仮説の対応関係が明示的に定義できていればそれを背景知識に組み込むことで、作業仮説を使った検証ではなく、その上位の仮説を検証することが可能である。

上記で述べた仮説「水の買いだめ行動を引き起こすひとは、地震への不安を持つひとである。」を用いる。「地震に対する不安」があることを「 $has_anxietyA(X)$ 」と表現して、これをもとに仮説を記述した場合以下ようになる。このとき目標概念の右にくる説明の部分はボディと呼ばれる。ここにくる述語は予めモード宣言による設定が可能である。

仮説 $W_overbuy(X) : - has_anxietyA(X)$.

（「地震に対する不安がある（すべての）X は、水の買いだめ行動を行った」というルール）

具体的に上記に加えてそれ以外の仮説も検証が可能なように、「情報不足に対する不安」「モノ不足に対する不安」「原子力に対する不安」「漠然とした不安」を設定し、それぞれ「has_anxietyB(X).」「has_anxietyC(X).」「has_anxietyD(X).」「has_anxietyE(X).」という述語で表現し、検証が可能なように has_anxietyA と同様にボディにくるように設定する。

加えて、仮説と作業仮説との対応付けを行う。これは対応を論理として明確に表現することで背景知識のルールとして記述が可能である。具体的には「地震への不安」を調査するために用意した下位概念「余震への不安」(anxiety_a_1 とする)および「今後の地震の増加への不安」(anxiety_a_2 とする)の二つについて、「余震への不安があるならば、地震への不安がある」、「今後の地震増加への不安があるならば、地震への不安がある。」というルールとして考え、以下のように記述する。「地震への不安」以外の変数の対応については補遺に記載した(補遺1)。このような概念の対応関係を論理によって記述するのは、ILP の利点である。なお、仮説と作業仮説の対応において利用するのはひとつの例である。この点については考察でもふれる。

背景知識 anxietyA(X) :- anxiety_a_1(X).

anxietyA(X) :- anxiety_a_2(X).

(「X に余震への不安があるならば、地震への不安がある」、「X に今後の地震増加への不安があるならば、地震への不安がある」というルール)

上記のような不安の項目に加えて、さらにより探索的にルールを調べるため、他にも仮説の候補として性別年齢および都道府県なども仮説のボディにくるように設定する。例えば「30代女性である」ことは述語を用いて「female30(X)」と表記し、同様に他の男女別年齢10歳刻み、さらに地域(都道府県)も仮説のボディにくることが可能なように設定を行う。

さて、その他のデータは各個人が回答した事実として、質問紙調査の結果を変換することで作成が可能である。正事例・負事例および、背景知識として以下のように記述する。ここで p1 および p2 は調査対象者個人に割り当てられたインデックスである。W_overbuy の正事例は水の買いだめ行動について1に回答したひと、負事例は1以外を回答したひとを割り当てた。またそれ以外の不安および性別年齢、地域についても各回答者が該当するものを下記のような述語で表現を行った。「・・・」は省略を表す。

正事例 W_overbuy(p1). . . .

負事例 W_overbuy(p2). . . .

背景知識 anxiety_a_1(p1). anxiety_a_2(p2). . . .

male20(p1). female30(p2). . . .

tokyo(p1). saitama(p2). . . .

以上のようにデータを作成することで、問題とそれに対する仮説の検証、すなわち ILP における目標概念とそれに対する仮説検証をすることができる。

3.4. 分析結果

3.3 の設定のもとで、ILP を実行し出力した結果は以下の四つのルールであった。

W_overbuy (X) : - female30(X), tokyo(X).

W_overbuy (X) : - anxietyA(X), anxietyC(X), anxietyD(X).

W_overbuy (X) : - anxietyB(X), anxietyC(X), anxietyD(X).

W_overbuy (X) : - anxietyA(X), anxietyC(X).

(「年齢女性 30 代でありかつ東京都に住んでいる X は、水の買いだめ行動を行った」)

(「地震に対する不安があり、かつモノ不足に対する不安かつ原子力に対する不安がある X は、水の買いだめ行動を行った」)

(「情報不足に対する不安、モノ不足に対する不安、原子力に対する不安がある X は、水の買いだめ行動を行った」)

(「地震に対する不安があり、かつモノ不足に対する不安がある X は、水の買いだめ行動を行った」)

上記のルールをもとにして統計的に検証した結果、F 値 9) は、0.17 であった。再現率は 0.68、適合率 0.10 であった。正事例のうち説明できたものは 103、説明できなかったものは 48、負の事例のうちルールで説明したものは 940、それに当てはまらなかったものは 909 となり、カイ 2 乗検定の結果はカイ 2 乗値は 16.19、自由度は 1、p 値は 5.708e-05 あった。ただし、12 個のサンプルについてはルールが形成されていなかった。

以上の結果はあらかじめ設定した不安が買いだめ行動に関連しているという仮説の一部が検証されたことを示す。加えて、情報不足に対する不安、モノ不足に対する不安、原子力に対する不安など別の不安が水の買いだめ行動に関係していることが明らかとなった。また年齢女性 30 代が関係しているのみならず、東京都に住んでいることも関係していることが明らかとなった。また、二つ以上の変数が論理積「かつ」で結ばれており、決定木などの手法でしか検出できない複雑な変数間の交互作用の影響も検出できている。事前に設定していなかった要素も抽出できていることがわかる。このほか、従来の手法（例えばロジスティック回帰）との差異については、補遺でも記述している（補遺 2）。

4. 考察および結語

以上のようにリサーチにおける問題に対して ILP を適用する方法を考案し、さらに仮説を検証することができた。これらをふまえて示唆されたことを述べたい。

まず分析法として示唆されることを述べたい。最大の利点として“事前知識”などを論理で記述することができる点が挙げられる。つまり、背後にある理論や知識とどのような関係にあるかを式として明確に記述することができ、それを利用したうえでルールを獲得することが挙げられる。これは3章の後半で述べた他の手法ではない特徴である。この点についてより詳細に考察する。単に変数変換を行うことで決定木による実行も可能と考えられるがILPが行っているのは関係データベースの分析を行っていることに注意されたい。関係データベースとは、複数の表とその関連を記述したものである。これらは統合して一つの表にすることも可能であるが、非常に大きなデータベースになる。その意味で、データベースの圧縮、そして単純に一つ一つのデータベースや変数間の関係を単純に記述するだけで分析が可能になるという利点があると考えられる。

次に、多くの社会調査では、予め与えられた仮説のみならず、その検証を通じて、あらかじめ設定された仮説以外の仮説が得られる可能性があることが多い。この点について、ひとつの仮説を設定するのではなく、バイアスを利用してより柔軟に設定することで自動的に探索・検証できる可能性を示した。

加えて、述語同士の複雑な構造を持った交互作用が検出できる点が既存の手法と違う点として挙げられる。3.4にあるように説明する要素として、AかつBというような二つ以上の条件が論理積で得られている。これは統計における交互作用と解釈できる。回帰分析など通常の統計分析においても単純な交互作用は抽出できるが、上記のように変数が多い場合、組み合わせが膨大になり、簡単に抽出するのは難しい。これを述語論理のような関係のレベルで自動的に検出できているが利点として挙げられる。

それ以外にもメリットとして「可読性」と「再利用性」が高いことが挙げられる。得られた式は、論理式となっているため、そのまま意味を理解することが可能である。また得られたルールを蓄積して、新たに演繹的な推論に再利用することも可能である。

ただし、本研究の主な目的は、下記で述べるようにリサーチ全体を帰納の枠組みとしてとらえなおすことであり、新しい手法を適用することのみではないことに注意されたい。

社会調査やマーケティング・リサーチの理論として示唆されることを述べたい。リサーチにおける検証が一部ILPで実行が可能であるということは、その元となっている帰納推論として議論ができるということである。具体的にはこれまで調査で利用されてきた「仮説」を数理論理上の命題として捉えて表現し、その表現のまま検証することが可能であることを示すものである。また曖昧であった仮説と作業仮説の関係が背景知識として記述することで議論できる可能性を示した。ただし、これはILPの応用の一つの事例である。「不安」測定した変数と上位の概念との対応関係は、検証的因子分析などを行って変数を作成する方法や、質問項目をつくり、その合計得点化をする方法など様々なものが想定されるが本研究ではILPと比較しやすくするためILPと同様に論理和としての対応、すなわち測定した変数のいずれかが成立していれば上位概念も成立しているという対応を採用している。

今後の課題について述べる。ILPで得られるルールは通常は「含意」としての関係であることが

挙げられる。多くの社会事象では「因果関係」を問題としている場合が多い。具体的に述べると、社会調査における問題は大きく分けて、「なぜか」のある対象についての因果を特定する「説明」と、「何か」というある対象についての「記述」の二つに分類できる(新, 2005)。今回検証に扱ったのは厳密には後者の問題である。因果関係については時間的順序など様々な基準があるが統計の分野においても未解明な部分が多く課題を残した分野であるため本研究の範囲を超えている9)。しかし、仮説において重要な位置を占めている。これについては、因果をひとつの述語として扱う方法や様相論理を用いた方法も考えられており、今後検討する必要がある(井上, 2010)。また、統計的因果推論をはじめ、実証分析と呼ばれる分野での扱われているような統計的に「因果」を扱った手法との融合も必要であろう。

続いて、人工知能における帰納的な手法の発展の大きな流れとして、数多くの機械学習やディープラーニングなどの手法がある(LeCun, Y., Bengio, Y., Hinton, G., 2015)。このような手法との対応関係や応用をどのようにとるのかということも課題である。本研究のILPは知識としての側面を扱っているため、人にとって得られた知識が利用しやすく、実際の意思決定に用いることが容易である。このような方法はホワイトボックスとされている。一方でディープラーニングやサポートベクターマシンなどの手法は内部の構造が非線形であったり、複雑になっているなど明確ではないものがあるためブラックボックスと呼ばれている。後者は仮にバグが発生した場合などのチューニングが難しい。しかし、ディープラーニングでは、膨大なデータをもとに被説明変数と説明変数の間に、ある種の複雑な中間的な変数を学習し、その結果として予測の精度が高いことも多い。そのため、背後の因果構造の推定や、さらには実世界のセンサーデータと概念の対応関係を推定するのに利用できる可能性がある。現状では研究途上であるが可読性の向上などが進めば、本研究との対応関係も明確になってくる。

社会調査における仮説生成についての議論はあまり行わなかった。しかし、これらは帰納推論における仮説の生成とも関連している。これは観測したものと規則(ルール)から事実を導き出す、発想推論としても捉えることが可能であり、すでに多くの仮説の生成方法が考えられている(Kakas, A.C., Kowalski, R.A., Toni, F., 1992)。類推と呼ばれる手法を用いて仮説を生成する研究もある(金城・尾崎・古川 ほか, 2014)。この部分の議論については今後研究を行う必要がある。同様に、「問題が明確でない」、つまり、あらかじめどのような問題があるか自体があまり明確でない問題発見の場合にも社会調査は行われる。このようなケースについての議論や形式化については本研究では扱うことができなかったため今後の課題としたい。

【注】

- 1) この一部に複数の推論が利用されていることは帰納推論が発想推論を利用していることや、演繹推論が利用されることにも関連している。
- 2) 作業仮説との区別から理論仮説と呼ぶこともある。
- 3) ホーン節というのは、例えば論理記号を使って書くと下記のようなもののみが該当する。

$A \wedge B \rightarrow C$

$\rightarrow D$

$E \wedge F \rightarrow$

さらにこれを論理プログラミングでは

C: $\neg A, B,$

D: \neg

: $\neg E, F,$

と記述する。なお第一式の C をヘッド、A, B をボディという。述語もふまえて記述した例としては 2 章で取り上げたものがあるが、より詳細な定義については (Kowalski, R.A., 1974, 古川・尾崎・植野, 2001) などを参考にされたい。

- 4) 例えば、Progol では逆伴意、すなわち $B \cup H \models O$ を書き換えた $B \cup \{\neg O\} \models \neg H$ に基づいた仮説生成を利用して最弱仮説を導出し、それを利用することで仮説空間の絞り込みを行う。
- 5) ここでは 1 章で述べたように「作業仮説」も多義的であるため、本研究では仮説の概念を測定可能なように抽象度を下げたものという定義を採用する。
- 6) 一般にはすべての概念について測定が可能なように概念とデータの対応関係をとる必要があるが、水の買いだめ行動やのちに述べる年齢や地域といった属性については直接、測定が可能なものであるため作業仮説において概念の抽象度を下げる作業を行わなかった。
- 7) 厳密には 2 章で述べたように $B \cup H \models O$ となる H を求めるという設定が広義の「問題」といえるが、定義で述べた狭義の「問題」を考えた場合、O に対する説明があるため、ILP におけるその設定にあたる目標概念を問題として対応づけた。
- 8) 真の値が正事例でうち予測で正のものは TP、予測で負のものは FN。真の値が負事例で予測で正のものは FP、予測で負のものは TN。適合度は、 $TP / (TP + FP)$ 。再現率は、 $TP / (TP + FN)$ 。F 値は $2 \times \text{適合率} \times \text{再現率} / (\text{適合率} + \text{再現率})$ である。
- 9) 含意関係と因果関係の違いについては、例えば、「高級マンションを買うならば金持ちである」は含意であり、「高級マンションを買うならば貧乏になる」は因果であり時間的な順序がある関係である。

謝辞

本研究は、株式会社サーベイリサーチセンターの協力と調査をもとにされました。また技術的な面については、古川康一慶應義塾大学名誉教授の助言をもとに作成しました。加えて、JSPS 科研費 16K17203 および、沖縄国際大学 特別研究費 (奨励) の助成を受けたものです。ここで感謝の意を表したいと思います。

参考文献

Birks, D. F., Malhotra, N. K. 2007 *Marketing research: an applied approach*, Pearson

Education UK.

- Breiman, L., Friedman, J.H., Olshen, J.H., Stone, C.J. 1984 *Classification and Regression Trees*. Chapman&Hall/CRC.
- Glaser, B. G., Strauss, A. L. 1. 1999 *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Aldine Transaction.
- Harvey, L., MacDonald, M. 1993 *Doing Sociology: A Practical introduction*, Palgrave Macmillan.
- Kakas, A.C., Kowalski, R.A. and Toni, F. 1992 Abductive logic programming, *Journal of Logic and Computation*, 2(6), 719-770.
- Karalić, A., Bratko, I. 1997 First order regression. *Machine Learning*, 26(2-3), 147-176.
- Kotler, P., Keller, K., Brady, M., Goodman, M., Hanser, T. 2012 *Marketing Management*, Harlow: Pearson Education.
- Kowalski, R.A. 1974 Predicate Logic as a Programming Language, *Proceedings of International Federation for Information Processing*, 556-574.
- LeCun, Y., Bengio, Y., Hinton, G. 2015 Deep learning. *Nature*, 521(7553), 436-444.
- Muggleton, S. 1995 Inverse Entailment and Progol, *New Generation Computing*, 13, 245-266.
- Plotkin, G.D. 1970 A note on inductive generalization, *Machine intelligence*, 5, 153-163.
- Punch, K.F. 2005 *Introduction to Social Research: Quantitative and Qualitative Approaches (Essential Resource Books for Social Research)*, Sage Publications Ltd, Second. (川合隆男 (監訳) 社会調査入門, 慶應義塾大学出版会 2005)
- Robinson, J.A. 1965 A Machine-Oriented Logic Based on the Resolution Principle, *Journal of the ACM*, 12:23-41.
- Sapiro, E.Y. 1982 *Algorithmic program debugging*, The MIT press.
- Srinivasan, A., Muggleton, S., King, R., Sternberg, M. 1994 Mutagenesis: ILP experiments in a non-determinate biological domain, *Proceedings of the 4th International Workshop on Inductive Logic Programming*, 217-232.
- Srinivasan, A. 1999 *The Aleph Manual*.
<http://www.comlab.ox.ac.uk/oucl/research/areas/machlearn/Aleph/>
- Wallace, W.L. 1971 *The logic of Science in Sociology*, Aldine Publishing Company.
- 新睦人 2005 社会調査の基礎理論、川島書店。
- 石川淳志 1980 社会調査の論理、社会労働研究、27(1), 49-71。
- 稲上毅 1980 現代社会学における理論と調査、季刊労働法、総合労働研究所、別冊、6, 20-29。
- 井上克己 2010 アブダクションとインダクション、人工知能学会誌、25(3), 389-399。
- 鹿又伸夫・長谷川計二・野宮大志郎 2001 質的比較分析、ミネルヴァ書房
- 川喜田次郎 1965 発想法—創造性開発のために、中央公論社。
- 栗木契 2007 構想としてのマーケティング・リサーチ、国民経済雑誌、195(5), 39-57。

金城敬太・尾崎知伸・相澤彰子 2010 調査データに基づく社会構造変化の抽出、人工知能学会論文誌, 25(3), 452-463。

金城敬太・尾崎知伸・古川康一・原口誠 2014 アナロジーを組み込んだルール発想推論によるスキル獲得支援。人工知能学会論文誌, 29(1), 188-193。

指田隆一 2008 社会学における理論構成の方法、四天王寺国際仏教大学紀要、45。

佐藤健二・山田一成（編）2009 社会調査論、八千代出版。

真鍋一史 2003 ファセット・アプローチにもとづく調査票設計とデータ解析の試み、行動計量学、30, 53-69。

見田宗介 1965 現代日本の精神構造、弘文堂。

古川康一・尾崎知伸・植野研 2001 帰納論理プログラミング、共立出版。

安田三郎・原純輔 1982 社会調査ハンドブック、第3版、有斐閣。

補遺1. 本研究で使用した調査票の質問文（利用箇所を記載。また「不安」についての具体的な回答と仮説の対応については回答文の右のカッコに記載してある。）

Q あなたは、今回の地震（東日本大震災）が発生した後、次のような商品を購入しましたか。（それぞれ回答はひとつ）

「水」

1. 通常よりも多めに買った
2. 通常よりも、多めに買ったかったが、買えていない
3. 通常と変わらない
4. 買ってない

Q 地震の後に不安なことについてお答えください。（回答はいくつでも）

1. なぜ被害の全体像がなかなか把握できないのが不安だ（情報不足）
2. まだ安否の確認できない人がたくさんいることが不安だ（情報不足）
3. 被災者に物資・支援が届いていないことが不安だ（モノ不足）
4. モノ不足で、必要なものが手に入らなくなるのが不安だ（モノ不足）
5. 余震がたくさん続いているので不安だ（地震不安）
6. 福島第一原子力発電所の今後の推移が不安だ（原子力事故）
7. 全国の原子力発電所の安全性が不安だ（原子力事故）
8. 日本が、地震の活動期に入ったのではないかと不安だ（地震不安）
9. なんとなく不安だ（漠然とした不安）
10. 不安なことはない

補遺2. ILP とロジスティック回帰との比較

ILP と他の分析手法の比較のために本研究で用いたデータに対してロジスティック回帰を用いて

分析した結果について述べる。変数については上記と同様に被説明変数として買いだめ行動の有無、説明変数として各種の不安、年齢、地域を設定し、それぞれダミー変数に変換して分析を行った。なお、本研究では、仮説と作業仮説の対応に関しては作業仮説のいずれかの項目に回答していた場合は仮説への回答を1、そうでない場合は0と設定した。具体的に述べる。「不安」測定した変数と上位の概念との対応関係は、検証的因子分析などを行って変数を作成する方法や、質問項目をつくり、その合計得点化をする方法など様々なものが想定されるが本研究ではILPと比較しやすくするためILPと同様に論理和としての対応、すなわち測定した変数のいずれかが成立していれば上位概念も成立しているという対応を採用している。また、ILPと同様に情報量規準を用いたステップワイズ法により変数選択を行った。

分析した結果は以下ようになった。

	係数	標準誤差	z値	P値	有意水準
定数	-3.172	0.634	-5.000	0.000	***
性別	0.264	0.179	1.476	0.140	
年齢	-0.031	0.007	-4.719	0.000	***
埼玉県	1.572	0.562	2.796	0.005	**
千葉県	1.956	0.550	3.559	0.000	***
東京都	2.039	0.547	3.727	0.000	***
神奈川県	1.751	0.556	3.149	0.002	**
静岡県	1.260	0.579	2.178	0.029	*
愛知県	1.414	0.571	2.478	0.013	*
大阪府	0.695	0.623	1.116	0.264	
広島県	0.218	0.680	0.320	0.749	
福岡県	0.812	0.611	1.329	0.184	
地震不安	0.340	0.191	1.782	0.075	.
Residual deviance:	990.05	on 1987 degr freedom			
AIC: 1016.0	***'	0.001'	**' 0.01'	*' 0.05'	' 0.1

補遺 図 ロジスティック回帰の結果

推定したロジスティック回帰を用いて、買いだめを行う確率を推測し、0.5以上を買いだめを行う、それ以下を行わないとして予測を行った。結果、F値は0（精度0、再現率1）であった。これは、もともと正事例の数が少なかったことも影響し、正事例のうち説明できたものはひとつもなく、一方、負の事例のうちルールで説明したものは1849で説明できたのは1849であったためである。カイ2乗値は1441.60、自由度は1、p値は2.2e-16であった。つまり、正事例の明確な予測ができなかった。

本研究で利用した調査データでは、ILPとロジスティック回帰を用いた結果では異なる結果を得

ることができた。また ILP ではロジスティック回帰では得ることができなかった地域と年齢と性別の関係、不安に関する複雑な構造を持っていることを把握することができた。特に、ILP で得られたルールに条件が二つ以上あるものについては、その条件が AND 条件つまり論理積であることを意味しており、これは統計における交互作用（3 項以上も含む）に該当するがこれを自動的に抽出することができたことも違いとして挙げられる。