

因果概念束にもとづく知識構造モデルの研究

鈴木 治

A Study of the Knowledge Structural Model Based on Causal Concept Lattice

SUZUKI Osamu

要 旨

形式概念分析は、概念とその概念構造に関する数学的理論である。本研究は形式概念分析を応用し、説明文等に含まれる因果関係を概念束に組み込み、その因果概念束の特性を明らかにする。本論文では、概念束に因果関係を組み込むための理論的考察を述べ、さらに説明文による 1,998 命題の因果概念束をもとに、演繹、帰納及びアブダクションなどの論理的推論への応用を試みる。

キーワード：形式概念分析、因果概念束、因果関係、知識構造モデル

Abstract

Formal concept analysis is a mathematical theory of concepts and conceptual hierarchies. In this study, we apply formal concept analysis to causal concept lattices which embed causal relationships involved in explanatory texts, and reveal the property of causal concept lattices. In this paper, we describe theoretical considerations of causal concept lattices applied to deductions, inductions and abductions using 1998 propositions of explanatory texts.

Keywords : formal concept analysis, causal concept lattice, causal relationship, knowledge structural model

1. はじめに ^{(5),(7),(9),(11),(12)}

形式概念分析は、概念束とよばれる概念とその概念構造に関する数学的理論である。その応用分野は、社会学、医学、生物学、言語学、情報工学など多岐にわたる。形式概念分析の特徴は、人間の思考に近い「概念」を形式化し、ものごとの概念化や概念の構造化及びその視覚化を容易にすることである。

近年、機械学習の進展は目覚ましく、画像や音声等の認識能力は人間を凌ぐ程になり、社会のあらゆる場面での応用が期待されている。しかし反面、新井紀子氏は著書の中で、論理的思考を要する場面では、確率と統計を基礎とする現在の人工知能には限界があり、一方、人間の子供が人工知能に負けないためには、推論などを含む7つの能力が必要であると述べている。裏返せば、これは現在の人工知能にはこれら7つの能力が欠けており、極論すれば、論理的な推論能力を持たない限り、今の人工知能は人間の子供の知能を超えられないことになる。人工知能のブームは、これまで推論やルールの研究からニューラルネットワークを用いた深層学習へと移った。しかし、本研究では、大量の数値化データのみ依存するのではなく、人間が持つ論理的な推論能力とそれを生み出す知識構造に着目し、人の思考により近い知能を模索する。

子供の学習による知能の発達著しく、その知識構造の形成過程には興味深いものがある。子供は日々獲得する知識を過去の知識と融合し、それらを組み合わせて新たな知識を生み出している。D.Albert と J.Heller らは、子供の学習過程の研究において、束構造を用いた知識空間を導入し、学習による知識構造の形成過程を説明した。本研究はこれをヒントに、因果概念束を知識構造と捉え、質問への回答を通して、知識の生成過程とその構造を視覚化する。具体的には、小学生国語の読解用説明文をもとに因果概念束を生成し、その因果概念束を知識構造とみなして、読解問題に対してどの程度人間に近い回答が得られるかを検証する。本研究は、データマイニング等による因果関係の発見ではなく、既知の因果関係を構造化し、それを推論等に应用することを目的とする。なお、本論文における引用文献の表記形式は、参考文献 (14) に示す学会誌に準拠する。

2. 形式概念分析と因果概念束の定義

2.1 形式概念分析 ^{(1),(2),(3),(4),(14)}

形式概念分析は文脈 (context) をもとに、外延と内包からなる形式概念 (formal concept : 概念と略す) を構成し、これを概念束 (concept lattice) として構造化

する分析手法である。ここで A_j を外延、 B_j を内包とすると、概念束には上位概念と下位概念が存在し、それらは *The Basic Theorem on Concept Lattices* (概念束の基本定理) により式 (1) と式 (2) を用いて次のように定義される。

上位概念 :

$$\bigvee_{j \in I} (A_j, B_j) = \left(\left(\bigcup_{j \in I} A_j \right)'' , \bigcap_{j \in I} B_j \right) \quad (1)$$

下位概念 :

$$\bigwedge_{j \in I} (A_j, B_j) = \left(\bigcap_{j \in I} A_j , \left(\bigcup_{j \in I} B_j \right)'' \right) \quad (2)$$

形式概念分析の *Attribute Implication* (属性の包含関係) は、Agrawal らが提唱した *Association Rules* と同等機能を持ち、概念束を応用したデータマイニングの一つである。これは、式 (3) の擬似内包を用いて L を求め、式 (4) の $L^*(X)$ によりルール抽出を行うアルゴリズムである。但し、 $L^*(X)$ は式 (5) を再帰的に計算して求められる。

$$L := \{P \rightarrow P'' \setminus P \mid P \text{ は擬似内包} \} \quad (3)$$

$$X^{L^*} := X \cup \bigcup \{C \mid A \rightarrow C \in L, A \subseteq X, A \neq X\} \quad (4)$$

$$X^{L^*L^*} := X^{L^*} \cup \bigcup \{C \mid A \rightarrow C \in L, A \subseteq X^{L^*}, A \neq X^{L^*}\} \quad \dots(5)$$

また、概念束が式 (6) を満たすとき、一般に概念束 $\underline{\mathcal{B}}_L$ の部分束 $\underline{\mathcal{B}}_M$ を部分概念束 (sublattice) とよぶ。

$$\begin{aligned} \emptyset \neq \underline{\mathcal{B}}_M \subseteq \underline{\mathcal{B}}_L \text{ のとき、} a, b \in \underline{\mathcal{B}}_M \text{ ならば、} \\ a \vee b \in \underline{\mathcal{B}}_M \text{ かつ } a \wedge b \in \underline{\mathcal{B}}_M \end{aligned} \quad \dots(6)$$

これに加え、概念束 $\underline{\mathcal{B}}_L$ の部分概念束ではない束を本研究では部分束とよぶ。

2.2 用語の定義 ^{(2),(14),(15)}

形式概念分析の用語定義に加え、本論文では以下の定義を用いる。

(1) 文脈と命題

一般に、説明文には経験的な事象が含まれる。本論文は説明文を「命題」と考え、 P_1 や $\{P_1 \rightarrow P_2\}$ (P_1 ならば P_2) などと表記する。ここで言う命題とは、論理的な真偽ではなく、事象の生起に関わる命題を指す。本研究における文脈は、命題の主部 (主語) と述部 (述語) の主要単語 (キーワード) を抽出した外延 (主部) と内包 (述部) により構成される。文脈の構成は複数と単数の場合があるが、本論文ではまとめて $K := (G, M, I)$ (K と

略す)と記す。ここで、 G は外延、 M は内包、 I は両者間の関係を示す(以下同様)。文脈を構成する外延と内包の間には、ガロア結合(Galois Connections)が存在するものとする。また、命題間における因果関係は文脈に引き継がれ、その因果関係には仮言命題や定言三段論法などが含まれる。

(2) 概念

文脈は形式概念分析の適用により因果概念(causal concept: 誤解のない限り、概念と略す)に写像される。概念は外延と内包の組で構成され、ここでは $\underline{B}(G, M, I)$ (\underline{B} と略す)と記す。本研究では推論などへの応用を考慮して、外延と内包に加え、概念には命題(説明文)や元の文脈(後述)及び概念発生頻度などの情報も含める。因果概念束(誤解のない限り、概念束と略す)は要素として概念を持ち、その概念は命題概念と、概念束のTopped \cap -Structure(\cap -構造と略す)により生成される帰納概念から構成される。概念束は、包含関係を順序とする完備束であり、概念と同様に $\underline{B}(G, M, I)$ (\underline{B} と略す)と記す。本研究では、概念束はデータベースを用いて永続化する。

3. 因果概念束による知識構造モデル

3.1 因果関係とその表現^{(2),(6),(10),(17)}

哲学における原因と結果の成立要件には、「規則性説」、「因果プロセス説」、「反事実的依存性説」、「決定説」、「確率上昇説」などの諸説がある。しかし、いずれの説においても、因果関係の表現方法としてニューロン・ダイアグラムが多く用いられる。このダイアグラムは、事象を円、発生済み事象を塗りつぶし円、そして事象間の因果関係を矢印により表現する簡易な図式である。ニューロン・ダイアグラムは、情報分野における「アローダイアグラム」や「PERT図」に類似する。因果関係の哲学的考察は本研究の対象外であるが、ニューロン・ダイアグラムは事象間の因果関係を曖昧なく表現できるため、本研究では命題(説明文)を文脈に変換するときこれをを用いる。

いま、命題 P_i と P_j において、 P_i の原因により P_j が生起する関係を $\{P_i \rightarrow P_j\}$ と表記する。例として、ここで式(7)に示す命題集合を考える。

$$P := \{P_1 \rightarrow P_2, P_2 \rightarrow P_5, P_3 \rightarrow P_4, P_4 \rightarrow P_5, P_5 \rightarrow P_6, P_5 \rightarrow P_7\} \quad \dots(7)$$

式(7)をニューロン・ダイアグラムにより図示すると図1になり、この図には命題間の関係が明確に示されている。

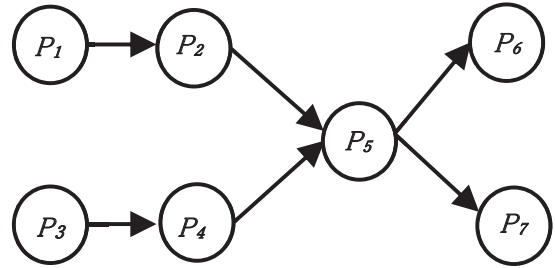


図1 ニューロン・ダイアグラム

本研究では、命題の因果関係は文脈 $K=(G, M, I)$ に組み込まれ、その文脈に対して形式概念分析が適用される。以下では、因果関係を文脈に組み込む方策を理論的に考察する。

形式概念分析におけるAttribute Implicationは、確信度100%のルール抽出であり、Association Ruleは確信度100%未満のルール抽出である。本研究では、因果概念束はAttribute Implicationと整合性を保つことを前提とする。概念束のAttribute Implicationは式(3)の擬似内包 P を用い、これに式(5)を再帰的に適用して式(4)を満たす属性集合を求める。この手法により、 $\{P \rightarrow C\}$ (P ならば C)なるルール(因果関係)が抽出されるが、式(3)の P'' は P の属性閉包(closure)であることから、概念束では P の概念は C の概念の下位に位置する。つまり、命題の因果関係 $\{P \rightarrow C\}$ は、概念束では下位概念から上位概念へと遡る関係と見ることができ。また、概念束において、原因となる下位概念は結果となる上位概念の内包を含んでおり、これは哲学における必然性の考え方と整合性が取れている。

以上の理論的考察をもとに、因果関係を明示的に概念束へ埋め込むためには、原因または前提となる概念を式(2)の下位概念に、結果を式(1)の上位概念にそれぞれ対応させればよい。

例えば、隣接する命題間の因果関係 $\{P_1 \rightarrow P_2\}$ の場合、それぞれの文脈を $K_1 := (G_1, M_1, I_1)$ 、 $K_2 := (G_2, M_2, I_2)$ 、その概念を $\underline{B}_1(G_1, M_1, I_1)$ 、 $\underline{B}_2(G_2, M_2, I_2)$ とすると、 \underline{B}_2 を \underline{B}_1 の隣接上位概念とする因果概念束は、式(8)に示す束論における垂直和(Vertical Sum)を用いて構成できる。

$$\underline{B}_1 \oplus \underline{B}_2 = \underline{B}(G_1 \cup G_2, M_1 \cup M_2, I_1 \cup I_2 \cup (G_1 \times M_2)) \quad \dots(8)$$

但し、因果関係が式(8)を満たすためには、式(9)及び式(10)による文脈間の変換が必要になる。

$$K_1 := (G_1, M_1, I_1) \mapsto K_1' := (G_1, M_1 \cup M_2, I_1 \cup (G_1 \times M_2)) \quad (9)$$

$$K_2 := (G_2, M_2, I_2) \mapsto K_2' := (G_1 \cup G_2, M_2, I_2 \cup (G_1 \times M_2)) \quad (10)$$

この文脈間の変換により、 K_1 と K_2 の外延と内包は元の

文脈とは異なるものになる。後で述べるが、元の文脈は問合せ質問の回答を得るのに必要となるため、概念がもつ情報の一つとして保持される。

次に、因果関係組込みの具体的方策を述べる。本研究では、式 (9)、式 (10) を用いた因果関係組込みのために、2つの指定子を導入する。それらは、結論（上位概念）を指定する「 φ 指定子」と前提（下位概念）を指定する「 ϕ 指定子」である。例えば、 $\{P_1 \rightarrow P_2\}$ の場合、それぞれの文脈 K_1 と K_2 のどちらかに対して、次式 (11) に示す φ 指定子または ϕ 指定子を適用する。

$$K_1 := (G_1, M_1, I_1, \varphi G_2) \quad \text{または} \quad K_2 := (G_2, M_2, I_2, \phi G_1) \quad \dots(11)$$

因果概念束の生成過程において、複雑な命題群は一旦、ニューロン・ダイアグラムに置き換えると述べた。文脈を作成する場合、ニューロン・ダイアグラムを参照して上位/下位概念指定子を適用するのは、因果関係を間違えずに文脈に組込むための手段である。

補足として、因果関係の適用について本研究における考え方を示しておく。本研究では、説明文に限らず、いろいろな分野の命題に因果関係を適用することにより、その応用範囲を広げたいと考えている。このため、「原因と結果」のみに限らず、本研究で扱う因果関係には「前提と結果」、「前事象と後事象」、「過去と未来」、「具体化と抽象化」なども含めることにする。

ここで、図 1 に示すニューロン・ダイアグラムを例にとって、具体的に因果概念束を説明する。この例では、図 1 に示した命題 P_1 の文脈を $K_1 := (G_1, M_1, I_1)$ 、その概念を $\underline{B}_1(G_1, M_1, I_1)$ (G_1 は外延、 M_1 は内包) とし、命題 P_2 の文脈 K_2 が、例えば、次のような外延と内包を持つ文脈であったとする。

$$\begin{aligned} K_1 &:= (\{g_1\}, \{m_1\}, I_1), K_2 := (\{g_2, g_{24}\}, \{m_2\}, I_2), \\ K_3 &:= (\{g_3\}, \{m_3\}, I_3), K_4 := (\{g_4, g_{24}\}, \{m_4\}, I_4), \\ K_5 &:= (\{g_5\}, \{m_5\}, I_5), K_6 := (\{g_6\}, \{m_6, m_{67}\}, I_6), \\ K_7 &:= (\{g_7\}, \{m_7, m_{67}\}, I_7) \end{aligned}$$

これらの文脈に対して、形式概念分析を適用すると、図 2 に示す因果概念束が得られる。図 2 において、「●」は命題概念を、「○」は帰納概念を示す（以下、同様）。図 1 と図 2 を比較すると、ニューロン・ダイアグラムに示される原因ノードは図 2 では下位に、また結果ノードは上位に配置されていることが確認できる。 $\underline{B}_1 \sim \underline{B}_7$ は図 1 と同型であり、その因果関係は概念束上で適切に構造化されている。また、例えば、図 2 の因果関係の連鎖 $\underline{B}_1 \rightarrow \underline{B}_2 \rightarrow \underline{B}_5 \rightarrow \underline{B}_6$ は、概念束の下位から上位に向かって概念が連鎖している。図 1 とこれを重ね合わせると、 \underline{B}_1 から上位方向に辿ることは演繹の推論に相当し、逆

に、 \underline{B}_6 から下位方向に辿ることはアブダクションに相当すると解釈できる。一方、図 2 には、図 1 に存在しない概念 \underline{B}_{67} と \underline{B}_{24} が表示されている。これらは「概念束の基本定理」(式 (1) 及び式 (2)) により生成されたもので、 \underline{B}_{67} は内包 \cap -構造、 \underline{B}_{24} は外延 \cap -構造によるものである。



図 2 図 1 の因果概念束

図 2 の \underline{B}_{67} と \underline{B}_{24} は、次のような外延と内包をもつ。

$$\begin{aligned} \underline{B}_{67} &:= (\{g_1, g_2, g_{24}, g_3, g_4, g_5, g_6, g_7\}, \{m_{67}\}) \\ \underline{B}_{24} &:= (\{g_{24}\}, \{m_2, m_4, m_5, m_6, m_{67}, m_7\}) \end{aligned}$$

本研究では、これらの概念を帰納概念とよび、命題概念と区別する。上記の外延と内包を命題の観点からみると、 \underline{B}_{67} は \underline{B}_6 と \underline{B}_7 の内包の積集合及び外延の和集合をもつ汎化概念、 \underline{B}_{24} は、 \underline{B}_2 と \underline{B}_4 の外延の積集合及び内包の和集合をもつ特化概念と考えられる。一方、図 2 の概念ノードの外延と内包は元の文脈とは異なっている。これは式 (9)、式 (10) による因果関係の組込みによって生じた差異であり、因果概念束における特徴の一つである。このように、因果関係が組込まれた因果概念束は、命題（文脈）の単なる集合ではなく、概念束全体が集合演算によって関係付けられた束構造を成している。後で述べるように、この束構造は本研究の範囲内において、命題（説明文）間の演繹や帰納そしてアブダクションなどへの応用に利用される。

3.2 因果概念束の生成^{(8),(14),(15)}

形式概念分析の概念束生成アルゴリズムには、「直観的アルゴリズム」、「Next Closure アルゴリズム」、「逐次更新アルゴリズム」がある。それぞれには一長一短があり、概念束がどのように利用されるかにより、これらを使い分ける必要がある。本研究では、因果関係を含む命題は一旦文脈に変換され、これに形式概念分析が適用

されるが、一般的なデータと異なり、命題を元にする文脈には、名詞や動詞、形容詞、副詞などの多種類の属性が含まれる（表1中段参照）。このため、因果関係のない概念間の Π -構造は、因果概念束上に離散して配置される。一方、因果関係を持つ概念間の Π -構造は、式(8)の包含関係により、局所的な概念束として因果概念束上に配置される。このような因果概念束の特性を考慮すると、因果概念束の生成・更新処理はR.Godinらが考案した「逐次更新アルゴリズム」が適している。その理由は、この方式では既存の概念束が永続化されており、新たに文脈が追加されても関連する概念だけが影響を受け、因果概念束の更新が局所的な処理に留まるからである。

以上をもとに、ここで因果概念束の生成・更新に関する理論的考察を述べる。いま、既成の因果概念束があるとして、これに新たな文脈を追加する場合を考える。但し、概念束には式(1)、式(2)の関係があることを念頭に置く。既成の概念束を $\underline{B}_L(G_L, M_L, I_L)$ 、追加する文脈を $K_X=(G_X, M_X, I_X)$ とすると、本研究では検索条件式(12)を用いて、 K_X の追加に必要な \underline{B}_L の部分束 $\underline{B}'(G', M', I')$ を抽出する。

$$(\bigvee_{i \in I} g_i) \vee (\bigvee_{j \in J} m_j) \quad (12)$$

但し、 $g_{i \in I} \in G_X$ 、 $m_{j \in J} \in M_X$ とする。

ここで、式(12)は \underline{B}_L を構成する概念の外延と内包に対して適用する。式(12)の検索により、外延要素 g_i を含む \underline{B}_L の部分束と内包要素 m_j を含む \underline{B}_L の部分束との和集合 $\underline{B}'(G', M', I')$ がデータベースから抽出され、その \underline{B}' に対して追加文脈 K_X による形式概念分析が適用される。このとき、 \underline{B}_L から \underline{B}' を除いた因果概念束は、 g_i 及び m_j を含まないため、 K_X による概念束更新による影響を受けない。つまり、この方法は、 \underline{B}_L 全体に対する K_X の形式概念分析の適用と等価になる。一方、大量の文脈をもつ K_X に対し、一括して形式概念分析を適用すると、 G_X 及び M_X の要素個数が多くなり、 \underline{B}' 抽出条件がSQL仕様を超える恐れがある。このため、本研究では文脈のまとまりである「段落」を \underline{B}_L の更新単位とした。段落は意味的にまとまりをもつ文脈集合であるため、段落を超えた文脈間の結びつきは比較的少ない。従って、 K_X を段落に分割して \underline{B}' の抽出を行えば、外延と内包の検索条件は緩和され、 \underline{B}' の読み込みと \underline{B}_L の更新は、 K_X の一括処理に比べ処理が簡素化される。但し、段落を超える因果関係を考慮して、段落はなるべく上位/下位概念指定子の有効範囲を超えないことが望ましい。

3.3 因果概念束の応用^{(2),(6)}

因果概念束を知識構造とみなすと、その知識要素となる概念の蓄積とその構造、概念個数及び概念束からの知

識抽出など、因果概念束には応用面での興味深い点が多く含まれる。以下では、生成された因果概念束の応用について考察する。

(1) 因果概念束への問合せ質問とその種類

因果概念束の特性から、大別して4つの問合せ質問への応用が考えられる。1つ目は、問合せ質問に対する直接的な回答を求める場合である。これは、因果概念束から質問回答に適する説明文をもつ概念を抽出すればよい。2つ目は、“原因”に対する“結果”を求める問合せの場合である。因果概念束には命題間の因果関係が組込まれており、結果となる命題は概念束の外延を用いて上位概念を辿ることにより求められる。これは命題における演繹的推論に相当する。3つ目は“結果”からその“原因”を求める問合せの場合である。これはアブダクションに相当し、因果概念束では概念束の内包を用いて結果概念から下位概念を辿ることにより、信頼度100%未満の原因命題を求めることができる。4つ目は、因果概念束で最も強力な帰納概念の利用である。因果概念束には、帰納概念として命題間の汎化概念と特化概念が存在する。例えば、複数命題が同じ述部を共有する場合、それらの概念は因果概念束に帰納概念としてまとめられている。例として、「空を飛ぶ」のような述部を持つ複数命題があった場合、因果概念束には「飛行機」「鳥」などを外延とする帰納概念が存在し、述部による命題汎化が行われる。さらに、命題汎化による帰納概念にはもう一つの特徴がある。それは因果関係による帰納概念には、その結果に至る原因の外延要素が含まれることである。つまり、ある述語属性をもつ帰納概念に着目すると、その外延には原因となった命題群の主部が含まれ、どんなオブジェクトが関与してその結果を招いたか、因果関係を個々に追わなくても把握できる。ある命題の原因を探す問合せの場合、全ての命題から原因を探し出すのは容易なことではないが、帰納概念を用いれば質問への回答は容易に得られる可能性がある。

(2) 問合せ質問に対応する部分束の抽出

問合せ質問に対して、どの様に因果概念束を用いて回答を導き出すか、ここではその方策について考察する。いま、因果概念束を $\underline{B}_L(G_L, M_L, I_L)$ 、問合せ質問の文脈を $K_Q=(G_Q, M_Q, I_Q)$ とする。本研究では、命題と同様に、質問も文脈によって表現する。問合せ文脈 K_Q が与えられたとき、本研究では式(13)を用いて、その回答を含む部分束 $\underline{B}'(G', M', I')$ を \underline{B}_L から抽出する。

$$((\bigwedge_{i \in I} g_i) \wedge (\bigvee_{j \in J} m_j)) \vee ((\bigvee_{i \in I} g_i) \wedge (\bigwedge_{j \in J} m_j)) \quad \dots(13)$$

但し、 $g_{i \in I} \in G_q$ 、 $m_{j \in J} \in M_q$ とする。

ここでは、式(13)は \underline{B}_I の因果関係組込み前の文脈に対して適用する。この検索により、 $K_q=(G_q, M_q, I_q)$ の上界(upper bound)と下界(lower bound)をもつ部分束 \underline{B} が抽出され、 $K_q=(G_q, M_q, I_q)$ と距離の近い概念の中から問合せ質問に最も適合する候補概念を求めることができる。本研究では、概念間の距離は外延間と内包間のJaccard係数を用い、外延と内包のユークリッド距離により計算する。この結果得られる候補概念は、ほとんどの場合質問に適合する。しかし、たとえ適合しない場合でも、候補概念はその近隣に存在するため、本研究では \underline{B} を表示して目視により候補概念を選択できるようにした。一方、質問内容が不適切な場合や曖昧な場合もある。この場合は、質問の都度、 \underline{B} の概念個数を表示し、それがゼロまたは個数過大の場合には、回答が存在しない不適切な質問として質問者に再考を促す。

(3) 部分束 \underline{B} の sublattice 化

一般に(2)で述べた \underline{B} は $K_q=(G_q, M_q, I_q)$ を含む部分束であり、必ずしも部分概念束 $\underline{B}'(G', M', I')$ ではない。つまり、 $\underline{B}_Y \subseteq \underline{B}'$ かつ $\underline{B}_Y \not\subseteq \underline{B}$ なる \underline{B}_Y が存在し得る。例えば、 $G_q = \phi$ かつ $M_q = \{a, b\}$ のとき、 $M' = \{\{a, x\}, \{b, x\}\}$ 、 $M'' = \{\{a, x\}, \{b, x\}, \{x\}\}$ の場合がそれであり、 $\{x\}$ は \underline{B} には含まれない。これは、 \underline{B} が式(6)に示す sublattice ではないことに起因する。そこで、本研究では、上記のような $\{x\}$ を含ませるために、 \underline{B} の抽出後、再度形式概念分析を適用して sublattice 化を図っている。

(4) 段落を超えた演繹的推論

同一命題が異なる段落の別の命題と因果関係を持つとき、段落を超えた因果関係が検出されない場合がある。例を用いてこれを説明する。例えば、段落A: $\{P_1 \rightarrow P_2 \rightarrow P_3\}$ 及び段落B: $\{P_2 \rightarrow P_4\}$ があったとする。このニューロン・ダイアグラムを図3に示す。

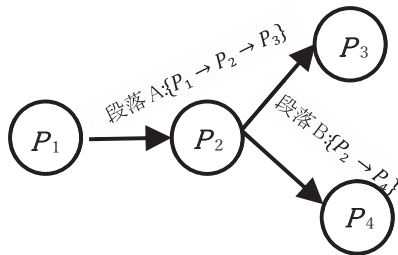


図3 段落の異なる因果関係

段落Aと段落Bは、別々に形式概念分析が適用される。このため、段落Aでは P_2 と P_3 は検出されるが、段落Bの P_4 は検出されない。これが段落Aでの $\{P_2 \rightarrow P_4\}$ 未

検出の原因である。段落Bについても、同様である。ところがこの場合、因果概念束においては $\{P_2 \rightarrow P_3\}$ と $\{P_2 \rightarrow P_4\}$ を汎化する帰納概念が生成される。本研究では、この帰納概念を用いて同一命題(「●」表示)を質問者に知らせる。問合せ質問において、この同一命題を示す帰納概念が検出された場合、式(13)を用いてその帰納概念を再読み込みすれば、 $\{P_2 \rightarrow P_3\}$ 及び $\{P_2 \rightarrow P_4\}$ を検出することができる。

(5) 因果概念束の部分束による簡潔な回答表示

問合せ質問の回答には、その内容により望ましい表現形式が求められる。例えば、演繹的推論やアブダクションなどを回答とする場合、因果概念束の命題概念のみを抽出した部分束は、回答となる命題(説明文)連鎖を簡潔に表現することができる。本研究では、このような部分束を命題概念部分束とよび、同様に、帰納概念のみによる部分束を帰納概念部分束とよぶことにする。

4. 因果概念束による知識構造の検証

4.1 知識構造としての因果概念束の検証^{(13),(16)}

これまで述べた因果概念束の生成・更新とその応用に関する考察をもとに、ここでは実際に説明文を用いて因果概念束を検証する。本研究では、小学生が使用する読解用国語ドリルの説明文をもとに因果概念束を生成し、ドリルに付随する読解問題を問合せ質問として使用する。今回、国語ドリルを使用した理由は、説明文を命題とする因果概念束が知識構造としてどの程度小学生の回答に迫れるか検証するためである。検証用ソフトウェアは今後の応用研究を考慮し、全て独自開発を行っている。(表1下段参照)

また、小学1年生から6年生までの説明文は、学年ごとに質、量ともに増加するが、その増加が因果概念束にどのような影響を与えるかも本研究の重要なテーマの一つである。

表1に今回使用したデータの諸元を示す。小学1年から6年までの命題数は1,998命題あり、学年毎の累計命題数は図4に示す通りである。図4において、4年生から命題増加率がやや下がっているのは、国語ドリルの出版社2社のうち1社が4年生以降の出版を行っていないためである。

表1の上段に示した命題に、形式概念分析を適用した結果を表1中段に示す。今回は、表1中段の帰納概念個数には、命題特化概念個数は含めていない。その理由は、表1上段に示したように、外延要素個数(平均)が1.06個と少なく、生成される命題特化概念の個数も少ないためである。また、表1の命題数と命題概念個数に差があ

るのは、1,998 命題中に同一命題が含まれているためである。

次に、命題数と概念総個数の関係を図5に示す。因果概念束においては、新たな文脈（命題）が追加されると全ての既存概念との集合演算が行われ、その結果、新規概念（一つとは限らない）が生成されると、それらは因果概念束に追加される。

しかし、図5を見る限りでは、命題数に対する概念総個数の増加は線形的である。これを知識構造に置き換えて考えると、知識の追加は知識量全体を線形的に増加させることを示唆する。この結果はやや意外であったが、しかし人間の場合でも、知識は年齢とともに増加するが、だからと言って、年齢を重ねても知識量が指数関数的に増加するとは考えにくいことと類似する。

また、概念総個数における帰納概念の占める割合を図6に示す。図6から、帰納概念の構成比率は命題数の増加または学年進級に関わらず、約0.6と一定であること

が分かった。帰納概念は文脈の内包汎化に関わる概念であり、関連する命題間の共通性を示すものである。つまり、帰納概念は、知識構造の観点から見れば、学習した複数知識の「まとめ」に相当するものである。この帰納概念の個数割合が一定であるということは、命題数が増加しても概念個数全体は急激な増加傾向を示さないことを意味する。

表1上段に示す命題（説明文）により得られた因果概念束全体の外観を図7に示す。同図は因果概念束の複雑性を示すものであり、小学6年生までの因果概念束全体のHasse図は視覚化がほぼ不可能であり、因果概念束の利用は、その局所化が不可欠であることを示している。

表1 データ諸元

諸元	詳細
命題 (説明文)	命題数:1,998 命題 外延要素個数(平均):1.06 個 内包要素個数(平均):4.21 個 命題平均文字数:28.6 文字 段落個数:262 段落 1段落当たりの命題数:7.6 命題
因果概念束	概念総個数:5,385 個 全オブジェクト数:549 個 全属性数:2,958 個 帰納概念個数:3,411 個 命題概念個数:1,974
問合せ	質問数:1,144 問
検証用 ソフトウェア	開発環境:Java1.7、ACCESS 2019 ソフトウェア構成: 形式概念分析処理、文脈変換処理、 概念束表示処理、問合せ処理(命題 概念部分束、帰納概念部分束等)、 ルール抽出処理、DB 処理など

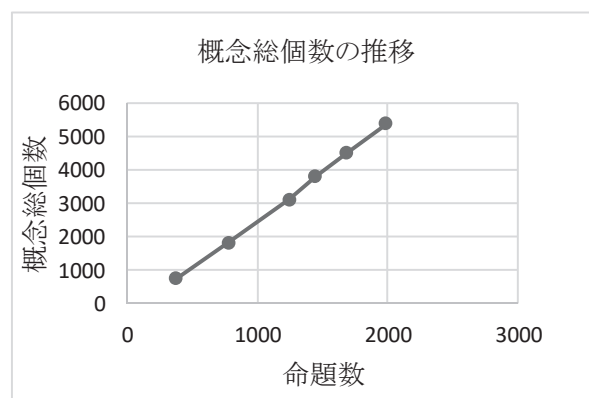


図5 命題数と概念総個数の関係

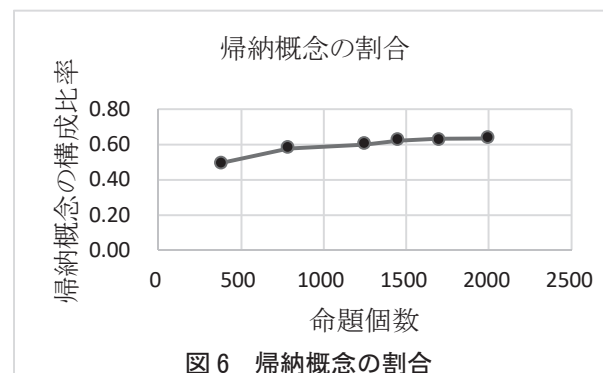


図6 帰納概念の割合

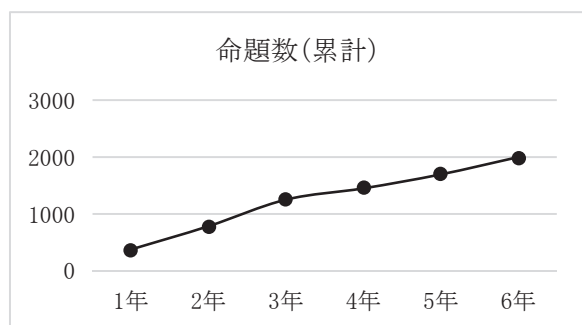


図4 学年毎の累計命題数

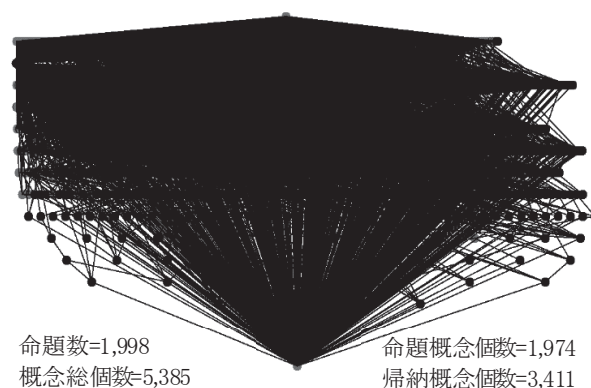


図7 小学6年生までの因果概念束

4.2 因果概念束への問合せ検証^{(13),(14),(16)}

本節では、因果概念束を知識構造とみなし、これに問合せ質問を行った結果を述べる。本研究では、問合せ質問に関する因果概念束の検証を優先したため、ユーザーインターフェースは最小限に留めている。(表1の下段を参照)

ここで、質問への直接回答、演繹的推論、帰納的推論及びアブダクションの4種類について因果概念束を検証する。1つ目は、質問キーワードに適合する概念を求めるもので、最も簡単な問合せである。2つ目の演繹的推論は、因果関係の連鎖を求める例であり、*Douglas Kutach* のいう因果プロセスに近いものである。3つ目の帰納的推論は、命題間の汎化や特化に関する問合せであり、ここでは帰納概念の命題汎化の事例を取上げる。4つ目は結果からその原因を求めるアブダクションに関する例であり、これには因果概念束の上位概念から下位概念への確率的推論を用いる。但し、1つ目は、式(13)を因果概念束に適用し、直接回答となる説明文を得る単純なものであるため、ここではその検証事例は省略する。以下では残り3つの問合せに対して、具体的データを示してその有用性を検証する。

演繹的推論と帰納的推論の検証

因果概念束に対して、次の問合せ質問を行ってみる。

「紙風船と板を同時に落とすとどうなりますか？」

この問合せの文脈（{紙風船, 板}, {落とす}）を入力して、因果概念束に問合わせると、図8に示す部分束が返ってくる。ここでは表示の煩雑さを避けるため、外延、内包ともに差分表現を用いる。図8に示す①は、この質問に対する回答には2系統の命題連鎖が含まれることを示している。

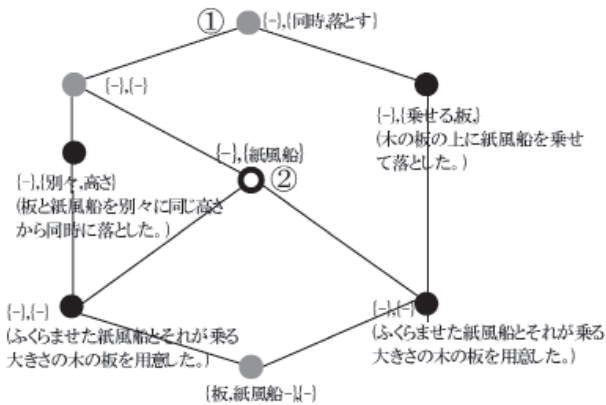


図8 問合せ質問の部分束事例

つまり、紙風船と板の実験には「板と紙風船を別々に落とす実験」と「板の上に紙風船を乗せて落とす実験」が

あり、①は実験に関する帰納的役割を持つ概念となっている。図8に示す②は、3.3節で述べた段落を跨ぐ因果関係であり、同一の原因から異なる結果への因果関係が存在することを示す。問合せ質問の回答を得るために、同図①(または②)を選択し、式(13)による再読込み後、その命題概念部分束を表示する。すると図9①に示される「落とし方により、落ちる速さは異なる」が回答として返される。図中の点線は指定概念の下界を示す(以下、同様)。この回答は質問の答としては正しいが、さらに回答の根拠が示されれば、納得が得やすい。同図には、その回答の根拠となる2つの演繹的推論が並置され、それらは②「別々に落とせば板が早く落ち」、③「板の上に風船を乗せて落とせば同時に落ちる」を含む。小学生も紙風船と板の落ち方を別々の知識として持つのではなく、両者を並置して理解すると思われる。これと同様に、因果概念束を用いたこの回答事例においても、根拠となる命題連鎖が並置して表示される。

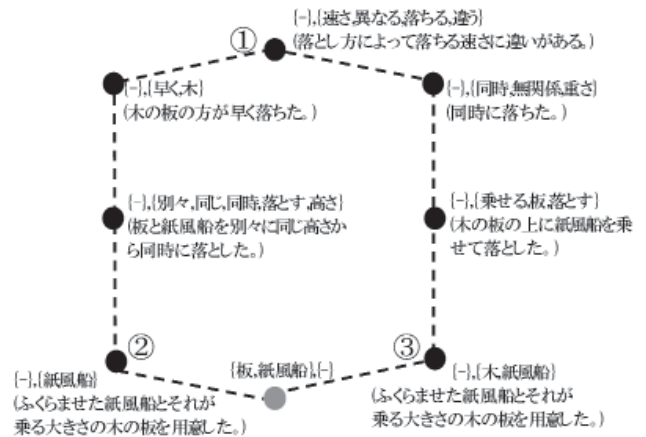


図9 演繹的推論と帰納的推論事例

アブダクションの検証

因果概念束に対して、次の問合せ質問を行ってみる。

「なぜ、胃は胃液によって消化されないか？」

これは、胃が胃液で消化されないことを結果とした場合、その理由を問う質問である。問合せ質問の文脈（{胃}, {消化されない}）を入力し、因果概念束に問合わせさせた結果の部分束を図10に示す。図10の①は、質問の問い直し概念であるため、回答となる原因を探るには、②に向かって部分束を辿れば、胃が消化されない理由に到達することができる。

ここで図10の部分束を含む因果概念束を表示すると、図11に示す興味ある結果が得られる。具体的には、質問への回答に加え、例えば、図11①の「胃という器官」の説明や、同図②の質問とは逆の「胃が消化される」胃潰瘍の因果関係など、同図からは質問に関連するいくつ

かの知識を得ることができる。質問に関連する知識は、回答の元となった因果概念束が保有しており、この概念束による命題間の結びつきは、人間の連想記憶と類似しているように思われる。

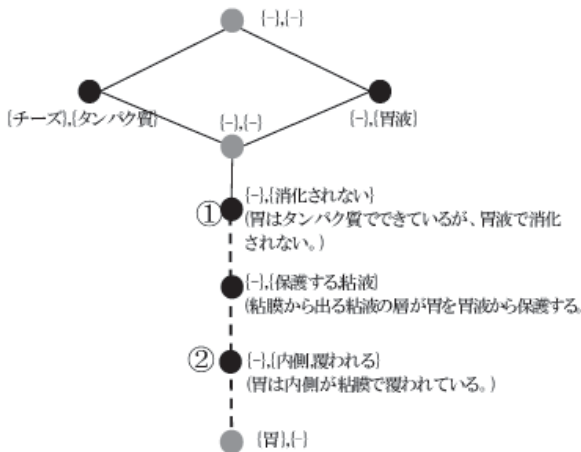


図 10 アブダクション事例

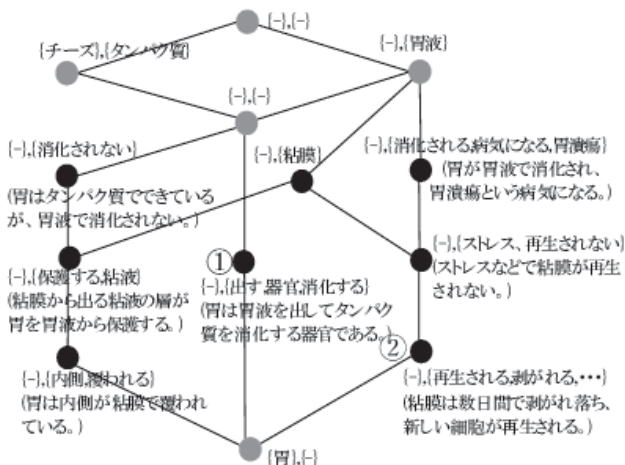


図 11 命題間の連想事例

5. 結論⁽¹⁵⁾

以上、因果関係が埋め込まれた因果概念束について、理論的考察と具体的な説明文による問合せ回答の検証を行った。以下では、これらの結果を結論としてまとめる。

(1) これまでの概念束と因果概念束との整合性

本研究では、命題に含まれる因果関係が、形式概念分析における *Attribute Implication* と整合性を保ち、かつ概念束の中に適切に組み込まれることを確認した。*Attribute Implication* は属性の包含関係であるが、因果概念束には属性の包含関係に加え、因果関係の包含関係も組み込まれている。本研究ではこの因果関係の包含を *Causal Implication* と呼ぶことにする。これら2つの *Implication* は相互に整合性を保つため、*Attribute*

Implication によっても因果関係のルール抽出は可能である。しかし、その結果抽出されるルール形式には式 (5) が適用されるため、因果関係の表現として適切ではない場合がある。また、本研究では、因果概念束には興味ある性質があることも明らかになった。ある概念を指定して式 (13) により因果部分束を生成すると、指定された概念を境にして上界 (*upper bound*) と下界 (*lower bound*) が形成され、上界には指定した概念の結果概念の集合、下界には原因概念の集合が形成される。問合せ質問による演繹的推論やアブダクションは、上界または下界の概念集合を探索することに他ならず、これは人間が思考範囲を狭め、集中して物ごとを推論することに相当する。

(2) 因果概念束における概念個数の数的特性

今回、本研究では1,998命題の説明文により、5,385個の因果概念が生成された。この結果、概念総個数は命題数に対して線形的に増加することが分かった。これは、命題数をさらに増加させても、その概念総個数は線形的にしか増加しないことを示唆する。例えば、10万件程度の命題の因果概念束を生成しても、その概念総個数は高々27万個程度であり、因果概念束の永続化には規模的な支障は生じない。また、本研究では、ノートPC (Intel Core i7-8700 3.20GHz) を用いて因果概念束の生成・更新を行っているが、1,998命題の一括処理に約10分11秒を要した。このことから10万件の命題の一括処理では、約8時間54分もの時間が必要とされることになる。この予測結果から、本研究における逐次更新アルゴリズムの採用は、命題数の増加に対し、処理時間の観点からも妥当な選択であったことが分かる。なぜなら、逐次更新アルゴリズムでは、段落入力たびに、段落に含まれる文脈 (命題) と既存概念との局所的な集合演算が行われるからである。つまり、段落ごとのインタラクティブな更新処理は、文脈全体の処理時間を分散化することを可能にしている。また、一般に *SQL* には検索条件個数の仕様制限が伴う。因果概念束は命題を扱うため、不特定多数の属性やオブジェクトが含まれ、命題数が増加すると、その検索条件は *SQL* の仕様を超えてしまう。一方、段落は意味のまとまりであり、離れた段落の文脈間の外延・内包の相関は比較的少ない。本研究では、この段落の特性を活かすことにより *SQL* における検索条件を緩和することができた。段落による因果概念束の生成・更新は、文脈の分散化とともに、概念束の更新処理の効率化にも寄与していることが分かった。また、段落を跨ぐ因果関係は、帰納概念を用いることによって、もれなく推論可能となることも確認できた。

(3) 的確な問合せ質問への回答

質問に適合する部分束を得るため、本研究では、当初、因果概念に対して検索条件式 (13) を適用した。しかしこれはうまくいかなかった。その理由は、因果関係が組込まれた概念には、束構造の情報が含まれているためである。これを解決するため、本研究では因果関係組込み前の“元の文脈”に着目し、これに検索条件を適用した。この結果、高い精度で問合せ質問に適合する初期概念を得ることができた。因果概念束においては、問合せ質問への回答プロセスに応じて、因果概念と因果関係組込み前の文脈を使い分けることが重要であった。また、命題の演繹的推論、帰納及びアブダクションに関わる質問では、帰納概念が重要な役割を持つことも確認した。一方、今回は命題間の因果関係を *Hasse* 図で表示したが、これは回答表現としては不十分である。本研究では、形式概念分析における *Attribute Implication* と *Association Rule* が実装済みであり、これらを因果概念束用に改修すれば、今後、日本語や英語等の文章表現による回答が可能になる。

本研究を通して、幾つかの課題も明らかになったので、ここでは3点取り上げて述べる。

(4) 説明文から文脈への変換プロセスの効率化

今回、本研究では手作業によって文脈を作成した。しかし、これには多くの時間を要したため、形態素解析等を用いて省力化すれば、この作業は軽減できると思われる。

(5) シソーラスを用いたキーワードの統一

因果概念束は命題のキーワードにより構成されるが、命題から文脈を作成するとき、キーワードに同義語が混在すると、同じ内容の命題であっても異なる概念として扱われてしまう。この問題を解決するためには、シソーラスによるキーワードの統一が不可欠である。命題から文脈への変換作業において、キーワードシソーラスを併用すれば、この問題は解決されると思われる。

(6) ニューロン・ダイアグラムの自動作成

今回は、命題のニューロン・ダイアグラムも手作業で作成し、因果関係を明確にした上で文脈作成を行った。しかし、この作業も相当な時間を要したため、今後はダイアグラム自動作成による省力化が必要であると思われる。

以上、因果概念束に関する考察とその検証結果を述べたが、最後に本論文の結論を総括する。本論文の序論に

において「人間は獲得した知識を過去の知識と融合し、それらを組み合わせて新たな知識を生み出す」旨を述べた。これに対して、因果概念束では、新たな命題（文脈）の追加により、既存の束構造の再編成が行われ、問合せ質問への回答能力を向上させる。この類似性は、因果概念束が人間の思考により近い知能であることを示すものである。

今後、この論文を起点として、本研究ではさらに高度な演繹的推論やアブダクションへの応用事例を検証し、機会があればその成果を報告したいと考えている。

6. 謝辞

本研究は、9年間の研究休止期間を経て、2019年4月から研究を再開した。このため、本論文は素案段階において、情報処理学会論文審査各位からご意見を頂き、それをもとに全面的な改定を行ったものである。ここに各位に感謝の意を表します。また、この1年半の期間、教育研究の機会を与えて頂いた常葉大学関係各位にも感謝の意を表します。

参考文献

- (1) G.Birkhoff, *Lattice Theory*, American Mathematical Society, 1967.
- (2) B.A.Davey, H.A.Priestley, *Introduction to Lattices and Order*, 2nd ed., Cambridge University Press, 2002.
- (3) B.Ganter, G.Stumme, R.Wille eds, *Formal Concept Analysis Foundation and Application*, LNAI, 3626, 2005.
- (4) B.Ganter, R.Wille, *Formal Concept Analysis Mathematical Foundations*, Springer, 1999.
- (5) D. Albert, T.Held, *Establishing knowledge spaces by systematical problem construction*, Dietrich Albert ed., Springer-Verlag, 1994.
- (6) Douglas Kutach, *Causation*、相松慎也・一ノ瀬正樹訳、岩波書店、2019。
- (7) J.Heller, *Semantic Structures*, Dietrich Albert ed., Springer-Verlag, 1994.
- (8) R.Godin, R.Missaoui, H.Alaoui, Learning algorithm using a Galois lattice structure, *Proc. of 1991 IEEE Int. Conf. on Tools for AI*, pp.22-29, 1991.
- (9) R.Wille, *Restructuring lattice theory, an approach based on hierarchies of concept*, I.Rival ed., Ordered Sets, D.Reidel, pp.445-470, 1982.
- (10) Stephen Mumford, Rani Lill Anjum, *Causation : A Very Short Introduction*、塩野直之・谷川卓訳、岩波書店、2017。

- (11) 新井紀子、東中竜一郎、『人工知能プロジェクト「ロボットは東大に入れるか」第三次 AI ブームの到達点と限界』、東京大学出版会、2018。
- (12) 新井紀子、『AI に負けない子供を育てる』、東洋経済新報社、2019。
- (13) 学研教育出版編、『文章読解（毎日のドリル小学1～6年）』、学研プラス、2018。
- (14) 鈴木 治、室伏俊明、『形式概念分析－入門・支援ソフト・応用－』、日本知能情報ファジィ学会、日本知能情報ファジィ学会誌 19(2)、pp.103-142、2007。
- (15) 鈴木 治、『形式概念分析とは何か－その理論的背景と応用事例－』、情報処理学会ソフトウェア工学研究会 Tutorial 講演資料、2010。
- (16) 成美堂出版編集部、『文章読解（1日10分小学ドリル1年～3年生）』、成美堂、2018。
- (17) 米盛裕二、『アブダクション 仮説と発見の論理』、勁草書房、2017。

