

形式概念解析ツールとデータ解析

Formal Concept Analysis Tool and Data Analysis

長田 博泰

1 はじめに

著者はこの数年形式概念解析およびその応用に取り組んできた(長田 2004, 2006a, 2006 b, 2007). 本稿の目的はその過程で試作している形式概念解析ツールの使い方を解説するとともに, ツールを用いた質的データ分析の実際的手法を提示することである. そして形式概念解析が広く人文・社会科学において, 例えば言語学における意味解析などさまざまな方面で利用されることを期待している.

Web の普及により入手できるデータは爆発的に増大した. Web サイトにアクセスする利用者のさまざまなデータを(その気なれば)収集・蓄積することはきわめて容易である. 他方, 種々のデータ解析パッケージも豊富になり誰でも気軽にデータを分析することができるようになった.

しかし, この状況が本当に好ましい事態か否かはきわめて疑わしい. その理由のひとつは, 調査・分析結果の単純さ加減である. 図表などの多さの割に立ち入った分析になっていないことなどに端的に現れている. とくにマスコミではどちらが多いか少ないか, あるいは少数の事柄の関連だけを分析した結果しか報道されず, その結果単純比較でものごとをわかったような気分を助長していると考えられるからである. さらに, データの取得とその数量的分析が容易になるとともに, 統計的数量的データ解析アプローチが万能であるかのような錯覚を引き起こす一方, 統計的扱い

に不向きなデータの解析を無視したり, 真剣に研究しようとししない傾向を助長していることである.

人自身も, また人がかかわる社会的現象も実にさまざまな要因が絡んでいると想像される. だとすれば, その要因すべてを一挙に詳らかにすることは不可能だとしても, せめていくつかの要因がらみで現象を説明する多面的, 複眼的切り口がほしいと考えるのは当然であろう.

形式概念解析は多数のことがらの絡み合いをそのまま解析しようとする方法である. このような方法が単純な切り口に慣らされているかもしれない現状に対し一石を投ずることを願ってやまない.

2 形式概念とは

形式概念はその名が示すように“概念”をある見地から定義しようとするものである. すなわち, 概念を“外延”, つまり概念に属するものを一つ一つ挙げてゆく定義と“内包”, つまり概念が満たすべき条件を述べる定義がピッタリと重なったとき, 一つの概念が定義されたとする見方である. 形式概念は概念に対するこの立場を形式的(数学的)に忠実に表現しようとするアプローチである. 以下, 直感的な理解を助ける解説を試みる. 詳細はGanter & Wille (1999)あるいはDavey & Priestley (2002)を参照されたい.

概念の外延(対象), 内包(属性)はそれぞれ表の行, 列で表わすことができる. 問題は

概念の外延と内包の一致が表の上でどのように表されているかである。具体例を用いて説明しよう[Janssen 2002 から借用]。いま図1に示す図形の集まりがあり、これを整理することを考えてみる。各図形の形と色に注目する。形の属性値は円・三角形・四角形であり、色の属性値は白・灰色・黒である（以下、このように属性と属性値を同じように扱う）。形式概念解析ではこれを表1のような属性表（正確には、「クロステーブル」という）で表す。

表1中の対象 o7, o8, o9 のもつ属性は三角形と白である。逆に、属性が三角形で白いものは o7, o8, o9 の3つである。外延の集合 {o7, o8, o9} と内包（属性）の集合 {三角形, 白} の対 ({o7, o8, o9}, {三角形, 白}) を概念とみなし、これを形式概念という。しかし、例えば ({o8, o9}, {三角形, 白}) の対は形式概念とみなさない。なぜなら、属性{三

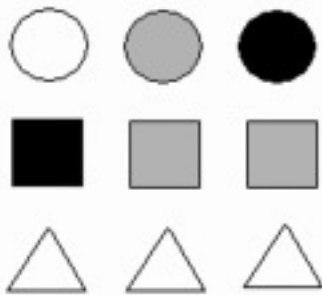


図1 図形の世界

表1 図1の属性表（ブール値）

ブール値	円	正方形	三角形	白	灰色	黒
o1	x				x	
o2	x			x		
o3	x					x
o4		x				x
o5		x			x	
o6		x			x	
o7			x	x		
o8			x	x		
o9			x	x		

角形, 白} を有する対象をすべて集めると, {o7, o8, o9} であり, {o8, o9} と異なるからである。このように形式概念は内包の各属性を有する対象をすべて含む外延をもたなければならないのである。直感的に言えば、属性を共有する外延の集合の中で最も大きな（極大な）矩形である。この定義にしたがえば, ({o5, o6}, {正方形, 灰色}), ({o4}, {正方形, 黒})などは形式概念である。表1のすべての概念を表2に示す。表2には異なる5種の網掛けおよび6種の色分け, 合計11個の形式概念が示されている。これらの形式概念が表3である。

概念の間に上位, 下位の関係があるように形式概念間にも同様の関係を定義することができる。2つの形式概念の間に外延の集合としての包含関係, つまり含むあるいは含まれる関係によって上位・下位関係を定義する。

表2 図1の形式概念

ブール値	円	正方形	三角形	白	灰色	黒
o1	⊗				⊗	
o2	⊗			⊗		
o3	⊗					⊗
o4		⊗				⊗
o5		⊗			⊗	
o6		⊗			⊗	
o7			x	⊗		
o8			x	⊗		
o9			x	⊗		

表3 形式概念

No	形式概念
1	{{黒}, {o3, o4}}
2	{{灰色}, {o1, o5, o6}}
3	{{白}, {o2, o7, o8, o9}}
4	{{正方形}, {o4, o5, o6}}
5	{{円}, {o1, o2, o3}}
6	{{三角形, 白}, {o7, o8, o9}}
7	{{正方形, 黒}, {o4}}
8	{{正方形, 灰色}, {o5, o6}}
9	{{円, 黒}, {o3}}
10	{{円, 灰色}, {o1}}
11	{{円, 白}, {o2}}

一方の外延が他方の外延を含む場合、前者を上位概念、後者を下位概念という。内包=属性の集合を用いて定義することも可能である。すなわち、一方の属性集合が他方のそれに含まれる場合、言い換えると対象を規定する属性の種類が少ない方が上位概念、多い方が下位概念になる。例えば、上の形式概念では $(\{1, 6, 7, 8\}, \{\text{白}\})$ は $(\{6, 7, 8\}, \{\text{三角形}, \text{白}\})$ の上位概念である。同様に、 $(\{3, 4, 5\}, \{\text{正方形}\})$ は $(\{4, 5\}, \{\text{正方形}, \text{灰色}\})$ の上位概念である。

上位・下位関係にある概念の関係は、上位概念を図の上方、下位概念を下方に描き、両者を線で結んで表すことができる。すべての形式概念をこの方法で描き、さらに図上の最上方、最下方にそれぞれ最上位概念、最下位概念として $(\{\text{全対象}\}, \{\})$, $(\{\}, \{\text{全属性}\})$ の形式概念を付け加えた図を Hasse 図という。表 1 の形式概念の Hasse 図を図 2 に示す。

Hasse 図は単に形式概念の上位・下位の関係だけを示しているのではない。表 1 に含まれるすべての情報が圧縮表現されている。Hasse 図の情報を読み取るため、以下では Hasse 図上の形式概念をノードと呼ぶことにする。最上位のノードに直接つながるノードを深さ 1 のノード、深さ 1 に直接つながる

ノードを深さ 2 のノード、以下同様である。

深さ 1 のノードは単一の属性（同じ外延を有すれば複数のこともある）を表し、図のノードにその属性（値）を記す。深さ 2 のノードは結合されている上位のノードのすべての属性を受け継ぐ。例えば、「正方形」と「黒」につながるノードは正方形で黒いことを示し、とくに属性を記入しない。論理式で表せば、「正方形 \wedge 黒」である。深さ 3 以下についても同じである。

深さ 2 のノードに現れた「三角形」とその上位概念「白」の関係はどうであろうか。三角形と記入されたノードは「白」属性を受け継ぐ。したがって、下から上に読んで「三角形は白い」ことがわかる。これは表 1 の行に戻れば一層容易に納得するであろう。論理式では「三角形 \rightarrow 白」と記し、このような関係を含意関係という。言い換えると、属性「三角形」と属性「白」は依存関係にあるといえる。この例では依存関係はこれしかないが、独立関係として正方形で白いものはないことがわかる。正方形を受け継ぐ深さ 2 の 2 つのノードは色の属性（値）として黒と灰色しか受け継いでいないことがわかるからである。完全を期すために導き出されるその他の独立関係を掲げれば、以下のとおりである。

1) 同時に灰色と黒であることはない。

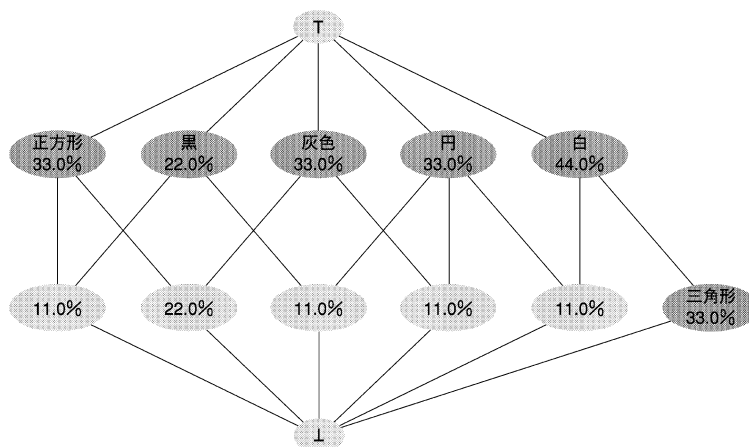


図 2 表 1 の Hasse 図

- 2) 同時に白は黒であることはない.
- 3) 同時に白は灰色であることはない.
- 4) 円は正方形ではない (あるいは, 正方形は円ではない).

以上のような含意関係は色, 形に関して人が有している常識からは当たり前のことであって何ら意味のないことのように思えるかもしれない. しかし, 属性間の関係が明確でない場合には新たに見出された関係として重要な意味を有するのである. つまり, 属性間の依存ないしは独立関係を発見することにつながるのである.

では, 下から上へはどうであろうか. これは対象を束ね, 上位概念を見出すことである. 例えば, 深さ 1 の「円」につながる深さ 2 の 3 つのノードから, 「円」には黒と灰色と白いものがあることがわかる. 論理的に表現すれば, $(円 \wedge 黒) \vee (円 \wedge 灰色) \vee (円 \wedge 白) = 円 \wedge (黒 \vee 灰色 \vee 白)$ である.

以上のようにクロステーブル, Hasse 図および含意関係で表される論理的表現は全く同じ情報を表現しているのである. しかし, データ量が多くなると属性表を眺めてもひと目でそこに隠れている関係を見出すのは容易ではない. したがって, 形式概念解析の主たる目標は, 属性表から Hasse 図を描く, あるいは含意関係を明示的に表現することによって隠れていた関係を炙り出すことにある. しかしながら, データ量が膨大になるとその関係を読み取ることはそれほど容易ではない. 次節以降では現実のデータを扱う際に有効な方法を提示しながら, 形式概念解析の手法とツールの解説をする.

3 形式概念解析をデータ解析に適用する

形式概念解析を用いたデータ解析とは一言でいえばデータに含まれるパターンの発見である. 文字どおりパターンというべきものが Hasse 図に認められる場合も無くはないが

それは稀有であろう. むしろ形式概念の外延に含まれる対象の個数や含意規則などからその特徴をさぐって行くことになる.

形式概念解析の一般の手順を示すと次のとおりである:

- 1) 属性表 (クロステーブル) を準備する.
- 2) 解析ツールを用いて形式概念解析を行い, 概念構造の特徴あるいは属性間の依存・独立性を発見する.
- 3) 概念数が多くてダイアグラム (Hasse 図) から概念間の関係が把握し難い場合, “支持度” を指定し, “粗い概念束” を描き, 特徴把握に努める.

ここでは以下の 4 つの方法を事例によって解説する.

- 1) 基本手順
- 2) 粗い概念束の利用—外延からクラスタリングする
- 3) 属性の絞込み—内包から概念構造を分析する
- 4) 決定表—含意規則の応用

3.1 基本手順

基本的手順を示すために表 4 に掲げる動物の属性表を分析しよう [Kohonen 1995 から借用]. 後述する解析ツールを用いると 29 個の形式概念と含意規則が得られ, その結果を Hasse 図で描く. しかし見通しのよい Hasse 図を描くのは必ずしも容易ではない. 見通しのよい Hasse 図を描くには種々の視点からデータを捉える必要があり, 以下これを支援する方法を述べる.

3.2 粗い概念束の利用—外延からクラスタリングする

概念数が多い, あるいは複雑な構造の場合など, 形式概念解析を適用しても必ずしも見通しのよい結果が得られない. その場合できる限り単純な概念, すなわち属性の個数が少なく, なおかつ外延に含まれる対象の個数が

表4 動物属性表

動物	小	中	大	二足	四足	羽	毛	飛	狩	走	泳	齧	蹄
ハト	x			x		x		x					
ニワトリ	x			x		x							
アヒル	x			x		x		x			x		
ガチョウ	x			x		x		x			x		
フクロウ	x			x		x		x	x				
タカ	x			x		x		x	x				
ワシ		x		x		x		x	x				
キツネ		x			x		x		x	x			
イヌ		x			x		x			x			
オオカミ		x			x		x		x	x		x	
ネコ	x				x		x		x	x			
トラ			x		x		x		x	x			
ライオン			x		x		x		x	x		x	
ウマ			x		x		x			x		x	x
シマウマ			x		x		x			x		x	x
ウシ			x		x		x						x

多い概念に注目することである。ここで以下の説明のために“支持度”と“粗い概念束”という用語を導入する。まず支持度をつぎのように定義する。

支持度 = (一つの形式概念の外延に属する対象の個数 ÷ 対象全体の個数) × 100

ある支持度以上のノードの集まりは厳密な意味では“束”¹⁾ではないが、最下位のノードを付加すれば“束”になるので“粗い概念束”と呼ぶことにする。

Hasse 図の構造が複雑で見通しの悪い場合、適当な支持度を与え、粗い概念束を描画することによって全体の大枠の構造を把握することができる。表4の動物属性を支持度0.4および0.3で描いたものを掲げる(図4, 5)。

支持度0.4の粗い概念束では属性「狩」、「小」が属性「四足・毛」と「二足・羽」にどのように関係するかが判然としない。さらに支持度を0.3にしてみると、この図で見る限り属性「狩」、「小」がそれぞれ「四足・毛」、「二足・羽」に関連しているらしいことがわかる。したがって、「四足・毛」、「二足・羽」を

中心に概念の構造を整理するのが適当であると推測される。実際、さらに支持度を0.2にするとこの見通しが適当であったことが一層明確になる。しかし、さらに支持度を下げてゆくと、「四足・毛」、「二足・羽」に二分して捉えた概念に「大」、「中」、「小」、「狩」などの付加的属性が絡んできて図3に示すようにHasse 図は次第に複雑になってゆく。

以上のように支持度に基づく粗い概念束は大枠の概念を捉えるのには適しているが、他の付加的概念が複数の大枠の概念に絡んでくると、見通しが悪くなる。その原因は上位の概念では分離されているにもかかわらず、下位では付加的属性を共有するためである。これを克服するのはある意味で簡単である。上位で分離された属性ごとにクロステーブルを作成し、各々を形式概念解析すればよい。これは単純作業ではあるが、あまり賢明な方法とも思えない。

そこで、分離する属性を指定するだけでクロステーブルを作り変えることなく行う方法をつぎに述べよう。

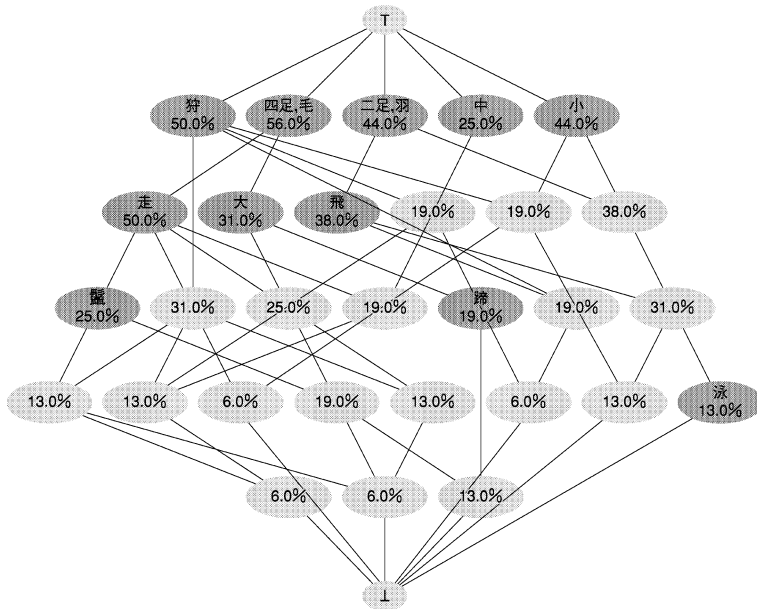


図3 解析ツール（後述）で描いた表4のHasse図

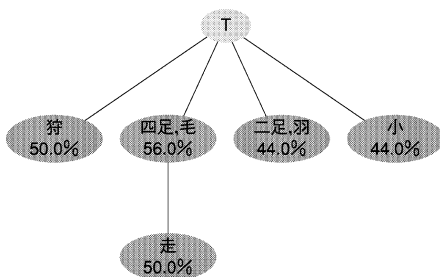


図4 支持度 0.4=40%

3.3 属性の絞込み—内包から概念構造を分析する

四足・二足のように互いに排他的な属性で分離する場合、排他的属性の一つを含んだ対象だけを属性表から取り出し、形式概念解析すれば、それぞれ概念束になる。それぞれの概念束を描き、全体に最上位、最下位ノードを付け加えると全体として概念束になる。表4をこの方法で描いた概念束が図6である。この図からわかるように図3で共通であった属性が排他的な概念束のそれぞれに描かれている故に、図3に比し属性間の関係が非常に読み取りやすくなっている。

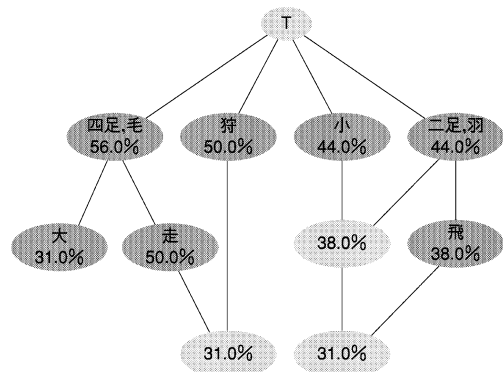


図5 支持度 0.3=30%

3.4 決定表—含意規則の応用

属性の絞込みを適用し、そこに見出される含意規則を意思決定のルールとみなすことも可能である。ここでは、メガネ屋が顧客に対してコンタクトレンズの使用をどのように勧めたかを示すデータ（表5）に形式概念解析を適用し、その決定ルールを調べてみよう。コンタクトレンズのハード、ソフトおよびコンタクトレンズなしの属性を絞り込むことによってつぎの含意規則（表6, 7, 8）を得る。

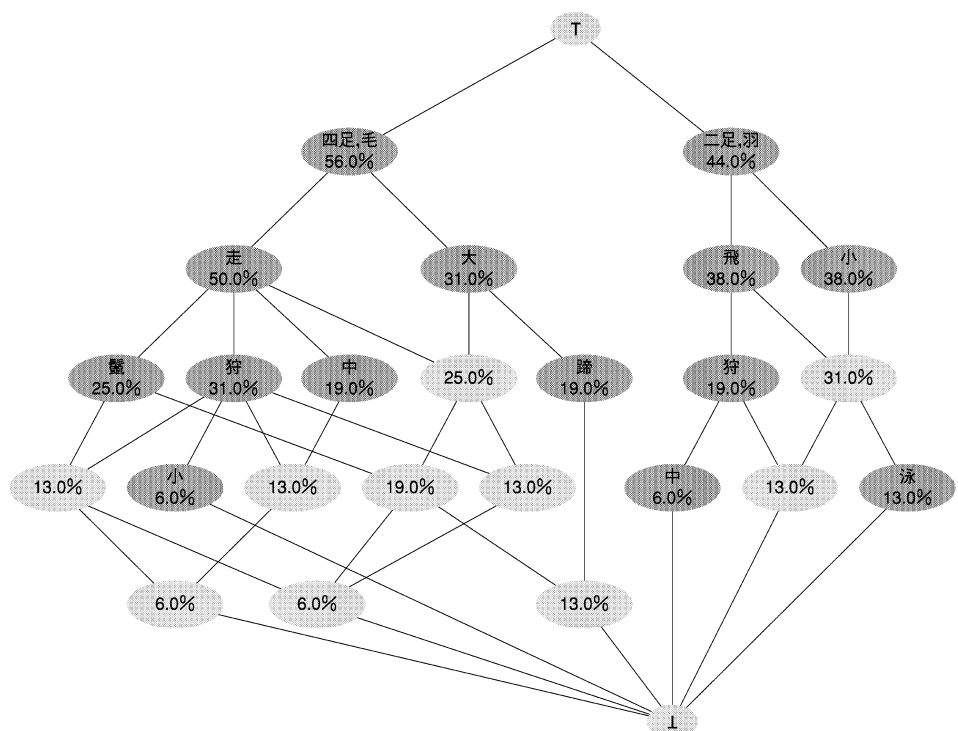


図 6 属性の絞込みから描かれた Hasse 図

表 5 コンタクトレンズ属性表^(*)

コンタクト	若い	初老	老齢	近視	遠視	乱視有	乱視無	涙産少	涙産正常	ハード	ソフト	なし
i1		若い		近視		乱視無		涙産正常		ハード		
i2		若い		遠視		乱視無		涙産正常		ハード		
i3		初老		近視		乱視無		涙産正常		ハード		
i4		老齢		近視		乱視無		涙産正常		ハード		
i5		若い		近視		乱視有		涙産正常		ソフト		
i6		若い		遠視		乱視有		涙産正常		ソフト		
i7		初老		近視		乱視有		涙産正常		ソフト		
i8		初老		遠視		乱視有		涙産正常		ソフト		
i9		老齢		遠視		乱視有		涙産正常		ソフト		
i10		若い		近視		乱視有		涙産少		なし		
i11		若い		近視		乱視無		涙産少		なし		
i12		若い		遠視		乱視有		涙産少		なし		
i13		若い		遠視		乱視無		涙産少		なし		
i14		初老		近視		乱視有		涙産少		なし		
i15		初老		近視		乱視無		涙産少		なし		
i16		初老		遠視		乱視有		涙産少		なし		
i17		初老		遠視		乱視無		涙産少		なし		
i18		初老		遠視		乱視無		涙産正常		なし		
i19		老齢		近視		乱視有		涙産少		なし		
i20		老齢		近視		乱視有		涙産正常		なし		
i21		老齢		近視		乱視無		涙産少		なし		
i22		老齢		遠視		乱視有		涙産少		なし		
i23		老齢		遠視		乱視無		涙産少		なし		
i24		老齢		遠視		乱視無		涙産正常		なし		

[横森・小林 1994 から借用] 注) 表 5 の作り方は 4.1 節で説明する。

表6 ハードコンタクトの含意規則

人数	前 件	後 件
4	{ハード}	→ {乱視無, 涙産正常}
3	{近視, 乱視無, 涙産正常}	→ {ハード}
2	{若い, 乱視無, 涙産正常}	→ {ハード}
1	{遠視, 乱視無, 涙産正常, ハード}	→ {若い}
1	{老齢, 乱視無, 涙産正常, ハード}	→ {近視}
1	{初老, 乱視無, 涙産正常, ハード}	→ {近視}

表7 ソフトコンタクトの含意規則

人数	前 件	後 件
5	{ソフト}	→ {乱視有, 涙産正常}
3	{遠視, 乱視有, 涙産正常}	→ {ソフト}
2	{初老, 乱視有, 涙産正常}	→ {ソフト}
2	{若い, 乱視有, 涙産正常}	→ {ソフト}
1	{老齢, 乱視有, 涙産正常, ソフト}	→ {遠視}

表8 コンタクトレンズなしの含意規則

人数	前 件	後 件
12	{涙産少}	→ {なし}
4	{若い, なし}	→ {涙産少}
3	{遠視, 乱視有, なし}	→ {涙産少}
3	{近視, 乱視無, なし}	→ {涙産少}
2	{乱視無, 涙産正常, なし}	→ {遠視}
2	{遠視, 涙産正常, なし}	→ {乱視無}
2	{老齢, 遠視, 乱視無}	→ {なし}
2	{老齢, 近視, 乱視有}	→ {なし}
2	{初老, 乱視有, なし}	→ {涙産少}
2	{初老, 遠視, 乱視無}	→ {なし}
2	{初老, 近視, なし}	→ {涙産少}
1	{乱視有, 涙産正常, なし}	→ {老齢, 近視}
1	{近視, 涙産正常, なし}	→ {老齢, 乱視有}
1	{初老, 涙産正常, なし}	→ {遠視, 乱視無}

(1) ハードコンタクトの含意規則

どのような場合にハードコンタクトを薦めたかを定める決定ルールを導くには表6の後件が属性ハードである含意規則を論理演算子 \vee で結合すればよい。

$(\text{近視} \wedge \text{乱視無} \wedge \text{涙産正常}) \vee (\text{遠視} \wedge \text{乱視無} \wedge \text{涙産正常}) \rightarrow \text{ハード}$

右辺を簡単にすれば、以下のようになる。

$(\text{近視} \vee \text{遠視}) \wedge (\text{乱視無} \wedge \text{涙産正常}) \rightarrow \text{ハード}$

(2) ソフトコンタクトおよびコンタクトレンズなしの含意規則

ソフトコンタクトおよびコンタクトレンズなしの場合も同様なので以下にそれぞれの結果だけを掲げる。

$(\text{遠視} \vee \text{初老} \vee \text{若い}) \wedge (\text{乱視有} \wedge \text{涙産正常}) \rightarrow \text{ソフト}$

この結果から、ハード・ソフトどちらのコンタクトを薦めるべきかは涙が正常に出て、乱視の無・有であることがわかる。コンタクトレンズなしを決めるルールは以下のとおりであるが、15人中12人、つまり80%は涙の出

が少ない場合であり，他の論理式は残り3人を決める例外的ルールと考えられる。

涙産少 $\vee((\text{老齡}\vee\text{初老})\wedge(\text{遠視}\wedge\text{乱視無}))$
 $\vee(\text{老齡}\wedge\text{近視}\wedge\text{乱視有})\rightarrow\text{なし}$

以上に示すように形式概念解析において，ある属性に焦点を当て含意規則に注目すると決定ルールの導出にも利用できることがわかる。

4 形式概念解析ツールを用いたデータ解析

上で述べたように形式概念解析によってデータを解析するには，属性表の準備，概念束・含意規則の計算および Hasse 図の描画，粗い概念束の描画および属性の絞込みなどを行わなければならない。ここではこれらの手順を試作中の形式概念解析ツール（以下，解析ツールということにする）を用いて行う方法を解説する。

まず解析ツールのイメージを与えるためにメインメニューとサブメニューの一部を図7，8に掲げ，メニュー項目とそのサブメニューの機能概略を表9に示す。分析手順に沿って各々の具体的な使い方を説明してゆく。

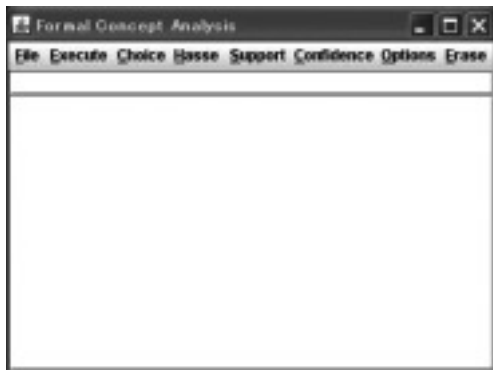


図7 メインメニュー画面

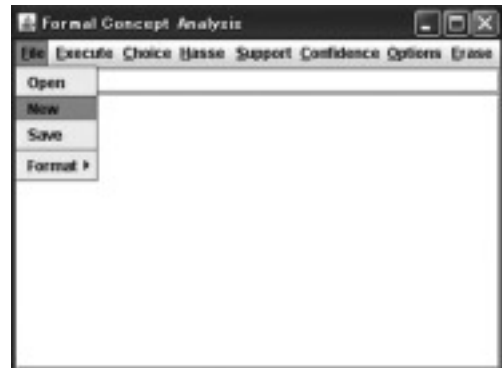


図8 サブメニュー「New」

4.1 属性表（クロステーブル）の作成

形式概念解析では対象がある属性を有するか否か，つまり二値（ブール値）を属性表で与えることを基本としている。しかし，実際のデータ処理では，ある属性が複数の属性値をとると考える方が自然である。したがって，実際に属性表を用意する場合，表1のようなブール値ばかりでなく表10に示すように複数の属性値を英数字以外の半角文字（ここでは，半角文字の縦棒「|」を用いている）で区切って書けばよい。属性の個数が多いときこの形式は便利であろう。この種の表を多値型（many-valued）という。なお，現在，ブール値と多値の混在は認めていない。属性表を作成する際，どちらであるかを明確に意識しておく必要がある。

表1あるいは表10の形式の属性表を作成するには基本的に2つの方法がある：

- 1) MS Excelのような表計算言語で属性表を作成し，CSV形式で保存する。
- 2) 解析ツールの「File」メニューのサブメニュー「New」をクリックする（図8）と図9に示す簡易クロステーブル作成画面が表示されるので，これを利用して属性表を作成することができる。

表 9 形式概念解析ツールメインメニュー

メニュー項目	サブメニュー項目		説 明
File	Open New Save Format	ファイルダイア ログボックス 別フレーム ファイルダイア ログボックス boolean many-valued	属性表をオープンする。 属性表を作成し、保存する。 属性表がブール値か多値かを指示する。標準＝ブール値 属性表はブール値である。 属性表は多値である。
Execute	Start Stop Restart Quit		形式概念を計算する。 オープンした属性表を形式概念解析する。 将来用。 将来用。 将来用。
Choice	Attr_List	attributes 属性リスト	属性を絞込む。 選択した属性を決定する。 多重選択可能な属性リスト
Hasse	New&Save Load&Edit	別フレーム 別フレーム	Hasse 図を編集・保存・ロードする。 新規に Hasse 図が描かれる。必要なら編集し保存する。 保存してある Hasse 図を開き、必要なら編集し保存する。
Support	入力フィールド 支持度リスト		粗い概念束を描く際の支持度を設定する。省略時＝0.0 0.0～100 までの任意の値を設定する。 0.0, 10.0, …, 90.0 のいずれかを選択する
Confidence	入力フィールド		連関規則を計算する際の信頼度 ^{注)} を設定する。省略時＝1.0 0.0～1.0 までの任意の値を設定する。
Options	CrossTable Lattice Implications Atoms Successor Predecessor		形式概念解析の出力情報を選択する。 入力(あるいはブール値に展開)された属性表を表示するか否か。省略時＝表示 解析された形式概念を表示するか否か。省略時＝表示 含意規則を表示するか否か。省略時＝表示 最下位ノードのすぐ上の概念(atom)内容を表示するか否か。省略時＝表示 ノードに結合する下位ノードを表示するか否か。省略時＝非表示 ノードに結合する上位ノードを表示するか否か。省略時＝非表示
Erase	Param_area Out_area		結果出力を消去する。 メニュー項目の真下のファイル情報を消去する。 出力された形式概念情報を消去する。

注) 連関規則 $X \rightarrow Y$ の信頼度^{def} $\frac{|(X \cup Y) \text{ の外延}|}{|X \text{ の外延}|}$ 、 $| \cdot |$ は個数を表すものとする。

表 10 図 1 の属性表 (多値)

図 1	円 正方形 三角形	白 灰色 黒
o1	円	灰色
o2	円	白
o3	円	黒
o4	正方形	黒
o5	正方形	灰色
o6	正方形	灰色
o7	三角形	白
o8	三角形	白
o9	三角形	白



図 9 簡易クロステーブル作成画面

4.2 概念束と含意規則の計算と Hasse 図の描き方

属性表が作成されると解析ツールによって形式概念と含意規則および Hasse 図を描く情報を計算することができる。その手順は以下のとおりである。

- 1) 形式概念解析ツールの「File」メニューのサブメニュー「Open」をクリックする(図 10)と図 11 に示すファイルダイアログボックスが表示されるので、準備した属性表の CSV 形式のファイルを選択し、「開く」をクリックする。開いたファイルのパス情報とファイル名がメインメニューのパラメータ領域(メインメニューのすぐ下のテキストフィールド)に表示される。なお、CSV ファイルが 4.1 節で述べた多値型の場合、「Open」に

先立ってサブメニュー「Format」をクリックし、「many-valued」を選択しておくかねばならない。

- 2) つぎに解析ツールの「Execute」メニューのサブメニュー「Start」をクリックする(図 12)と形式概念などが計算され、結果がメインメニューのテキスト領域に表示される(図 13)。どのような結果を表示するかは「Options」メニューで指定することができる。詳細は表 9 を参照されたい。
- 3) 形式概念の計算が完了すると、「Hasse」メニューのサブメニュー「New&Save」をクリックして Hasse 図を描くことができる(図 14)。

現在、解析ツールが描く Hasse 図は必ずしも見通しがよいものではない。そのため使用者自身が自由に Hasse 図を編集できるよう

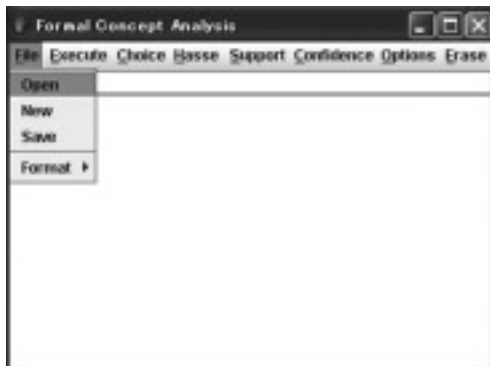


図 10 CSV ファイルの「Open」



図 12 形式概念の計算「Start」

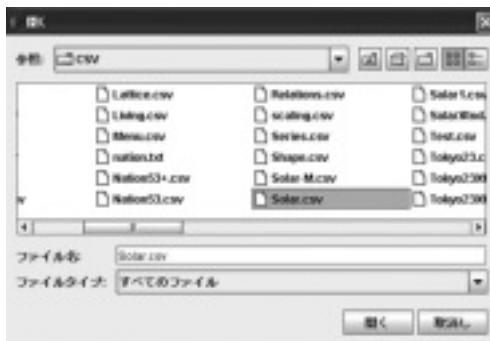


図 11 ファイルダイアログボックス

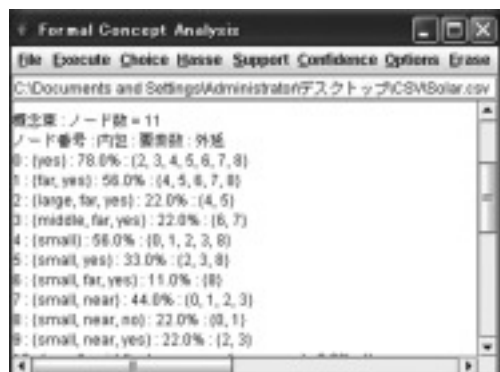


図 13 形式概念解析結果

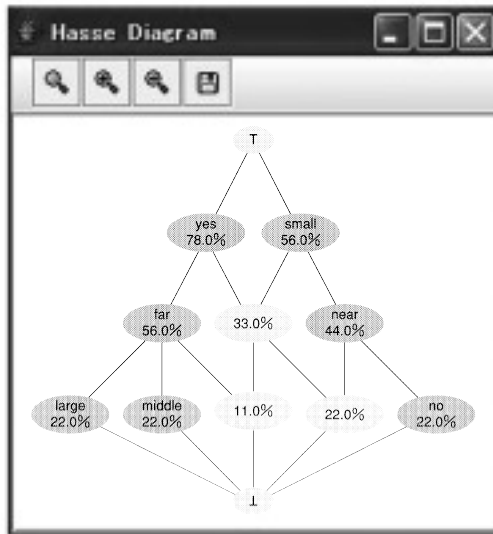


図 14 Hasse 図

になっている。図の編集は解析ツールの基本的な使い方と独立なので、その方法は 4.5 節で述べる。

4.3 粗い概念束の描き方

3.2 節で述べたように概念数が多いあるいは複雑なデータ構造の場合、Hasse 図を描いても直ちに全体の特徴を把握することができるわけではない。解析ツールではつぎの手順によって粗い概念束を描くことができる。

- 1) 「Support」メニューのサブメニューの「入力フィールド」に 0.0~100.0 の任意の値を入力し「Enter」キーを押すか、あるいは 0.0, 10.0, ..., 90.0 が設定されている（ラジオ）ボタンのいずれかをクリックする（図 15 参照）。なお、指定しない場合、0.0 になっている。
- 2) ついで Hasse 図を描く。「Hasse」メニューのサブメニュー「New&Save」をクリックすれば「Support」メニューで指定された支持度にしがった粗い概念束を描くことができる。図 4, 5 は支持度 40.0, 30.0 を与えて描いた粗い概念図（を若干編集したもの）である。

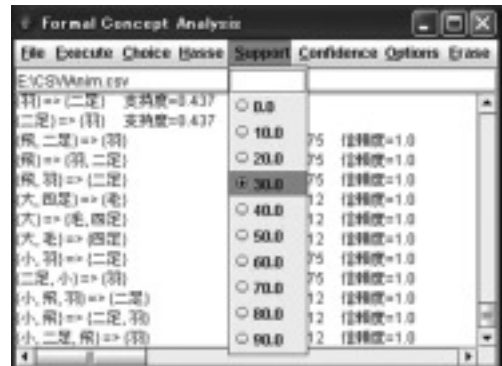


図 15 支持度の設定

- 3) 上記の手順を支持度を変化させながら、適度な見通しが得られるまで試行錯誤で続ける。

支持度を設定した後、形式概念の計算、つまりツールの「Execute」メニューのサブメニュー「Start」から計算しなおす必要がない。既に計算されている形式概念から支持度以上ののノードを選択するだけだからである。

4.4 属性の絞込み

3.3 節で述べた排他的な属性の集合に焦点を当てて概念図を描きたい時の手順はつぎのとおりである。

- 1) 「Choice」メニューのサブメニュー「Attr_List」の追加メニューの属性リストから属性を選択する。属性リストにはクロステーブルの属性（値）がすべて表示されているので、これらの中から選択する。複数の属性を選択可能である。1 番目の選択はクリックだけで指定できるが、2 番目以降はコントロールキー（ctrl）を押したまま、クリックする（図 16）。
- 2) すべての選択が終了したなら、「属性リスト」メニューの真上にある「attributes」をクリックする（図 17）。
- 3) つぎに、形式概念を再度計算しなおす必要がある。必要があるので、「Execute」メニューのサ

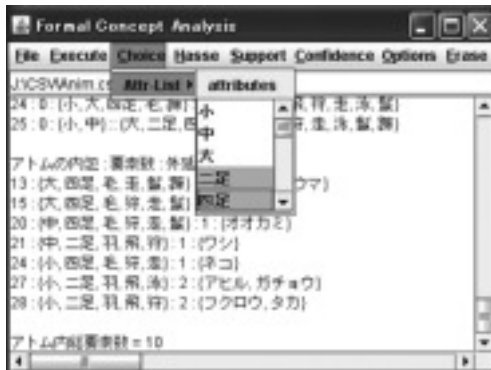


図 16 属性選択

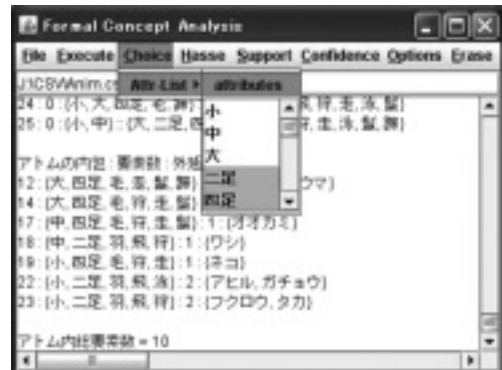


図 17 属性決定

ブメニュー「Start」をクリックする。

- 4) 「Hasse」メニューのサブメニュー「New&Save」をクリックする。

表 4 の動物属性表の形式概念解析結果に対し、2つの排他的属性「四足」と「二足」を選択して描いた Hasse 図を若干編集した図が図 18 である。

属性の選択には以下に掲げるいくつかの注意点、制限がある。

- 1) 属性選択は属性表の形式概念解析を一度

行ったあとでなければならない。

- 2) 排他的でない属性の集まりを選択することは可能であるが、この場合、排他属性を深さ 1 のレベルにもつが、排他的でない概念が下位に現れ、分離された Hasse 図にならない。共有部分だけを別な Hasse 図とし、全体を描けば見通しのよい Hasse 図がえられるが、現時点ではこれを実現していない。

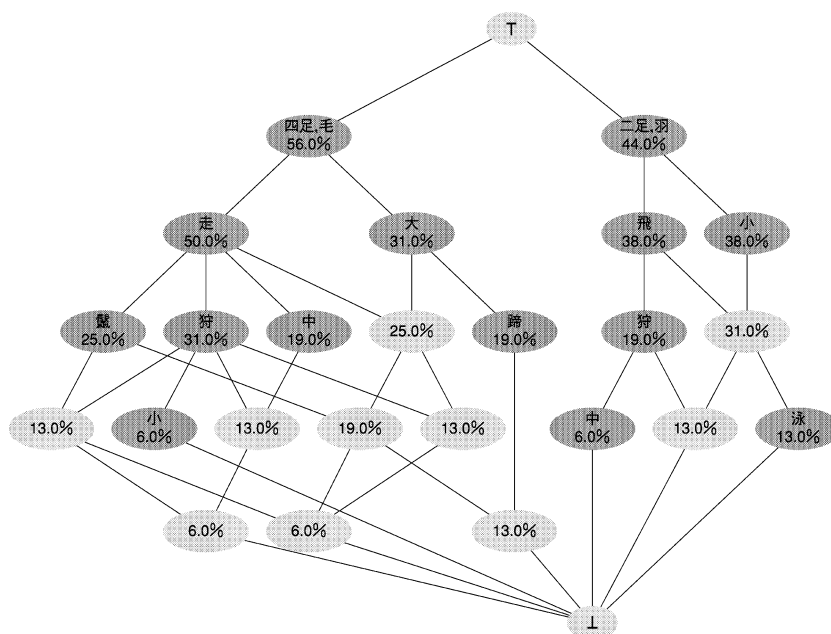


図 18 四足・二足を選択して描いた Hasse 図

4.5 Hasse 図の編集等とその他補足事項

ここでは Hasse 図の編集・保存・ロード、説明し残した事項および実際に解析ツールを用いてデータを分析する際に心得ておくると便利な事項を補足する。

4.5.1 Hasse 図の編集・保存・ロード

Hasse 図はオープンソースの JGraph を利用し描いているので、目的に応じて自由に編集することが可能である。Hasse 図の編集方法および Hasse 図の再編集可能な保存とその復旧方法は以下のとおりである。

(1) Hasse 図の編集

Hasse 図が最初に描画されたとき、描かれた図全体が移動対象の状態になっている。移動したければ、図内にカーソルを移し、ドラッグすれば自由に移動することができる。移動の必要が無ければ、図が描かれていないところでクリックし図全体の移動対象の状態の指定を解除する。その後は以下に説明する方法で図を編集することができる。

・ノードの移動

移動したいノードをクリックするとノード

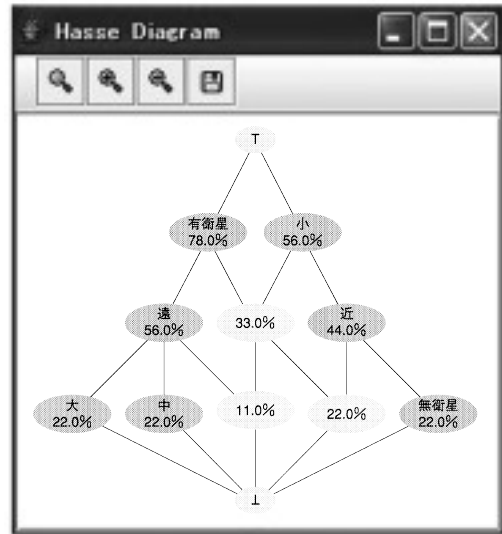


図 20 領域指定解除後の Hasse 図

の領域を示す枠が表示されるので、このノードを自由にドラッグすることができる。このノードに結合されているすべての線分（辺）も自動的に伸び縮みする。

・ノードグループの移動

ノードや矢印のない領域でクリックすると十字記号（+）が画面上に現れるので、十字記号をドラッグしながらグループとしたい

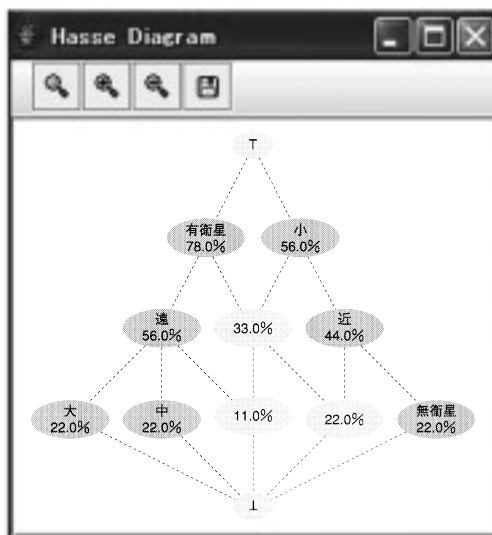


図 19 描画直後の Hasse 図

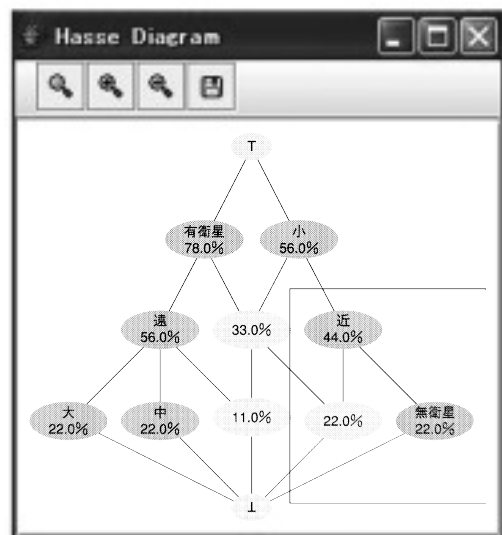


図 21 移動したいノード群を囲む

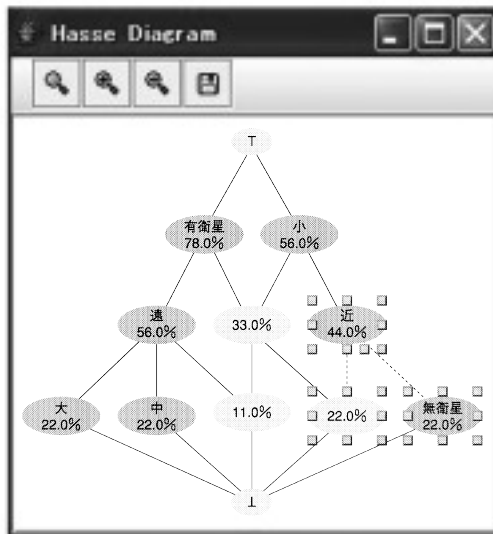


図 22 移動したノード群の決定

ノードを矩形で囲み、ドラッグを停止する。矩形内の領域にカーソルをおいてドラッグすればグループ全体を自由に移動することができる。

・ノードの編集

ノードの拡大・縮小

ノードをクリックすると、ノードの領域を示す矩形が表示されるので、左右上下および斜めへの拡大縮小を可能にする 8 個の四角の取手が現れる。この取手をドラッグしてノードを自由に拡大・縮小することができる。

ノード見出し（ラベル）の編集

ノードをダブルクリックすると、ノードの見出し（ラベル）を編集できるテキストボックスが現れるので、ラベルの追加・削除・変更が可能である。ラベル間に半角スペースを挟むと改行し表示される。

(2) Hasse 図の保存とロード

Hasse 図を保存するには静止画としての保存も考えられるが、これでは図を自由に編集できずあまり意味がない。したがって、編集可能な形式で保存するようにしている。Hasse 図では Java の XML 形式で保存する

機能を利用し、再編集が可能である。その使い方は極めて簡単である。

・Hasse 図の保存

解析ツール「Hasse」メニューのサブメニュー「New&Save」をクリックすると、別フレームに Hasse 図が描かれる。必要なら Hasse 図を編集し、Hasse 図フレームのツールバーの保存 (Save) アイコンをクリックすると、ファイルダイアログボックスが現れるので適当なところにファイルとして保存するだけである (図 23, 24)。なお、ファイルの内容は XML 形式で書かれているので、識別子として XML をつけておく方がよい。

・Hasse 図のロード

解析ツールを起動し、メインメニューにある「Hasse」のサブメニュー「Load&Edit」

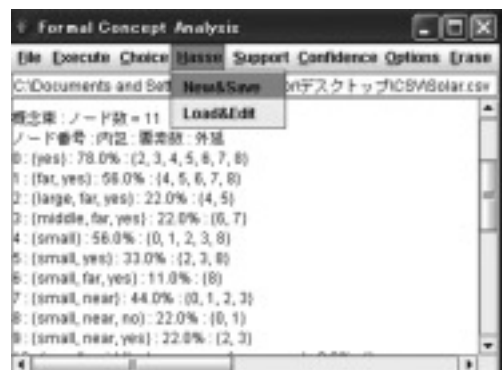


図 23 Hasse 図の保存



図 24 ファイルダイアログボックス

をクリックすると、ファイルダイアログボックスが現れるので、XML ファイルとして保存されているファイルを開くだけで Hasse 図が描かれる (図 25, 26)。その後は Hasse 図の編集を行い、保存することができる。

4.5.2 その他補足事項

(1) テキスト領域の保存

メインメニューのテキスト領域に出力された形式概念に関するデータを Excel 等で図

表として整理または分析するため、次の 2 つの方法でファイルに保存することができる。

- ・「File」メニューから「Save」を選択する

図 24 と同じファイルダイアログボックスが表示されるので、適当な場所を指定すればテキストとして保存することができる。したがって、必ずしもその必要はないが、識別子 TXT を付けたほうが、エディタで編集しやすいだろう。

- ・コピー (あるいはカット)&ペースト

保存したい領域をドラッグして矩形で囲



図 25 Hasse 図のロード

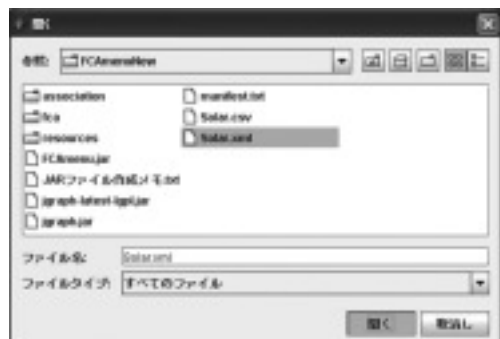


図 26 ファイルを指定する

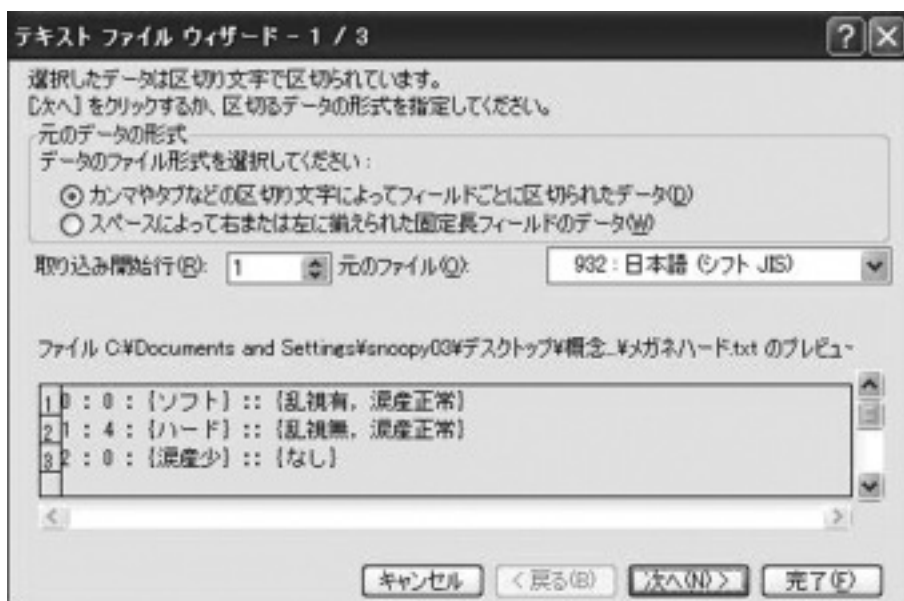


図 27 Excel で区切り記号の要求

み、コントロール (ctrl) キーを押しながら、(コピー) または X (カット) キーを押す、その後、適当なエディタまたはワードプロセッサにコントロール (ctrl) キーを押しながら、V キーを押せば、ペーストすることができる。

(2) Excel 等の利用

概念数が多くなれば一般に含意規則の個数も多くなるので、属性間の依存関係を見通すことは容易ではない。その場合上で述べ方法で保存したファイルを Excel 等の入力ファイルとして含意規則が成り立つ対象の個数の大小関係で並べ替えると見通しを得やすい。

形式概念解析の結果を保存したテキストファイルを Excel で開くと、データを区切る記号を尋ねてくる (図 27)。「カンマやタブなどの区切り文字によってフィールドごとに区切られたデータ」をクリックし、「次へ(N)」をクリックする。次の画面 (図 28) で「その他」をクリックし、空の記号欄に半角コロン(:)を入力し「完了(E)」をクリックすればフィールドに分割される。その後は必要なフィール

ドを指定し並べ替えすればよい。表 6, 7, 8 はこのようにして作成したものである。

5 おわりに

上で述べてきた形式概念解析ツールは、形式概念を計算し Hasse 図を描くものが必要だと思い自分のために作っているうちに、少しずつ自分以外の利用を意識するようになり次第に形をなしてきたものである。したがって、使い勝手のよくないところや盛り込むべき機能も多々あるにちがいない。しかし、冒頭で述べたように形式概念解析を多くの人に知ってもらうと同時に実際にも使ってみて欲しいのである。データの質的分析を行いたい人にも以下の理由から是非利用をお勧めしたい。

- 1) 形式概念解析はデータ解析の前処理として有益な示唆を与える。なぜなら数量的分析を行う前に属性の関連からデータ全体の傾向を把握しておくことによって分析の焦点を見定めることができる。

テキスト ファイル ウィザード - 2 / 3

フィールドの区切り文字を指定してください。[データのプレビュー] ボックスには区切り位置が表示されません。

区切り文字

☒ タブ(T) ☐ セミコロン(M) ☐ カンマ(Q) 文字列の引用符(Q): "

☐ スペース(S) ☒ その他(Q): :

☐ 連続した区切り文字は 1 文字として扱う(B)

データのプレビュー(Q)

0	0	{ソフト}	{乱視有, 涙産正常}
1	4	{ハード}	{乱視無, 涙産正常}
2	0	{涙産少}	{なし}

< >

キャンセル < 戻る(B) 次へ(N) > 完了(F)

図 28 区切り記号の設定

- 2) 形式概念解析が示す傾向と照合することによって数量的分析結果を（例えば、因子分析結果など）一層適切に解釈することができる。

研究あるいは非営利目的で利用したい方には解析ツールを配布いたします。ご連絡ください（連絡先：nagata@sgu.ac.jp）。

注

- 1) O を集合 L における関係とする。 O がつぎの性質をもつとき、 O を順序関係という。

- (1) 反射律 xOx
- (2) 推移律 $xOy, yOz \Rightarrow xOz$
- (3) 反対称律 $xOy, yOx \Rightarrow x=y$

ある集合 L において一つの順序関係 O が与えられたとき、 (L, O) を順序集合という。たとえば、自然数の間の通常の大小関係 \leq は、自然数の集合 N における一つの順序関係であり、 (N, \leq) と記す。

順序集合 L において、任意の二元 a, b に対して $\{a, b\}$ の上限および下限がいつも L の中に存在するとき、 L を束（「そく」と読む—Lattice）という。順序集合 L の空でない任意の部分集合 A に対してその上限 $\sup A$ および下限 $\inf A$ が存在するとき、 L を完備束という。ここで、 A ($\forall L$) の上限（最小上界） $\sup A$ とはつぎの条件をみたす要素である：

- (1) $a \in A \Rightarrow a \leq \sup A$
- (2) $(a \leq A \Rightarrow a \leq x) \Rightarrow \sup A \leq x$

$\sup A$ と双対的に、すなわち上の二つの条件の \leq を \geq に置き換えて、 A の下限（最大下界） $\inf A$ が定義される。

謝 辞

まず誰よりも葛西俊治本学人文学部教授に感謝申し上げます。全く面識のなかった先生がインターネットで調べて小生の形式概念解析に関心をお示し下さったばかりでなく、実際に研究に利用するためいろいろ注文を出して下さらなければ、ツール開発の進行は遅々

としていたであらうし、このような解説文書を書く機会もなかったでしょう。また社会情報学部研究会の席で質問・意見・コメントをし活発に議論して下さいました本学教員の皆様にお礼申します。そしていつものことながら、10年以上続いている研究会で種々議論し、形式概念についてコメントを下された田中一北海道大学名誉教授および千葉正喜本学社会情報学部教授にお礼申し上げます。

引用文献

- Davey & Priestley (2002): *Introduction to Lattices and Order* (2nd ED.), Cambridge University Press.
- Ganter, B and Wille, R. (1999): *Formal Concept Analysis*, Springer.
- Janssen, M (2002): *SIMuLLDA: a Multilingual Lexical Database Application using a Structured Interlingua* PhD-thesis. Utrecht: Universiteit Utrecht.
- Kohonen, T (1995): *Self-Organizing Maps*, Springer.
- 長田博泰 (2004): 「形式概念にもとづく質的分析」, 『社会情報』, Vol.14, No.1, pp.19-37, 札幌学院大学社会情報学部。
- (2006a): 「社会情報解析への一寄与：形式概念によるデータ解析」, 『社会情報』, Vol. 15, No.2(103-118), 札幌学院大学社会情報学部。
- (2006b): 「形式概念を用いたデータ解析：幼児発達行動の分析」, 『社会情報学研究』, Vol.11, No.2(71-81), 日本社会情報学会 (JSIS)。
- (2007): Web メニューデザインの形式的な方法—情報アーキテクチャへの形式概念解析の応用, 『社会情報』, Vol.16, No.2(1-17), 札幌学院大学社会情報学部。
- 横森貴・小林聡 (1994): ラフ集合と意思決定, リレー連載「ラフ集合」第3回, 数理科学, 9, pp.76-83.