

原著論文

階層的可視化手法を用いたアソシエーション分析によるプロファイリング

伊藤 晃*, 吉川 大弘*, 古橋 武*, 光松 佐和子**

* 名古屋大学, ** 名古屋経済大学短期大学部

Profiling by Association Analysis using Hierarchical Visualization Method

Akira ITO*, Tomohiro YOSHIKAWA*, Takeshi FURUHASHI* and Sawako MITSUMATSU**

* Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

** Junior College of Nagoya Keizai University, 4-25-13 Meieki Nakamura-ku, Nagoya 450-0002, Japan

Abstract : In the field of marketing, questionnaires are often carried out and the acquired data are analyzed in order to design a marketing strategy. Association analysis is one of the most powerful techniques for the analysis of questionnaire data. The purpose of this analysis is to extract important rules, e.g., {Sex=male \Rightarrow Evaluation of product A=3}, from the data, called association rules. It is an interesting approach to apply the association analysis to profiling in questionnaire data. However, it is difficult for the analyst to grasp the relations between rules and to find interesting rules from a large set of rules. This paper proposes a visualization method of relations between association rules with hierarchical graph structure to support the association analysis. This paper applies the proposed method to an actual questionnaire data on the impression of clothes for men's suits, and shows that the proposed method supports users to grasp relations between rules and to find interesting rules.

Keywords : Analysis of Questionnaire Data, Association Rule, Visualization, Hierarchical Graph Structure, Profiling, Cloth for Men's Suits

1. はじめに

アンケートは、企業が市場の動向を調査するための重要な手段の一つであり、アンケートデータを解析し、得られた知見に基づいてマーケティング戦略の立案を行うことがしばしば行われている[1, 2]。例えば、新製品の販売を企画する際に、その製品に関する印象調査を行い、得られたデータを解析して製品に対する印象・評価や購買層に関する知見を得ることで、コマーシャル戦略の立案、販売価格の設定、購買規模の予測、製品デザインへのフィードバックなどを行う。アンケートデータの解析には一般的に様々な多変量解析手法が用いられており、代表的な解析手法として主成分分析やクラスター分析、コンジョイント分析などがある。主成分分析は、変数（質問項目）を合成し、新たな統合指標を作り出す手法であり、各回答者を統合指標によって特徴付けることや、各質問項目の回答への寄与度などを分析することができる[3]。クラスター分析は、データを類似度に基づいて分類する手法であり、似通った傾向をもつ回答者をグルーピングし、解析する[4]。またコンジョイント分析は、評価対象に対する嗜好に影響を与える要素（色や価格など）や、その度合いを分析することができる[5]。

一方、アソシエーション分析[13, 14]は、POSデータ[6]などの巨大なデータベースから、データの要素間の相関を、ルール（規則性）の形で抽出することが可能であり、データに内在する価値のある項目間の関係を把握できる。アソシエーションルールは $\{A \Rightarrow B\}$ の形式で表現され、Aは条件部、Bは結論部と呼ばれる。アソシエーション分析には、結

論部を特定の項目に固定するルール抽出法と、固定しないルール抽出法があり、前述のPOSデータの要素相関解析などは、結論部を固定しない方法が有効である。しかしアンケートデータ解析において、結論部を固定しない方法を用いた場合、例えば $\{20代 \Rightarrow 女性\}$ 、 $\{既婚 \Rightarrow 会社員\}$ など、回答者の属性間の相関は得られるが、マーケティング戦略立案の上では重要ではないルールが多数抽出されてしまう。一方、対象製品の購入希望に対する回答など、解析の目的に沿った質問項目の回答を結論部に固定することで、 $\{20代 \Rightarrow 購入希望層\}$ 、 $\{既婚 \Rightarrow 非購入希望層\}$ など、ターゲット層のプロファイリングとして用いることが可能となる。こういったプロファイリングには、数量化Ⅲ類やコレスポネンス分析などが用いられることが多い[7, 8]が、これらは購入希望層等と関係性が高い項目（女性、会社員など）を抽出し、またそれらは相対的な関係の強さを表したものであるため、アソシエーションルールとして抽出された関係とあわせて解析することで、より有益な知見の獲得が期待できる。

しかし一般的に、アソシエーション分析において抽出されるルール数は膨大であり、特徴的なルールを見出すことや、ルール間の関係性を把握することは困難である。この問題に対して、抽出されたルールを可視化することで、ルールの特徴把握を支援する手法が提案されている[16-18]。しかし、これらはいずれも、結論部を固定しないルール抽出法を前提とした可視化手法であるため、ある特定の結論部を持つルール間における条件部の関係性の把握を行うことは難しい。また、アソシエーション分析以外にも、データからルールを抽出する手法として決定木[9]やラフ集合合理論[11]を用いるものなどがあり、それらにおいても可視化手法がいくつか

提案されている [10, 12] が、これらの手法においても、同一の条件が分散しうること、条件部の階層的関係、例えば {20代⇒購入希望層} と {20代∧女性⇒購入希望層}、{20代∧女性⇒購入希望層} と {20代∧会社員⇒購入希望層} といった関係を把握することが困難となることがある。

そこで本論文では、アンケートデータにおける、アソシエーション分析によるプロファイリングの支援を目的として、階層グラフ構造を用いることで、アソシエーションルールにおける条件部間の関係性を可視化する手法を提案する。提案手法では、ルール抽出の際の評価指標の選択や閾値の設定を、解析者であるユーザがインタラクティブに行うことが可能であり、また、ルール条件部の階層構造を能動的に展開していくことや、不要なルールの削除を行うことが可能である。さらにアンケートデータでは、データ数が十分にとれないことで、属性の偏りが生じることがある。提案手法では、この結果実質的には同じ意味となる（例えば回答者に会社員かつ既婚が多い場合、{会社員⇒購入希望層} と {既婚⇒購入希望層} のルールは独立ではない。）ルール間の従属関係も、あわせて把握することが可能である。提案手法により、ユーザが条件部の階層的構造やルール間の関係性を把握しながら、特徴的なルールの探索を行うことを支援することが可能となる。本論文では、メンズスーツ用の布地に関するアンケート調査データに提案手法を適用し、従来の表形式で表されている場合には困難であった、アソシエーションルール間の関係性や特徴的なルールの把握が行えることを示す。

2. アソシエーション分析

アソシエーション分析とは、データに潜む要素どうしの相関を、ルール（規則性）の形で抽出する手法である。アソシエーションルールは、{A⇒B} の形式で表現され、Aを条件部、Bを結論部と呼ぶ。通常、ルールの重要性を表す評価指標に基づき、閾値などの形で指標の条件をユーザが指定することによって、条件を満たすルールを抽出する。多く用いられる評価指標である *support*, *confidence*, *lift* を以下に示す [20]。

$$support = \frac{N(A \wedge B)}{N_r} \tag{1}$$

$$confidence = \frac{N(A \wedge B)}{N(A)} \tag{2}$$

$$lift = \frac{confidence}{P(B)} \tag{3}$$

N_r は全データ数（アンケートデータでは、全回答者数）、 $N(A)$ は条件Aを満たすデータ数、 $P(B)=N(B)/N_r$ である。例えば表1において、{ $(Q2=1) \wedge (Q3=1) \Rightarrow (Q4=1)$ } の *support* は $2/4=0.5$ 、*confidence* は $2/2=1.0$ 、*lift* は $1.0/(2/4)=2.0$ である。一般的には、この3つの指標を用いてルール抽出を行うことが多いが、これらがそれぞれの程度の値を満たせば重要なルールであると判断できるのかという明確な基準はない。

これに対し、統計学で用いられる χ^2 値を指標として用いる場合もある [21]。{A⇒B} において、表2のような分割表を考える。表中の各セルは観測度数を示しており、例えば、条件Aと結論Bをともに満たす回答者 ($N(A \wedge B)$) は a 人いることを表す。この分割表に対して、以下のような式を定義する。

$$T = \frac{N_r(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \tag{4}$$

式 (4) において、 T はAとBとの関連性を表し、 T が自由度1の χ^2 分布に従うことが知られている。これにより、統計で用いられる有意水準が、ルール抽出における一つの閾値の指標として用いることができる。例えば、有意水準を5%とすると $T > 3.84$ 、1%とすると $T > 6.63$ を閾値として、重要なルールを抽出することが可能である。また表2からは、{A⇒B} (ポジティブルール) だけでなく、{A⇒notB} (ネガティブルール) などとも式 (1) ~ (4) を指標に抽出することができる。

3. ルールの可視化

3.1 従来手法

1. はじめに述べた通り、アソシエーションルールの特徴把握支援を目的とした可視化手法はいくつか提案されている。しかしこれらの多くは、結論部を固定しない場合でのルールの可視化手法であるため、本論文で対象とするようなアンケートデータでのプロファイリングを目的として、結論部を固定した上で、ルール間の条件部の関係性の把握を行うことは困難である。図1に、Rule graph [19] によるアソシエーションルールの可視化結果の例を示す。図では、アルファベット A~D が、ルールの条件部または結論部を表しており、これと矢印を組み合わせることでルールを表現している。矢印の太さ（丸で囲まれた数字）は、*support* または *confidence* の大きさを表している。これにより例えば、アソシエーションルール {A⇒B} の *support/confidence* が0.4であること、結論部がBのルールの条件部はA、D、 $D \wedge C$ があることなど

表1 アンケートデータ例

回答者No.	質問項目			
	Q1	Q2	Q3	Q4
1	1	0	1	0
2	0	1	1	1
3	1	1	1	1
4	0	1	0	0

表2 {A⇒B} についての分割表

	B	notB	Σ_{row}
A	a	b	$a+b$
notA	c	d	$c+d$
Σ_{col}	$a+c$	$b+d$	N_r

がわかる。図の例では、条件部と結論部あわせて4つしかないため、DとD∧Cなど、条件部間の関係性も把握しやすいが、条件の数が多い場合や、ルール数が増えた場合に、可視化結果が煩雑になり、それらの関係性の把握が困難になるという問題点がある。また図2に、決定木 [10] によるルールの可視化の例を示す。図2において、例えば(outlook=sunny)という条件が、複数の箇所に分散して表示されていることがわかる。これにより、例えばoutlookとwindyとの関係など、ルールの条件部間における関係性を把握することが困難になると考えられる。

3.2 提案手法

提案手法では、階層グラフ構造を用いることで、アソシエーションルールにおける条件部間の関係性を可視化する。ここでは、提案手法について説明する。アソシエーションルールの例を表3に、これらのルールの可視化イメージを図3にそれぞれ示す。ここで表3のpは結論部である。図3の長方形で表されているのは結論部pであり、円形で表された各ルールの条件部をノードと呼ぶ。また、ノード間やノード-結論部間を結んでいる線をリンクと呼ぶ。ノードは条件の数により階層構造を持ち、色分けされて表示される。

3.2.1 縦リンク

異なる階層にあるノード間を結ぶリンクは条件部の包含関係を表している。リンクで結ばれたノードにおいて上の階層にあるものを上位ノード、下の階層にあるものを下位ノードと呼ぶ。下位ノードほど、条件が絞り込まれることでconfidenceやliftの値は大きくなるが、解析においては上位

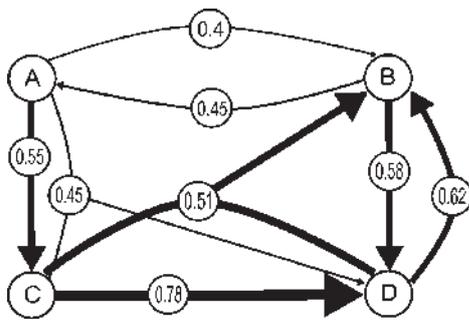


図1 Rule graphによる可視化

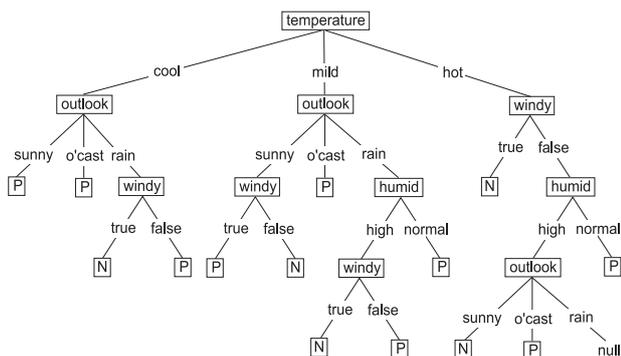


図2 決定木による可視化

ノードほどユーザにとって直感的にわかりやすいため、このように階層的に条件部を表示することは、解析を行う上で有益であると考えられる。また、上位ノードが存在しない(ルール抽出の条件を満たさない)ノードは結論部と直接リンクが結ばれる。このような上下階層でのリンクを以降「縦リンク」と呼ぶ。これにより、共通の上位ノードに基づいて階層的にルールをクラスタリングすることができる。また、上位ノードと下位ノードの評価指標を知ることで、条件が加えられた場合の各指標の変化を把握できる。

3.2.2 横リンク

例えば、2つのルール $\{a \wedge b \Rightarrow p\}$, $\{a \wedge c \Rightarrow p\}$ が抽出され、かつ条件bとcの間に強い関連があった場合、これら2つのルールは本質的に同じ意味を持つことになり、独立したルールとして扱われるべきではないと考えられる [15]。そこで、同じ階層にあるノード(条件部)どうしの組み合わせにおいて、2.で述べたルールの評価指標のうち、関連性を表す指標である式(3)のlift、式(4)のTを算出し、それらの値が閾値を超えるものを、関係が強い条件として、それらのノード間をリンクで結ぶ。liftとTのどちらか、あるいは両方を用いるか、またその閾値については、ユーザが指定することができる。ただし、条件部どうしが共通の条件を含む場合は、その条件を除いた上で指標を求める。例えば $\{a \wedge b \Rightarrow p\}$ と $\{a \wedge c \Rightarrow p\}$ の場合はbとcの組み合わせ、 $\{a \wedge b \Rightarrow p\}$ と $\{c \wedge d \Rightarrow p\}$ の場合は $(a \wedge b)$ と $(c \wedge d)$ の組み合わせにおいて関連性指標を求める。図3において、ノード $(a \wedge c)$ と $(a \wedge d)$ がリンクで結ばれている。このことは、 $\{c \Rightarrow d\}$ (または $\{d \Rightarrow c\}$) というルールに対するliftもしくはTの値が大きいことを意味し、対象となるデータ内では、 $\{a \wedge c \Rightarrow p\}$ と $\{a \wedge d \Rightarrow p\}$

表3 ルール例

ルール
$a \Rightarrow p$
$b \Rightarrow p$
$a \wedge c \Rightarrow p$
$a \wedge d \Rightarrow p$
$f \wedge g \Rightarrow p$
$a \wedge c \wedge e \Rightarrow p$
$a \wedge c \wedge f \Rightarrow p$

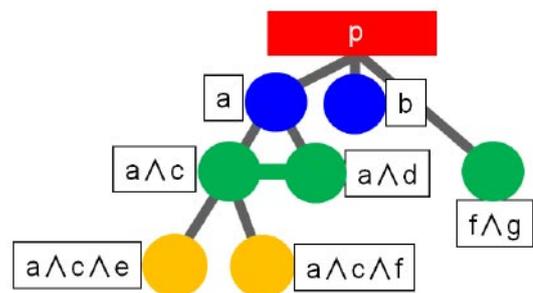


図3 提案手法のイメージ

という2つのルールが本質的にはほぼ同一の意味を持つことを表している。このような同一階層内でのリンクを以降「横リンク」と呼ぶ。

3.2.3 実装システム

提案手法を実装したシステムのインターフェースを図4に、出力されたルールの可視化の例を図5にそれぞれ示す。初めに、(A)のウィンドウで、アソシエーションルールの抽出条件を設定する。ここでは、ルールの結論部、式(1)～(4)のsupport, confidence, lift, Tに対する閾値や、条件の数、およびそれらの組み合わせにより条件を設定できる。これにより抽出されたルールは、表形式で(B)のウィンドウに表示されるとともに、メインウィンドウに図5のように呈示される。図の(C)では、後に述べるノードの呈示の大きさを決める指標の選択ができるとともに、(D)のように選択されたノード(ルール)の詳細が表示される。

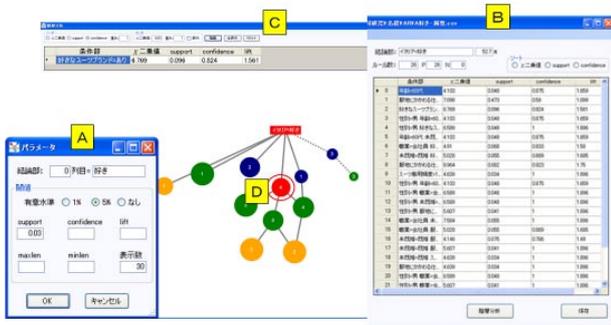
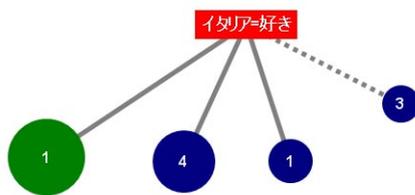
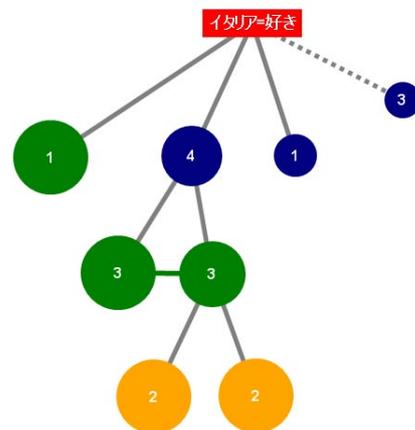


図4 実装システムのインターフェース



(a) 初期ルール表示



(b) 探索中盤

図5 可視化結果の例

図5において、実線の縦リンクはポジティブルール ($\{A \Rightarrow B\}$) を表し、破線の場合はネガティブルール ($\{A \Rightarrow \text{not} B\}$) を表している。ノードの大きさは、図4 (C) で指定したルールの評価指標の大きさを表しており、横リンクの太さは式(4)で表される指標Tの大きさを表している。これにより、ルール間におけるユーザの選択した指標の大小が視覚的に把握できる。また、各ノードに描かれている数字は、そのノードの下位につながるノード数を表している。

煩雑なルール表示は、ユーザによる特徴的なルール探索の妨げとなるため、図5 (a) に示すように、初期画面に表示するノードは結論部に直接つながるノードのみとする。ユーザは、図5 (b) のように、注目したノードを階層的に展開していくことができるが、この際、現れる下位ノードを、ノードの数や各指標の大きさを制限することができる。また、不要と判断したノードを削除することもできる。このように、条件部間の関係性や評価指標に基づくルールの重要性を把握しながら、ユーザが能動的に特徴的なルールの探索を行えることが本システムの特徴である。

4. 実験と考察

4.1 アンケートデータ

本実験では、メンズスーツ用の布地に関するアンケートデータを解析対象とする。本アンケート調査では、回答者に2種類の布地を実際に触ってもらい、各布地に対する印象をSD法により評価してもらった。アンケートデータは布地に対する評点データと、年齢、性別などの属性データ(表4)で構成される。評価対象である2種類の布地として、イタリア製のスーツ布地と日本製のスーツ布地を用いたが、回答者にはスーツ布地に関するアンケートであることのみを伝え、イタリア製/日本製であることや、どちらの布地が日本製であるかなどの情報は与えていない。回答者数は146名であった。

布地の印象評価の質問に、各布地について「1:嫌い~5:好き」という直接的に布の好き嫌いを問う質問があることから、本実験ではこの質問に注目、すなわちアソシエーション分析における結論部として、どのような属性を持った人がイタリア製の布地を好むのかを調べた。なお、今回のアンケートは比較的小規模なものであったため、5段階評価の評点を

表4 属性に関する質問項目

属性
性別
年齢
職業
未婚/既婚
服地にかかわる仕事経験あり/なし
スーツ着用頻度
好きなスーツのブランドあり/なし
スーツ選びで重視するところ

そのまま用いると、各評点に該当する回答者数が少なくなってしまう。そこで、5段階評価の1,2の回答を「嫌い」、3を「どちらでもない」、4,5を「好き」と置き換えた3段階の評点について解析を行った。

4.2 ルール抽出

提案手法により、アンケートデータからアソシエーションルールの抽出を行う。4.1で述べた通り、結論部は「好き（イタリア製）」に固定した。ルール抽出の条件として、 χ^2 値における有意水準5% ($T > 3.84$)とし、また、ごくわずかにしか当てはまらないルールは重要ではないと考え、ポジティブルール、ネガティブルールともに $support \geq 0.04$ (5人以下にしか当てはまらないルールは抽出しない)とした。

抽出されたポジティブルールを表形式にしたものを表5に示す。表5は $confidence$ の値でソートされている。条件の数は最大で5つとなっているが、それ以上のものは $support$ の閾値により抽出されていなかった。従来のアソシエーション分析では、表5のように表形式にされたルールを注目した指標でソートし、値の大きいルールから見ていく、という方法が多くとられる。本アンケートでは、回答者数が少ないため、抽出されたルール数が少なくなっており、表5からでも特徴的なルールの発見は可能であるが、各ルールが独立に扱われることで、条件部の関係性が把握しにくいと考えられる。また、 T や $support$ の閾値を少し下げたとき、抽出されるルール数は爆発的に増えるため、表5のような形で表されたルール群から、特徴的なルールや条件部の関係性を把握するに

は、提案手法のように可視化による支援が特に効果があると考えられる。

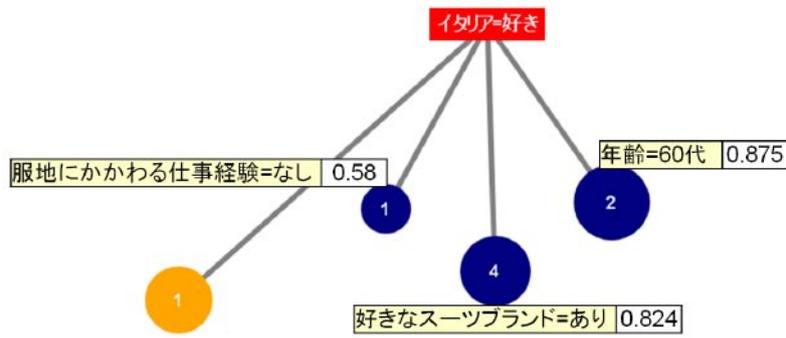
4.3 結果と考察

4.2で抽出されたルールに対して、提案手法を適用した結果を図6, 7に示す。横リンクは、条件部どうしの組み合わせによるルールにおいて、 $lift > 1$ かつ χ^2 値が有意水準0.1%を下回っている場合 ($T > 10.82$)に結ばれている。初めに、システム画面には図6 (a)が表示される。ここでは、ノードの大きさは $confidence$ の大きさを表しており、ノード付近にルールの条件部と $confidence$ の値を表記している。ここで、指標の中で $confidence$ に注目したのは、結論部を決定づける確率を表す $confidence$ の値が大きい条件を把握することが、実際のマーケティング戦略立案につながりやすいと考えたためである。図6 (a)より、ポジティブルールは主に「好きなスーツブランド=あり」、「年齢=60代」および「服地に関わる仕事経験=なし」を条件部に含むルールであることがわかる。この図6 (a)から、各ルールを階層的に展開し、特徴的なルールの探索を行っていく際、上位ノードに比べて $lift$ が0.1以上増加してない下位ノードは、冗長なルールと考えここでは削除した。

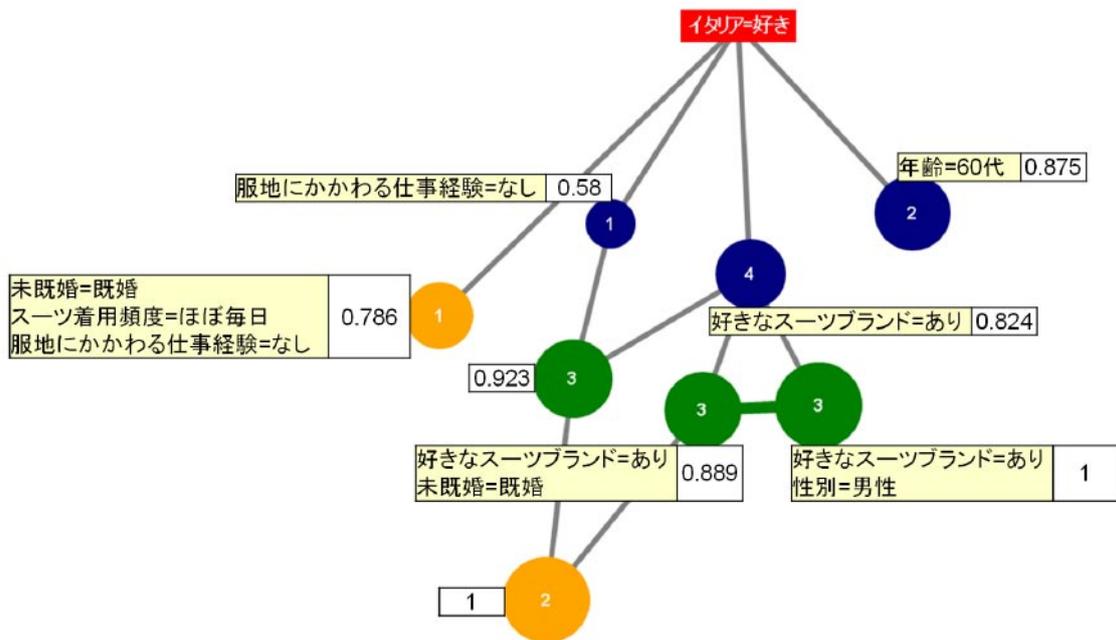
ユーザによるルール探索後のルールの可視化結果を図6 (b)に示す。縦リンクから、「好きなスーツブランド=あり」という条件に、「未既婚=既婚」、「服地に関わる仕事経験=なし」、「性別=男性」という条件を加えることで、それぞれ $confidence$ が大きくなっていることがわかる。特に、「性別

表5 抽出されたポジティブルール

条件部	T	$support$	$confidence$	$lift$
性別=男 好きなスーツブランド=あり	6.589	0.048	1	1.896
性別=男 職業=会社員 好きなスーツブランド=あり	6.589	0.048	1	1.896
性別=男 未既婚=既婚 好きなスーツブランド=あり	6.589	0.048	1	1.896
性別=男 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
職業=会社員 未既婚=既婚 好きなスーツブランド=あり	7.584	0.055	1	1.896
未既婚=既婚 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
性別=男 職業=会社員 未既婚=既婚 好きなスーツブランド=あり	6.589	0.048	1	1.896
性別=男 職業=会社員 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
性別=男 未既婚=既婚 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
職業=会社員 未既婚=既婚 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
性別=男 職業=会社員 未既婚=既婚 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.607	0.041	1	1.896
服地にかかわる仕事経験=なし 好きなスーツブランド=あり	8.964	0.082	0.923	1.75
未既婚=既婚 好きなスーツブランド=あり	5.028	0.055	0.889	1.685
職業=会社員 服地にかかわる仕事経験=なし 好きなスーツブランド=あり	5.028	0.055	0.889	1.685
年齢=60代	4.103	0.048	0.875	1.659
性別=男 年齢=60代	4.103	0.048	0.875	1.659
年齢=60代 未既婚=既婚	4.103	0.048	0.875	1.659
性別=男 年齢=60代 未既婚=既婚	4.103	0.048	0.875	1.659
職業=会社員 好きなスーツブランド=あり	4.91	0.068	0.833	1.58
好きなスーツブランド=あり	6.769	0.096	0.824	1.561
未既婚=既婚 服地にかかわる仕事経験=なし スーツ着用頻度=ほぼ毎日	4.146	0.075	0.786	1.49
性別=男 未既婚=既婚 服地にかかわる仕事経験=なし スーツ着用頻度=ほぼ毎日	4.146	0.075	0.786	1.49
服地にかかわる仕事経験=なし	7.098	0.473	0.58	1.099



(a) 初期ルール表示



(b) 探索後

図6 ポジティブルールの可視化結果

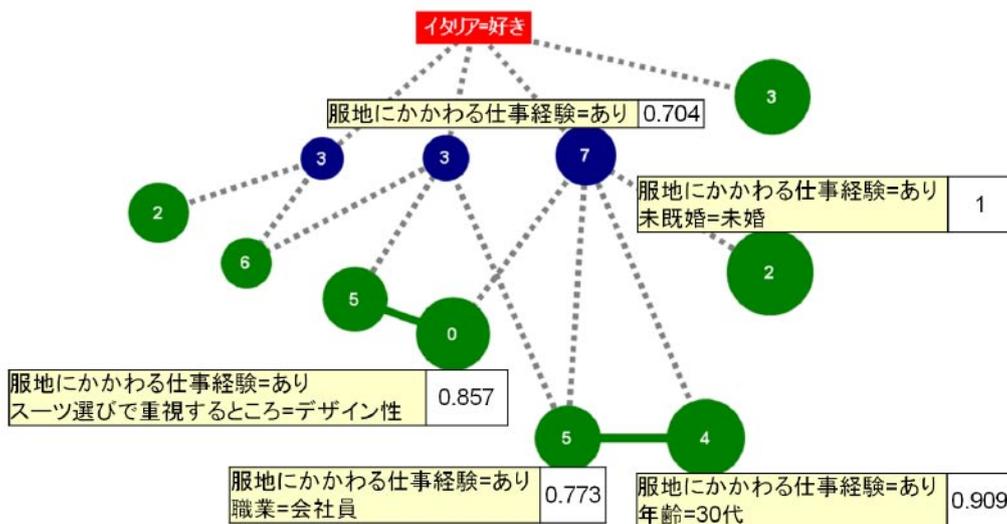


図7 ネガティブルールの可視化結果

＝男性」を加えることで、*confidence*は1.0にまで達している。すなわち、本調査の回答の範囲内では、好きなスーツブランドがある男性は全員、イタリア製のスーツ布地（回答者には知らされていない）を「好き」と回答していることがわかる。また、横リンクから、{好きなスーツブランド＝あり∧未既婚＝既婚⇒好き}と{好きなスーツブランド＝あり∧性別＝男性⇒好き}は、実質的にはほぼ同一のルールを異なった条件で表現しているだけであることがわかる。

一方、ネガティブルールについて2層目まで表示すると図7のようになった。「服地に関わる仕事経験＝あり」という条件を含むルールが多く出現し、また、*confidence*も大きくなっていることがわかる。すなわち、服地に関わる仕事経験がある回答者の多くは、イタリア製のスーツ布地を「好き」とは回答していないことを表している。背景として、イタリア製の布地は柔らかく、暖かみを感じられるものの、しわにもなりやすく、服地に関わる仕事を経験した人にとっては、経験のない人と比較して、必ずしも好まれないことが挙げられる。このような背景から、今回示したこれらのルールを服地マーケティングの専門家に呈示したところ、極めて興味深いものであるという意見が得られた。

提案手法により、アンケートデータにおける、アソシエーション分析によるプロファイリングが容易になることを評価するため、主観評価実験を行った。本実験では、12名の20代の理系学生に対し、表形式（表5）で表されたものと、提案手法により可視化形式で表されたものとでそれぞれプロファイリングを行ってもらい、分析の行いやすさを評価してもらった。被験者には、2.で示したアソシエーションルールの各評価指標の意味を説明した上で、表形式のものでは、各指標ごとにルールのソートが可能であること、および提案手法については3.2で示した説明を行った。評価は7段階（-3～+3）で、正の値が大きいほど可視化形式のものが良く、負の値が大きいほど表形式のものが分析を行いやすかったことを表している。また、呈示順による評価への影響を考慮し、半数の被験者は表形式を先に、もう半数の被験者は可視化形式を先に呈示した。その結果、12名の被験者による評価の平均値は1.25（標準偏差1.23）となった。この結果に対してt検定を行ったところ、1%の有意水準で有意差が認められた。

5. おわりに

本論文では、アソシエーション分析において、階層グラフ構造を用いることで、アソシエーションルールにおける条件部間の関係性を可視化する手法を提案した。メンズスーツ用の布地に関するアンケート調査データに提案手法を適用し、ユーザが条件部の階層的構造やルール間の関係性を把握しながら、特徴的なルールの探索を行うことを支援することが可能となることを示した。本実験において、専門家にとっても興味深いルールの抽出・発見を行うことができた。

今後、アソシエーションルールの抽出における閾値の検討

や、ユーザが特徴的なルールの探索を行った履歴に基づき、ユーザのルールに対する嗜好を知識として抽出する手法の検討などを行っていく予定である。

謝辞

本研究の一部は、文部科学省科学研究費（基盤研究（C）、No.22500088）の補助を得て遂行された。

参考文献

- [1] 木下祐介, 井上勝雄, 酒井正幸: 携帯電話機デザインの男女差の調査分析, 感性工学会研究論文集, vol.7, no.3, pp.449-460, 2008.
- [2] 関庸一, 宮崎晴夫, 十代田三知夫: 携帯電話機デザインの男女差の調査分析, 評点カテゴリーおよび評定対象間の類似度による評定空間の構成とその商品ブランドイメージの数量化への応用, 日本経営工学会誌, vol.38, no.5, pp.334-339, 1987.
- [3] 上田尚一: 主成分分析-講座<情報をよむ統計学8>, 共立出版, 2003.
- [4] 上田尚一: クラスタ分析-講座<情報をよむ統計学7>, 共立出版, 2003.
- [5] 君山由良: コンジョイント分析, データ分析研究所, 2006.
- [6] 阿部誠, 近藤文代: マーケティングの科学-POSデータの解析, 朝倉書店, 2005.
- [7] 松村真宏: 影響の普及モデルに基づくオンラインコミュニティ参加者のプロファイリング, 人工知能学会論文誌, vol.18, no.4, pp.165-172, 2003.
- [8] 内田治: すぐわかるSPSSによるアンケートのコレスポネンダ分析, 東京図書, 2006.
- [9] Quinlan, J.R.: Induction of decision trees, Machine Learning, vol.1, no.1, pp.81-106, 1986.
- [10] Kirkby, R.: Improving Hoeffding Trees, PhD thesis, Department of Computer Science, University of Waikato, 2008.
- [11] 森典彦, 田中英夫, 井上勝雄: ラフ集合と感性, 海洋堂出版, 2004.
- [12] 戸本裕太郎, 大平知美, 中村剛士, 加納政芳, 伊藤英則: 多値決定グラフのIf-Thenルール可視化への適用, 日本感性工学会論文誌, vol.9, no.2, pp.259-268, 2010.
- [13] Zhang, C., Zhang, S.: Association Rule Mining: models and algorithms, Lecture Notes In Artificial Intelligence, vol.2307, p.238, 2002.
- [14] Hang, J., Pei, J., Yin, Y.: Mining frequent pattern without candidate generation, Proc. of ACM SIGMOD, pp.1-12, 2000.
- [15] 岡田孝, 元田浩: 相関ルールとその周辺, オペレーションズ・リサーチ: 経営の科学, vol.47, no.9, p.565, 2002.

- [16] Bruzzese, D., Davino, C.: Visual Mining of Association Rules, Lecture Notes in Computer Science, vol.4404, pp.103-122, 2008.
- [17] Celgar, A., Roddick, J., Calder, P., Rainsford, C.: Visualising hierarchical associations, Knowledge and Information Systems, vol.8, no.3, pp.257-275, 2005.
- [18] Blanchard, J., Pinaud, B., Kuntz, P., Guillet, F.: A 2D-3D visualization support for human-centered rule mining, Computer&Graphics, vol.31, no.3, pp.350-360, 2007.
- [19] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., Verkamo, A.I.: Finding interesting rules from large sets of discovered association rules, Proc. of 3rd Int'l Conf. on Information and Knowledge Management, pp.401-407, 1994.
- [20] Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules, Proc. of the 20th VLDB Conf., pp.487-499, 1994.
- [21] Brin, S., Motwani, R., Silverstein, C.: Beyond marketbaskets: generalizing association rules to correlations, Proc. of ACM SIGMOD, pp.265-276, 1997.



伊藤 晃 (非会員)

2010年名古屋大学工学部電気電子・情報工学科卒業。同年名古屋大学大学院工学院研究科博士前期課程計算理工学専攻入学，現在に至る。主として感性工学に関する研究に従事。



吉川 大弘 (正会員)

1997年名古屋大学大学院博士課程修了。同年カリフォルニア大学バークレー校ソフトコンピューティング研究所客員研究員。1998年三重大学工学部助手。2005年名古屋大学大学院工学研究科COE特任准教授。2006年10月同研究科准教授。現在に至る。主としてソフトコンピューティングとその応用に関する研究に従事。博士（工学）。



古橋 武 (正会員)

昭和60年名古屋大学大学院工学研究科博士後期課程電気系専攻修了。工学博士。平成16年名古屋大学大学院工学研究科計算理工学専攻教授。現在に至る。ソフトコンピューティング、感性工学に関する研究に従事。平成8年日本ファジィ学会論文賞受賞。IEEE、日本知能情報ファジィ学会、電気学会等の会員。



光松 佐和子 (正会員)

お茶の水女子大学大学院家政学研究科修了。現在、名古屋経済大学短期大学部教授。衣服の着用快適性に関する研究。布地の感性価値に関する研究。日本家政学会、日本繊維製品消費科学会、日本衣服学会などの会員。