

電子上の社会ネットワーク研究 データ・マイニングを用いたルール発見

濱岡豊研究会 3期生 島村哲生*

平成 17 年 9 月 15 日

1 はじめに

まずはじめに、本論文の背景を 2 つの節で解説する。1 つはインターネットとマーケティングの関係であり、もう 1 つはコミュニティに関するものである。企業や消費者の環境の変化によるマーケティングの文脈の変化を、インターネットとマーケティングの関係から説き、今回焦点を当てているコミュニティについてふれている。消費者志向といわれるマーケティングにおいて消費者との関係性を変える技術（インターネット）と新たなマーケティングの可能性にふれられれば幸いである。

1.1 インターネットとマーケティングの関係 関係性マーケティングというフレームを用いて

関係性マーケティングとは、潜在需要を前提としないマーケティングである。つまり、従来のマネジリアルマーケティングでは、潜在需要への適合を戦略の目的としていたが、関係性マーケティングにおいては企業と消費者の相互作用を目的としている点が重要である。

潜在需要が見つからないという局面においては、企業と消費者が相互作用を繰り返すことで需要を創造するというものであり、すなわち、関係性マーケティングの中核概念は企業と消費者による価値共創であるといえる。

この関係性マーケティングを実行に移している企業がいくつか見受けられる。たとえば、無印良品¹のウェブサイト (<http://www.muji.net/>) に商品開発や改良など商品の製造に関するものから、商品の使い方や活用法をユーザー同士が共有

*慶應義塾大学 理工学部 管理工学科 4 年

¹株式会社良品計画が展開するブランド

しあうようなものまで、消費者間および消費者と企業間のコミュニケーションを高めようという努力が見られる。

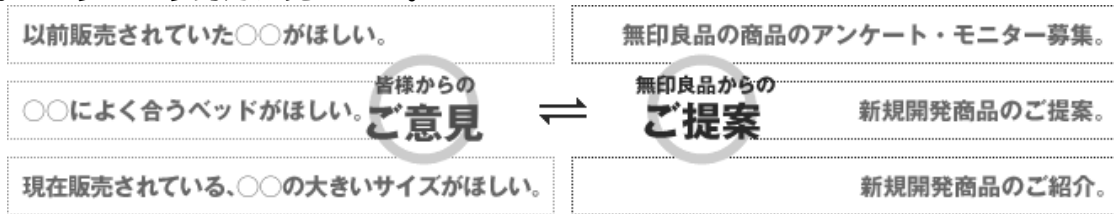


図 1: 無印良品のモノづくりネットコミュニティ 出所 [MUJI net]

このように、現在のマーケティングはこれまでのマーケティングの文脈から変化しているように感じられることと、それに対して適応しようとする企業努力が見られることが分かる。上述の企業の試みが表しているのは、これまで以上に消費者とのコミュニケーションが重要になっていることと、ただ情報を与えるだけでなく、消費者と意見のやりとりをする中で消費者の意向をうまく製品に反映させていく必要性が高まっていることが挙げられる。

1.2 コミュニティについて

これまでの多くの議論から、コミュニティというのは「構成員相互の交流」「共通の目標・関心事等の絆の存在」「一定の地理的範囲を伴うこと」が一般要件だと定義されてきた¹。また、「コミュニティとはある価値体系を共有し、その価値観のもとに協力する関係を持っている集団」とも定義される²。インターネットという仮想空間上のコミュニティと実空間上のコミュニティの大きな違いは時間と空間の依存性で示される。インターネット上のコミュニティでは時間と空間による行為者の制限が取り払われる。インターネットコミュニティの特性を「双方向性」、多くの利用者内でも特定少数間の情報流通の可能性を示す「脱大衆化」、情報利用の時間的制約からの脱却を示す「無同期性」とする論もある³。また、電子的な空間において各員は純粋に情報によって結ばれた共同体を形成している⁴ということも可能である。

つまり、インターネットコミュニティの特徴は、コミュニティに参加するものにとって地理的な影響や時間的な影響、また対面しないということで仮想世界における安易な接触可能性によって、情報の取得が圧倒的に容易になったこと表現できる。また、コミュニティによっては匿名性が維持されるものもあり、実社会のコミュニティでは実現しないほどのフランクな情報のやり取り⁵や、接点を持ちに

² 國領二郎 『オープン・アーキテクチャ戦略』ダイヤモンド社, 1999

³ E.M.Rogers. 『コミュニケーションの科学 マルチメディア社会の基礎理論』共立出版, 1992

⁴ Norbert Bolz 『ゲーテンベルクの銀河系の終焉 新しいコミュニケーションのすがた』, 1999

⁵ FPN (Future Planning Network) ニュースコミュニティ <http://www.future-planning.net/x/modules/news/article.php?storyid=662> 『ソーシャルネットワーキングと社会の構造改革』という記事がある

くい専門家と関わるなどすれば高度な情報のやり取り⁶も期待できる。

以上のことを考慮すると、インターネットという技術によって情報取得コスト⁷が著しく低下したことを受けて、従来のクチコミよりも情報の伝達量や速度が向上したと考える。そのため、これまで地理的制約の大きかったクチコミというメディアに比べて、インターネット上のコミュニケーションによる情報伝達は実社会のクチコミ以上の影響力を持つ可能性がある⁸。

それと同時に、情報の還流速度が増すということによる情報の影響力の増大は各個人の影響力が大きくなるということ⁹につながる。重大な影響を与えうる情報は速やかにコミュニティ内を還流し、関わるものへ影響を与えたと考えられる。そのため、インターネットコミュニティ上では実社会のコミュニティに比べて個人の持つ影響力は大きい。そのため、各個人が自立的に、そして同時に各要素と相互作用することで活性化していくと考えられる。

したがって、より個人の影響力を考慮し、また、要素間(この場合、各個人)の相互作用を念頭に置いた方法を用いてインターネットコミュニティを解釈しようとする。

2 目的

前述の通り、マーケティングにおける消費者自身の意向を知ることと、ウェブ上のコミュニティという新たなコミュニケーションの形態とを混合することが重要であると考え、インターネットコミュニティに参加している人の傾向を記述することを目的とする。これまでクチコミという情報伝播は消費者同士の地理的制約を受ける事が多い一方、インターネット上のコミュニティでは地理的制約が取り払われたという変化があった。しかしながら、その変化によって各人がどのような行動を新たにとるようになったかは一考の価値があると思われる。たとえば、インターネット上で自ら情報を発信する人や、情報を発信する人とうまくコミュニケーションする人の行動パターンを把握することは新たな創造活動に応用できるかもしれない。

⁶今回使用した統計パッケージソフト R では、筑波大学の岡田 昌史さんという研究者(筑波大学講師)を中心に wiki が展開され、利用法などのユーザー間で情報のやりとりがある。また、その wiki から『The R Book - データ解析環境 R の活用事例集』 九天社,2004 という本も出版されている。

⁷情報へのアクセスが容易になったことを指す。

⁸モルドバ共和国出身の O-ZONE による『DISCO-ZONE ~恋のマイアヒ~』というアルバムがオリコン・アルバム総合チャート1位(2005年8月25日付)を獲得した。この曲は名古屋のFM局でオンエアされて以来名古屋周辺の人の口コミで広まり、匿名掲示板 2ちゃんねるを中心に大ブレイクした。詳しくは <http://blog.livedoor.jp/gilcrow/archives/28147424.html>

⁹インターネット上ではマイノリティの意見を切り捨てる傾向が低い。詳しくは <http://www.umechando.com/view/13.htm>

そこで今回は特定の、目的のはっきりしているコミュニティ¹⁰を対象とせず、様々な目的をもったコミュニティを対象とする。ここでいう様々な目的をもったコミュニティとは、集まること自体に重点を置き、目的意識はその都度形成するようなものを指し、一般的には社会ネットワーク¹¹と呼ばれるものを今回は対象とする。

したがって、今回はコミュニティに参加する各個人の行動を記述し、その傾向をつかむことを目的とする。傾向をつかむことによって、コミュニティに対し、また、コミュニティの参加者に対して適切な行動を起こすことができ、それによって、新たな情報流を生み出すことが可能になる。

また、同時に、各人が持つ個人情報の発信についても考慮したい。ここでは特に職業という社会的に情報の価値が高いものを考慮しようと思う。というのは、その人がどんな仕事をしているかということはその人が発信する情報に対する信頼性に大きく影響を与えると考えられるし、同時に、職業が分かっていると、その人に対するアクションというのも変化するためである。つまり、人材発掘を目的としている人ならばほしい人材に対して適切なアクションを起こしやすいし、コミュニティにおける発言等についても、発言者の従事しているものが分かれば当該テーマについて議論を進めるなどしてお互いの知見を深めることも容易になるかもしれない。そういった意味で、個人の嗜好以上にコミュニケーション上で大きな役割を担うと思われる。

3 仮説について

今回扱うソーシャルネットワーキングサイトでは、様々な目的を持ち、また、様々な理由でそのサイトを利用している人たちがいると考えられる。そういったコミュニティサイトの中で、情報を受けたり、発信したりするという選択を各人が行っている。自ら情報を発信したり、個人情報を公開するという、他人に対して『自分を出す』という行動をとっている人達を分類することで、そのサイトの中にいる人たちの行動を記述したいと考える。

つまり、ある共通の要因があって同じ行動を引き起こしている人たちをまとめられるのではないかと考えている。

今回の仮説としては、『情報を発信する人は同じ行動様式をとる』とする。

¹⁰前述の無印良品のネットコミュニティ等。自社の製品を開発・活用という意味で目的がはっきりしている。

¹¹日本国内でソーシャルネットワーキングサイトとして有名なものは登録者数100万人を超えたmixiやGREEがある。

4 手法紹介

今回の分析では一般的なデータ・マイニングの手法2つを用いる。1つは相関ルール (association rule)¹²であり、もう1つは決定木 (decision tree) 分析である。これらの手法はデータ・マイニングの手法としてよく知られ、コミュニティ内の参加者の行動の特徴を示すのに適切であると判断した。以下にデータ・マイニングおよび、今回使用した2つの手法について紹介する。

4.1 データ・マイニングについて

データ・マイニング (data mining) について簡単な説明を行う。データ・マイニング、データからの知識発掘とは、一般的に大規模なデータベースから有用なパターンやルールを発見し、新たな知識¹³を発見・学習するプロセスのことである (矢田 2004)。データ・マイニングによって得られる新たな知識 (情報) は、パターンやルール発見の結果としてだけでなく、有用性¹⁴を持たなければならない。

そのため、データ・マイニング研究では3つの主な領域があり、それぞれ、知識表現、評価基準とアルゴリズムに大別される。知識表現 (knowledge representation) は膨大なデータを人間が理解できる形態に一定の手続きに基づいて変換する表現の枠組み (矢田 2004) である¹⁵。評価基準とは、大量なデータから得られる大量のルールやパターンを効率的に発見するために必要な判断基準のことである¹⁶。アルゴリズムとは、様々なパターンやルールの発見に必要なものである。適切なアルゴリズムによって発見のプロセスが効率化されることに対する研究も需要である¹⁷。

4.2 相関ルールについて

今回使用した相関ルール (association rule) について詳しく述べる。相関ルールはID付POSデータのデータ・マイニングなどで使用される分析手法の1つである。相関ルールを使用すると、バスケット分析のような同じバスケット (買い物かご) に入れられやすい商品の組を得るように、ある行為 (選択) をする際に同時に行われやすい行為 (選択) を選び出すことができる。このことを定式化すると以下のようなになる。

$$X \Rightarrow Y$$

¹² 連関規則とも呼ばれる。

¹³ データ (data) と対比として情報 (information または intelligence) と考えるとわかりやすい。

¹⁴ 得られた知識が利用可能性を持つという意味で。

¹⁵ 今回用いた決定木もまた知識表現の1つである。

¹⁶ 今回用いた相関ルールにおけるサポートおよびコンフィデンスも評価基準の1つである。

¹⁷ 今回用いた決定木発見のC4.5アルゴリズムなど。

であり、このとき X, Y は

$$\phi \quad X \subseteq I, \phi \quad Y \subseteq I, X \cap Y = \phi$$

を満たす部分集合である。ここで I は全てのアイテムを含む集合である。以下に簡単な例を示すと、

例 簡単な例として、毎食後に飲む飲料について考えてみる。ここでは選択肢の集合を以下のようにし、

$$I = \{ \text{紅茶, コーヒー, 緑茶, ジュース, 水, アルコール} \}$$

X は第 n 回目の飲み物とし、 Y は第 $n + 1$ 回目の飲み物とし、前回飲んだものは一切飲まないという仮定をおくと、 $X \cap Y = \phi$ を満たす。何人かのデータを収集したところ、以下のようなレコード (記録) が得られたとする。8 人のサンプルに対して以下のようなレコード (表 1) が得られたとする。

表 1: 得られたサンプル (仮想)

ID	n 回目の選択	$n + 1$ 回目の選択
1	コーヒー	アルコール
2	紅茶	緑茶
3	コーヒー	水
4	緑茶	水
5	紅茶	緑茶
6	紅茶	アルコール
7	水	コーヒー
8	紅茶	緑茶

ここで、相関ルールの結果が

$$X = \{ \text{紅茶} \}, Y = \{ \text{緑茶} \}$$

となったとすると、つまり、

$$\{ \text{紅茶} \} \Rightarrow \{ \text{緑茶} \}$$

となったということは、紅茶を飲んだ次の回には緑茶を飲むというルールがあるということである。

4.2.1 評価基準について

ここでは、上記のようなルールの評価基準について述べる。例で示したようなルールが有用であるかどうかをはかる指標について。相関ルールでは次の2つの基準が設けられている。1つはサポート (support) であり、もう1つは確信度 (confidence) である。

1. サポート (support) について

サポートとは次のようなものである¹⁸。

$$\text{support}(X \Rightarrow Y) \equiv P(X \cup Y)$$

したがって、サポートはそのルール自体の出現率を表している。

先ほどの飲み物の例を振り返ると、得られたルール { 紅茶 } \Rightarrow { 緑茶 } のサポートは $3/8$ である。これを他のものについて考えてみると、たとえば { コーヒー } \Rightarrow { アルコール } では、 $1/8$ であり、得られたルールの方がサポートが大きい。このように、ルール間の比較が行える。

2. 確信度 (confidence) について

確信度とは次のようなものである。

$$\text{confidence}(X \Rightarrow Y) \equiv P(Y|X)$$

したがって、確信度はルールの条件 (X) が発生したときに、結論 (Y) が起こる割合を示している。

同様に先ほどの飲み物の例で考えてみると、得られたルール { 紅茶 } \Rightarrow { 緑茶 } の確信度は $3/4$ となり、サポート同様に他のルールと比較することができ、たとえば、{ コーヒー } \Rightarrow { アルコール } の場合では確信度が $1/2$ となる。

3. 2指標を用いて以上の2指標を用いて各ルールについて比較を行うことができ、また、目的にあったルールを選定することもできる。ルールを選定する際には、2指標を用いて以下のように行う。

$$Pr(X \cup Y) \geq \text{min}_{\text{support}}$$

であり、

$$Pr(X|Y) \geq \text{min}_{\text{confidence}}$$

となるようなルールを“興味深い”ルールとして選定する。先ほどの例を用いると、次のようになる¹⁹。

表 2: 例ルール評価

¹⁸@ IT 情報マネジメント用語事典 を参考にした。

¹⁹support の下限値を $1/8$ とし、confidence の下限値を $1/2$ とした。

	サポート	確信度	興味深いかな否か
{ 紅茶 } ⇒ { 緑茶 }	3/8	3/4	
{ コーヒー } ⇒ { アルコール }	1/8	1/2	×

このようにして、ルールを選定し、有用な相関ルールを発見する。

4.3 決定木分析について

決定木分析もまた、相関ルールと同様にルールを選定する分析方法といえる。与えられたサンプルデータ群をその属性変数の値から分類し、その繰り返しによってデータ全体を樹形モデルで表現する手法をいう。決定木分析は特にデータ・マイニングにおいては、分析行程は一定の規則（アルゴリズム）によって自動的にデータを分類していくものである。一般に、決定木分析は2分木構造を作成する。つまり、1つのノードにおいてはyesかnoかの分岐を生成するため、属性変数に応じて、対象を分析することができる。

つまり、目的変数（被説明変数）と説明変数の間のルール抽出することができるのである。決定木分析の特徴の1つは特定の分布に従うという仮定を必要としない点である。そのため、説明変数・目的変数ともに質的変数・量的変数の双方を扱うことができるし、もちろん線形性を考慮しないものも扱える。

決定木を用いることで、樹形“図”の形で示されるため、視覚的に理解しやすい。

5 分析・結果

5.1 データについて

今回はあるソーシャルネットワーキングサイトにおける登録者の内、500人分のデータを採取²⁰し明らかに不自然な者を取り除き、その後予備分析においてはずれ値となった者を取り除いてルールを選定することとした。

また、今回の分析には統計パッケージソフトR²¹および、機械学習パッケージソフトWEKA²²を使用した。

²⁰項目毎に{表示・非表示}の2分表現にしたため、個人の特定は不可能。

²¹R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

²²Waikato Environment for Knowledge Analysis Version 3.4.5 (c)1999-2005 University of Waikato New Zealand

5.2 アルゴリズムについて

5.2.1 相関ルール

相関ルールではアプリアリ (Apriori) アルゴリズムを用いた。

5.2.2 決定木分析

決定木分析では C4.5 アルゴリズムを用いた。

5.3 分析結果

5.3.1 データの概要

データの概要を説明する。各個人の公表状態を記述するためにパラメーターを以下のように設定した (表 3)。

表 3: 変数説明表

変数名	説明	値
address	住所を公開しているか	2 値 (0(非公開), 1(公開))
sex	男性か女性か	2 値 (0(女性), 1(男性))
birthday	誕生日を公開しているか	2 値 (0(非公開), 1(公開))
age	年齢がいくつか	公開者のみその値
blood	血液型を公開しているか	2 値 (0(非公開), 1(公開))
hometown	出身地を公開しているか	2 値 (0(非公開), 1(公開))
hobby	趣味はいくつあるか	個数
job	職業を公開しているか	2 値 (0(非公開), 1(公開))
affiliation	所属先を公開しているか	2 値 (0(非公開), 1(公開))
self.introduction	自己紹介を記述しているか	2 値 (0(非公開), 1(公開))
favorite	好きなものの登録状況	個数 (0 ~ 3 個)
photo	写真を公開しているか	2 値 (0(非公開), 1(公開))
friend	友人の数	人数
community	所属コミュニティの数	所属数
comment	友人からの本人に対するコメント	2 値 (0(非公開), 1(公開))
review	書籍や CD などのレビューを付けているか	2 値 (0(非公開), 1(公開))
diary	日記を公開しているか	2 値 (0(非公開), 1(公開))

データクリーニング前の各値の状況は次の図 (図 2) の通り。

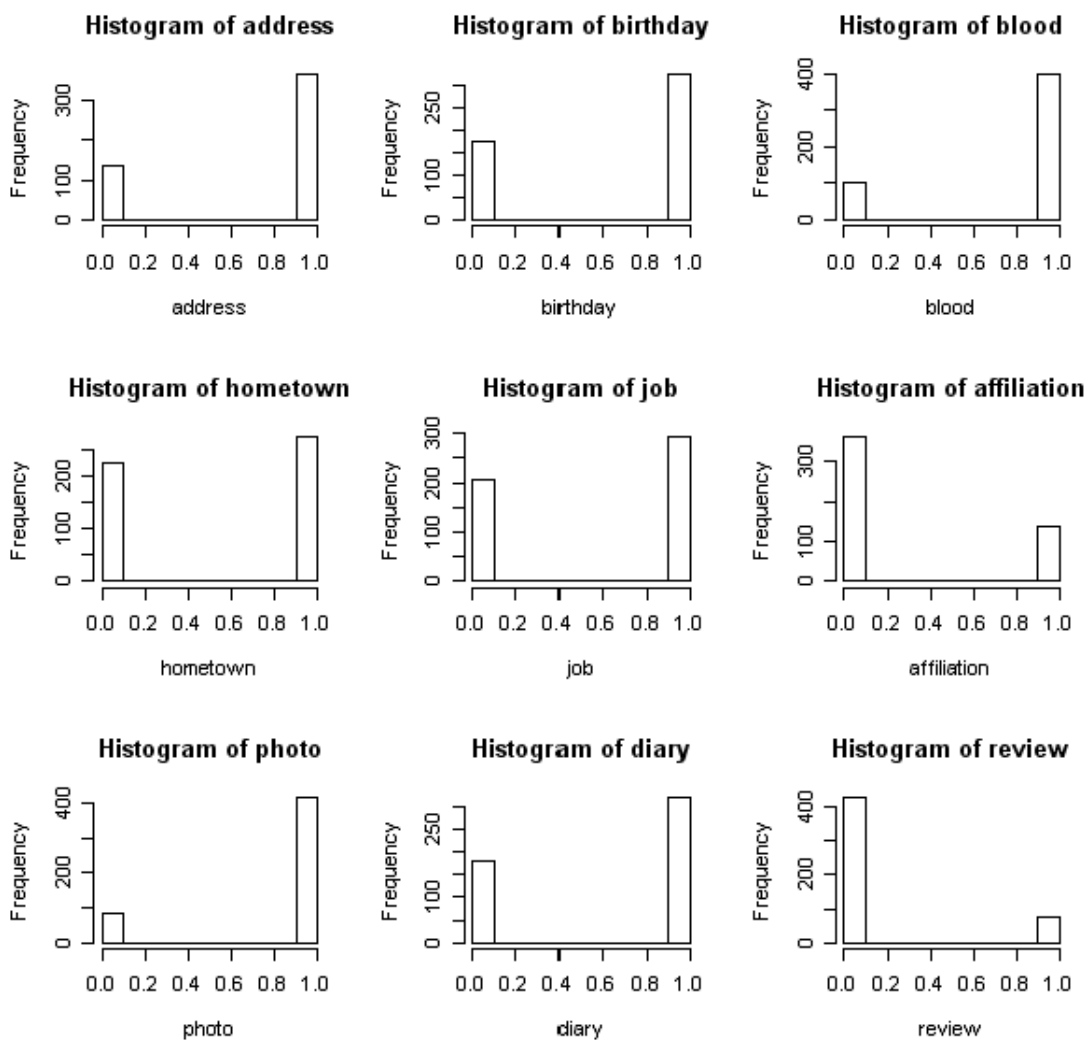


図 2: 2 値のものの公開状況をヒストグラムで

住所や誕生日、血液型や写真などの航海者が多いことが分かる。こういったいわゆる個人情報²³でも公開している人の方が多いことが分かる。一方少ないのは、所属先とレビューである。

²³個人情報の保護に関する法律(日本)第1章 総則 2 定義(2条)によると、個人情報とは『生存する個人に関する情報(識別可能情報)』となっている。

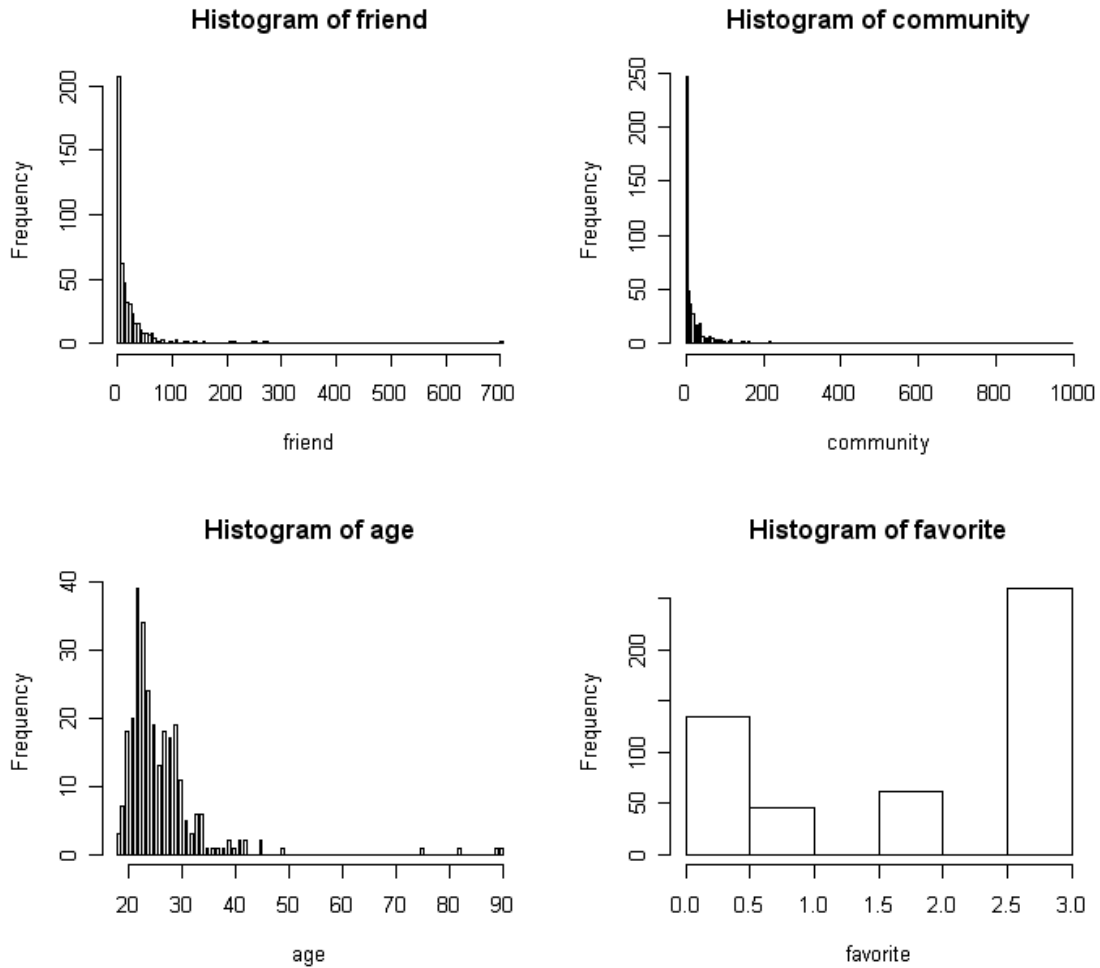


図 3: 2 値でないものの公開状況をヒストグラムで

データクリーニング前の場合，上図(図 3)の友人数・年齢・コミュニティ数ではごく一部の人がきわめて大きな数の登録数を持つ者が混ざっている。たとえば，年齢のグラフで 90 歳付近の者が混ざっているが現実的でないのが自然である。友人数でも同様に，700 名もの登録を行っている者が混ざっている。そこで，はずれ値をのぞいたデータクリーニング後のヒストグラムを次に示す(図 4)。

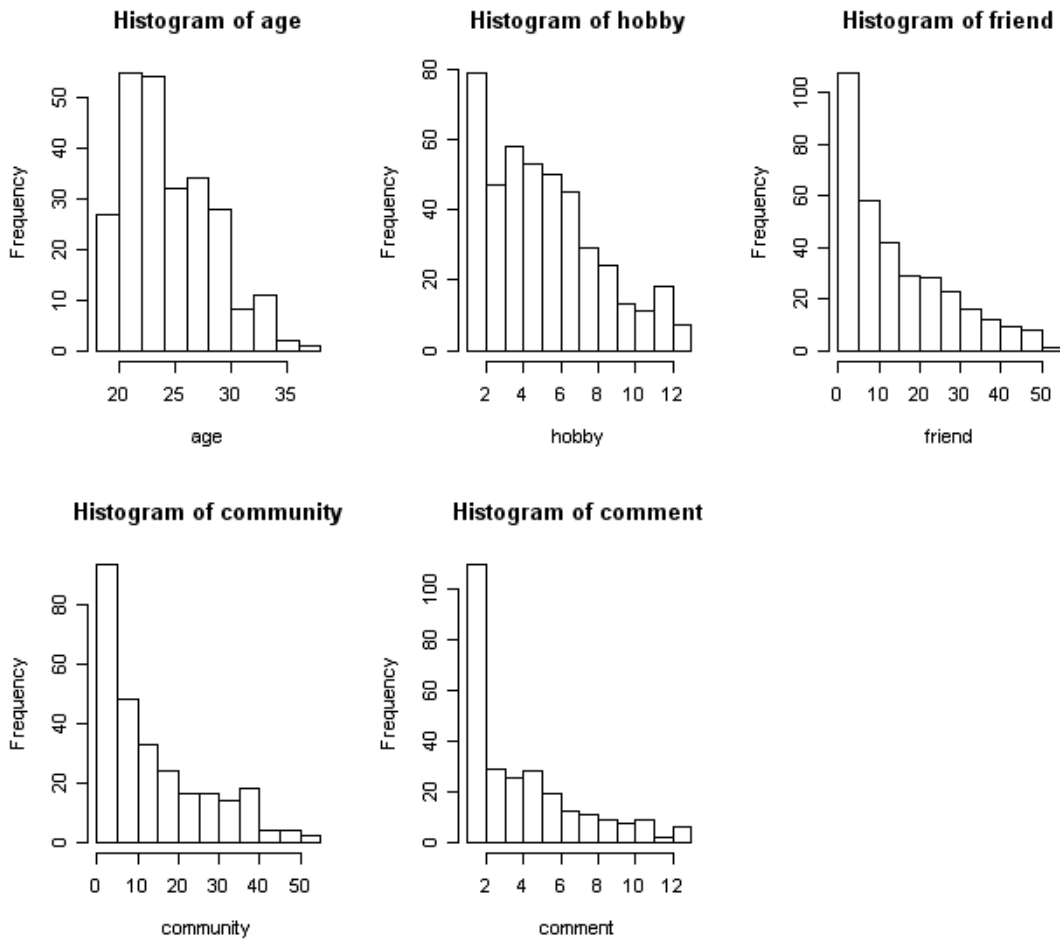


図 4: データクリーニング後をヒストグラムで

はずれ値をのぞくことで、こういった分布をしているかがより分かるようになった。年齢の分布は20代前半をピークに、正規分布に近い分布になっており、また、他の属性では右下がりとなっており、hobby に関してはやや正規分布に近いものの、他の3属性についてはポアソン分布になっているように見える。

以降ではクリーニング後のデータ²⁴を用いて、分析を行った。

5.3.2 相関ルール

相関ルールを用いた分析では、ルールの結論部にレビュー・日記・友人からのコメント・職業(所属先)をもつものを対象に行った。これは、各個人について他の人に対する影響力(受ける側としても与える側としても)を検討したく、そういった影響力を持っている個人の公開状況にルールがあるかを検討した。また、一方で、人材発掘の観点から、どのような人が職業や所属先を公開しているかを検討

²⁴クリーニングによってレコード数(サンプル数)は500から467となった。

することも考慮した。

全ての検討はアプリアリ (Apriori) アルゴリズムを用いた。以下のような設定 (図 5) で分析した。

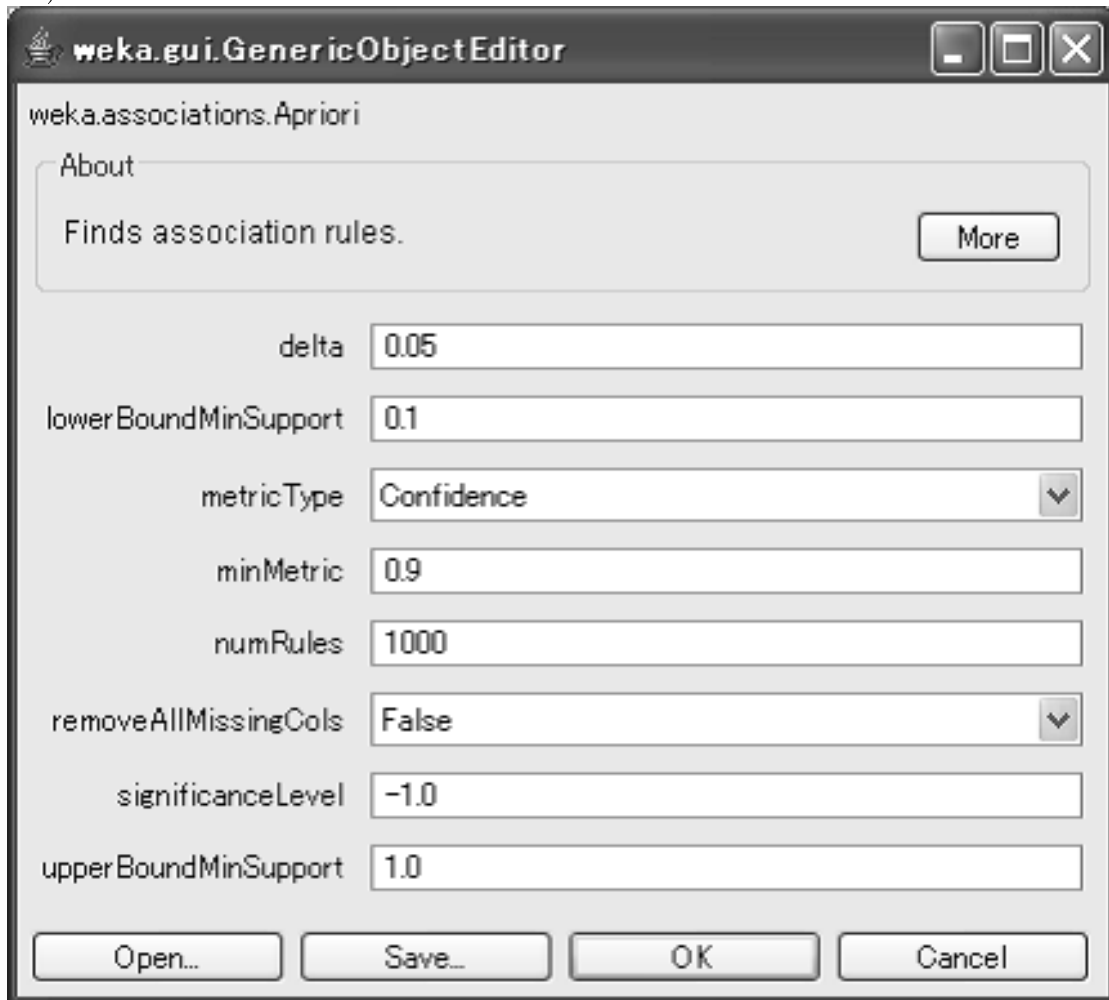


図 5: 相関ルールの検索方針

設定の説明 delta – Iteratively decrease support by this factor. Reduces support until min support is reached or required number of rules has been generated.

lowerBoundMinSupport – Lower bound for minimum support.

metricType – Set the type of metric by which to rank rules. Confidence is the proportion of the examples covered by the premise that are also covered by the consequence.

minMetric – Minimum metric score. Consider only rules with scores higher than this value.

numRules – Number of rules to find.

removeAllMissingCols – Remove columns with all missing values.

significanceLevel – Significance level. Significance test (confidence metric only).
 upperBoundMinSupport – Upper bound for minimum support. Start iteratively decreasing minimum support from this value.

次の表(表 4)にこれから注目していく項目の表示率を示す。また、友人からのコメント数に関しては、4分位点を基準に(少数・やや少数・やや多数・多数)の4つに分割した。467レコードのうち、年齢や好きなもののような個数や数を指定できる項目についても同様に、はずれ値をのぞいた部分を4分位点を基準に(少数・やや少数・やや多数・多数)の4つに分割したあと、はずれ値を(非常に多い)という項目にし分割した。

表 4: 項目毎の表示率

注目する項目	表示者数	非表示者数	表示者率 (%)
レビュー	67	400	14
日記	296	171	63
職業	272	195	58
所属先	126	341	27

1. レビューについての相関ルール

まず、レビューに関する検討を行う。

設定(図 5)にしたがって、コンフィデンスが0.9以上となるようなアソシエーションルール1000個を生成し、そのうち検討すべきルールを抽出しそれを表の形にした。以下の各項目についての検討でも同様に行った。まず、レビューについては結論にレビューを含むものを列挙した。それを以下の表に示す。それは次の表(表 5)の通り。表中の条件、および結論の末尾の値はそのような条件であるようなレコード数を表し、コンフィデンスについては、conf:()の括弧内の値である。

表 5: レビューを結論に含むルール

参照番号	条件		結論	コンフィデンス
878	好きなもの=非表示 125	⇒	レビュー=非表示 118	conf:(0.94)
748	誕生日=非表示	⇒	レビュー=非表示 118	conf:(0.95)
	年齢=非表示			
	所属先=非表示 124			
637	出身地=非表示	⇒	レビュー=非表示 118	conf:(0.96)
	自己紹介(自由文)=表示			
	職業=非表示			

	所属先=非表示 123			
865	年齢=非表示	⇒	レビュー=非表示 119	conf:(0.94)
	職業=非表示			
	所属先=非表示 126			
846	友人からのコメント=非表示	⇒	レビュー=非表示 120	conf:(0.94)
	所属先=非表示 127			
982	誕生日=非表示	⇒	レビュー=非表示 122	conf:(0.94)
	自己紹介(自由文)=表示			
	出身地=非表示 130			
631	出身地=非表示	⇒	レビュー=非表示 122	conf:(0.96)
	職業=非表示			
	所属先=非表示 127			
969	年齢=非表示	⇒	レビュー=非表示 123	conf:(0.94)
	自己紹介(自由文)=表示			
	出身地=非表示			
	所属先=非表示 131			
947	誕生日=非表示	⇒	レビュー=非表示 126	conf:(0.94)
	出身地=非表示 134			
801	誕生日=非表示	⇒	レビュー=非表示 126	conf:(0.95)
	自己紹介(自由文)=表示			
	年齢=非表示 133			
938	年齢=非表示	⇒	レビュー=非表示 127	conf:(0.94)
	出身地=非表示			
	所属先=非表示 135			
786	誕生日=非表示	⇒	レビュー=非表示 130	conf:(0.95)
	年齢=非表示 137			
569	自己紹介(自由文)=表示	⇒	レビュー=非表示 130	conf:(0.97)
	コミュニティ=非表示 134			
562	コミュニティ=非表示 137	⇒	レビュー=非表示 133	conf:(0.97)
874	所属先=非表示	⇒	レビュー=非表示 135	conf:(0.94)
	自己紹介(自由文)=表示			
	日記=非表示 143			
842	所属先=非表示	⇒	レビュー=非表示 138	conf:(0.95)
	日記=非表示 146			
794	自己紹介(自由文)=表示	⇒	レビュー=非表示 145	conf:(0.95)

	友人からのコメント=非表示 153			
782	友人からのコメント=非表示 157	⇒	レビュー=非表示 149	conf:(0.95)
939	自己紹介(自由文)=表示	⇒	レビュー=非表示 158	conf:(0.94)
	日記=非表示 168			
925	日記=非表示 171	⇒	レビュー=非表示 161	conf:(0.94)
785	コミュニティ=非表示 137	⇒	自己紹介(自由文)=表示	conf:(0.95)
			レビュー=非表示 130	

ある商品や物に対してレビューを付けるという行動は、他者への影響力を測る1つの指標と考えられそうであるが、今回の相関ルールからは、レビューを表示するルールを生成できなかった。各ルールを見ていくと、条件部には様々な項目が挙がっているのだがどの項目でも非表示が目立つ。総じて、各種項目のいずれかを非表示にしている人はレビューも非表示であるというルールを生成しているように見える。一方、唯一表示されている項目は自己紹介(自由文)である。レビューを表示しているレコードは1割程度と生起確率が低いため、8割以上生起している非表示の項目含むレコードが選ばれている。

2. 日記についての相関ルール

続いて、同様のことを日記について行う。1000ルールの内、日記を含むルールを抽出すると結論部に日記を含むルールを生成することができなかった。生成したルールの内、条件部に日記を含むルールを以下の表に示す(表6)。なお、条件部に日記を含むルールの内コンフィデンスが1であるルールを表に記す。

表 6: 日記を条件に含むルール (コンフィデンス1)

参照番号	条件		結論	コンフィデンス
61	日記=表示	⇒	自己紹介(自由文)=表示 139	conf:(1)
	性別=男性			
	血液型=表示 139			
59	日記=表示	⇒	自己紹介(自由文)=表示 139	conf:(1)
	好きなもの=3.0			
	職業=表示 139			
58	日記=表示	⇒	自己紹介(自由文)=表示 139	conf:(1)
	血液型=表示			
	好きなもの=3.0			

	住所=表示 139			
57	日記=表示 好きなもの=3.0 レビュー=非表示 140	⇒	自己紹介(自由文)=表示 140	conf:(1)
55	日記=表示 血液型=表示 好きなもの=3.0 誕生日=表示 140	⇒	自己紹介(自由文)=表示 140	conf:(1)
47	日記=表示 誕生日=表示 好きなもの=3.0 写真=表示 145	⇒	自己紹介(自由文)=表示 145	conf:(1)
45	日記=表示 住所=表示 好きなもの=3.0 写真=表示 146	⇒	自己紹介(自由文)=表示 146	conf:(1)
42	日記=表示 好きなもの=3.0 誕生日=表示 149	⇒	自己紹介(自由文)=表示 149	conf:(1)
38	日記=表示 好きなもの=3.0 住所=表示 151	⇒	自己紹介(自由文)=表示 151	conf:(1)
37	日記=表示 性別=男性 写真=表示 151	⇒	自己紹介(自由文)=表示 151	conf:(1)
24	日記=表示 性別=男性 163	⇒	自己紹介(自由文)=表示 163	conf:(1)
23	日記=表示 血液型=表示 好きなもの=3.0 写真=表示 164	⇒	自己紹介(自由文)=表示 164	conf:(1)
17	日記=表示 好きなもの=3.0 血液型=表示 169	⇒	自己紹介(自由文)=表示 169	conf:(1)
11	日記=表示	⇒	自己紹介(自由文)=表示 179	conf:(1)

	好きなもの=3.0			
	写真=表示 179			
8	日記=表示	⇒	自己紹介(自由文)=表示 185	conf:(1)
	好きなもの=3.0 185			
132	日記=表示	⇒	自己紹介(自由文)=表示 123	conf:(1)
	性別=男性			
	誕生日=表示 123			
125	日記=表示	⇒	自己紹介(自由文)=表示 124	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	出身地=表示 124			
124	日記=表示	⇒	自己紹介(自由文)=表示 124	conf:(1)
	誕生日=表示			
	血液型=表示			
	好きなもの=3.0			
	住所=表示 124			
119	日記=表示	⇒	自己紹介(自由文)=表示 125	conf:(1)
	好きなもの=3.0			
	写真=表示			
	レビュー=非表示			
	血液型=表示 125			
117	日記=表示	⇒	自己紹介(自由文)=表示 126	conf:(1)
	レビュー=非表示			
	性別=男性 126			
113	日記=表示	⇒	自己紹介(自由文)=表示 126	conf:(1)
	出身地=表示			
	好きなもの=3.0			
	写真=表示 126			
111	日記=表示	⇒	自己紹介(自由文)=表示 126	conf:(1)
	血液型=表示			
	職業=表示			
	好きなもの=3.0			
	写真=表示 126			
106	日記=表示	⇒	自己紹介(自由文)=表示 127	conf:(1)
	住所=表示			

	誕生日=表示			
	好きなもの=3.0			
	写真=表示 127			
101	日記=表示	⇒	自己紹介(自由文)=表示 128	conf:(1)
	住所=表示			
	性別=男性			
	写真=表示 128			
97	日記=表示	⇒	自己紹介(自由文)=表示 129	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	職業=表示 129			
96	日記=表示	⇒	自己紹介(自由文)=表示 130	conf:(1)
	好きなもの=3.0			
	出身地=表示 130			
94	日記=表示	⇒	自己紹介(自由文)=表示 130	conf:(1)
	レビュー=非表示			
	好きなもの=3.0			
	血液型=表示 130			
92	日記=表示	⇒	自己紹介(自由文)=表示 131	conf:(1)
	好きなもの=3.0			
	誕生日=表示			
	住所=表示 131			
87	日記=表示	⇒	自己紹介(自由文)=表示 132	conf:(1)
	性別=男性			
	血液型=表示			
	写真=表示 132			
78	日記=表示	⇒	自己紹介(自由文)=表示 134	conf:(1)
	レビュー=非表示			
	好きなもの=3.0			
	写真=表示 134			
77	日記=表示	⇒	自己紹介(自由文)=表示 134	conf:(1)
	住所=表示			
	血液型=表示			
	好きなもの=3.0			
	写真=表示 134			

74	日記=表示	⇒	自己紹介(自由文)=表示 136	conf:(1)
	職業=表示			
	好きなもの=3.0			
	写真=表示 136			
73	日記=表示	⇒	自己紹介(自由文)=表示 136	conf:(1)
	誕生日=表示			
	血液型=表示			
	好きなもの=3.0			
	写真=表示 136			
69	日記=表示	⇒	自己紹介(自由文)=表示 138	conf:(1)
	住所=表示			
	性別=男性 138			

各ルールを見ると、全て結論部が自己紹介(自由文)=表示であることが分かる。条件部には様々な項目があるものの、全ての帰結先が自己紹介となっている。この自己紹介が表示となっている状況は、先ほどのレビューに関しても見られる傾向であった。

3. 友人からのコメントについての相関ルール

続いて、同様のことを友人からのコメントについて行った。

友人からのコメントについても、日記同様に結論部に含まれるルールが生成されず、以下の表には友人からのコメントを条件部に含むルールを列挙した(表7)。なお、今回のルールはコンフィデンスが1とは限らず、生成された全てのルールを表に記した。

表7: 友人からのコメントを条件に含むルール

参照番号	条件		結論	コンフィデン
846	友人からのコメント=非表示	⇒	レビュー=非表示 120	conf:(0.94)
	所属先=非表示 127			
794	友人からのコメント=非表示	⇒	レビュー=非表示 145	conf:(0.95)
	自己紹介(自由文)=表示 153			
782	友人からのコメント=非表示 157	⇒	レビュー=非表示 149	conf:(0.95)
585	友人からのコメント=非表示	⇒	自己紹介(自由文)=表示 123	conf:(0.97)
	所属先=非表示 127			
553	友人からのコメント=非表示	⇒	自己紹介(自由文)=表示 145	conf:(0.97)
	レビュー=非表示 149			

546	友人からのコメント=非表示 157	⇒	自己紹介(自由文)=表示 153	conf:(0.97)
-----	-------------------	---	------------------	-------------

友人からのコメントを含むルールを見てみると、生起確率の高いレビュー=非表示と、自己紹介=表示が結論部になっている。特に、794番のルールについては友人からのコメント=非表示と自己紹介=表示を条件としてレビュー=非表示となっており、その構造を表しているようにも見える。また、友人からのコメントについては全てが友人からのコメント=非表示であった。

4. 職業および所属先についての相関ルール

続いて、職業および所属先についての分析を行った。

(a) 職業についての相関ルール

まず、職業についてルールを見てみる。以下の表では職業を含むルールの抽出を調べた(表8)。職業を含むルールの内、結論部に職業を含むものは生成されなかった。そこで、職業を条件部を含むルールをまとめた。以下の表についてはコンフィデンス1のものを載せた。

表 8: 職業を条件に含むルール

参照番号	条件		結論	コンフィデンス
134	職業=表示	⇒	自己紹介(自由文)=表示 122	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	出身地=表示 122			
182	職業=表示	⇒	自己紹介(自由文)=表示 240	conf:(1)
	血液型=表示 241			
29	職業=表示	⇒	自己紹介(自由文)=表示 159	conf:(1)
	好きなもの=3.0			
	血液型=表示 159			
129	職業=表示	⇒	自己紹介(自由文)=表示 123	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	レビュー=非表示 123			
36	職業=表示	⇒	自己紹介(自由文)=表示 152	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	写真=表示 152			

111	職業=表示	⇒	自己紹介(自由文)=表示 126	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	写真=表示			
	日記=表示 126			
97	職業=表示	⇒	自己紹介(自由文)=表示 129	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	日記=表示 129			
193	職業=表示	⇒	自己紹介(自由文)=表示 221	conf:(1)
	血液型=表示			
	写真=表示 222			
205	職業=表示	⇒	自己紹介(自由文)=表示 206	conf:(1)
	住所=表示			
	血液型=表示 207			
65	職業=表示	⇒	自己紹介(自由文)=表示 138	conf:(1)
	血液型=表示			
	好きなもの=3.0			
	住所=表示 138			
85	職業=表示	⇒	自己紹介(自由文)=表示 132	conf:(1)
	住所=表示			
	血液型=表示			
	好きなもの=3.0			
	写真=表示 132			
162	職業=表示	⇒	自己紹介(自由文)=表示 117	conf:(1)
	出身地=表示			
	好きなもの=3.0			
	住所=表示 117			
187	職業=表示	⇒	自己紹介(自由文)=表示 226	conf:(1)
	住所=表示 227			
46	職業=表示	⇒	自己紹介(自由文)=表示 146	conf:(1)
	好きなもの=3.0			
	住所=表示 146			
56	職業=表示	⇒	自己紹介(自由文)=表示 140	conf:(1)
	住所=表示			

	好きなもの=3.0			
	写真=表示 140			
149	職業=表示	⇒	自己紹介(自由文)=表示 119	conf:(1)
	住所=表示			
	好きなもの=3.0			
	写真=表示			
	日記=表示 119			
135	職業=表示	⇒	自己紹介(自由文)=表示 122	conf:(1)
	住所=表示			
	好きなもの=3.0			
	日記=表示 122			
207	職業=表示	⇒	自己紹介(自由文)=表示 205	conf:(1)
	住所=表示			
	写真=表示 206			
163	職業=表示	⇒	自己紹介(自由文)=表示 117	conf:(1)
	住所=表示			
	性別=男性			
	血液型=表示 117			
89	職業=表示	⇒	自己紹介(自由文)=表示 132	conf:(1)
	住所=表示			
	性別=男性 132			
158	職業=表示	⇒	自己紹介(自由文)=表示 118	conf:(1)
	住所=表示			
	性別=男性			
	写真=表示 118			
121	職業=表示	⇒	自己紹介(自由文)=表示 125	conf:(1)
	住所=表示			
	誕生日=表示			
	血液型=表示			
	好きなもの=3.0 125			
148	職業=表示	⇒	自己紹介(自由文)=表示 119	conf:(1)
	住所=表示			
	誕生日=表示			
	血液型=表示			
	好きなもの=3.0			

	写真=表示 119			
98	職業=表示	⇒	自己紹介(自由文)=表示 129	conf:(1)
	住所=表示			
	誕生日=表示			
	好きなもの=3.0 129			
128	職業=表示	⇒	自己紹介(自由文)=表示 123	conf:(1)
	住所=表示			
	誕生日=表示			
	好きなもの=3.0			
	写真=表示 123			
170	職業=表示 272	⇒	自己紹介(自由文)=表示 271	conf:(1)
191	職業=表示	⇒	自己紹介(自由文)=表示 222	conf:(1)
	レビュー=非表示 223			
15	職業=表示	⇒	自己紹介(自由文)=表示 172	conf:(1)
	好きなもの=3.0 172			
88	職業=表示	⇒	自己紹介(自由文)=表示 132	conf:(1)
	好きなもの=3.0			
	レビュー=非表示 132			
22	職業=表示	⇒	自己紹介(自由文)=表示 165	conf:(1)
	好きなもの=3.0			
	写真=表示 165			
122	職業=表示	⇒	自己紹介(自由文)=表示 125	conf:(1)
	好きなもの=3.0			
	写真=表示			
	レビュー=非表示 125			
74	職業=表示	⇒	自己紹介(自由文)=表示 136	conf:(1)
	好きなもの=3.0			
	写真=表示			
	日記=表示 136			
59	職業=表示	⇒	自己紹介(自由文)=表示 139	conf:(1)
	好きなもの=3.0			
	日記=表示 139			
179	職業=表示	⇒	自己紹介(自由文)=表示 245	conf:(1)
	写真=表示 246			
210	職業=表示	⇒	自己紹介(自由文)=表示 204	conf:(1)

表 8 中の各ルールに目を向けてみると、全てのルールの結論が自己紹介=表示であることが分かる。また、全てのルールに含まれる職業の項目は全て表示であることも分かる。つまり、職業を表示している人は自己紹介も表示しているということがみてとれる。さらに、ほとんどのルールが条件部に複数の項目を要するものである。これは、職業の表示には他の項目の表示も同時に行われ、そこで自己紹介が表示されるというルールが多いということとも考えられる。

(b) 所属先についての相関ルール

続いて、所属先についてのルールをみていく。以下の表では職業を結論部に含むルールをまとめた(表 9)。

表 9: 所属先を結論に含むルール

参照番号	条件		結論	コンフィデンス
877	職業=非表示	⇒	所属先=非表示 118	conf:(0.94)
	出身地=非表示			
	自己紹介(自由文)=表示			
	レビュー=非表示 125			
611	職業=非表示	⇒	所属先=非表示 119	conf:(0.97)
	レビュー=非表示			
	年齢=非表示 123			
827	職業=非表示	⇒	所属先=非表示 122	conf:(0.95)
	レビュー=非表示			
	出身地=非表示 129			
828	職業=非表示	⇒	所属先=非表示 122	conf:(0.95)
	自己紹介(自由文)=表示			
	年齢=非表示 129			
802	職業=非表示	⇒	所属先=非表示 126	conf:(0.95)
	年齢=非表示 133			

表 9 中のルールをみていくと、全てのルールの条件部に職業=非表示が含まれ、結論部は所属先=非表示となっていることが分かる。これは直感でも理解できるとおり、職業を表示しないのだから、所属先も表示しない、ということといえる。一方、全てのレコードを見ると、ごく少数であるが、職業を表示せずに所属先を表示する例があった。しかしながら、一般的な人の行動パターンとしては、職業を表示しないから所属先も表示しない、ということと考えられる。ここでも、自己紹介=表示以外の項目は非表示である項目が多い。

引き続き，所属先についてのルールを見ていく。所属先を結論部に含む場合，全て所属先 = 非表示であったので，今度は条件部に所属先を含むものを集めて見ていくことにする。それを以下の表にまとめた（表 10）。

表 10: 所属先を条件に含むルール

参照番号	条件		結論	コンフィデン
118	所属先=表示 126	⇒	自己紹介(自由文)=表示 126	conf:(1)
591	所属先=表示 126	⇒	自己紹介(自由文)=表示 写真=表示 122	conf:(0.97)
593	所属先=表示 126	⇒	写真=表示 122	conf:(0.97)
592	所属先=表示 自己紹介(自由文)=表示 126	⇒	写真=表示 122	conf:(0.97)
138	所属先=表示 写真=表示 122	⇒	自己紹介(自由文)=表示 122	conf:(1)

表 10 を見ると，全てのルールが所属先 = 表示を含むものであることが分かる。所属先を表示している人は，自己紹介 = 表示，または写真 = 表示といえるようである。さらによく見ると全てのルールを構成している項目が所属先 = 表示，写真 = 表示，自己紹介 = 表示の 3 つだけであるという事に気づく。

ここで，これまでのルールの観測から，1つの項目が目立っていることに気づく。自己紹介(自由文) = 表示という項目が結論部で使われていることが多い，または条件部にまざっていることが多いということが気になるので，改めて，この自己紹介(自由文)について生起確率を確認してみる。すると，自己紹介(自由文)を表示している人が 462 レコードで，非表示の人が 5 レコードと，極端に表示者の比率が多い(生起確率が高い)ことが分かる。つまり，これまでの分析では生起確率の高い自己紹介(自由文) = 表示が多く採用される傾向にあったことが前述のような状況を招いたと考えられる。

そこで，以下の分析では自己紹介(自由文)に関する項目を削除し，他の項目でのみ前回と同じ条件でルールを生成させた。自己紹介の項目を取り除き，特筆すべき結果が得られたもののみを載せる。ここで特筆すべき結果が得られたのは，職業および所属先についてのルールにおいてである。

(c) 自己紹介(自由文)を除いた職業についての相関ルール

まず，職業について改善された部分について確認する。以下の表には，職業を結論部に含むルールをまとめた（表 11）。

表 11: 職業を結論に含むルール

参照番号	条件		結論	コンフィデンス
713	住所=表示	⇒	職業=表示 70	conf:(0.95)
	誕生日=表示			
	血液型=表示			
	出身地=表示			
	写真=表示			
	所属先=表示 74			
694	誕生日=表示	⇒	職業=表示 71	conf:(0.95)
	血液型=表示			
	レビュー=非表示			
	所属先=表示 75			
697	住所=表示	⇒	職業=表示 71	conf:(0.95)
	誕生日=表示			
	レビュー=非表示			
	所属先=表示 75			
649	誕生日=表示	⇒	職業=表示 73	conf:(0.95)
	血液型=表示			
	出身地=表示			
	写真=表示			
	所属先=表示 77			
653	住所=表示	⇒	職業=表示 73	conf:(0.95)
	誕生日=表示			
	血液型=表示			
	出身地=表示			
	所属先=表示 77			
628	誕生日=表示	⇒	職業=表示 74	conf:(0.95)
	写真=表示			
	レビュー=非表示			
	所属先=表示 78			
591	誕生日=表示	⇒	職業=表示 76	conf:(0.95)
	血液型=表示			
	出身地=表示			
	所属先=表示 80			
900	誕生日=表示	⇒	職業=表示 77	conf:(0.94)

	出身地=表示			
	写真=表示			
	所属先=表示 82			
902	住所=表示	⇒	職業=表示 77	conf:(0.94)
	誕生日=表示			
	出身地=表示			
	所属先=表示 82			
558	誕生日=表示	⇒	職業=表示 78	conf:(0.95)
	レビュー=非表示			
	所属先=表示 82			
820	誕生日=表示	⇒	職業=表示 81	conf:(0.94)
	出身地=表示			
	所属先=表示 86			

職業に関するルールを見てみると、結論部で全て職業 = 表示となっていることがまず分かる。レコード数は80前後とやや少ないものの、コンフィデンスが0.95前後で興味深いルールである。さらに見ていくと、職業を表示しているルールはそれぞれ、条件部に複数の項目を要していることが分かる。また、全てのルールの条件部に所属先 = 表示が含まれている。条件部に含まれる項目のほとんどが表示されている項目であり、それらは住所や誕生日、出身地であるような個人情報である。つまり、個人情報を多く表示している人ほど、職業を表示しているといえそうである。

(d) 自己紹介（自由文）を除いた所属先についての相関ルール

その関連で、所属先を結論に含むルールも見てみる（表 12）。前述、職業を結論部に含むルールでは、表示するルールが得られ、かつ、条件には所属先 = 表示を含んでいたことから、今回も所属先と職業の関係を確認できるのではないかと考えた。

表 12: 所属先を結論に含むルール 2

参照番号	条件		結論	コンフィデンス
50	レビュー=非表示	⇒	所属先=非表示 77	conf:(0.99)
	コミュニティ=非表示			
	職業=非表示 78			
45	コミュニティ=非表示	⇒	所属先=非表示 80	conf:(0.99)
	職業=非表示 81			
86	日記=非表示	⇒	所属先=非表示 79	conf:(0.98)

	年齢=非表示			
	レビュー=非表示			
	職業=非表示 81			
82	年齢=非表示	⇒	所属先=非表示 80	conf:(0.98)
	日記=非表示			
	職業=非表示 82			
186	年齢=非表示	⇒	所属先=非表示 119	conf:(0.97)
	レビュー=非表示			
	職業=非表示 123			
141	誕生日=非表示	⇒	所属先=非表示 70	conf:(0.97)
	年齢=非表示			
	レビュー=非表示			
	日記=非表示			
	職業=非表示 72			
133	誕生日=非表示	⇒	所属先=非表示 71	conf:(0.97)
	出身地=非表示			
	レビュー=非表示			
	日記=非表示			
	職業=非表示 73			
135	誕生日=非表示	⇒	所属先=非表示 71	conf:(0.97)
	年齢=非表示			
	日記=非表示			
	職業=非表示 73			
128	誕生日=非表示	⇒	所属先=非表示 72	conf:(0.97)
	出身地=非表示			
	日記=非表示			
	職業=非表示 74			
113	年齢=非表示	⇒	所属先=非表示 74	conf:(0.97)
	出身地=非表示			
	レビュー=非表示			
	日記=非表示			
	職業=非表示 76			
109	年齢=非表示	⇒	所属先=非表示 75	conf:(0.97)
	出身地=非表示			
	日記=非表示			

	職業=非表示 77			
217	出身地=非表示	⇒	所属先=非表示 83	conf:(0.97)
	レビュー=非表示			
	日記=非表示			
	職業=非表示 86			
211	出身地=非表示	⇒	所属先=非表示 84	conf:(0.97)
	日記=非表示			
	職業=非表示 87			
201	性別=女性	⇒	所属先=非表示 87	conf:(0.97)
	レビュー=非表示			
	職業=非表示 90			
293	日記=非表示	⇒	所属先=非表示 100	conf:(0.96)
	職業=非表示 104			
284	誕生日=非表示	⇒	所属先=非表示 102	conf:(0.96)
	年齢=非表示			
	職業=非表示 106			
272	年齢=非表示	⇒	所属先=非表示 103	conf:(0.96)
	出身地=非表示			
	レビュー=非表示			
	職業=非表示 107			
337	友人からのコメント=非表示	⇒	所属先=非表示 73	conf:(0.96)
	レビュー=非表示			
	職業=非表示 76			
321	誕生日=非表示	⇒	所属先=非表示 74	conf:(0.96)
	レビュー=非表示			
	日記=非表示			
	職業=非表示 77			
285	誕生日=非表示	⇒	所属先=非表示 76	conf:(0.96)
	日記=非表示			
	職業=非表示 79			
287	友人からのコメント=非表示	⇒	所属先=非表示 76	conf:(0.96)
	職業=非表示 79			
245	レビュー=非表示	⇒	所属先=非表示 80	conf:(0.96)
	好きなもの=非表示			
	職業=非表示 83			

416	誕生日=非表示	⇒	所属先=非表示 90	conf:(0.96)
	年齢=非表示			
	出身地=非表示			
	レビュー=非表示			
	職業=非表示 94			
403	誕生日=非表示	⇒	所属先=非表示 93	conf:(0.96)
	年齢=非表示			
	出身地=非表示			
	職業=非表示 97			
388	性別=女性	⇒	所属先=非表示 94	conf:(0.96)
	職業=非表示 98			
340	誕生日=非表示	⇒	所属先=非表示 97	conf:(0.96)
	年齢=非表示			
	レビュー=非表示			
	職業=非表示 101			
341	日記=非表示	⇒	所属先=非表示 97	conf:(0.96)
	レビュー=非表示			
	職業=非表示 101			
395	誕生日=非表示	⇒	所属先=非表示	conf:(0.96)
	年齢=非表示		レビュー=非表示 70	
	日記=非表示			
	職業=非表示 73			
362	誕生日=非表示	⇒	所属先=非表示	conf:(0.96)
	出身地=非表示		レビュー=非表示 71	
	日記=非表示			
	職業=非表示 74			
317	年齢=非表示	⇒	所属先=非表示	conf:(0.96)
	出身地=非表示		レビュー=非表示 74	
	日記=非表示			
	職業=非表示 77			
251	年齢=非表示	⇒	所属先=非表示	conf:(0.96)
	日記=非表示		レビュー=非表示 79	
	職業=非表示 82			
723	出身地=非表示	⇒	所属先=非表示 122	conf:(0.95)
	レビュー=非表示			

	職業=非表示 129			
672	年齢=非表示	⇒	所属先=非表示 126	conf:(0.95)
	職業=非表示 133			
717	性別=女性	⇒	所属先=非表示 70	conf:(0.95)
	写真=表示			
	職業=非表示 74			
685	年齢=非表示	⇒	所属先=非表示 72	conf:(0.95)
	レビュー=非表示			
	写真=表示			
	職業=非表示 76			
658	出身地=非表示	⇒	所属先=非表示 73	conf:(0.95)
	コミュニティ=非表示			
	レビュー=非表示 77			
638	レビュー=非表示 78	⇒	所属先=非表示 74	conf:(0.95)
	写真=非表示			
639	出身地=非表示	⇒	所属先=非表示 74	conf:(0.95)
	コミュニティ=非表示 78			
603	誕生日=非表示	⇒	所属先=非表示 75	conf:(0.95)
	年齢=非表示			
	出身地=非表示			
	レビュー=非表示			
	日記=非表示 79			
612	年齢=非表示	⇒	所属先=非表示 75	conf:(0.95)
	好きなもの=非表示			
	レビュー=非表示 79			
613	住所=非表示	⇒	所属先=非表示 75	conf:(0.95)
	レビュー=非表示			
	職業=非表示 79			
615	写真=非表示 79	⇒	所属先=非表示 75	conf:(0.95)
590	誕生日=非表示	⇒	所属先=非表示 76	conf:(0.95)
	年齢=非表示			
	出身地=非表示			
	日記=非表示 80			
539	誕生日=非表示	⇒	所属先=非表示 79	conf:(0.95)
	出身地=非表示			

	レビュー=非表示			
	日記=非表示 83			
529	誕生日=非表示	⇒	所属先=非表示 80	conf:(0.95)
	年齢=非表示			
	レビュー=非表示			
	日記=非表示 84			
514	誕生日=非表示	⇒	所属先=非表示 81	conf:(0.95)
	出身地=非表示			
	日記=非表示 85			
515	誕生日=非表示	⇒	所属先=非表示 81	conf:(0.95)
	年齢=非表示			
	日記=非表示 85			
517	住所=非表示	⇒	所属先=非表示 81	conf:(0.95)
	職業=非表示 85			
498	年齢=非表示	⇒	所属先=非表示 82	conf:(0.95)
	出身地=非表示			
	レビュー=非表示			
	日記=非表示 86			
485	年齢=非表示	⇒	所属先=非表示 83	conf:(0.95)
	出身地=非表示			
	日記=非表示 87			
705	誕生日=非表示	⇒	所属先=非表示 88	conf:(0.95)
	レビュー=非表示			
	日記=非表示 93			
665	誕生日=非表示	⇒	所属先=非表示 91	conf:(0.95)
	日記=非表示 96			
597	誕生日=非表示	⇒	所属先=非表示 94	conf:(0.95)
	出身地=非表示			
	レビュー=非表示			
	職業=非表示 99			
581	職業=非表示	⇒	所属先=非表示	conf:(0.95)
	コミュニティ=非表示 81		レビュー=非表示 77	
479	出身地=非表示	⇒	所属先=非表示	conf:(0.95)
	日記=非表示		レビュー=非表示 83	
	職業=非表示 87			

895	年齢=非表示	⇒	所属先=非表示 108	conf:(0.94)
	出身地=非表示			
	職業=非表示 115			
952	レビュー=非表示	⇒	所属先=非表示 166	conf:(0.94)
	職業=非表示 177			
969	年齢=非表示	⇒	所属先=非表示 75	conf:(0.94)
	コミュニティ=非表示 80			
934	年齢=非表示	⇒	所属先=非表示 76	conf:(0.94)
	好きなもの=非表示 81			
904	出身地=非表示	⇒	所属先=非表示 77	conf:(0.94)
	好きなもの=非表示			
	レビュー=非表示 82			
802	好きなもの=非表示	⇒	所属先=非表示 82	conf:(0.94)
	職業=非表示 87			
826	誕生日=非表示	⇒	所属先=非表示 97	conf:(0.94)
	出身地=非表示			
	職業=非表示 103			
989	誕生日=非表示	⇒	所属先=非表示	conf:(0.94)
	日記=非表示		レビュー=非表示 74	
	職業=非表示 79			
995	写真=非表示 79	⇒	所属先=非表示	conf:(0.94)
			レビュー=非表示 74	
962	誕生日=非表示	⇒	所属先=非表示	conf:(0.94)
	年齢=非表示		レビュー=非表示 75	
	出身地=非表示			
	日記=非表示 80			
849	誕生日=非表示	⇒	所属先=非表示	conf:(0.94)
	年齢=非表示		レビュー=非表示 80	
	日記=非表示 85			
799	年齢=非表示	⇒	所属先=非表示	conf:(0.94)
	出身地=非表示		レビュー=非表示 82	
	日記=非表示 87			

表 12 を見ると、全てのルール結論部には所属先 = 非表示となっていることが分かる。こちらは表 11 とは対照的に、所属先 = 非表示となるルールを見ていくことになる。それぞれのルールを見ると、個人情報以外にも様々なものを非表示にしている人は所属先も非表示であるというルールが得られているというのが分かる。また、ほとんどのルールでは条件部に職業 = 非表示が含まれており、職業 = 非表示が所属先 = 非表示

を導くものとして重要な役割を果たしているということが改めて確認できたと思う。

自己紹介（自由文）の項目を除去した分析では，特に所属先と職業について知見を深めるに足る，結果が得られたと考えられる。所属先とともに個人情報を詳細に公表している人は職業も表示していることがルールとして認められそうであるが，一方で，所属先を結論に含むルールとしては所属先を表示にするルールは生成されず，結果として，様々なものを非表示にしている人は所属先も非表示にしているというルールが認められそうである。

(e) その他の結果（参考までに）

表 13 はレビューについてのもので，表 14 は日記についてのものであり，表 15 は友人からのコメントについてである。

表 13: レビューを結論に含むルール（抜粋）

参照番号	条件		結論	コンフィデンス
33	好きなもの=非表示	⇒	レビュー=非表示 70	conf:(1)
	友人からのコメント=非表示 70			
64	誕生日=非表示	⇒	レビュー=非表示 70	conf:(0.99)
	年齢=非表示			
	職業=非表示			
	所属先=非表示			
	日記=非表示 71			
65	誕生日=非表示	⇒	レビュー=非表示 70	conf:(0.99)
	年齢=非表示			
	出身地=非表示			
	職業=非表示			
	日記=非表示 71			
68	所属先=非表示	⇒	レビュー=非表示 70	conf:(0.99)
	好きなもの=非表示			
	日記=非表示 71			

表 14: 日記を条件に含むルール（自己紹介（自由文）を除き，コンフィデンス1）

参照番号	条件		結論	コンフィデンス
27	血液型=表示	⇒	写真=表示 71	conf:(1)

	出身地=表示			
	所属先=表示			
	日記=表示 71			
31	血液型=表示	⇒	写真=表示 70	conf:(1)
	所属先=表示			
	好きなもの=3.0			
	日記=表示 70			
2	血液型=表示	⇒	写真=表示 91	conf:(1)
	所属先=表示			
	日記=表示 91			
8	血液型=表示	⇒	写真=表示 80	conf:(1)
	職業=表示			
	所属先=表示			
	日記=表示 80			
10	住所=表示	⇒	写真=表示 80	conf:(1)
	血液型=表示			
	所属先=表示			
	日記=表示 80			
25	住所=表示	⇒	写真=表示 71	conf:(1)
	血液型=表示			
	職業=表示			
	所属先=表示			
	日記=表示 71			
30	住所=表示	⇒	写真=表示 71	conf:(1)
	出身地=表示			
	所属先=表示			
	日記=表示 71			
5	住所=表示	⇒	写真=表示 88	conf:(1)
	所属先=表示			
	日記=表示 88			
16	住所=表示	⇒	写真=表示 76	conf:(1)
	職業=表示			
	所属先=表示			
	日記=表示 76			
20	住所=表示	⇒	写真=表示 74	conf:(1)

	誕生日=表示			
	血液型=表示			
	所属先=表示			
	日記=表示 74			
11	住所=表示	⇒	写真=表示 80	conf:(1)
	誕生日=表示			
	所属先=表示			
	日記=表示 80			
26	住所=表示	⇒	写真=表示 71	conf:(1)
	誕生日=表示			
	職業=表示			
	所属先=表示			
	日記=表示 71			
23	出身地=表示	⇒	写真=表示 74	conf:(1)
	所属先=表示			
	日記=表示 74			
18	所属先=表示	⇒	写真=表示 75	conf:(1)
	レビュー=非表示			
	日記=表示 75			
17	所属先=表示	⇒	写真=表示 76	conf:(1)
	好きなもの=3.0			
	日記=表示 76			
1	所属先=表示	⇒	写真=表示 101	conf:(1)
	日記=表示 101			
6	職業=表示	⇒	写真=表示 87	conf:(1)
	所属先=表示			
	日記=表示 87			
9	誕生日=表示	⇒	写真=表示 80	conf:(1)
	血液型=表示			
	所属先=表示			
	日記=表示 80			
19	誕生日=表示	⇒	写真=表示 74	conf:(1)
	血液型=表示			
	職業=表示			
	所属先=表示			

	日記=表示 74			
29	誕生日=表示	⇒	写真=表示 71	conf:(1)
	出身地=表示			
	所属先=表示			
	日記=表示 71			
4	誕生日=表示	⇒	写真=表示 88	conf:(1)
	所属先=表示			
	日記=表示 88			
12	誕生日=表示	⇒	写真=表示 79	conf:(1)
	職業=表示			
	所属先=表示			
	日記=表示 79			

表 15: 友人からのコメントを条件に含むルール

参照番号	条件		結論	コンフィデンス
33	友人からのコメント=非表示	⇒	レビュー=非表示 70	conf:(1)
	好きなもの=非表示 70			
49	友人からのコメント=非表示	⇒	レビュー=非表示 77	conf:(0.99)
	所属先=非表示			
	日記=非表示 78			
39	友人からのコメント=非表示	⇒	レビュー=非表示 83	conf:(0.99)
	出身地=非表示			
	所属先=非表示 84			
34	友人からのコメント=非表示	⇒	レビュー=非表示 89	conf:(0.99)
	日記=非表示 90			
83	友人からのコメント=非表示	⇒	レビュー=非表示 80	conf:(0.98)
	コミュニティ=非表示 82			
73	友人からのコメント=非表示	⇒	レビュー=非表示 89	conf:(0.98)
	出身地=非表示 91			
145	友人からのコメント=非表示	⇒	レビュー=非表示 70	conf:(0.97)
	誕生日=表示			
	血液型=表示 72			
99	友人からのコメント=非表示	⇒	レビュー=非表示 77	conf:(0.97)
	年齢=非表示			
	所属先=非表示 79			
222	友人からのコメント=非表示	⇒	レビュー=非表示 83	conf:(0.97)

	年齢=非表示 86			
264	友人からのコメント=非表示 住所=表示 108	⇒	レビュー=非表示 104	conf:(0.96)
336	友人からのコメント=非表示 職業=非表示 所属先=非表示 76	⇒	レビュー=非表示 73	conf:(0.96)
286	友人からのコメント=非表示 職業=非表示 79	⇒	レビュー=非表示 76	conf:(0.96)
247	友人からのコメント=非表示 誕生日=表示 83	⇒	レビュー=非表示 80	conf:(0.96)
619	友人からのコメント=非表示 157	⇒	レビュー=非表示 149	conf:(0.95)
699	友人からのコメント=少数 75	⇒	レビュー=非表示 71	conf:(0.95)
636	友人からのコメント=非表示 住所=表示 血液型=表示 78	⇒	レビュー=非表示 74	conf:(0.95)
533	友人からのコメント=非表示 住所=表示 所属先=非表示 84	⇒	レビュー=非表示 80	conf:(0.95)
741	友人からのコメント=非表示 性別=男性 91	⇒	レビュー=非表示 86	conf:(0.95)
747	友人からのコメント=非表示 所属先=非表示 127	⇒	レビュー=非表示 120	conf:(0.94)

5.3.3 決定木分析

決定木分析におけるパラメーター設定については以下の図の通りである (図 6)。

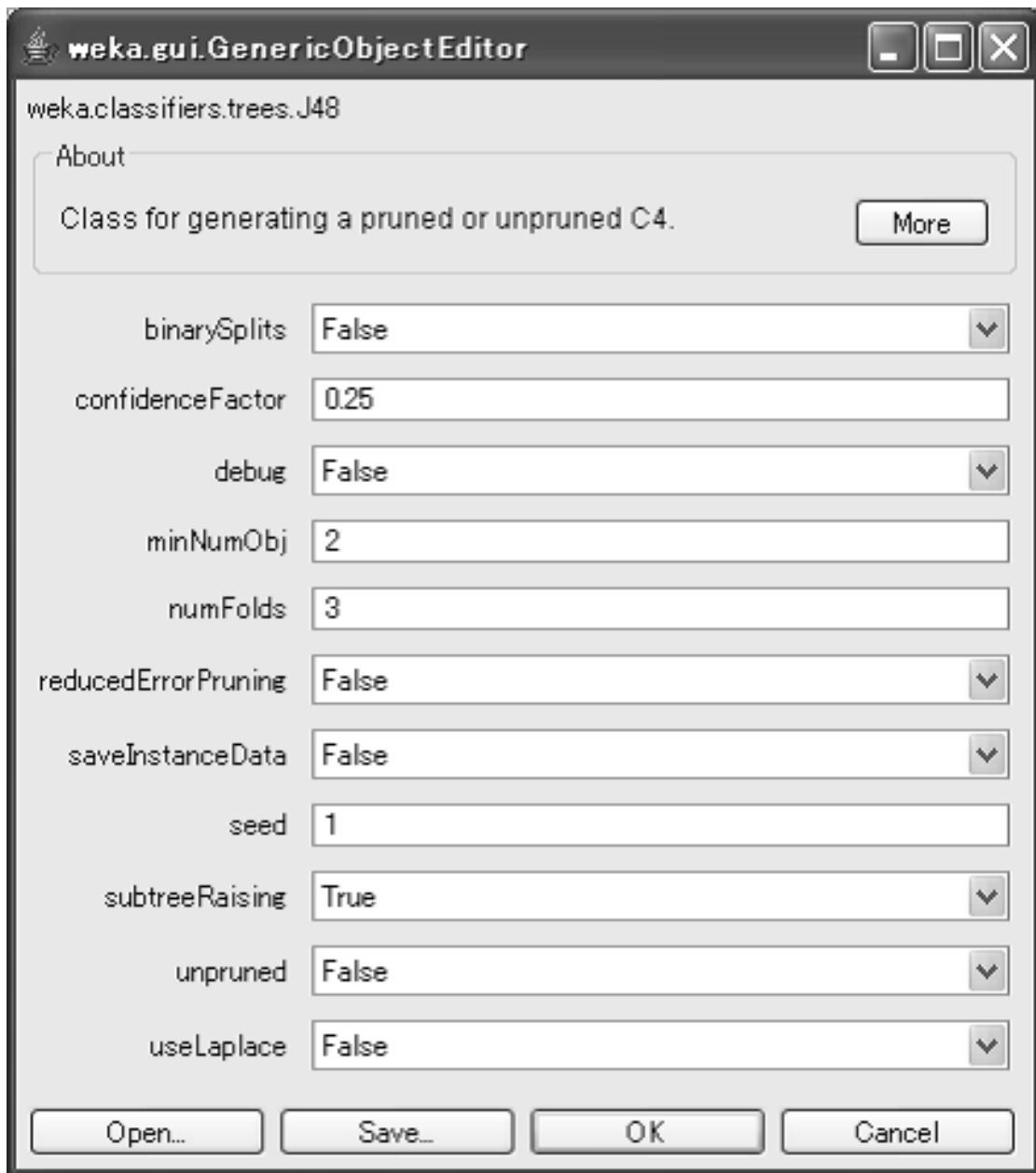


図 6: 決定木分析の検索方針

設定の説明 binarySplits – Whether to use binary splits on nominal attributes when building the trees.

confidenceFactor – The confidence factor used for pruning (smaller values incur more pruning).

debug – If set to true, classifier may output additional info to the console.

minNumObj – The minimum number of instances per leaf.

numFolds – Determines the amount of data used for reduced-error pruning. One

fold is used for pruning, the rest for growing the tree.

reducedErrorPruning – Whether reduced-error pruning is used instead of C.4.5 pruning.

saveInstanceData – Whether to save the training data for visualization.

seed – The seed used for randomizing the data when reduced-error pruning is used.

subtreeRaising – Whether to consider the subtree raising operation when pruning.

unpruned – Whether pruning is performed.

useLaplace – Whether counts at leaves are smoothed based on Laplace.

決定木分析においても相関ルールと同じ内容のものを検討する。表 3 から、各項目に対して該当するものを被説明変数とし、被説明変数以外の項目から自己紹介(自由文)をのぞいたものを説明変数とした。

1. レビューについての決定木分析

まず、レビューについてであるが、初期設定のままでは樹形図を生成することができなかつたため、confidenceFactor の部分を操作し、初期設定の 0.25 から 0.35 へと変更したところ初めて木が描けた。以下にその結果を示す。

レビューの場合の分析概要

J48 pruned tree

```
コミュニティ = 非常に多い
|  好きなもの = 表示: 非表示 (3.0/1.0)
|  好きなもの = 3.0
|   |  友達 = 非常に多い
|   |   |  誕生日 = 表示: 表示 (10.0/1.0)
|   |   |  誕生日 = 非表示: 非表示 (4.0)
|   |   |  友達 = やや少数: 非表示 (1.0)
|   |   |  友達 = 1人: 非表示 (0.0)
|   |   |  友達 = やや多数
|   |   |  誕生日 = 表示: 非表示 (6.0/1.0)
|   |   |  誕生日 = 非表示
|   |   |   |  住所 = 表示: 非表示 (3.0/1.0)
|   |   |   |  住所 = 非表示: 表示 (2.0)
|   |   |   |  友達 = 少数: 非表示 (1.0)
|   |   |   |  友達 = 多数
```

```

|   |   |   住所 = 表示: 非表示 (12.0/2.0)
|   |   |   住所 = 非表示: 表示 (3.0)
|   好きなもの = 2.0: 非表示 (6.0)
|   好きなもの = 非表示
|   |   出身地 = 表示: 非表示 (4.0/1.0)
|   |   出身地 = 非表示: 表示 (2.0)
コミュニティ = やや多数
|   血液型 = 表示
|   |   出身地 = 表示
|   |   |   誕生日 = 表示: 非表示 (30.0/6.0)
|   |   |   誕生日 = 非表示
|   |   |   |   好きなもの = 表示: 非表示 (2.0)
|   |   |   |   好きなもの = 3.0: 表示 (3.0)
|   |   |   |   好きなもの = 2.0: 表示 (0.0)
|   |   |   |   好きなもの = 非表示: 表示 (0.0)
|   |   出身地 = 非表示: 非表示 (23.0)
|   血液型 = 非表示
|   |   誕生日 = 表示: 表示 (4.0)
|   |   誕生日 = 非表示: 非表示 (4.0)
コミュニティ = 少数: 非表示 (81.0/5.0)
コミュニティ = 多数: 非表示 (65.0/13.0)
コミュニティ = やや少数: 非表示 (61.0/10.0)
コミュニティ = 非表示: 非表示 (137.0/4.0)

```

Number of Leaves : 26

Size of the tree : 39

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	394	84.3683 %
Incorrectly Classified Instances	73	15.6317 %
Kappa statistic	0.0153	
Mean absolute error	0.2374	
Root mean squared error	0.3656	

```
Relative absolute error          96.1229 %
Root relative squared error      104.295 %
Total Number of Instances        467
```

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.98	0.97	0.858	0.98	0.915	非表示
0.03	0.02	0.2	0.03	0.052	表示

```
=== Confusion Matrix ===
```

```
  a   b  <-- classified as
392  8  |  a = 非表示
 65  2  |  b = 表示
```

サマリーの部分を見ると、正しく分類されたレコードが全体の84%で、非表示と分類されたもので状態が非表示であったものは85.8%である一方、表示と分類されたものの内、状態が表示であったものはわずか20%であった。これは、ほとんどのレコードが表示に分類されていることを示している。実際、表示であるレコード67の内65レコードが非表示と分類されている。このように分類された決定木を図で確認してみることにする(図7)。

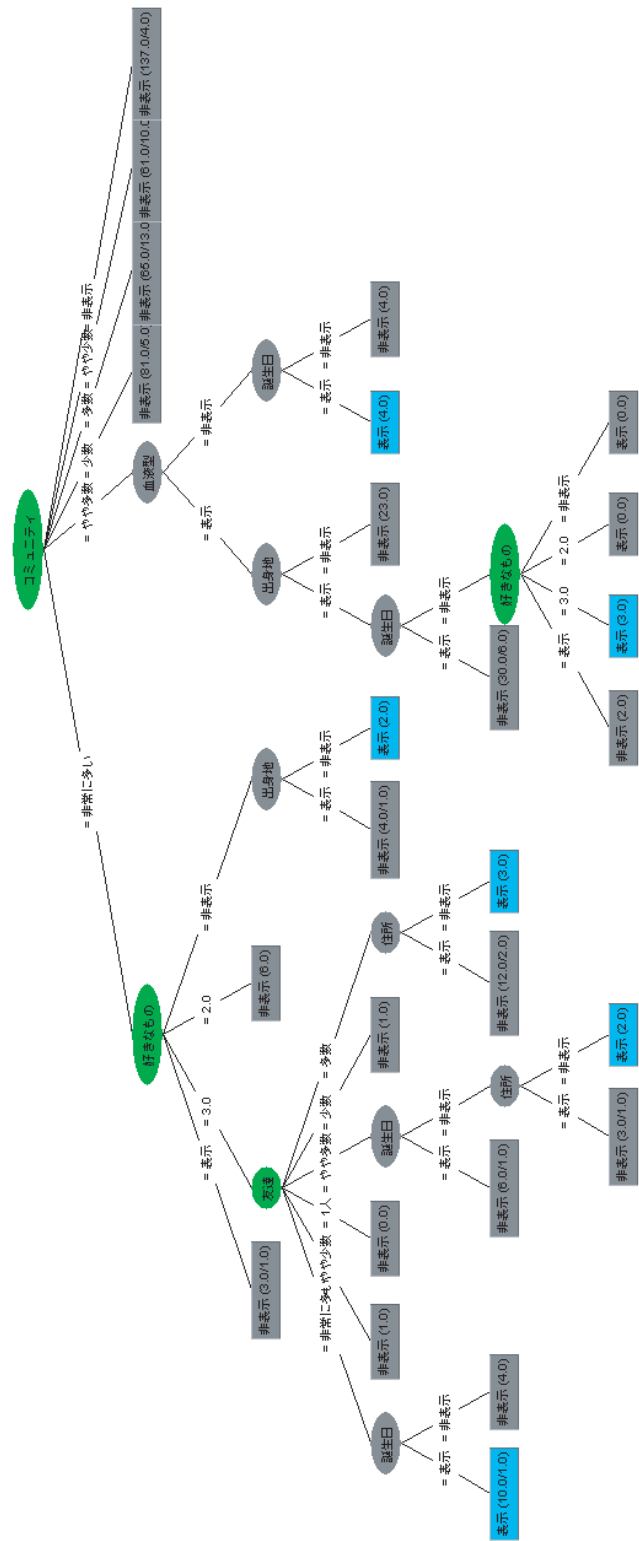


図 7: レビューの決定木

図の見方は丸型ノードが説明変数ノードで、四角型ノードが被説明変数ノード

ドである。それぞれ、ノード内の文字が示しているのは、丸型ノードの場合説明変数名であり、四角型ノードは被説明変数の状態である。また、四角型ノード内の()内の値については該当するレコード数である。()内をスラッシュ(/)で分けている場合、左の値が正しく分類されているレコード数、右の値が誤って分類されているレコード数である。さらに、ノードとノードを結ぶアークについては上の階層から下の階層へと伸び、アーク上の文字が分類ルールである。

まず、頂上のノードがコミュニティ数であり{少数, 多数, やや少数, 非表示}の場合、レビューも非表示である。一方、コミュニティ数 = やや多数の場合、血液型が非表示で誕生日が表示されている場合、全てのレコードがレビューを表示している。と、このようにして各ノードおよび分岐について検討していくことになる。

まず、コミュニティ数 = やや多数の人(達)は血液型の公開状態によって項目の持つ意味が変わってくる。血液型 = 非表示の人の中、誕生日 = 表示の場合全てのレコードでレビューが表示されている一方、血液型 = 表示で誕生日 = 表示の場合(出身地の表示非表示如何によらず)ほとんどのレコードでレビューが非表示となる。所属しているコミュニティが一般的にやや多数である場合、各項目の持つ意味合いが異なるということがいえる。

一方、コミュニティ数 = 非常に多いと分類される場合、レビューを表示しているのは

- 好きなもの = 非表示, 出身地 = 非表示の人
- 好きなもの = (最大の)3個表示, 友達 = 多数, 出身地 = 非表示の人
- 好きなもの = (最大の)3個表示, 友達 = やや多数, 誕生日 = 非表示, 住所 = 非表示の人
- 好きなもの = (最大の)3個表示, 友達 = 非常に多い, 誕生日 = 表示の人

である。上記の4タイプの内、第1タイプの人所属コミュニティは人並みはずれて多いが非表示の項目が多い人であり、他の3タイプの人から共通していると思われる点を取り出すと、好きなものを3つ最大まで提示し、個人の嗜好を伝える性行がある点、また、友達が多く、いくつか公開する項目に個人情報の部分がある点である。

このようにして、レビューをする人を分類していくことができる。

2. 日記についての決定木分析

続いて、日記についての決定木分析。

日記の場合の分析概要

J48 pruned tree

写真 = 表示

| 住所 = 表示

| | 友達 = 非常に多い

| | | 性別 = 男性: 表示 (23.0)

| | | 性別 = 女性

| | | | 誕生日 = 表示: 表示 (7.0/2.0)

| | | | 誕生日 = 非表示: 非表示 (2.0)

| | 友達 = やや少数: 表示 (51.0/12.0)

| | 友達 = 1人

| | | 職業 = 表示

| | | | 血液型 = 表示

| | | | | レビュー = 非表示

| | | | | 趣味 = 少数: 非表示 (3.0/1.0)

| | | | | 趣味 = やや少数: 非表示 (6.0/1.0)

| | | | | 趣味 = やや多数

| | | | | | 所属先 = 非表示: 非表示 (4.0)

| | | | | | 所属先 = 表示: 表示 (2.0)

| | | | | | 趣味 = 多数: 表示 (4.0)

| | | | | | 趣味 = 非表示: 非表示 (0.0)

| | | | | | レビュー = 表示: 表示 (2.0)

| | | | | 血液型 = 非表示: 表示 (2.0)

| | | | 職業 = 非表示: 非表示 (7.0)

| | 友達 = やや多数: 表示 (55.0/2.0)

| | 友達 = 少数

| | | 血液型 = 表示: 表示 (41.0/6.0)

| | | 血液型 = 非表示

| | | | 好きなもの = 表示: 表示 (1.0)

| | | | 好きなもの = 3.0: 表示 (2.0)

| | | | 好きなもの = 2.0: 非表示 (0.0)

| | | | 好きなもの = 非表示: 非表示 (6.0/1.0)

| | 友達 = 多数: 表示 (67.0/13.0)

| 住所 = 非表示

| | レビュー = 非表示

| | | 年齢 = 18~22: 非表示 (12.0/5.0)

```

|   |   |   年齢 = 23 ~ 24: 表示 (7.0/1.0)
|   |   |   年齢 = 25 ~ 28: 非表示 (4.0/1.0)
|   |   |   年齢 = 29 ~ 37: 表示 (3.0/1.0)
|   |   |   年齢 = 非表示
|   |   |   |   誕生日 = 表示: 表示 (7.0/1.0)
|   |   |   |   誕生日 = 非表示
|   |   |   |   |   血液型 = 表示
|   |   |   |   |   |   趣味 = 少数: 非表示 (12.0/4.0)
|   |   |   |   |   |   趣味 = やや少数
|   |   |   |   |   |   |   友達 = 非常に多い: 非表示 (0.0)
|   |   |   |   |   |   |   友達 = やや少数: 表示 (2.0)
|   |   |   |   |   |   |   友達 = 1人: 表示 (1.0)
|   |   |   |   |   |   |   友達 = やや多数: 非表示 (1.0)
|   |   |   |   |   |   |   友達 = 少数: 非表示 (3.0)
|   |   |   |   |   |   |   友達 = 多数: 非表示 (0.0)
|   |   |   |   |   |   |   趣味 = やや多数: 表示 (9.0/1.0)
|   |   |   |   |   |   |   趣味 = 多数
|   |   |   |   |   |   |   性別 = 男性: 表示 (3.0/1.0)
|   |   |   |   |   |   |   性別 = 女性: 非表示 (10.0/1.0)
|   |   |   |   |   |   |   趣味 = 非表示: 非表示 (1.0)
|   |   |   |   |   |   |   血液型 = 非表示: 非表示 (11.0/1.0)
|   |   |   |   |   |   レビュー = 表示
|   |   |   |   |   |   性別 = 男性: 表示 (7.0)
|   |   |   |   |   |   性別 = 女性
|   |   |   |   |   |   |   血液型 = 表示: 表示 (8.0/2.0)
|   |   |   |   |   |   |   血液型 = 非表示: 非表示 (2.0)
写真 = 非表示
|   |   |   |   出身地 = 表示
|   |   |   |   |   所属先 = 非表示: 表示 (23.0/10.0)
|   |   |   |   |   所属先 = 表示: 非表示 (4.0)
|   |   |   |   |   出身地 = 非表示: 非表示 (52.0/6.0)

```

Number of Leaves : 43

Size of the tree : 66

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	334	71.5203 %
Incorrectly Classified Instances	133	28.4797 %
Kappa statistic	0.3537	
Mean absolute error	0.3509	
Root mean squared error	0.4629	
Relative absolute error	75.5744 %	
Root relative squared error	96.0903 %	
Total Number of Instances	467	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.845	0.509	0.742	0.845	0.79	表示
0.491	0.155	0.646	0.491	0.558	非表示

=== Confusion Matrix ===

```
  a   b  <-- classified as
250  46 |  a = 表示
 87  84 |  b = 非表示
```

サマリーを見ると正しく分類されたレコードが71.5%で、表示・非表示についての正確さも双方ともに70%前後と同程度に分類されていることが読み取れる。それでは、図を見て分類について検討する(図8)。

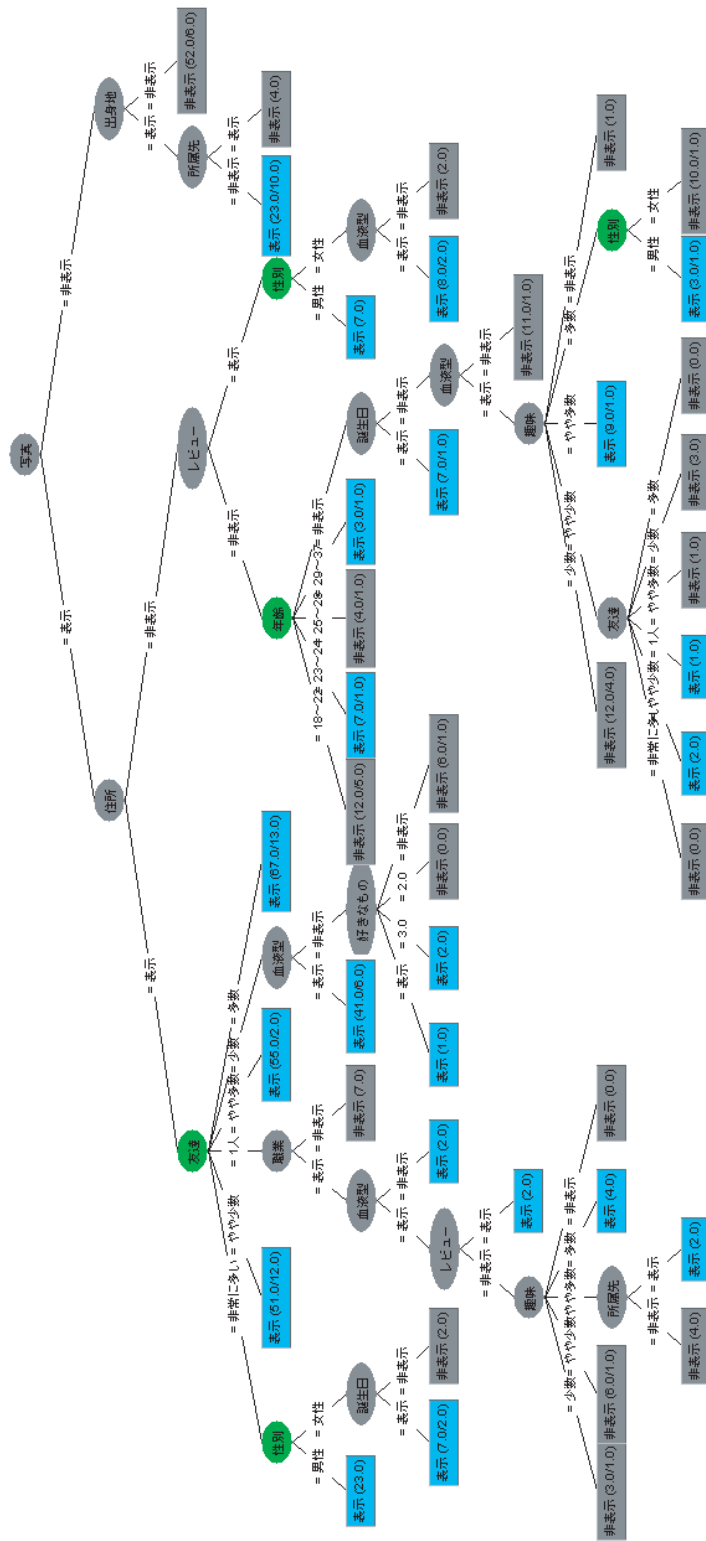


図 8: 日記の決定木

頂上のノードは写真である。写真 = 非表示となっている場合, 出身地 = 表示,

所属先 = 非表示の人が日記を付けている。こちらの構造はやや理解しやすいのであるが、写真 = 表示となっている場合はやや難解である。ここで、キーポイントとなると思われるノードを挙げると、性別ノードと年齢ノード（住所ノードの直下の）友達ノードである。それぞれを見ていくことにする。

まず、性別ノードであるが、この分岐の直後では、男性が表示するというを示している箇所が多い。また、女性と分岐した後も表示項目の条件をいくつか挟んだ後、表示となる場合もある。たとえば、友達 = 非常に多い、男性は日記を表示する。また、レビュー = 表示の男性も日記を表示する。

続いて、年齢ノードであるが、住所 = 非表示、レビュー = 非表示の場合、年齢で分類されている。そのようなレコードの内、23、24歳および29歳以降は表示する傾向がある10代の場合は日記を表示しているレコードが少ない。

最後に（住所ノードの直下の）友達ノードに関してであるが、このノードの分岐後、下層に辿っていけばどの分岐でも表示となるレコードは現れているものの、第3層目での分岐で多くの表示者を分類することができる点では特徴的なノードといえる。つまり、住所・写真・友達数を調べて該当すれば、日記の有無を判断することができる。

日記の決定木分析では、特徴的なノードというものが確認された。

3. 友人からのコメントについての決定木分析

続いて友人からのコメントについての決定木分析。誤分類率が5割前後と高く、改善されなかったため、最も誤分類率の低かった結果を以下に記す。

友人からのコメントの場合の分析概要

J48 pruned tree

友達 = 非常に多い

| 血液型 = 表示: 非常に多い (40.0/12.0)

| 血液型 = 非表示: 多数 (5.0/1.0)

友達 = やや少数

| 好きなもの = 表示

| | 写真 = 表示

| | | 年齢 = 18 ~ 22: 少数 (1.0)

| | | 年齢 = 23 ~ 24: 少数 (0.0)

| | | 年齢 = 25 ~ 28: やや多数 (2.0)

| | | 年齢 = 29 ~ 37: 少数 (0.0)

| | | 年齢 = 非表示: 少数 (3.0/1.0)

| | 写真 = 非表示: 非表示 (2.0)
 | 好きなもの = 3.0
 | | コミュニティ = 非常に多い: 少数 (1.0)
 | | コミュニティ = やや多数: 非表示 (9.0/4.0)
 | | コミュニティ = 少数
 | | | 職業 = 表示: やや多数 (7.0/3.0)
 | | | 職業 = 非表示: やや少数 (6.0/2.0)
 | | コミュニティ = 多数: やや少数 (6.0/3.0)
 | | コミュニティ = やや少数
 | | | 職業 = 表示
 | | | | 所属先 = 非表示: やや少数 (3.0/1.0)
 | | | | 所属先 = 表示: 非表示 (2.0/1.0)
 | | | 職業 = 非表示: 少数 (5.0/2.0)
 | | コミュニティ = 非表示
 | | | 性別 = 男性
 | | | | 誕生日 = 表示: やや多数 (2.0/1.0)
 | | | | 誕生日 = 非表示: 非表示 (2.0)
 | | | 性別 = 女性: 少数 (2.0)
 | 好きなもの = 2.0: 少数 (11.0/7.0)
 | 好きなもの = 非表示
 | | 血液型 = 表示: 非表示 (15.0/8.0)
 | | 血液型 = 非表示: 少数 (4.0/1.0)
 友達 = 1人
 | 好きなもの = 表示
 | | 日記 = 表示: 少数 (2.0)
 | | 日記 = 非表示: 非表示 (6.0/1.0)
 | 好きなもの = 3.0
 | | 誕生日 = 表示
 | | | 出身地 = 表示: 少数 (13.0/4.0)
 | | | 出身地 = 非表示: 非表示 (6.0/1.0)
 | | 誕生日 = 非表示: 少数 (6.0/1.0)
 | 好きなもの = 2.0: 非表示 (8.0/1.0)
 | 好きなもの = 非表示: 非表示 (47.0/5.0)
 友達 = やや多数: やや多数 (80.0/47.0)
 友達 = 少数: 非表示 (90.0/33.0)
 友達 = 多数: 多数 (81.0/52.0)

Number of Leaves : 32

Size of the tree : 48

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	216	46.2527 %
Incorrectly Classified Instances	251	53.7473 %
Kappa statistic	0.313	
Mean absolute error	0.2095	
Root mean squared error	0.3511	
Relative absolute error	78.9274 %	
Root relative squared error	96.4021 %	
Total Number of Instances	467	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.393	0.102	0.344	0.393	0.367	多数
0.107	0.077	0.211	0.107	0.142	少数
0.764	0.232	0.625	0.764	0.688	非表示
0.472	0.154	0.358	0.472	0.407	やや多数
0.063	0.072	0.121	0.063	0.082	やや少数
0.651	0.04	0.622	0.651	0.636	非常に多い

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
22	0	1	20	4	9	a = 多数
3	8	46	9	8	1	b = 少数
5	15	120	11	6	0	c = 非表示
15	6	5	34	8	4	d = やや多数
10	8	20	19	4	3	e = やや少数
9	1	0	2	3	28	f = 非常に多い

誤分類率が5割強と高いものの、非表示であるものや非常に多いと分類されるものについても正確性が60%強と他の水準よりも高く、また、やや多数、

多数の場合も3割強の正確性となっているので、この決定木を用いると、非表示であるか、非常に多くのコメントがついているレコードを識別することができると考えられる。その反面、少数であるようなものを正しく判別できるわけではないようである。

confusion matrixを見ると、少数であるレコードの内ほとんどが非表示と分類され、やや少数であるレコードは非表示またはやや多数と分類されている。多数であるレコードは多数と正しく分類されるのと同程度の精度でやや多数と分類されている。この点を考慮すると、水準を正確に分類できてはいないものの、程度として大きく誤分類しているわけではなく²⁵、推測には十分使えると考える。

それでは、以下の決定木を見て分類のルールを見ていくことにする(図9)。

²⁵4分位点で分類したのでその影響が考えられる。

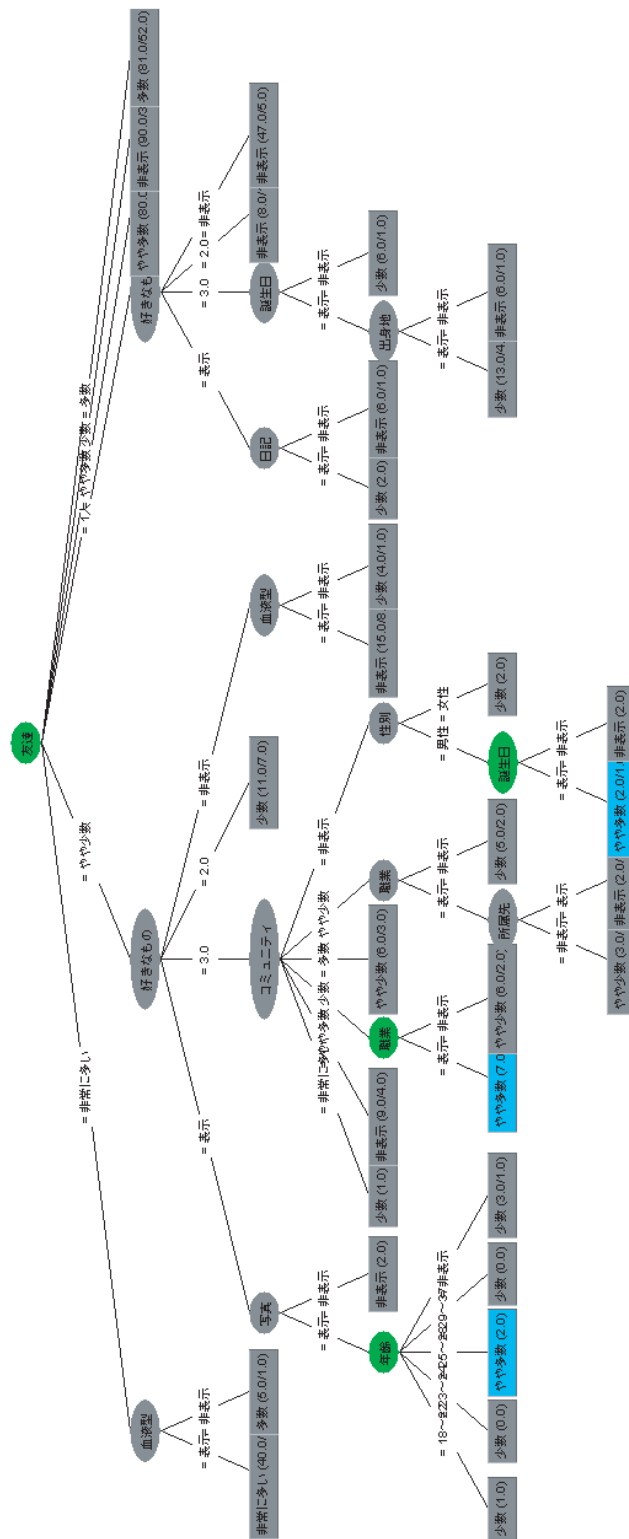


図 9: 友人からのコメントの決定木

図を見ると，やはり，頂点は友達数であった。友人の数が多ければコメント

を付けてくれる人が多いということになっているようである。友人の数とコメントの数には正の相関があるのかもしれない。ここで、注目すべき点は友達 = 1人およびやや少数に関してである。友達 = 1人である場合、表示項目が多い場合にはたった1人、唯一の友人からコメントがついている。一方、友達 = やや少数の場合、特徴的なノードは年齢ノード、コミュニティ = 少数の直下の職業ノード、誕生日ノードである。

それぞれ見てみる。年齢ノードについては20代後半になるとコメントがつきやすくなっており、他の年代では軒並み少数のコメントしかついていない。これは20代後半が社会人数年目という人が多く、他者との関わり・コミュニケーションネットワークへの意識が高まっていることが要因の1つに考えられる。つまり、新たな職場環境に入り、多くの未知の人と知り合うため、より他者に関わっていこうという意図があるのかもしれない。

続いて、職業ノードについては表示非表示双方ともコメントはついているものの、職業 = 表示となっている場合、友人からのコメント = 多数となる。職業 = 表示となることで、コメントを付ける側が付けやすい状況となっているのかもしれない。たとえば、学生である人にコメントを付ける場合「学校では のような人」といったように、人柄を記述するのに奏功する場所があるのではないかと思う。

最後に、誕生日ノードであるが、誕生日が表示されているとコメントが付き、非表示であるとコメントがつかない。このことが表しているのは、友達がやや少数でコミュニティに参加していない男性である人が誕生日のような個人情報を表示していると、コメントがつきやすいということである。つまり、個人情報をより公表する傾向のある人に対しては、コメントがつくという形で他者が関わりを持つ機会が増えるということである。

このように、3つのノードを取り上げて友人からのコメントについて考えてみたが、他者との関わりについては年齢のような世代間の差や職業や誕生日といったその人の行動様式による差があるように考えられるということがいえそうである。

4. 職業についての決定木分析

続いて職業についての決定木分析を行う。まず、概要を見る。

職業の場合の分析概要

J48 pruned tree

誕生日 = 表示


```

|  所属先 = 非表示
|  |  住所 = 表示
|  |  |  好きなもの = 表示: 表示 (20.0/5.0)
|  |  |  好きなもの = 3.0: 表示 (85.0/19.0)
|  |  |  好きなもの = 2.0: 表示 (21.0/6.0)
|  |  |  好きなもの = 非表示: 非表示 (34.0/13.0)
|  |  住所 = 非表示
|  |  |  性別 = 男性: 表示 (12.0/4.0)
|  |  |  性別 = 女性
|  |  |  |  コミュニティ = 非常に多い: 表示 (1.0)
|  |  |  |  コミュニティ = やや多数: 非表示 (4.0)
|  |  |  |  コミュニティ = 少数: 非表示 (4.0)
|  |  |  |  コミュニティ = 多数: 表示 (2.0)
|  |  |  |  コミュニティ = やや少数: 表示 (4.0/1.0)
|  |  |  |  コミュニティ = 非表示: 非表示 (5.0)
|  所属先 = 表示: 表示 (108.0/10.0)
誕生日 = 非表示
|  年齢 = 18~22
|  |  写真 = 表示: 表示 (8.0/2.0)
|  |  写真 = 非表示: 非表示 (2.0)
|  年齢 = 23~24: 表示 (9.0/2.0)
|  年齢 = 25~28: 表示 (4.0)
|  年齢 = 29~37
|  |  日記 = 表示: 表示 (4.0/1.0)
|  |  日記 = 非表示: 非表示 (3.0)
|  年齢 = 非表示
|  |  所属先 = 非表示: 非表示 (124.0/22.0)
|  |  所属先 = 表示: 表示 (13.0/4.0)

```

Number of Leaves : 20

Size of the tree : 30

Time taken to build model: 0.08 seconds

```

=== Stratified cross-validation ===
=== Summary ===

```

Correctly Classified Instances	332	71.0921 %
Incorrectly Classified Instances	135	28.9079 %
Kappa statistic	0.4018	
Mean absolute error	0.3587	
Root mean squared error	0.458	
Relative absolute error	73.7366 %	
Root relative squared error	92.872 %	
Total Number of Instances	467	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.768	0.369	0.744	0.768	0.756	表示
0.631	0.232	0.661	0.631	0.646	非表示

=== Confusion Matrix ===

```

  a   b  <-- classified as
209  63 |  a = 表示
 72 123 |  b = 非表示

```

誤分類率は3割弱，表示および非表示双方とも，正確性は7割前後となっている。confusionmatrixを見ても，表示・非表示の両方ともが同程度の精度で分類されていることが分かる。

それでは，決定木を試してみる（図 10）。

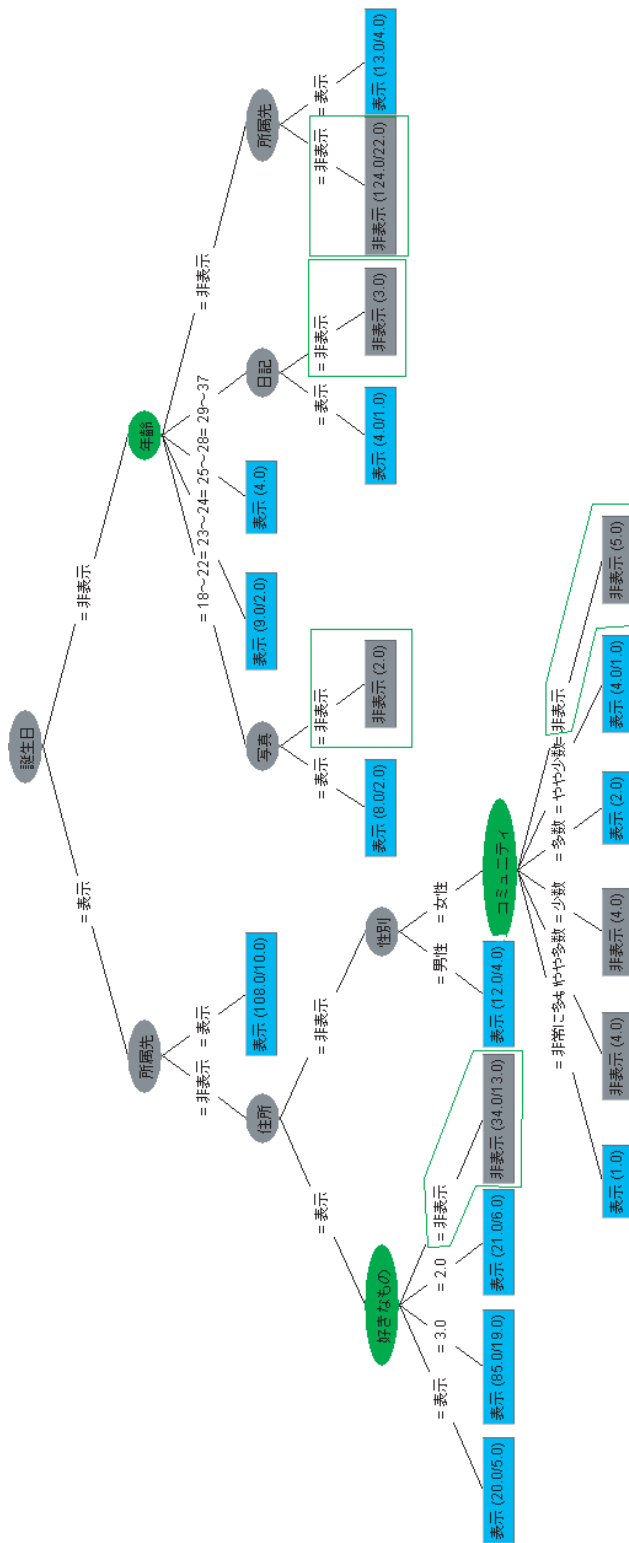


図 10: 職業の決定木

注目するノードは3つ。年齢ノード・好きなもののノード・コミュニティノード

ドである。各ノードにおいて非表示となっている場合の職業の状態が特徴的なのでその点にふれる。

まず、年齢ノードについてであるが、20代中盤・後半については職業が表示されているものの、他の層ではさらに条件が加わって表示されている。若い層については、写真 = 表示、30歳前後では日記 = 表示、年齢 = 非表示の場合には所属先 = 表示が必要である。それぞれの年代の生活スタイルによってこのような差があるのかもしれない。

その点について言及すると、若年層における写真の表示と職業の表示の関係とはどのようなものかということ、若年層における職業とは大方学生ということになると思われる点が重要だと考える。専門学校の学生であったり、大学・短期大学の学生である場合が多いと考える²⁶と、学生であることを示すことと、写真を表示することは大差のないことなのかもしれない。学生生活の中での、職業と写真とは自分を示すものであること以外の価値がほとんど無いのかもしれない。

一方、定職に就いている人の多いと思われる30歳前後については、日記 = 表示が必要となる。これは、日記を表示することと職業を明かすことの相乗効果があるのではないかと考えられる。つまり、近年盛り上がりを見せているブログ（ウェブログ）のように、このという職に就いている人の日記と銘打っている場合も考えられるし、逆に、そのような職に就いていながら考えていること、というように、職業を表示することで記述している記事自体への信憑性やその人への信頼感を構築するのに役立っているかもしれない。

このように、各年代について考えると、職業というものの持っている各人の意味合いというものが異なってくるのが分かってくる。このことが職業を表示することに対して何らかの影響を与えているということがいえそうである。

続いて、好きなものノードであるが好きなもの = 非表示となっている場合のみ職業が表示されていない。このパスについては誕生日、住所が表示されているものの好きなものや職業、所属先が非表示となっているものである。その人の個人的属性²⁷は表示するものの、社会的属性²⁸を非表示としているタイプといえそうである。つまり、ネットワークを円滑にするために必要な個人情報の積極的な開示は行うものの、できるだけ社会性の低いものを選ぼうとする傾向にあるといえそうである。

最後に、コミュニティノードである。個人情報の公開状況のあまりない女性については、多くのコミュニティに所属している場合は職業を公開する傾向にあるといえる。つまり、コミュニティ数というのは他者との関わりの強さ

²⁶もちろん、フリーターである場合や就職し定職に就いている場合もあった。

²⁷たとえば年齢であるとか住所であるとか、個人情報の内でも存在の部分を示すものとする。

²⁸職業などの社会的地位を示すものとする。

の1評価指標になると考えることができる²⁹ので、多くのコミュニティに所属し、同時に職業を表示することでコミュニケーションを円滑にしようと試みているといえそうである。また、多くのコミュニティに所属するため、その人の発言やコミュニティへの関わりを明示する働きが職業を表示することで得られると考えられる。

また、ターミナルノード³⁰になると、全体に共通していえることがある。それは、「非表示」のエッジ（弧）の場合、職業に関しても非表示となるのである。つまり、ターミナルノード前のノードに関して表示でない場合は職業は表示されない。このことは、職業を表示する人は他者に対して何らかの情報を提示し、その上で職業を表示しているということである。コミュニケーションをよりとろうとする個人の傾向（性向）は職業という形や他の属性の表示という形で現れているといえる。

職業という個人を表現する媒体の表示には、その個人の他者とのコミュニケーションへの性向の強さが表れているということが考えられる。上記のように、年代での差はあるものの、自分を表現するという部分と、他者と関わるということに対して、何らかの結びつきがあるといえる。

5. 所属先についての決定木分析

所属先については、分岐数が多くなり、図が煩雑になってしまうのを避けるため、誤分類率を見ながら、樹形図が描けるところで誤分類率が最も低くなったものを以下に示す。

所属先の場合の分析概要

J48 pruned tree

職業 = 表示

| 誕生日 = 表示

| | 友達 = 非常に多い: 表示 (26.0/8.0)

| | 友達 = やや少数

| | | 友人からのコメント = 多数: 非表示 (0.0)

| | | 友人からのコメント = 少数: 表示 (8.0/1.0)

| | | 友人からのコメント = 非表示

| | | | 年齢 = 18 ~ 22: 表示 (4.0)

| | | | 年齢 = 23 ~ 24: 非表示 (1.0)

²⁹多くのコミュニティに所属することは興味関心が外向的であるといえると考えた。

³⁰各枝の最後のノード：それ以降は分岐しないノードのこと。

| | | | 年齢 = 25 ~ 28: 非表示 (0.0)
 | | | | 年齢 = 29 ~ 37: 非表示 (0.0)
 | | | | 年齢 = 非表示: 非表示 (3.0)
 | | | 友人からのコメント = やや多数: 非表示 (10.0/3.0)
 | | | 友人からのコメント = やや少数: 非表示 (8.0/2.0)
 | | | 友人からのコメント = 非常に多い: 表示 (1.0)
 | | 友達 = 1人: 非表示 (31.0/7.0)
 | | 友達 = やや多数
 | | | 趣味 = 少数
 | | | | 性別 = 男性: 表示 (8.0)
 | | | | 性別 = 女性: 非表示 (3.0/1.0)
 | | | 趣味 = やや少数: 非表示 (11.0/2.0)
 | | | 趣味 = やや多数: 非表示 (12.0/4.0)
 | | | 趣味 = 多数
 | | | | 出身地 = 表示: 非表示 (8.0/2.0)
 | | | | 出身地 = 非表示: 表示 (3.0)
 | | | 趣味 = 非表示: 非表示 (0.0)
 | | 友達 = 少数: 非表示 (39.0/15.0)
 | | 友達 = 多数
 | | | コミュニティ = 非常に多い: 表示 (7.0/3.0)
 | | | コミュニティ = やや多数: 非表示 (10.0/1.0)
 | | | コミュニティ = 少数: 非表示 (1.0)
 | | | コミュニティ = 多数
 | | | | 好きなもの = 表示: 非表示 (1.0)
 | | | | 好きなもの = 3.0: 表示 (13.0/3.0)
 | | | | 好きなもの = 2.0: 非表示 (2.0)
 | | | | 好きなもの = 非表示: 表示 (0.0)
 | | | コミュニティ = やや少数
 | | | | レビュー = 非表示: 非表示 (5.0/1.0)
 | | | | レビュー = 表示: 表示 (3.0)
 | | | コミュニティ = 非表示: 表示 (3.0/1.0)
 | 誕生日 = 非表示: 非表示 (51.0/10.0)
 職業 = 非表示: 非表示 (195.0/18.0)

Number of Leaves : 32

Size of the tree : 43

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	339	72.591 %
Incorrectly Classified Instances	128	27.409 %
Kappa statistic	0.0336	
Mean absolute error	0.3825	
Root mean squared error	0.4447	
Relative absolute error	96.9488 %	
Root relative squared error	100.1788 %	
Total Number of Instances	467	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.977	0.952	0.735	0.977	0.839	非表示
0.048	0.023	0.429	0.048	0.086	表示

=== Confusion Matrix ===

```
  a  b  <-- classified as
333  8  |  a = 非表示
120  6  |  b = 表示
```

サマリーを見ていくと、誤分類率が3割弱で、非表示と分類する精度が73.5%である一方、表示と分類する精度が42.9%である。また、confusion matrixを見ると、ほとんどのレコードが非表示に分類されているのが分かる。極端に非表示に分類する決定木といえる。それでは、決定木の図を見て検討する(図11)。

ド、年齢ノード、性別ノード、レビューノードである。

(a) 職業ノードについて

職業ノードは所属先の状態に大いに関係がある。職業を非表示で所属先を表示するレコードも僅かにあるものの、多くの場合、職業 = 表示を必要とする。所属先が職業と関連しているため、サークルや地域組織を指しているというわけでもないようである。つまり、所属先とは職業における所属先である認識がなされているようである。

(b) 友達ノードについて

気づく点は友達 = 1人、および少数である場合所属先 = 非表示となる点である。所属先を表示する人は誕生日を表示していても、友達が少ないと所属先は表示しない。また、逆に友達 = 非常に多い場合、所属先 = 表示となっている。友人数で測る場合、交友範囲が広いほど所属先を表示しているということがいえそうである。関わる人が多いと所属先を表示する必要性に駆られるのであろうか。

(c) コミュニティノードについて

友達 = 多数の場合のコミュニティノードについてである。コミュニティ = 非常に多い、または非表示となっている場合は即座に所属先 = 表示となる。コミュニティ = 非常に多い場合、所属先を表示することがコミュニティでの発言やコミュニティへの関わりを円滑にしているのであろうか。一方、コミュニティ = 非表示の場合、所属先を表示するというのは友人達との関係上必要なものなのかもしれない。

(d) 年齢ノードについて

友人からのコメント = 非表示の場合の年齢ノードについては、所属先を表示しているのは年齢 = 18 ~ 22 である。友人からのコメントが非表示の場合、年齢が若い人のみ所属先を表示している。友達 = やや少数であり友人からのコメント = 非表示となっている若年層の人は所属先を表示する。つまり、友人との関係性がそのような場合、若年層のみ所属先を表示する傾向にあるということがいえそうである。

(e) 性別ノードについて

男性に特異な状況がある。友達 = やや多数で趣味 = 少数の男性は所属先を表示している。友達でみる交友範囲が広い人で、趣味の少ない男性は所属先を公開している。逆に女性は所属先を表示しない。このようなところで性差が見られる結果となった。

(f) レビューノードについて

所属しているコミュニティ数がやや少数である場合、レビューを表示していると、所属先を表示する。友達 = 多数である人なので、友達数でみ

る交友範囲は広いものの、所属コミュニティ数は少数である。その中でレビューを表示する人というのは、コミュニティに参加することで外部との交流を図るということをあまりせず、多くの友人にはレビューという形で情報を発信する傾向にあるといえる。そういった状況で所属先を表示するのは友人との関係上必要であるのかもしれない。

6 考察

6.1 相関ルールについての考察

分析結果をまとめると以下ようになる。

1. レビューに関して

- レビューを表示しているのは全体の14%である。(表4)
- 住所や年齢といった個人情報を非表示にしているレコードはレビューも非表示となる。(表5)
- コミュニティや友人からのコメントといったコミュニケーション状況を非表示にしているレコードはレビューも非表示となる。
- 好きなもの = 非表示を条件に含む場合レビュー = 非表示となるルールがある。

2. 日記に関して

- 結論部に日記を含むルールは生成されなかった。(表6)
- 日記とともに個人情報を表示していると、写真を表示している。(表14)
- 日記とともに好きなものを表示していると、写真を表示している。
- 日記 = 表示のみを条件に含むルールはコンフィデンス1ではなかった。

3. 友人からのコメントに関して

- 結論部に友人からのコメントを含むルールは生成されなかった。(表7)
- 友人からのコメント = 非表示を条件に含む場合、レビュー = 非表示を結論に持つ。
- 友人からのコメント = 非表示、自己紹介(自由文) = 表示、レビュー = 非表示の3つでほとんどのルールが生成されている。
- 友人からのコメント = 非表示で個人情報を表示とすると、レビュー = 非表示となる。(表15)

- 友人からのコメント = 非表示とともに所属先やコミュニティなど社会的な項目を非表示とすると、レビュー = 非表示となる。
- 全てのルールで友人からのコメント = 表示となる項目を含むものはなかった。

4. 職業に関して

- 職業 = 表示を条件に含む場合、自己紹介（自由文） = 表示が結論となる。（表 8）
- 所属先とともに個人情報を表示している場合、職業を表示している。（表 11）
- 所属先・個人情報を表示しているときにレビュー = 非表示となっても、職業 = 表示となる。

5. 所属先に関して

- 職業の非表示を条件に含む場合は所属先も非表示となる。（表 9）
- 所属先の表示を条件に含む場合は写真を表示する。
- 職業 = 非表示の場合、所属先 = 非表示となる。（表 12）
- 職業 = 非表示であり、他の項目が非表示となっていると所属先 = 非表示となる。その場合、非表示となる項目は個人情報でも社会的な項目でもかまわない。
- 上記の場合、所属先の他にレビューも非表示になる場合がある。

情報の受発信・コミュニケーションについて

まず、レビュー・日記・友人からのコメントを通して、情報の発信・他者とのコミュニケーションについて考察していく。レビューは他者に対する情報の積極的な発信の象徴であり、友人からのコメントは他者とのコミュニケーションの象徴であり、日記についてはその両者の中間に位置すると考えられるので、3項目について考慮することで、その人の情報の発信・他者とのコミュニケーションについての知見を深められる可能性がある。

レビューと日記は何かを発信したいという性質上、同時に非表示となっている項目が少ない場合はそれらも非表示になりやすく、個人情報のようなものでも表示している場合、同時に表示されやすい傾向にある。しかしながら、レビューについては表示するルールが生成されなかったため、表示されやすいか否かについては判断しきれない。また、両者に共通していることは、同時に必要な項目として好きなものが挙がっている。これは、双方とも発信者の嗜好が重要であるということが背景にあるからかもしれない。

それから、日記と友人からのコメントについて。双方とも他者からの反応を受信する性質がある³¹。そこで各ルールを見ていくと、日記の場合は日記のみを表示するルールは無く、同時に様々なものを表示することが求められているようであるし、友人からのコメントがつかない場合、レビューが非表示となっているルールが目立つ。他者からの反応を得るには、それなりの個人の素性を、住所などの個人情報あれ、好きなもののような個人の嗜好であれ、明らかにしておくことが必要なのかもしれない。

また、友人からのコメントとレビューについては、友人からのコメントがつかない条件の下では、結論がレビュー = 非表示となっているルールがある。これは、友人からのコメントがつかないようなコミュニケーション状況の場合、他者に対する情報発信も行っていないという状況とも考えられる。つまり、ネットワーク上のノードというよりはスタンドアロン型のノードであることが考えられる。そういった参加者もネットワーク上には存在しているというのは興味深い点である。

上記のことより、社会ネットワーク上では他者とのコミュニケーションに対して、また、情報の受発信に対して、ある程度の個人情報の開示が求められる用と考えられる。“Give And Take”という言葉があるが、むしろ“Give And Given”という言葉の方が当てはまるというのはいずれも。他者に対して何か与えたら、何かを得てよいというわけではなく、何かを為せば、何かを為されるというのが、インターネットの特性なのかもしれない。行為の主体者は他者であるということが現実の社会コミュニケーションよりも重要であると考えられる。

職業関連

続いて、職業および所属先について考察する。目的でも述べたとおり、職業はコミュニケーション上で大きな役割を担う可能性がある。

職業に関して、結論が職業 = 表示となったものは条件に所属先 = 表示を含むものだけであった。この結果から、所属先の表示に対して職業の表示が重要であるということが分かる。これは当然のことと思える。職業を明かさなくても所属先を明かすことはどういった仕事をしているかばかりかどのような会社にいるかといった情報も含んでいるため、情報の量が多い。つまり、公開する情報量を抑えたいという力が働けば所属先を公開する前に職業を公開する。同様に、職業 = 非表示である場合所属先 = 非表示となる。

一方、所属先とともに個人情報を表示し、レビューの非表示が条件に含まれているときにでも職業 = 表示となったルールが確認された。また、職業を非表示にしているとレビューも非表示になるというルールも確認された。レビューを表示するということは、所属先を表示するの同等かそれ以上の負担(コスト³²)があるのかもしれない。ここでは、レビューの表示と所属先の表示の関係について考えてみる。職業の非表示はレビュー = 非表示に対して前提条件であったが、レビュー =

³¹日記については記事に対する返信、友人からのコメントについてはその人自身についての意見や紹介等。

³²精神的なコストか時間的なコストかどういったものであるかは推し量れない。

表示は職業 = 表示の前提条件にはならなかった。一方，職業 = 表示の前提条件は所属先 = 表示であった。つまり，所属先を表示していれば職業を表示しているが，職業を表示しているといってもレビュー = 表示が期待できるわけではない。

職業と所属先の表示 / 非表示についての考察を進めると，レビューに対する表示 / 非表示についての知見を得られた。レビューというのは様々な個人情報の表示がある場合に期待できるものであったが，レビューの表示者率が低いことも影響し，最も前提条件が厳しいものであることが分かる。たとえば，職業であれば，所属先が表示されていれば表示されるが，レビューについてはそれだけでは足りないようである。

6.2 決定木分析についての考察

分析結果をまとめると以下ようになる。

1. レビューに関して

- 所属しているコミュニティ数がやや多数であるか非常に多い場合以外はレビューを表示しない。
- レビューを表示している人は所属コミュニティ数が非常に多く表示している項目数の少ない人。
- レビューを表示している人は好きなものを最大の3つ表示し，友人数が多い人である。
- レビューを表示する人というのは自分の嗜好を表示する傾向にあり，それらを伝える事のできる多くの友人もしくは多くのコミュニティに関係している。

2. 日記に関して

- 日記を表示しているある種の人には個人情報を表示するものの，所属先・写真が非表示な外部との交流を目的としないスタンドアローン型の人である。
- 日記を表示する人には性差がある。同一条件においては男性の方が表示している。女性については追加的な条件が必要となる。
- 日記を表示する人には友人数に違いがある。友人数が少ないと追加的な条件が重要となる。
- 友人数が少数である場合，個人情報か，嗜好を表すものの表示が必要となる。

- 友人数が1人である場合、項目毎に表示する項目があるほど日記も表示される傾向にある。

3. 友人からのコメントに関して

- コメント数については、コメントを付けてくれる友人数に大きく左右される。
- 友人数がやや少数である場合は追加的な条件によって状況が変わる。
 - － 年齢による差がある。25～28歳の場合はコメントが多くつきやすい。
 - － 職業の表示 / 非表示による差がある。表示にするとコメントが多くつきやすい。
 - － 誕生日の表示 / 非表示による差がある。表示にするとコメントがつく。

4. 職業に関して

- 職業の表示については年齢による差が見られる。20代中盤ではすぐに表示されるが他の年代には追加的な条件が必要となる。
- ターミナルノードへのエッジが{ = 非表示 }のものは職業も非表示となる。
- 職業の表示が持つ意味合いが各年代によって異なる可能性がある。
- 個人的属性は表示するものの、社会的属性を非表示としているタイプの人がいる。
- 多くのコミュニティに所属する女性は、コミュニケーションを円滑にするために職業を表示しているかもしれない。

5. 所属先に関して

- 所属先の表示には職業の表示が必要。
- 友人数が多いほど所属先を表示する。
 - － 友人数 = 非常に多いの場合、所属先を表示する。
 - － 友人数 = 多数の場合、コミュニティ数やレビューの状況により異なる。所属コミュニティや交友範囲によるレビューの持つ意味合いが異なっているからと考えられる。
 - － 友人数 = やや多数の場合、性差が見られる。趣味 = 少数の場合、男性は所属先を表示しているが女性は非表示。
 - － 友人数 = やや少数の場合、世代間差が見られる。友人からのコメント = 非表示の場合は若年層のみ所属先を表示している。

相関ルールのとおり同様に以下の2つについて考察をする。

情報の受発信・コミュニケーションについて

レビューを表示する人と層でない人の違いを外部との交流状況をコミュニティ数で判断する場合、レビューを表示する人がコミュニティに多く所属することからも、外部との接点を多く持ち、影響力を持ちたい、または、行使したいと考えることができるかもしれない。実際に、レビュー表示者は多くのコミュニティに所属し、比較的多くの友人を持つ。つまり、レビューという形で外部に影響を与えることができる（環境下にある）。また、レビュー表示者の特徴的な部分としては、外部との関わりを表す表示、つまり、コミュニティ数や友人数等は広く表示されているものの、個人情報については他の人に比べて非表示であることがある。レビューを表示するというのは外部への影響に大きく力点を置いたもので、自分自身についての情報をあまり曝さないものなのかもしれない。

一方、日記について考察すると、レビューと同じく外部に対する影響という面もあるのではあるが、個人情報を表示し外部との接点をあまり持たない人もいる。つまり、レビューは他者への積極的な影響力の行使である面が強いが、日記については自己完結する場合があるということである。そのため、友人数が少ない人でも日記を表示している場合がある。しかしながら、そういった人の場合は個人情報等を追加的に表示しなければならないようである。友人数が多い場合はレビューと同様の役割、つまり、他者への情報発信に力点が置かれるものの、友人数が少ない場合はそうではないようである。また、表示者率がレビューに比べて圧倒的に高く、心理的なハードルが低いものと思われる。また、性差があり、男性の方が比較的表示しやすいというのが注目すべき1つの点である。

友人からのコメントについてはある程度の友人数を必要とするものの、それ以外は年齢差や表示項目での差がある。友人からのコメントが表示されるにはコメントを付けてくれる友人数が強く効いてくるということが確認された。

職業関連

決定木分析でも職業と所属先の強い関係が見られた。所属先に関しては職業の表示が必要な条件であった。また、所属先の表示に友人数が関わっていることが確認された。所属先の表示が友人に向けたものである可能性がある。友人数の大小によって追加的な条件が変わることから、それらの条件によって分類できるような人たちにとっては所属先を表示することは意味が異なっているのかもしれない。たとえば、友人数 = 非常に多いの場合、所属先を表示するのは多くの友人達に対しての社会的コミュニケーション上に必要なアイデンティフィケーションなのかもしれないし、友人数 = やや少数の場合、追加的に友人からのコメントが必要となるが、友人からのコメント = 非表示においても若年層で表示されている所属先はもっと個人的なプロフィールの表示である性質が強いなのかもしれない。

つまり、所属先の意味合いが各者各様であるため、このような結果になったと思われる。それにもかかわらず、所属先と職業の強い関係性は維持されており、各人にとって意味合いが変わるものと変わらないものがあることが確認できる。

職業に関しては各者各様である性質がより強まる。所属先は職業との関係が強かったが、職業に関してはそういったものがない。強いて言うならば決定木におけるターミナルノードの存在であるが、それが各人の性質を表しているとは考えにくい³³ので、これは考慮しない。

そのような状況を考慮した上で考察すると、性差と年齢差があるように思われる。外部とのコミュニケーションを積極的に行おうとしている女性は職業を男性よりも表示するようであるし、各年代にとっての職業の意味合いが異なる可能性があるため年齢差が生じているようである。

7 終わりに

最後に全体を通した考察を記す。今回用いたデータ・マイニングの2手法を通して、共通の行動をとる人々をまとめて、各群について、また、各ルールについて眺めることで行動様式に対する知見を深めた。相関ルールを用いた分析では因果関係を考慮せずに同時に起こるイベントをまとめることからその行為者の性質を判断しようと試み、決定木分析でも同じように因果関係を考慮せずに各レコード（各人）を分類して同様の試みを行った。

当初の目的であった『情報を発信する人は同じ行動様式をとる』仮説については、相関ルールにおいては情報発信の項目で考察したとおり、表示する、すなわち情報を発信するルールがなかなか見つからなかったため行動様式が同じものであるかは判断できなかった。一方、決定木分析では、項目毎の分岐を通して当該項目における表示／非表示の各パスが描けたため、それはそれで意味のあるものであった。

目的に関しては、やや当てはまらないと考える。たしかに、同じく情報を発信する人を分類することによって行動様式が同じであるかどうかを判断することはできるとは思うが、結果的には同じ行動をとると言っても各人にとっての意味合いが異なるためにそれを1つの同じ情報発信とする考え方にはやや問題があるように感じる。広く情報を発信するのと、友人間にだけ情報を発信するのを区別することが必要となるかもしれない。実際、各人にとっての各項目は同じ意味合いを持っているとはかぎらず、様々な要因や場合分けが存在するように感じる。

また、こういった因果関係を考慮しないデータ・マイニングの手法を無意味と思う向きもないわけではない。ただ生起確率が高いからといってルールが存在するというのは多少乱暴な面があることは否めない。しかし、それを考慮して考えると、今回の結果からは各項目の個人にとっての意味合いが異なるとはいえ、分類することから次の分析・研究への課題と示唆が示せたと思う。

データを洗うことやルールや分類の検索によって、いくつかの課題が抽出されたと思う。これを次回に活かしたいと考えている。

³³決定木分析に用いたアルゴリズムが単純な分類のアルゴリズムであり、因果関係については考慮したものでないため。