

比率規則マイニングシステムの開発と評価

濱本 雅史[†] 北川 博之^{†,††}

[†] 筑波大学大学院システム情報工学研究科

^{††} 筑波大学計算科学研究センター

E-mail: [†]hamamoto@kde.cs.tsukuba.ac.jp, ^{††}kitagawa@cs.tsukuba.ac.jp

あらまし 比率規則は数値データ中における線形関係を表したものである。われわれはこれまでの研究において、比率規則を線分とその周辺領域として表現し、サポートや確信度といった概念を導入してきた。これによりユーザはパラメータを適当に与えることで、各ユーザの意図を汲んだ線形関係を比率規則として抽出可能となった。抽出された比率規則はデータの特徴を表すだけでなく未知データの予測などの応用が可能である。本論文では比率規則を抽出し応用するためのシステム RRMiner の構築について述べ、比率規則の応用手法を提案する。また試作した本システムを示すとともに、実データを用いた実験で比率規則の応用手法の妥当性を明らかにする。

キーワード 比率規則, 線形関係, 数値予測, データマイニング

Development and Evaluation of a Ratio Rule Mining System

Masafumi HAMAMOTO[†] and Hiroyuki KITAGAWA^{†,††}

[†] Graduate School of Systems and Information Engineering, University of Tsukuba

^{††} Center for Computational Sciences, University of Tsukuba

E-mail: [†]hamamoto@kde.cs.tsukuba.ac.jp, ^{††}kitagawa@cs.tsukuba.ac.jp

Abstract Ratio Rules represent linear relationships which are included in numeric data. In the previous works, we have defined a Ratio Rule as a line segment and its neighborhood, and introduced concepts such as support and confidence. These concepts enable us to extract linear relationships which reflect on user's intentions. Extracted Ratio Rules do not only reveal characteristics of data, but they are applicable to many kinds of purposes such as unknown data prediction. This paper describes development of a mining and application system for Ratio Rules, RRMiner, and some application methods for Ratio Rules are proposed. We show a demonstration system and evaluate adequacy of the application methods using real data.

Key words Ratio Rules, Linear Relationships, Numerical Prediction, Data Mining

1. はじめに

近年、大量のデータから重要な情報を抽出するデータマイニング手法は、扱うデータの種類により様々なものが検討されている。たとえば、相関ルールマイニング、クラスタリング、分類、テキストマイニング、時系列マイニング、Web マイニングなどがあげられる [3]。本論文では特に数値データを対象として比率規則 [8] を抽出する問題を考える。比率規則は論文 [8] において「『客は典型的に、パン：牛乳：バターに対し 1：2：5 の比率でお金を費やす』といった形のもの」という定義がなされているが、本研究は属性間で成り立つ線形関係を表したものと総じて比率規則と呼ぶ。

具体例として表 1 のように、各タプルが一人の学生を表し、それぞれ「身長」と「体重」の 2 つの数値属性を持つ学生デー

タを考える。各タプルをそれぞれの属性で張られる 2 次元空間へ射影したものが図 1 である。身長と体重は比例関係を持つので、全体的に黒い直線で表されたような線形関係の周辺に分布する。逆に言えばタプルの分布から黒い直線のような関係、すなわち比率規則を見つけることで、身長と体重の比例関係を明らかにすることができる。この例では 1 つの比率規則のみ示されているが、一般的には比率規則は複数同時に存在しうる。

主成分分析を用いた比率規則抽出手法 [8] や線形回帰分析などの既存の線形関係抽出手法は、比率規則を直線として表現する。それに対しわれわれは論文 [9] において、比率規則を線分およびその周辺領域内のデータが満たす性質として定義し、領域内に含まれているタプルはその比率規則に従うと定義した。線分として定義したことで大域的に成り立つ線形関係だけでなく、局所的に成り立つ線形関係を表現することが可能となった。

表 1 身長と体重の 2 属性を持つ学生データ例. いずれの属性も欠損値はないものとする

学生 ID	身長 (cm)	体重 (kg)
S0001	157	51.1
S0002	174	68.0
S0003	164	60.7
...

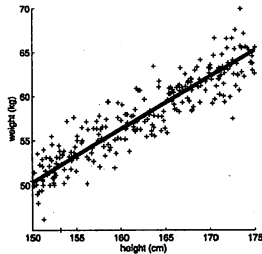


図 1 表 1 のデータに対する比率規則の例. 実線が比率規則を表す

また比率規則の周辺領域を評価することで、線形関係からの分散度合や従うタプル数による特徴付けが可能となった。具体的には、サポートや確信度といった相関ルールマイニングと対応づけられた諸概念により特徴付けされる。

このような先行研究を行ってきた一方、ユーザが比率規則を活用するための検討は不足している。得られた比率規則の評価も先行研究においては、与えたタプルの分布に沿った比率規則であるかといった点が主であった。そこで本論文では、比率規則を抽出し活用するためのシステム RRMiner の開発について述べる。本システムを用いることで、比率規則の抽出を GUI による操作のみで行うことができるほか、比率規則を可視化して表示したり未知のデータに対して数値予測をするなどの応用が可能となる。

本論文の以下は次のように構成される。2 章で線形関係の抽出に関する関連研究について触れ、3 章において本システムが扱う、われわれの先行研究の比率規則について述べる。4 章では比率規則マイニングシステム RRMiner について、その構成および各機能について述べる。5 章にて試作した本システムを示すとともに、本システムの数値予測機能について線形回帰分析と比較しその性質を明らかにする。最後にまとめと今後の課題について述べる。

2. 関連研究

数値データより線形関係を抽出する問題は、線形回帰分析をはじめとする多変量解析の対象として幅広い検討がなされてきた [4]。これらの目的は主に数値の予測であり、その代表の線形回帰分析では、目的変数の誤差が最小になるような単一の直線を抽出する。

これに対し、「身長 : 体重 = 3 : 2」のように数値属性間で成り立つ増分の比率に注目したものが比率規則 [7] である。比率規則の抽出手法として提案されているものとして主に Korn らによる手法 [7], [8] と Hu らによる手法 [5], [6] がある。

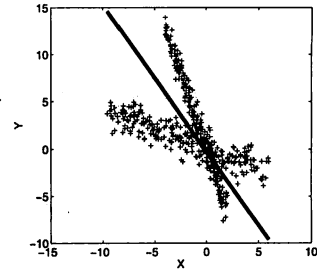


図 2 主成分分析を用いた手法ではうまくいかない例. 直線は得られた比率規則を表す. なお線形回帰分析でもほぼ同様の結果となる

Korn らの手法は主成分分析を用い、全体の分布を最大にする軸である主成分ベクトルのうち寄与率が一定以上のものを比率規則として抽出する。ただし主成分分析の意味を考慮すると、第一主成分に対応する比率規則は本質的に線形回帰とほぼ同じ結果である。この手法は図 2 のように複数の線形関係が混在する場合、第一主成分に対応する比率規則として図の直線のような結果が得られ、いずれの線形関係も直接的にとらえることができない。

Hu らの手法は与えられたデータが非負の実数で表され、かつ比率規則が負の相関を持たないことを仮定している。このとき与えられた各タプルが非負値からなるベクトルの線形和として表されていると仮定し、そのベクトルを比率規則として、非負行列分解を用いて抽出する。この手法では各比率規則は互いに直交するという制約はないが、原点を通るという制約を持っている。

これら既存の手法に対して、われわれが論文 [9] にて提案した比率規則の特徴として、以下の点があげられる。

- 比率規則を線分およびその周辺領域内のデータが満たす性質として定義した点。線分として表すことで局所的に成り立つ線形関係を抽出できる。また周辺領域に分布するタプルが比率規則に従うかどうかを考慮することで、個々の比率規則に特徴を与えている。
- 相関ルールマイニングと対応付けてサポートと確信度の概念を導入した点。相関関係の強弱や分布の疎密などにタプルの分布の仕方により、各比率規則のサポートや確信度が変化する。したがってユーザが最小サポートと最小確信度を与えることで、ユーザの意図に沿った比率規則が抽出可能となる。
- 条件を満たす比率規則をすべて列挙可能である点。回帰分析も含め、既存の手法では得られる線形関係の数に制約を持つが、本手法では最小サポートと最小確信度を満たす比率規則をもれなく列挙できる。

本論文は、先行研究では比率規則がタプルの分布を適当に捉えているかを重視していたのに対し、ユーザが比率規則を活用することに重点をおいている。そのうえで比率規則をマイニングし、予測などの応用を行うシステムの開発について述べている。

3. 比率規則

この章ではわれわれが先行研究 [9] にて提案した比率規則について、その定義と諸概念の概略を述べる。それに基づき、比率規則を抽出することがどのような意味を持つのかについて考察する。この比率規則は 2 属性間におけるものを対象としている。以下抽出対象の 2 属性を X, Y とし、それぞれの属性値を x, y と表す。また各属性値は区間 $[-0.5, 0.5]$ をとるよう正規化されているとする。

3.1 比率規則の定義

本研究で扱う比率規則は 1 章で述べたように、属性間の線形関係を表したものである。もし与えられたデータが全区間にわたり均一の線形関係を持つならば、単純に直線 $y = ax + b$ ($a, b \in \mathcal{R}$) のような形として比率規則を扱うことが考えられる。

しかし、この定義には 3 つの問題がある。1 点目は、一般的にはデータ全体的にわたり線形関係が成り立つとは限らず、局部的に成り立つ線形関係が表現できないことである。2 点目は y 軸に平行あるいはそれに近い線形関係を表す場合、パラメータ a, b がいずれも無限大に発散してしまうことで、この結果計算機中での取り扱いが困難になる。3 点目は、ユーザが知りたいことはタブルの分布が近似的に直線 $y = ax + b$ で表されるということである。

そこで 1 点目の問題については、線形関係が成り立つ部分空間をパラメータとして与え、その部分空間内の直線、すなわち線分として比率規則を表す。部分空間の与え方として 1 属性のみ [9] と 2 属性両方 [10] の 2 種類が考えられるが、本論文では前者を用い属性 X の区間 I として部分空間を表す。2 点目の問題については、Hough 変換 [1] により有限区間の変数へ変換を行う。Hough 変換を用いると直線 $y = ax + b$ は $\rho = x \cos \theta + y \sin \theta$ (ただし $\rho = b \sin(\tan^{-1}(-1/a))$, $\theta = \tan^{-1}(-1/a)$) と表現される。3 点目の問題に対して、パラメータに対する許容幅を設定し、許容幅内の任意の線分上に存在するタブルは同一の比率規則に従うとすることで、線分の近傍領域を考慮する^(注1)。

以上より、比率規則は次のように定義される。以下この定義に従う比率規則を誤解のない限り $RR_I(\rho, \theta)$ と省略して表現する。

タブル $t(x_t, y_t)$ ($x_t \in I, I \subseteq [-0.5, 0.5]$) が以下の式を満たす値 ϵ_t を持つとき、 t は比率規則 $RR_{x \in I}(\rho \pm \epsilon, \theta)$ に従う。
$$\rho + \epsilon_t = x_t \cos(\theta) + y_t \sin(\theta) \text{ ただし } |\epsilon_t| \leq \epsilon$$

3.2 比率規則の諸概念

比率規則 $RR_I(\rho, \theta)$ は単に線形関係を表すのではなく、区間 I 中でのタブルの振る舞い方を表し、各タブルが比率規則に従うかどうか厳密に区別される。そこでより直感的な尺度を与え

るため、数値属性に対する相関ルールマイニング [2] と対応付けし、以下のような諸概念を定義する。

- 区間 I に対するサポートは属性値 x が区間 I に含まれるタブルの、全タブルに対する割合とし $support(I)$ と表す。
- 比率規則 $RR_I(\rho, \theta)$ に対する確信度は区間 I に含まれるタブルのうち比率規則 $RR_I(\rho, \theta)$ に従うタブルの割合とし $conf(RR_I(\rho, \theta))$ と表す。これはサポートで表すと $support(RR_I(\rho, \theta)) / support(I)$ となる。
- 抽出される比率規則に対し、ユーザから与えられる最低限満たすべきサポートおよび確信度をそれぞれ最小サポート、最小確信度と呼ぶ。以下ではそれぞれ $minsup$, $minconf$ と表す。

これらサポートと確信度の概念を用いて、次の 2 種類の比率規則を定義する。

- 最適確信度比率規則: $support(I)$ が $minsup$ を満たし、かつ $conf(RR_I(\rho, \theta))$ が $minconf$ を満たしたうえで最大となるような比率規則 $RR_I(\rho, \theta)$ 。最大値を与える区間 I を最適確信度区間と呼ぶ。
- 最適サポート比率規則: $conf(RR_I(\rho, \theta))$ が $minconf$ を満たし、かつ $support(I)$ が $minsup$ を満たしたうえで最大となるような比率規則 $RR_I(\rho, \theta)$ 。最大値を与える区間 I を最適サポート区間と呼ぶ。

サポートは分析対象の区間 I に含まれるタブル数であり、確信度はその区間内においてタブルが比率規則に従う確率を与えると解釈できる。そのため多少の誤差はあってもなるべく多くのタブルを分析した比率規則が欲しい場合は、最小確信度を小さくし最小サポートを大きくして、最適サポート比率規則を求めればよい。一方少ないタブルでもある区間では確実に従っているような比率規則が欲しい場合は、最小サポートを小さくし最小確信度を大きく設定して、最適確信度比率規則を抽出すればよい。

4. 比率規則マイニングシステム: RRMiner

前章にて述べた比率規則をマイニングし、その結果を応用するためのシステム RRMiner を説明する。まずシステムの構成について述べ、続いて各機能の説明を行う。

4.1 システム構成

本システムの目的は比率規則を単に抽出するだけでなく、その結果を用いた応用も行うことである。そのためシステムの構成としては大まかに、比率規則の抽出を行うマイニング部と、得られた比率規則の応用を行うアプリケーション部の 2 つに分けられる。これに加え、ユーザとの入出力のやりとりを行う GUI を備えたものが RRMiner の構成である (図 3)。

マイニング部では、マイニング機能によりユーザより与えられたデータとパラメータから比率規則を抽出し、その結果を XML ファイルとして書き出す。アプリケーション部では抽出された比率規則を様々な用途に応用する。本システムでは比率規則の応用として、比率規則を可視化して表示する機能と、数値の予測を行う機能を持つ。次節から各機能について説明する。

実際に実装された RRMiner のスクリーンショットは図 4 で

(注1): 許容幅は ρ と θ いずれにも設定が可能であり、先行研究ではそれぞれ ϵ, δ という許容幅を設定していたが、本論文ではより単純に線分からのユークリッド距離を表す ϵ のみを用いる

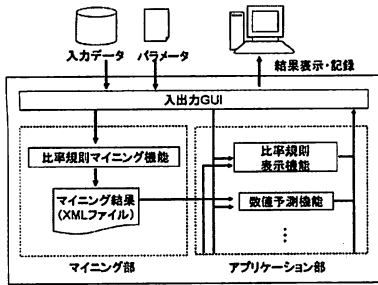


図 3 RRMiner のシステム構成図

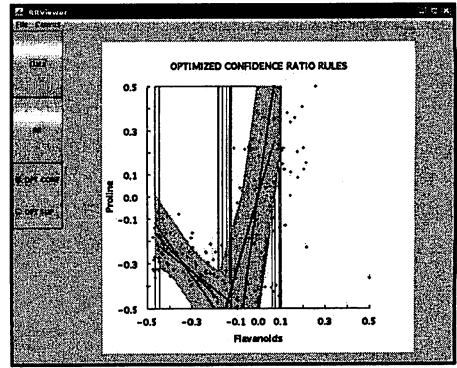


図 5 可視化機能画面

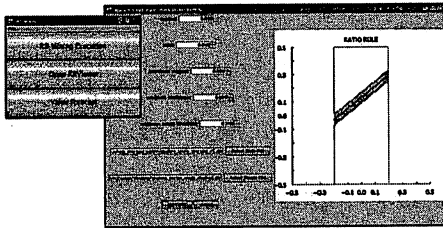


図 4 RRMiner の実行画面。左側の機能選択ボタンにてマイニング機能を選択したところである

ある。GUI 部分およびアプリケーション部は Java(JDK 6.0)、マイニング部は C 言語で実装されている。OS は現在のところ、WindowsXP と Linux(Fedora Core 5) にて動作を確認している。

4.2 マイニング機能

マイニング機能ではユーザが入力データとパラメータを与えることで、パラメータを満たす比率規則をもれなく抽出しその結果を XML ファイルに出力する。アルゴリズムの詳細は論文 [9] で述べているので、ここでは概要のみの説明にとどめておく。

このアルゴリズムはパラメータ ρ と θ をそれぞれ離散値として扱い、その上で各 (ρ_i, θ_j) の組み合わせに対し最適比率規則を調べる。もし最小サポートと最小確信度の条件を満たす比率規則が存在する場合は、目的の比率規則の種類に合わせサポートあるいは確信度が最大となる区間 I を抽出する。個々の最適比率規則は以上により求められるが、このまま出力すると多数の比率規則が得られた際、同一の線形関係を表す比率規則群とそうでない比率規則群の区別が困難になる。そこで類似した比率規則は比率規則集合としてグループ化して出力する。アルゴリズムの計算量は、与えたタプル数に対して線形時間で求められる。

4.3 可視化機能

マイニング機能の結果は XML ファイルとして出力されるが、そのままでは直感的にどのような結果が得られたのか理解するのは難しい。この機能では与えられたデータと得られた比率規則を 2 次元上に表示することで、タプルの分布に対しどのような比率規則が得られたのかを理解する。具体的な画面を図 5 に

示す。

中央の図において、黒の点は各タプルを表し、濃い色の線分は得られた各比率規則の許容幅を含まない形 $(RR_{x \in I}(\rho \pm \theta))$ 、薄い色の領域は許容幅も含め得られた各比率規則が成り立つ領域を示す。色は比率規則集合ごとに変わっており、同一の比率規則集合に含まれる比率規則はすべて同様に描画している。各比率規則を囲うように描かれている長方形の幅は、比率規則の分析対象となる区間 I である。すなわち、全タプル数に対する長方形内に含まれるタプル数が区間 I のサポートであり、長方形内に含まれるタプル中で比率規則に従うタプル数が比率規則の確信度となる。

4.4 数値予測機能

この機能では比率規則を抽出する際に与えたデータと似た振る舞いを未知のデータもすると仮定し、得られた比率規則から未知の数値を予測する。

具体的には属性 X (比率規則 $RR_I(\rho, \theta)$ の区間 I を与える属性) を既知とし、属性 Y を未知として予測を行う。このとき与えられた属性値 x と比率規則 $RR_I(\rho, \theta)$ において $x \in I$ が成り立つならば、その比率規則を用いて予測を行う。最終的にひとつも予測可能な比率規則が存在しない場合、予測不可を表す値を出力する。区間 I に x を含む比率規則が複数存在する場合、複数の候補が得られるが、本システムでは最も確信度が高い比率規則による予測値を出力する。

予測値は単一の値ではなく、出現する可能性がある範囲とその範囲に出現する確率を与える。これは比率規則に従うと仮定したとき取るべき値と、その比率規則の確信度である。属性値 x と比率規則 $RR_I(\rho, \theta)$ について、属性値 y の取りうる範囲 $[\min(y), \max(y)]$ は次のように求められる。

$$\begin{aligned} \min(y) &= \min((\rho - \epsilon - x \cos \theta) / \sin \theta, (\rho + \epsilon - x \cos \theta) / \sin \theta) \\ \max(y) &= \max((\rho - \epsilon - x \cos \theta) / \sin \theta, (\rho + \epsilon - x \cos \theta) / \sin \theta) \end{aligned}$$

この機能は図 6 のようなインタフェースを持ち、対話的に予測ができるほかファイルを読み込み一括予測することも可能である。

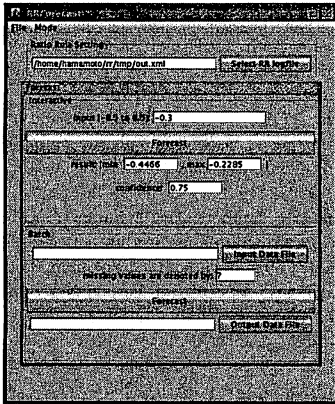


図6 数値予測機能画面

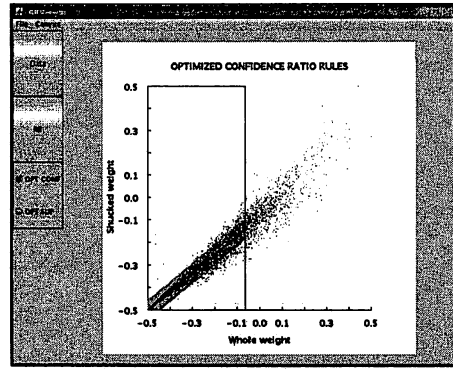


図7 アワビデータより抽出された最適確信度比率規則

5. 実験

本実験では RRMiner を用いて得られた比率規則が、どれだけ未知のデータの線形関係を予測できるか検証する。具体的には、3-fold Cross Validation, すなわち与えられたデータセットを3つのグループに分け、1つのグループ(テストデータ)を残り2つのグループ(学習データ)から予測することを全組み合わせについて行う。

実験には UCI Machine Learning Repository^(注2)で配布されているデータを使用した。

5.1 評価方法

本システムの目的は属性の絶対値の予測ではなく、ユーザの意図に沿った線形関係を抽出することである。そのため評価はテストデータがどれだけ学習データの比率規則に従っているかで行う。評価項目としては予測可能なタブルの全タブル数に対する割合 N と、予測可能なタブル中で比率規則に従うタブルの割合 NF を調べる。ここで N はサポート、 NF は確信度に対応する。3章で説明したように比率規則には最適サポートと最適確信度の2種類の比率規則があるが、今回は最適確信度比率規則のみを用いて予測を行う。

比較対象として、線形回帰分析で得られた回帰直線を用いた予測を行う。あるタブルが回帰直線に従うかどうかは比率規則と同様に、直線から許容幅 ϵ 内に含まれるかどうかで判定する。ここで回帰直線は全区間に渡って予測可能であるので、予測可能なタブルの割合 N は常に1となる。

5.2 実データ1: アワビデータ

このデータにはアワビの体長や性別などの計測データが記録されている。今回は連続値で表される属性のうち、貝全体の重さ(Whole weight)と身の重さ(Shucked weight)の2属性を用いた。全タブル数は4177個である。

図7は全タブルを対象として最適確信度比率規則を抽出した例である。パラメータは $\epsilon = 0.035$, $minsup = minconf = 0.8$ とした。タブルの分布はこの例からも分かるとおり、全体的に

表2 アワビデータに対する実験結果

	N	NF
比率規則 ($minsup = minconf = 0.8$)	0.821	0.847
比率規則 ($minsup = 0.5, minconf = 0.9$)	0.521	0.952
線形回帰	1.0	0.839

一種類の線形関係を持つが属性値が小さい部分では強い線形関係を持っており、属性値が大きくなるにつれタブルが分散していく。そのため例では属性値が小さい部分の線形関係のみを捉えている。

評価実験では最適確信度比率規則を抽出する際のパラメータとして $\epsilon = 0.035$ を固定し、最小サポートと最小確信度は $minsup = minconf = 0.8$ と $minsup = 0.5, minconf = 0.9$ の2種類を与えた。

実験結果は表2である。比率規則では強い線形関係が成り立ち分散が小さい線形関係の部分のみが得られるため、比率規則に従うタブルの割合 NF が線形回帰よりも大きくなっている。また比率規則の結果はユーザが与えた最小サポートと最小確信度の値に応じて変化しており、このことからパラメータによってユーザの意図を反映させることができると考えられる。

5.3 実データ2: ワインデータ

このデータは3つの異なる品種のワインについて、アルコールやリンゴ酸など13項目が調べられた化学分析データである。本実験では“Flavanoids”と“Proline”の2属性を用いた。全タブル数は178個である。

最適確信度比率規則を抽出する際のパラメータは $\epsilon = 0.10$, $minsup = 0.375$, $minconf = 0.725$ とした。本データは疎密の差があまりないためこの一種類のみを用いる。

実験結果は表3のとおり、予測対象のタブルに対して比率規則に従うタブルの割合 NF が大きく異なる。この理由は図8と9で示されているように、ワインデータが2種類の線形関係を含んでいるためであると考えられる。比率規則の場合はその2種類の比率規則をいずれも抽出できるのに対し、回帰分析では無理に一直線で表そうとしているため、結果としていずれの線形関係とも異なる直線になってしまっている。

以上2つの実データに対する実験より、本システムではユーザの意図に沿った線形関係を抽出できると考えられる。

(注2): <http://www.ics.uci.edu/~mllearn/MLRepository.html>

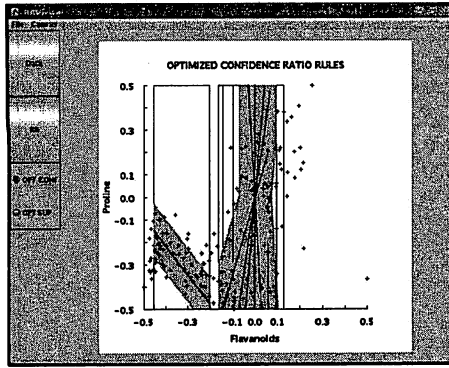


図 8 ワインデータ全体より抽出された最適確信度比率規則. パラメータは $\epsilon = 0.075$, $minsup = 0.35$, $minconf = 0.75$ である

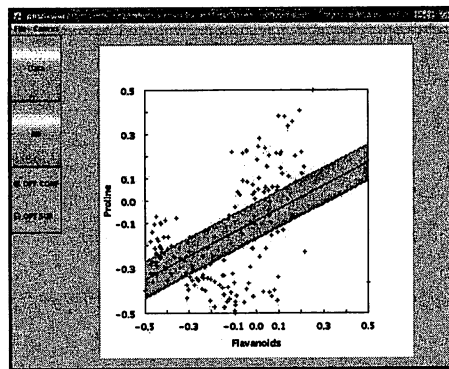


図 9 ワインデータ全体より抽出された回帰直線と, 図 8 に対応して設定された許容幅

表 3 ワインデータに対する実験結果

	N	NF
比率規則	0.837	0.758
線形回帰	1.0	0.433

6. おわりに

本論文では, 比率規則の抽出と応用を行う比率規則マイニングシステム RRMiner について述べた. 本システムはわれわれの先行研究である, サポートと確信度の概念を持った比率規則を GUI 上での操作によって抽出する. また単に比率規則を抽出するだけでなく, 本システムは比率規則の応用として可視化する機能と数値予測を行う機能を持つ. 本論文ではこの機能について 2 種類の実データに適用した結果を示し, その妥当性を確認した.

今後の課題として, 本論文で述べた以外の比率規則の応用や様々な実データに対する検証のほか, よりユーザが利用しやすくなるよう実装の改良を行うことがあげられる.

謝辞

本研究の一部は, 科学研究費補助金特定領域研究 (#19024006) による.

- [1] R. Duda and P. Hart, "Use of the hough transformation to detect lines and curves in pictures," Communications of the ACM, vol.15, no.1, pp.11-15, 1972.
- [2] T. Fukuda, Y. Morimoto, S. Morishita, and T. Tokuyama, "Mining optimized association rules for numeric attributes," Journal of Computer and System Sciences, vol.58, no.1, pp.1-12, 1999.
- [3] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2001.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, 2001.
- [5] C. Hu, Y. Wang, B. Zhang, Q. Yang, Q. Wang, J. Zhou, R. He, and Y. Yan, "Mining quantitative associations in large database," Proc. 7th Asia-Pacific Web Conference, Shanghai, China, pp.405-416, 2005.
- [6] C. Hu, B. Zhang, S. Yan, Q. Yang, J. Yan, Z. Chen, and W.Y. Ma, "Mining ratio rules via principal sparse non-negative matrix factorization," Proc. 4th IEEE International Conference on Data Mining, Brighton, U.K., pp.407-410, 2004.
- [7] F. Korn, A. Labrinidis, Y. Kotidis, and C. Faloutsos, "Ratio rules: A new paradigm for fast, quantifiable data mining," Proc. 24th International Conference on Very Large Data Bases, New York, pp.582-593, 1998.
- [8] F. Korn, A. Labrinidis, Y. Kotidis, and C. Faloutsos, "Quantifiable data mining using ratio rules," VLDB Journal, vol.8, pp.254-266, 2000.
- [9] 濱本雅史, 北川博之, "サポートと確信度をもとにした比率規則による線形関係抽出," 情報処理学会論文誌: データベース, vol.47, no.SIG19(TOD32), pp.54-71, 2006.
- [10] 濱本雅史, 北川博之, "対称比率規則の抽出手法," 電子情報通信学会第 18 回データ工学ワークショップ (DEWS2007), 2007.