

地域環境問題の総合モデルを用いた分析と シミュレーション技法の開発（概要書）

北陸先端科学技術大学院大学 知識科学研究科 河野小夜子

2002年8月

地域住民の環境に対する感性・感覚というソフトデータがハードデータを補完する役割を担うものであるかどうかを検討し、定性的・準定量的なソフトデータを総合モデルに組み込むための手法を考察する。最終的な研究の目的は、客観的性質を持つデータに加えて、個人や組織、社会が持つ知識を統合することで、地域環境問題の現状把握と将来予測を行なうことである。この基盤として総合モデルを用いている。

住民意識調査の結果とその分析を行うことによって、人間の意識というあいまいな性質を持ったデータの分析手法を提案した。住民データが既存の数値データと合致する、あるいは数値データで得られない知見が住民データから得たことで、主観的性質を持つ感性データが環境評価に対して有効であることを示唆することができた。

地域環境問題の総合モデルを用いた分析とシミュレーション技法の開発（報告書）

1 研究の背景と目的

環境問題に対する総合的なアプローチの必要性は以前より指摘されている。近年では国・世界レベルの地球温暖化問題対策において、総合モデルを用いた政策提言を目的とする研究が盛んに行われている（[1][2]他）。しかし、自治体レベルの地域環境管理にこうした手法を適用した研究事例は少ない。

環境問題は、これを構成する要素が非常に多く、要素間の関係が複雑であるという特性が起因となって、全体を捉えることが困難である。また、ほとんどの研究で利用されるハードデータ、すなわち数値で表される客観的な実測値は問題の分析に必要な要素に対して存在しておらず、また観測地点の数・場所の特性による不備などの欠点も指摘されている。ハードデータに対して、人間の感覚や感性、知識と言った主観的なデータを総称して、ソフトデータと呼ぶ。日々の生活において多かれ少なかれ環境に接して生きている我々人間の持つ知識というデータを利用すれば、ハードデータだけでは分析できない問題を扱えると考えている。

そこで本研究では、地域住民の環境に対する感性・感覚というソフトデータがハードデータを補完する役割を担うものであるかどうかを検討し、定性的・準定量的なソフトデータを総合モデルに組み込むための手法を考察する。最終的な研究の目的は、客観的性質を持つデータに加えて、個人や組織、社会が持つ知識を統合することで、地域環境問題の現状把握と将来予測を行なうことである。この基盤として総合モデルを用いている [3]。

地域環境問題と一口に言っても、その範囲は非常に大きい。ここでは、石川県で最も重要な問題とされている水質汚濁問題に注目し、2000年に石川県加賀地方（石川県の南半分であり、河北郡以南の市町村が含まれる）を対象にしたアンケート、及び2001年に石川県小松市、石川郡鶴来町を対象に行ったアンケート結果をそれぞれ利用して、ハードデータとの比較及び知識抽出を行っている。これを踏まえた人間活動と環境の関係を結びつけるモデリング作業については現時点では未完成であり、今後継続して行う。

環境評価における主観的なソフトデータの利用については、1980年代後半に行われた環境指標に関する研究が代表的である [4, 5, 6, 7]。これらの研究は、デー

タを分析する際に回帰分析、主成分分析を利用している。指標の作成は行っていないが、水域の環境を住民意識調査に基づいて分析した研究としては、連想法を用いた研究 [8, 9] がある。

以下では、ソフトデータからデータ・マイニング手法、数量化Ⅲ類を用いて知識抽出を行った研究を紹介する。

2 SC-Optimality アルゴリズムを用いた知識抽出

本節は [13] を抜粋したものであり、詳細（一部説明、ルールの表）は略している。

2.1 SC-Optimality アルゴリズム

2000年のアンケートは評価者が900、属性が10近くと、かなり大きなデータベースとなっている。ここから、環境評価に関わっていると考えられる属性間の関係を示す、ルールベース型モデルを得る。既存のデータ解析手法で精度の高いモデルを得るのが困難であったことから、データ・マイニング手法を用いている。KDD (Knowledge Discovery in Databases=大量のデータから以前には知られていない、有用な知識を引き出すプロセス全体)の中で、隠された知識や新しい規則を発見する段階がデータ・マイニングであり、「データの要素間の関係とそのルール（パターン）」が知識と定義されている。

ルールは $A \rightarrow C$ (もし A を満たしているならば C を満たしている) と表され、前提である A を前件部、結論である C を後件部と呼ぶ。ルールの評価には主として支持度と確信度という指標が用いられる。支持度 (support value) とは、対象とする全データの中でその条件を満たすデータ割合を示す。例えば、 $sup(A)$ は対象とする全データの中で A を満たすデータの割合を、 $sup(A \rightarrow C)$ は対象とする全データの中でルール $A \rightarrow C$ 、すなわち A と C の両方を満たしているデータの割合を表す。確信度 (confidence value) とは、前件部が A であるデータセットの中で後件部が C とな

るデータの割合を示し、

$$\text{conf}(A \rightarrow C) = \frac{\text{sup}(A \rightarrow C)}{\text{sup}(A)} \quad (1)$$

と表す。

SC-Optimality アルゴリズムは、制約条件を満たすルールを抽出した後に、SC-order という考え方に基いてそれらのルールを順序づける手法である。SC-Optimality アルゴリズムを用いるのは、ルールの多さが真に重要なルールを見出すことを困難にする場合があること、環境評価を政策策定に反映させていくことを考慮するとより多くの住民が感じている意見・知識を絞り込んでいきたいと考えていることが理由である。

SC-Optimality アルゴリズムは、以下のステップを踏む。

1. 対象の中から、制約条件を満たすルールを抽出する（ここでは Dense-Miner アルゴリズム [14] を用いている）。
2. SC-order を用いてルールを順序づける。ルール r が、最適なルールの集合であると仮定したルールの集合 R の全てのルールと比較され、 r が R の各ルールより良いルールであるか比較できないルールであるとされた場合には r が R に加わり、 R の中で r より劣るとされたルールは R から除去される。これを全てのルールについて繰り返し、 R の中に残ったルールが全て選択される。

支持度と確信度の両方を用いてルールの順位を付ける SC-order では、ルールの関係は \leq_{sc} (SC-order) を使って表される。今、二つのルール r_1 、 r_2 が $r_1 \leq_{sc} r_2$ を満たすのは、

- $\text{sup}(r_1) \leq \text{sup}(r_2)$ and $\text{conf}(r_1) < \text{conf}(r_2)$ または
- $\text{sup}(r_1) < \text{sup}(r_2)$ and $\text{conf}(r_1) \leq \text{conf}(r_2)$

が成り立つ場合であり、このとき r_2 は r_1 より良いルールとなって、 r_2 が抽出される。

$r_1 = r_2$ と表せるのは、

- $\text{sup}(r_1) = \text{sup}(r_2)$ かつ
- $\text{conf}(r_1) = \text{conf}(r_2)$

が成り立つ場合のみである。

もし下記の関係が成り立つ場合には、 \leq_{sc} によってルールを比較することはできず、どちらのルールとも選択される。

- $\text{sup}(r_1) < \text{sup}(r_2)$ and $\text{conf}(r_1) > \text{conf}(r_2)$ または
- $\text{sup}(r_1) > \text{sup}(r_2)$ and $\text{conf}(r_1) < \text{conf}(r_2)$

2.2 2000 年住民意識調査

住民への意識調査は、2000 年 12 月に石川県加賀地方の 4 市 13 町 5 村（金沢市、小松市、加賀市、松任市及び江沼郡、能美郡、石川郡、河北郡の全町村）の住民を対象に郵送方式で実施した。回答者数/調査票送付数は、1081/3000、回収率は 36%、有効回答数は 900 となった。

回答者に身近な水辺を思い浮かべてもらい、その水辺に対して質問（表 1）に回答してもらい、この回答をモデリングに利用している。表 1 中、Q1 から Q7 を個別（質問）項目、Q8 を総合評価と呼ぶ。

どの質問も 5 段階評価の中から該当するものを選択してもらおう。選択肢は表 2 の意味を持つ。

表 2: 評点の意味

質問	評価	意味
Q1	1	そう思わない
	2	あまりそう思わない
	3	どちらとも言えない
	4	ややそう思う
	5	そう思う
Q8	1	汚れている
	2	やや汚れている
	3	ふつう
	4	ややきれい
	5	きれい

2.3 ルール抽出

2.3.1 ハードデータによるクラスタ分析

意識調査を行った加賀地方全体は、居住環境の異なる地域が混在している。そこで、石川県内の各市町村を分割し、分割した地域毎に環境評価モデル構築の対象とする。分割の方法として、社会的・地理的特性を示す既存のハードデータを用いてクラスタリングを行い、得られたクラスタ毎にソフトデータによるルールを抽出するものとする。

表 1: アンケート調査質問一覧 (2000 年)

[身近な水辺空間について]

- Q1. 鳥や魚などの生き物を見かける
- Q2. 水遊び (泳ぐ、ボート、釣りなど) ができる
- Q3. 釣れた魚が食べられる
- Q4. 水が茶色い
- Q5. 水辺でバーベキューやキャンプができる
- Q6. 近隣の流域に汚濁の原因 (工場や生活排水の流入など) がある
- Q7. ヨシや水草などの植物をみかける
- Q8. 現在の水質はどうか

表 3: ハードデータによるクラスタリング結果 (石川県加賀地方)

クラスタ 1	金沢市
クラスタ 2	松任市、根上町、寺井町
クラスタ 3	小松市、加賀市、山中町、辰口町 鶴来町、津幡町、高松町
クラスタ 4	河内村、鳥越村、宇ノ気町
クラスタ 5	七塚町、内灘町
クラスタ 6	吉野谷村、尾口村、白峰村

クラスタリング結果は、3 に示すものを採用している。クラスタリング結果の地理的状况を判断し、結果の意味付けが容易であったためである (1 参照)。

クラスタ 1 は石川県で最大の都市である金沢市のみに構成されている。人口が県内で最も多く、県内一の繁華街がある一方で、広大な山林をも持っている。クラスタ 2 (松任市他) は金沢市から西南の臨海の市町が集まっており、工業と農業が共に発達している。クラスタ 3 (小松市他) は、地理的には金沢市を挟んでいる、発展途中の市町が集まっている。また、県内 2 番目の市も含まれる。クラスタ 4 (鳥越村他) はクラスタ 3 の隣に位置する郊外部の町村、クラスタ 5 (内灘町他) は金沢市のベッドタウンとなる臨海の町、クラスタ 6 (尾口村他) は白山山系山麓の村である。

2.3.2 ソフトデータによるルール作成

2.1 節で紹介した SC-Optimality アルゴリズムを用いて、2.3.1 節で分けた地域毎にルールを抽出する。こ



図 1: 石川県加賀地方のクラスタリング結果

のとき、Q1 から Q7 の個別項目を条件属性、これら条件属性によって決定される属性である決定属性を Q8 の総合評価とすることで、水質を判断する材料となっていると考えられる個別項目と総合評価との関係を示すルールを得る。ルールの前件部は条件属性とその評価値、後件部は決定属性の評価値で構成される。

ルールは、「水遊びを 1 (=水遊びをしたいと思わない)、かつ茶色を 5 (=水が茶色い)、かつ汚濁源を 5 (=近隣に汚濁源があると思う) と答えたならば総合評価は 1 (=汚い) である。ルールの支持度は 12.7%、確信度は 75% である」という形である。4 では、各クラスタ毎に個別項目の出現数の、そのクラスタで得られた全ルール数に対する割合を示している。例えば、クラスタ

表 4: 個別項目の出現回数の比較 (単位は [%])

クラスタ/項目	生き物	水遊び	魚	茶色	キャンプ	汚濁源	植物
クラスタ 1	48.3	37.9	34.5	62.1	44.8	62.1	41.4
クラスタ 2	31.6	15.8	47.4	31.6	21.1	42.1	26.3
クラスタ 3	35.7	35.7	14.3	46.4	53.6	57.1	42.9
クラスタ 4	52.9	29.4	29.4	29.4	29.4	41.2	41.2
クラスタ 5	7.7	23.1	15.4	38.5	46.2	15.4	15.4

1で「生き物がいる(生き物={1, 2, …, 5})」という項目は、29(総合評価1~5のルール数の合計)のルール中14のルールに現れているので、 $14/29 \times 100 = 48.3\%$ となる。一つのルールは複数のルールで構成されることがほとんどであるため、同じクラスタ内で個別項目の割合を合計した値は100%にはならない。

2.4 結果と地域毎の考察

[クラスタ1(金沢市)]

一つのルールを構成する項目数が他クラスタに比べると多い傾向にあり、その中でも「汚濁源の有無」「水が茶色かどうか」の出現数が6割以上を占めている。その一方、「水遊びの可・不可」「魚を食べることの可・不可」といった、水辺への親しみやすさを表していると考えられる項目は出現数が少ない。「茶色」「汚濁源」はどの段階の総合評価にもルールに含まれており、しかも総合評価によってこれらの項目の評点も異なっている。これに対して、「植物の有無」「生き物の有無」は、総合評価の高低に関わらず評点が高い傾向にある。このクラスタでは、植物や生物の有無は総合評価にあまり影響を及ぼしていないことが推測できる。

金沢市は繁華街を含む一方で広大な山林も含んでおり、水辺空間を取り巻く状況が様々である。そのため、このクラスタでは、水質そのものと共に水辺空間及びその近隣に関連がある項目が総合評価に影響を与えている、あるいはこうした項目の評価が似通う傾向があると考えられる。

[クラスタ2(松任市他)]

「魚を食べられることの可・不可」が各総合段階のルールに異なる評点で現れていることから、重要な項目となっていることが推測できる。また、総合評価

が4(=ややきれい)または5(=きれい)であっても「汚濁源=4(=汚濁源がややある)」という条件属性が含まれていることが特徴である。「水遊びの可・不可」「キャンプの可・不可」といったレクリエーションに関連する項目は他項目に比べると出現数が少ない。特に総合評価が4(=ややきれい)や5(=きれい)と高い場合に「水遊び」が全く現れておらず、これらの段階では「水遊び」に対する評価にばらつきがあると考えられる。

このクラスタに属する市町はいずれも海に隣接しており、また魚が取れることで有名な川を含んでいることから「魚」が、農業地帯であることに加えて近年工業地帯としても発展してきていることから「汚濁源」がそれぞれ重視されている、すなわち地域属性を顕著に反映した総合評価がされていると考えられる。

[クラスタ3(小松市他)]

クラスタ1で挙げた3つの属性(汚濁源、キャンプ、茶色)の出現数が多い。また、総合評価がどの段階においても「生き物または植物の有無」の評点が高い傾向にある。これはクラスタ1と同様の結果であり、このクラスタがクラスタ1に次ぐ規模の市を含んでいることから、規模の大きな市町に共通する性質ではないかと考えられる。

総合評価5(=きれい)においては「キャンプの可・不可」の評点が分かれている。また、他の項目に比べると「魚を食べることの可・不可」がルールに含まれる頻度が少ない。これらは、このクラスタ内には比較的BOD値の低い、すなわち水質のきれいな河川と共に汚染の進んだ河川・湖沼もあり、見ている水辺の違いがルールに反映されていること、水質と水辺でのレクリエーションに対する評価が別となっていること、汚染のある場所で魚を釣って食べようとは到底思わないこと、などを示唆している可能性がある。

[クラスタ 4 (鳥越村他)]

1 町 2 村を含むクラスタで、その位置は 2 村が隣り合い 1 町は大きく離れている。1 町は総合評価が 2 (= やや汚い) または 3 (= ふつう)、1 村は 4 (= ややきれい) または 5 (= きれい) しか付けられておらず、残りの 1 村の総合評価が 1~5 と全段階に分布している。ルールについては、総合評価が 2 の段階で得られるルールでも条件属性の評価値は比較的良く (例えば、「茶色=1 (水は茶色くない)、キャンプ=5 (キャンプができる)」)、個別項目の評価に対して総合評価が低くなる傾向が見られる。このクラスタは比較的水質の良い河川を含んでいるにも関わらず、住民は水質を厳しく評価していることが窺える。

このクラスタでは、「生き物の有無」がルールに使われる割合が、同じクラスタの他項目、他クラスタの「生き物」の割合と比較しても多い。逆に、極端に少ない項目はなく、どの項目も均等に近い割合で評価に使われていることがわかる。

[クラスタ 5 (内灘町他)]

総合評価 4 (= ややきれい) 及び 5 (= きれい) を支持するデータが 1 ずつだったため、総合評価 1~3 のルールのみを挙げている。ルールを構成する項目の数が少なく、同じ総合評価においても個別項目の評価が分かれていることがわかる。このクラスタは県内で最も汚染が進んでいる湖沼を含むとともに海域にも接しており、住民が最も近い水辺としてどの水域を評価したかが現れていると考えられる。総合評価 2 (= やや汚い) では、「茶色=4 (= 水がやや茶色い)」以外は比較的水辺の環境が良いことを意味する項目でルールが構成されており、このクラスタでは水の色という要素が総合評価を下げる主要因となっていると推測できる。

2.5 まとめ

地域環境問題、水質環境についてのソフトデータから、どのような個別項目が組み合わせられて水質全体が評価されているかを表すルール抽出を行った。社会的特性によって対象地域をクラスタリングし、その地域毎に SC-Optimality アルゴリズムによって最適なルールを抽出した。得られたルールは地域毎にその特性を反映しており、ルール抽出アルゴリズムが有効であることと同時に、環境に対する住民の評価が環境評価を行う上で意味のあるものであることがわかった。

3 数量化 III 類を用いた知識抽出

本節は [15, 16] を一部を抜粋したものであり、詳細はここでは略している。

3.1 はじめに

3.2 あいまいさを考慮した数量化 III 類

住民が観察する対象地点間の評価及び尋ねた質問項目の評価を、BOD 値が存在する、すなわちハードデータを持つ水辺について分析し、次に BOD 値が存在しない水辺も対象として数量化 III 類 (correspondence analysis) によって分析する。ソフトデータである人間の感性評価が対象の特性をどれだけ把握できるのかを、分散を取ることで住民の評価のばらつき具合に注目しながら分析している。また、全ての項目を用いた分析結果をデータ空間上にプロットし、その結果から項目を選択して減らしていくことで、水自体に関わる質問と地点の属性に関わる質問の中でどの個別の評価項目が水辺の全体評価に寄与しているのかを検証している。

数量化 III 類とは多変量解析手法の一つで、目的変数、説明変数という区別を持たないデータに対して、それらの相互の関係を調べることによって類似性やポジショニングを明らかにする、主成分分析と類似した手法である [17]。今回、用いる調査データが 1 から 5 の数字によって評点が付けられていること、説明変数、目的変数の区別なしに個別の要素を等しく扱って分類を行うことから数量化 III 類を適用している。提案する理論の有用性、新規性として、ソフトデータによる分析結果をハードデータと対応させて比較することが可能な点、水辺の快適性ではなく水質そのものをソフトデータから把握しようとしている点が挙げられる。本稿では以下のようにデータを取り扱う。

- 評価対象 $m \in \{1, \dots, M\}$: 評価地点 (水辺空間)
- 評価項目 $n \in \{1, \dots, N\}$: 質問項目
- 評価者 $k \in \{1, \dots, K\}$: 調査回答者
- 評価対象集合 O_k : 評価者 k が評価した対象の集合
- 評価者集合 E_m : 評価対象 m を評価した評価者の集合
- $\delta_{mnk} \in \{1, \dots, 5\}$: 評価者 k による対象 m の評価項目 n に関する評点

各データを評価者 k について平均して以下の δ_{mn} 、 p_{mn} を求める。

$$\delta_{mn} = \frac{1}{|E_m|} \sum_{k \in E_m} \delta_{mnk} \quad (2)$$

$$p_{mn} = \delta_{mn} / \sum_{m=1}^M \sum_{n=1}^N \delta_{mn} \quad (3)$$

但し、 $|E_m|$ は集合 E_m の要素数を表す。これから、 p_{mn} を要素とした相関表を用意する。

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MN} \end{pmatrix} \quad (4)$$

次に、以下の $p_{m\bullet}$ 、 $p_{\bullet n}$ を定義する。

$$p_{m\bullet} = \sum_{n=1}^N p_{mn}, \quad p_{\bullet n} = \sum_{m=1}^M p_{mn} \quad (5)$$

評価対象、評価項目にそれぞれ以下の数量を与える。

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_M)^t \quad (6)$$

$$\boldsymbol{y} = (y_1, y_2, \cdots, y_N)^t \quad (7)$$

但し、 $()^t$ は転置を表す。

数量化 III 類では、 \boldsymbol{x} と \boldsymbol{y} の相関が最大になる数量を求めることを目的としている。 \boldsymbol{x} と \boldsymbol{y} の分散と共分散を以下のように定義する。

$$\sigma_x^2 = \sum_{m=1}^M (p_{m\bullet} x_m^2) - \left(\sum_{m=1}^M p_{m\bullet} x_m \right)^2 \quad (8)$$

$$\sigma_y^2 = \sum_{n=1}^N (p_{\bullet n} y_n^2) - \left(\sum_{n=1}^N p_{\bullet n} y_n \right)^2 \quad (9)$$

$$\sigma_{xy} = \sum_{m=1}^M \sum_{n=1}^N p_{mn} x_m y_n - \sum_{m=1}^M p_{m\bullet} x_m \sum_{n=1}^N p_{\bullet n} y_n \quad (10)$$

これらの式から相関係数 ρ_{xy} を求める。

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (11)$$

以下は数量化 III 類の解法に従って、固有値問題に帰着させていく。式 (11) で \boldsymbol{x} と \boldsymbol{y} の相関を最大にするために次式を解く。

$$\frac{\partial \rho_{xy}}{\partial \boldsymbol{x}} = 0, \quad \frac{\partial \rho_{xy}}{\partial \boldsymbol{y}} = 0 \quad (12)$$

求めた固有値

$$\lambda_1, \cdots, \lambda_M; \lambda_1 = 1, \lambda_2 \geq \cdots \geq \lambda_M$$

のうち、 λ_1 を除いて次に大きい 2 つの固有値 λ_2, λ_3 に対する固有ベクトルから求められる $\tilde{\boldsymbol{x}}_i$ 、 $\tilde{\boldsymbol{y}}_i$ を

$$\tilde{\boldsymbol{x}}_i = \begin{pmatrix} \tilde{x}_{i1} \\ \tilde{x}_{i2} \\ \vdots \\ \tilde{x}_{iM} \end{pmatrix}, \quad \tilde{\boldsymbol{y}}_i = \begin{pmatrix} \tilde{y}_{i1} \\ \tilde{y}_{i2} \\ \vdots \\ \tilde{y}_{iN} \end{pmatrix}, \quad i = 2, 3 \quad (13)$$

とする。

ここで、 $(\tilde{x}_{2m}, \tilde{x}_{3m}), (\tilde{y}_{2n}, \tilde{y}_{3n})$ という座標を考える。各座標の第 2 固有ベクトルの成分である \tilde{x}_{2m} 、 \tilde{y}_{2n} を横軸に、第 3 固有ベクトルの成分である \tilde{x}_{3m} 、 \tilde{y}_{3n} を縦軸に取って 2 次元平面上にプロットし、それぞれの軸の意味を解釈する。

以上が数量化 III 類の手法である。これに加えて、本稿では $\tilde{\boldsymbol{x}}_i$ 、 $\tilde{\boldsymbol{y}}_i$ の分散を求めることで、評価のあいまい性についても考慮する。

評価者 k について i に無関係な以下の数量を定義する。

$$\begin{aligned} \tilde{x}_{mk} &= \tilde{x}_m + \frac{1}{N} \sum_{n=1}^N (\delta_{mnk} - \delta_{mn}), \quad m \in O_k(14) \\ \tilde{y}_{nk} &= \tilde{y}_n + \frac{1}{|O_k|} \sum_{m \in O_k} (\delta_{mnk} - \delta_{mn}), \quad \forall n \quad (15) \end{aligned}$$

これらの分散を次式で定義する。

$$u_m^2 = \frac{1}{|E_m|} \sum_{k \in E_m} (\tilde{x}_{mk} - \tilde{x}_m)^2 \quad (16)$$

$$v_n^2 = \frac{1}{K} \sum_{k=1}^K (\tilde{y}_{nk} - \tilde{y}_n)^2 \quad (17)$$

式 (16)(17) を式 (14)(15) を用いて書き直すと、次の式が得られる。

$$u_m^2 = \frac{1}{|E_m|N^2} \sum_{k \in E_m} \sum_{n, n'=1}^N (\delta_{mnk} - \delta_{mn})(\delta_{mn'k} - \delta_{mn'}) \quad (18)$$

$$v_n^2 = \frac{1}{K} \sum_{k=1}^K \frac{1}{|O_k|^2} \sum_{m, m' \in O_k} (\delta_{mnk} - \delta_{mn})(\delta_{m'nk} - \delta_{m'n}) \quad (19)$$

評価対象及び評価項目に対する評価の相対的あいまい性 s_m 、 t_n を、定数 c_s 、 c_t を用いて次のように定義する。

$$s_m = c_s u_m, \quad t_n = c_t v_n \quad (20)$$

ここで、 s_m 、 t_n が

$$\sum_{m=1}^M s_m^2 = \sum_{n=1}^N t_n^2 = 1 \quad (21)$$

を満たすようにすれば、定数 c_s 、 c_t は以下の式で表される。

$$c_s = \left(\sum_{m=1}^M u_m^2 \right)^{-\frac{1}{2}}, \quad c_t = \left(\sum_{n=1}^N v_n^2 \right)^{-\frac{1}{2}} \quad (22)$$

3.3 2001年住民意識調査

分析したデータは、2001年11月に石川県小松市及び石川郡鶴来町の住民を対象に行ったアンケート調査結果である [18]。回答者数/調査票送付数は148/500であった。全質問が選択回答方式で、回答者属性以外の質問には5段階で評点が付けられている。評点の意味は質問に対して1が「そう思わない」、3が「ふつう」、5が「そう思う」である。

質問内容は、水辺空間2点の水質及びその近隣の環境、回答者の居住地の近隣の環境、回答者属性である。回答者は近隣あるいはなじみのある水辺空間を、1点は石川県が水質の評価規範であるBOD値(生物化学的酸素要求量: Biochemical Oxygen Demand [mg/l])を測定している地点から、もう1点はこれを測定していない地点から選択してもらうものとした。

分析に用いた水辺空間に関する質問を表5に示している。

3.4 水質に関わる対象項目を用いた分析結果

3.4.1 BOD値が測定されている地点

まず、評価対象にBOD値が測定されている8地点(評価者75、データ数75)を、評価項目には表5の中で最も水質の特徴を表しているであろうと推測した質問2~6をそれぞれ選択して分析する。

結果は図2となった。8地点の位置が3つの河川のグループに分かれており、対象項目が分布する位置からも地点の特徴がほぼ正しく捉えられている。

図2において、縦軸はその水辺でのレクリエーション活動のしやすさ、親しみやすさを、横軸は水そのものの澄み具合やきれいさを表していると解釈できる。項目のあいまいさについては、「水が澄んでいる」がやや小さく「水遊びができる」がやや大きい以外は項目間に大きな差はない。

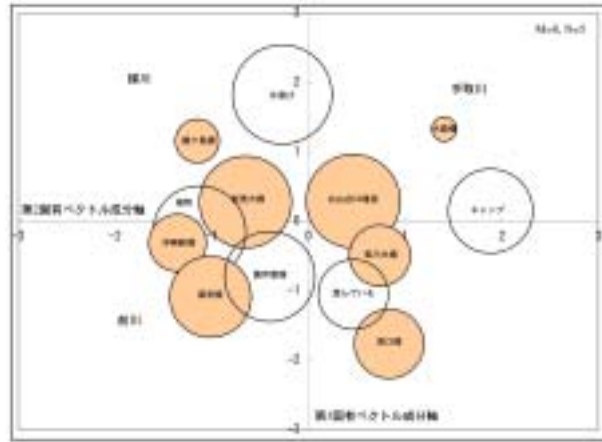


図 2: 8 地点 5 項目の分析結果

3.4.2 BOD 値が測定されていない地点

BOD値を持つ地点をうまく分類できるならば、BOD値を持たない地点についてもソフトデータから特性に従って分類し、BOD値やその水辺空間の性質を推測できる可能性がある。そこで、前節で行った分析の対象をBOD値が測定されていない地点に広げる。地点数は前節で分析した8地点を含めて合計23(評価者125、データ数165)対象項目は図2と同様に表5の質問2~6で、分析結果として図3が得られる。

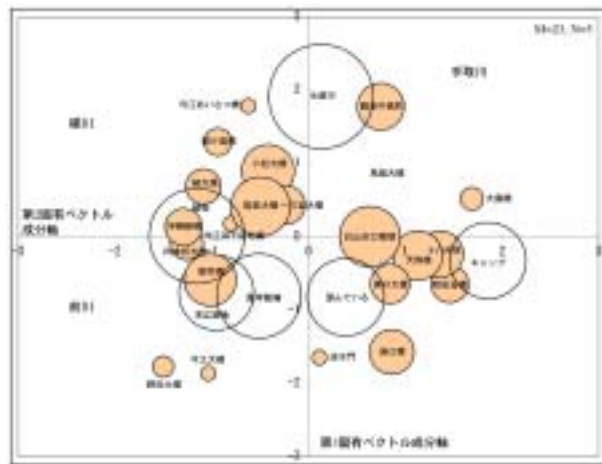


図 3: 23 地点 5 項目の分析結果

図3の縦軸、横軸の意味は、縦軸は水そのものに接触して遊べるかあるいは水辺でのみのレクリエーション活動ができるのかどうかを、横軸は水そのものの澄み具合やきれいさを表していると解釈できる。評価の

表 5: アンケート調査質問一覧 (2001 年)

[水辺空間について]	[水辺空間の近隣について]
1. 鳥や魚などの生き物を見かける	9. 山林がある
2. ヨシや水草などの植物を見かける	10. 田畑がある
3. 水遊び (泳ぐ、ボート、釣りなど) ができる	11. 交通量の多い道路がある
4. 水辺でバーベキューやキャンプができる	12. 街中である
5. 護岸整備がよくなされている	13. 汚濁の原因 (工場や生活排水の流入など) がある
6. 水が澄んでいる	
7. 水の流れる音がよく聞こえる	
8. 水の流れが急である	

あいまいさは、項目に関しては「水遊びができる」以外ほぼ同じである。

項目の分布位置は対象の実状に合っており、分類が対象としているデータによって相対的に変化することを考慮すると、本節で得た結果は良好であり、住民の主観的評価は河川の特徴を捉えられていると言える。

3.5 地点の地域属性を示す対象項目を加えた分析結果

23 地点全てを対象として、表 5 中でまだ使われていない質問、とりわけ水辺の近隣を尋ねる質問を用いることで、水質と近隣の状況の二要素を合わせて地点の特性を踏まえた地点の分類ができるのかどうかを調べている。そのために、評価項目の数を変えながら計算を行って、うまく河川を特徴づけて分類できる項目を見出すことを試みる。このとき、できるだけ必要最小限に近い数の項目を用い、かつそれらが相互に独立した関係に近いこと、項目自身が持つあいまいさは小さいことが望ましい。

対象項目の数を変えて得られた結果のうち、対象項目 10 及び縦軸は横軸を意味するところ、横軸が図 3 同様に水そのものの澄み具合やきれいさ、レクリエーション活動をしたいと感じるかかどうかであると推測できるのに対して、縦軸は「街中である」「田畑がある」といった地点自身が持つ属性が強く出ている。今回対象とした地点は、植物が多い場所は近隣に田畑が多く有り護岸整備がされている、という一件矛盾した項目間に関係が見られることがわかる。実際に数地点を見回ったところ、この印象通りの水辺が多く、ソフトデータからこうした性質が見出せたことは現状をきちんと把握できていると言える。

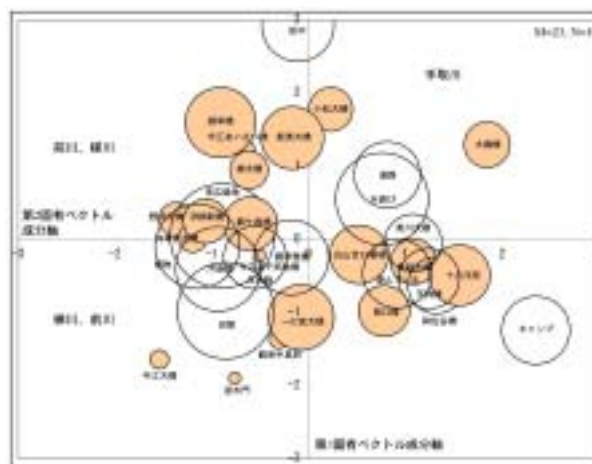


図 4: 23 地点 10 項目の分析結果

図 3 の結果から、水質に直接的に関わる項目だけでも十分な分類をすることはできているが、地域属性としてその地点がどういう場所にあるかを示す項目を入れれば、よりその地点の現状が把握しやすくなると言える。

3.6 まとめ

数量化 III 類を用いて水質・水辺空間についての住民の主観的評価データを、評価のあいまい性にも考慮して分析した。河川の観察結果を検討した結果、得られた結果は似通った性質を持つ地点毎にデータ空間上で分類されており、住民の主観データである人間の感性が対象の実状を客観的データ同様に、あるいは客観的データからわからない水辺の現状もソフトデータによって浮き彫りになることが検証できた。対象の評価

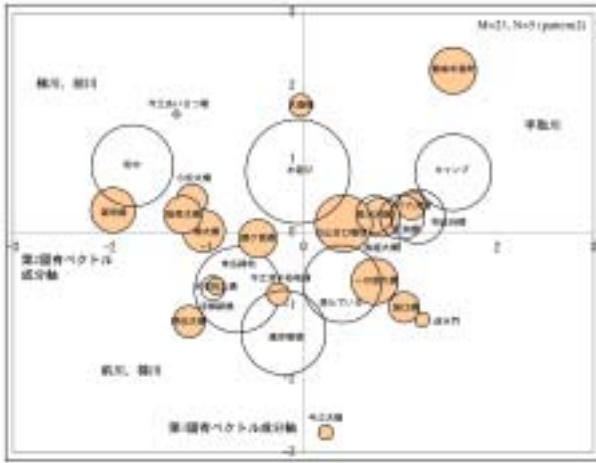


図 5: 23 地点 5 項目 (パターン 2) の分析結果

のあいまいさについては、対象の地形あるいは地理的屬性が評価に影響を与えることがわかった。評価項目のあいまいさについては、水そのものを観察する項目では大きくなるもの、周辺の状況を尋ねたり水に直接触れてレクリエーションができるかどうか尋ねた項目については大きくなる傾向が見られた。

4 まとめと今後の展開

本研究では、住民意識調査の結果とその分析を行うことによって、人間の意識というあいまいな性質を持ったデータの分析手法を提案した。住民データが既存の数値データと合致する、あるいは数値データで得られない知見が住民データから得たことで、主観的性質を持つ感性データが環境評価に対して有効であることを示唆している。

データマイニングの手法を用いた研究からは、社会的性質の似通った地域ごとに水質を判断するのに重要な属性が異なっていること、比較的人や人工物の多い都市部においては「汚濁源がある」「水が茶色い」「キャンプができる」という属性が重視されていることがわかった。数値化Ⅲ類を用いた研究では、研が水質を測定している地点で住民データがうまく分類されていることから、水質を測定していない地点に対しても分析を行い、住民データが実状を表していること、地点の特性を表している項目を抽出した。

今後は、こうしたソフトデータからの知見を踏まえて、人間活動と環境の関係を示すこと、どのように人間が活動すれば環境状態を改善していけるのかを提案

することが課題となる。例えば水質の場合、各種統計データと水質の関係を比較すると、下水道普及率、人口の伸びが水質と関連があるだろうと推測できる。大気質の場合は、自動車普及率、自動車通行台数が鍵となっている。より詳しくこれらの関係を観察し、どの程度の社会基盤の整備が水質や大気質、さらには廃棄物問題の解決に必要なのかを提案すること、あるいは提案に繋げる基盤を構築することが研究の最終目的である。

謝辞

公益信託エスベック地球環境研究・技術基金平成 13 年度助成金受けた 1 年間 (2001 年 8 月から 2002 年 7 月) に発表した論文は、本文中に引用した [13, 15, 16] の他、[19, 20, 21, 22, 23] (投稿中のものを含む) がある。2001 年のアンケート調査の実施及び分析、そして研究成果の発表を行うに当たって、助成金は本当に役に立ちました。エスベック及び基金の関係者の皆さまに心より御礼申し上げます。ありがとうございます。

なお、本研究を進める上で重要なデータとなった住民意識調査に協力して下さった石川県各市町村の住民の皆さまにも、ここに謝意を表します。

参考文献

- [1] Y.Matsuoka, M.Kainuma. and T.Morita., Scenario Analysis of Global Warming Using the Asian Pacific Integrated Model (AIM), Energy Policy, Vol.23, No.4/5, pp.357-371, 1995.
- [2] W.D.Nordhaus, The DICE Model, Cowles Foundation Discussion Paper, Yale Univ, 1992.
- [3] 河野小夜子・領家美奈・中森義輝, 総合モデルを用いた地域環境問題のソフト分析, 第 29 回環境システム研究論文発表会 講演集, pp.311-316 (2001) .
- [4] 森田恒幸・野田清敏・堀内葉子: 都市住民の意識に基づく環境指標の算定 - 北九州市の環境管理を対象にして - , 日本都市計画学会学術研究論文集, No.20 , pp.133-138 , 1985
- [5] 森田恒幸・西岡秀三・原科幸彦・内藤正明: 住民意識に基づく環境指標の総合化/新たな重みづけ

- 手法の提案, 昭和 61 年度日本不動産学会学術講演会梗概集, pp.125-128, 1986
- [6] 原科幸彦・東林知隆: 環境観察調査に基づく快適環境指標の試作 - 川崎市におけるケーススタディ, 第 2 回環境研究発表会論文集, pp.17-18, 1988
- [7] 日本計画行政学会(編): 「環境指標」の展開, 学陽書房, 1995
- [8] 須賀伸介・大井紘・宮本定明・阿部治・内藤正明・原沢英夫: 生活環境に関する自由記述式アンケート調査データの解析 - 水辺に関する住民の認知, 第 1 回環境研究発表会論文集, Vol.17, No.1, pp.38-43, 1988
- [9] 須賀伸介・大井紘: 自由連想法による東京湾についての意識調査とその解析, 環境システム研究, Vol.22, pp.31-41, 1994
- [10] J.Brian: Mining Air Pollution Data, *3rd ERCIM Environmental Modeling Group Workshop on Air Pollution Modelling and Measuring Processes* (1998).
- [11] S. Kawano, M. Shakato, M. Ryoike and Y. Nakamori, Soft Data Analysis for Evaluating Regional Environment; *Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pp.116-121 (2001).
- [12] 石川貴士他: 小学校校歌にみる福岡の環境イメージ, 環境システム研究, Vol.17, pp.257-263, 1993
- [13] 河野小夜子・Huynh Van Nam・領家美奈・中森義輝, ソフトデータに基づく環境評価モデル, システム制御情報学会論文誌, Vol.15, No.12, 2002 (印刷中) .
- [14] R.J.Bayardo, R.Agrawal and D.Gunopulos, Constraint-Based Rule Mining in Large, Dense Databases; *Data Mining and Knowledge Discovery*, Vol.4 (2/3), pp.217-240 (2000).
- [15] 河野小夜子・中森義輝, 数量化 III 類を用いた主観評価データの分析, 第 12 回ソフトサイエンス・ワークショップ講演論文集, pp.127-130, 2002 .
- [16] 河野小夜子・中森義輝, 評価のあいまい性を考慮した感性データの分析, 日本感性工学会誌 感性工学研究論文集, 2002 (投稿中) .
- [17] 林知己夫: 数量化の方法, 東洋経済新報社, 1974 .
- [18] 武田雄一・河野小夜子・Van Nam Huynh・中森義輝: コンテキストモデルを用いた住民意識と環境問題の關係の抽出, 第 12 回ソフトサイエンス・ワークショップ講演論文集, pp.123-126, 2002 .
- [19] 河野小夜子・領家美奈・中森義輝, 総合モデルを用いた地域環境問題のソフト分析, 第 29 回環境システム研究論文発表会講演集, pp.311-316, 2001 .
- [20] Sayoko Kawano, Van Nam Huynh, Mina Ryoike and Yoshiteru Nakamori, A Context-Dependent Knowledge Model for of Regional Environment, *Environmental Modeling and Software*, 2001 (submitted).
- [21] Sayoko Kawano, Yuichi Takeda, Van Nam Huynh and Yoshiteru Nakamori, Incorporating Hard and Soft Data Into a Model for Evaluation of Local Environment, *Proceedings of the IASTED International Conference Modelling and Simulation*, USA, pp.202-207, 2002.
- [22] Sayoko Kawano, Van Nam Huynh, Mina Ryoike and Yoshiteru Nakamori, A Context-Dependent Knowledge Model for Evaluation of Regional Environment, *IFIP Workshop Group 7.6 Workshop on Virtual Environments for Advanced Modeling*, Japan, pp.34-35, 2002.
- [23] Sayoko Kawano, Mina Ryoike and Yoshiteru Nakamori, Evaluation of Local Environments by Soft Data, *IFAC Workshop on Modeling and Control in Environmental Issues (Preprints)*, Japan, pp.63-66, 2001.

[問い合わせ先 (2003 年 3 月まで)]

〒 923-1292 石川県能美郡辰口町旭台 1-1
 北陸先端科学技術大学院大学 知識科学研究科
 河野 小夜子 (skawano@jaist.ac.jp)
 2003 年 4 月以降は sayoko_k@gem.hi-ho.ne.jp へ