主観的な情報空間における知的探索のためのデータ操作

Data operations for intellectual exploration on an information space consisting of subjective relations

林 幸雄 (北陸先端科学技術大学院大学)
Yukio HAYASHI (Japan Advanced Institute of Science and Technology, Hokuriku)

Abstract

We have proposed a new data model which dynamically generates a subjective relational structure among information flagments from a viewpoint. In this paper, to get an inspiration from the generated relational structures, we introduce two data operations. Through the operations, a user is able to select useful information flagments from interpolative viewpoints or discovery viewpoints on a range of view in his/her vague exploratory intention.

1 はじめに

一般に研究や企画立案などの非定形業務においては, 曖昧な意図から徐々に切り口 (情報探索の観点や視点) を絞り込んで, 大まかな目標に関連する情報を探すことが多い. その一方で, 奇抜なアイデアなどを触発しそうな意外な情報にめぐり会うことを秘かに期待する. 例えば, 既存の解決方法ではうまくいきそうにない場合や, 何か新しい要素が望まれる場合では, 後者のような発見的な探索も要求される. 本論文では, 曖昧な意図と意外性に関する情報探索を行なう手段として, 関係分布モデル [3][4] における 2 つのデータ操作を提案する.

2 知的探索のためのデータ操作

関係分布モデルは, 利用者の視点に応じて動的に情報の関係構造を生成するデータモデルである. このモデルでは, 情報間の直接の関係構造 (密な体系) を考えずに, 視点と情報との間の関連性の強弱 (関連度) のみで関係構造を規定して, 構造化の労力を大幅に低減している. すなわち, 構造破壊と再構造化を柔軟に行なうために, 緩やかで疎な関係のみを規定すれば良い仕組みを考えている.

さて, 関係分布モデルの最も基本的なデータ操作は, 予めモデルに規定された n 個の各視点 i に対する情報 x の関連度 $p_i(x)$ を, 利用者がその都度指定できる視点の重要度パラメータθの値に従って重み付けて組合せて, 新たな関連度 $p(x;\ \theta)$ を生成することである. この操作で生成された, θで重み付けられた複合的な視点に対する関連度 $p(x;\ \theta)$ の値の大きさに従って, 選択的に情報探索ができる.

ここで, m 個の情報に対する関係分布モデルは, $\mathcal{X} = \{1, 2, \ldots, m\}$ 上の指数型分布族 $S = \{p(x;\ \theta)\}$,

$$p(x;\ \theta) = \exp\left[C(x) + \sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)\right], \quad (1)$$

などによって具現化できる $(n < m - 1)$. 但し, $n + 1$ 個の $p_0(x)\ \sim\ p_n(x)$ はそれぞれ異なる確率分布で, $C(x) = \log p_0(x)$, $F_i(x) = \log p_i(x) - \log p_0(x)$, $\psi(\theta) = \log\{\sum_x \exp[C(x) + \sum_i \theta^i F_i(x)]\}$とする. より一般的には, 情報幾何学の$\alpha$-分布族 [1] で具現化でき , それは多様体の接続パラメータ$\alpha \in R\ (\alpha \neq 1)$ の値に従った min-max の中間のファジイ平均化作用 [2] を持つことが知られている [3]. 特に , $\alpha \to 1$ の場合が指数型で , 幾何平均に相当する.

視点の重要度パラメータθの値の連続的な変化は, 多様体 $S = \{p(x;\ \theta)\}$ 上の, ある 1 つの経路を特定する. そこで, このような経路上で利用者の意図に応じて情報探索を行なうような, 関係分布モデルのデータ操作として以下の 2 つを考える.

視点の揺さぶり操作: 利用者の (曖昧な) 意図の範囲を表すパラメータ空間の 2 点$\theta(p)$ と$\theta(q)$ を直線

補間する各点$\theta(s)$ の経路上の操作[1]. 通常は, 高々n'個 $(n' = |I| < n)$ の視点のみを非零の重要度の値$\theta^k(p) \neq 0$, $\theta^{k'}(q) \neq 0$ $(k, k' \in I)$ に選ぶ.

意外性に関する操作: 利用者が当初気付いてない視点 (重視して指示してない視点) の重みを, 先の操作とは別の経路上の各点$\theta(r)$ によって, モデルが自律的に大きい値に設定することで, 思考のはまり込みに陥らずにアイデアを触発する為の操作.

具体的には, それぞれのデータ操作は, 指数型分布族の多様体 S上の e-測地線上の$\theta(s)$, 及び, S上の m-測地線を$\theta^{k''}(k'' \notin I)$ で張られる Sの部分多様体上へ射影した経路上の$\theta(r)$ によって規定する[2].

同様に , α-分布族による関係分布モデルでは, 双対な$\pm\alpha$-測地線で両操作を規定できるが , Sの em-測地線は次のような点から都合が良い . まず , e-測地線は $[\theta^i]$ 空間の直線に相当した自然な経路である . 一方 , m-測地線は双対な $[\eta_i]$ 空間の直線に相当するが , これは関連度 $p_i(x)$ の対数 $(F_i(x)$ が感覚量) の期待値 $\eta_i = \sum_x F_i(x)p(x;\ \theta)$ を表す . さらに , n 次元多様体 Sと (有限な全ての確率分布からなる) $m-1$ 次元の離散分布族の em-測地線の自己平行性や , m-測地線上における $p(x;\ \theta)$ の線形性[3] などの理論的性質, 線形性を利用した測地線の効率的な計算法などの長所を持つ [5]. 紙面の都合上, この線形性を利用したデータ操作のアルゴリズムに関しては, 別の機会に報告したい.

3 実験例

文献 [4] と同様に, 地球環境問題を題材とした $m = 125$ 個の情報 $x \in \mathcal{X}$に対して, $n = 9$ 個の各視点 i から関連度 $p_i(x)$ を規定した. 具体的には, 9 つの視点 : 1. 身近さ, 2. 対策への関心, 3. 未来への関心, 4. 温暖化, 5. 酸性雨, 6. オゾンホール, 7. 食糧危機, 8. エネルギー問題, 9. 化学汚染, のそれぞれについて, 125 個の各情報内容の関連性の強弱 (感覚量 $\log p_i(x)$) を 5 段階で評価して, 関連度を定めた.

さて, 1 つの例として, 利用者の情報探索の意図が, 環境問題への何らかの対策を探すことと, 食糧危機の問題の深刻さを知ることにあったとする. 具体的には, 両者に関わる視点 2 と視点 7 を重視するよう, パラメータ空間上の範囲を表す 2 点 PQ として, 視点 2 軸上の点: $\theta^2(p) = 3$, $\theta^k(p) = 0$ $(k \neq 2)$ と視点 7 軸上の点: $\theta^7(q) = 3$, $\theta^k(q) = 0$ $(k \neq 7)$ を選んだとする. 以下に 2 つのデータ操作の結果 (関連度 $p(x;\ \theta)$ に従った順位表やパラメータθの値の変化) を示す.

図 1は, 視点の揺さぶり操作における順位表のベスト 10 の情報の入れ替わりの時間的変化を示すもので, 視点 2:対策案から視点 7:食糧危機に関するものへとランキングが大きく変わっている. 一方, 図 2は, 意外性に関する操作におけるベスト 10 の変化を示すもので, 前者よりも変動が小さい. ここで , βの値はパラメータ空間の 2 点$\theta(p)$ と$\theta(q)$ を結ぶ直線 (e-測地線に相当) を 5 等分した際の分割点の番号を示す .

次に, 両操作におけるこの傾向を, 順位表の時間的変遷として表 2と表 3 にそれぞれ示す (表 1と表 4は端点 PQ の場合). ここで, 意外性に関する操作における表 3では, 視点 1: 身近さ, 及び, 視点 8: エネルギー問題に関する情報が最も多くランキングされているのが特徴的である.

意外性に関する操作から利用者が「新しいメッセージ」を得る機会を想定して, 図 4のθ^kのグラフが極大値をとる$\beta = 2$ の時 (利用者が指示した視点から, もっとも異なる視点をモデルが重視した場合) の結果を見てみよう. この時の順位表 (表 3の上から 2 番目) では, 酸性雨による被害や, エネルギー問題に関連する情報が主にランキングされている. このことは, 地球環境問題への直接的な対策と食糧危機の問題しか脳裏になかった利用者にとっては意外な, 酸性雨による植物への生育被害, 発電や自動車による二酸化炭素の増加が招く生態系の破壊, 道路建設による農地や牧草地の破壊などが, 間接的に食糧危機に結び付くこと等を気付かせるものと考えられる.

4 おわりに

複合的な視点に応じて動的に関係構造を生成するデータベースモデルを考え, 利用者の曖昧な意図を補間したり, 意外な発見を促したりするための情報探索法をそれぞれ提案した. また, 実験例を通じて応用上有効なものを実現し得ることを示した .

本研究の一部は, (財) 人工知能研究振興財団, 及び, 文部省科学研究費 10878045 の援助を受けている.

[1] パラメータ空間の各点 P, Q, R, S に対応する座標ベクトルをそれぞれ, $\theta(p)$, $\theta(q)$, $\theta(r)$, $\theta(s)$ で表すものとする.

[2] 初めから意図した視点 $k \in I$の方が支配的にならないように, 意図してなかった視点のみで張られる部分多様体上に制限して探索を行なう (つまり, $\theta^k(r) = 0, k \in I$とする). これは m-測地線の部分多様体への射影なので, 幾何学な意味を失ってはいない.

[3] これは, S上の m-測地線が確率分布の関数空間においては直線に相当することを意味する.

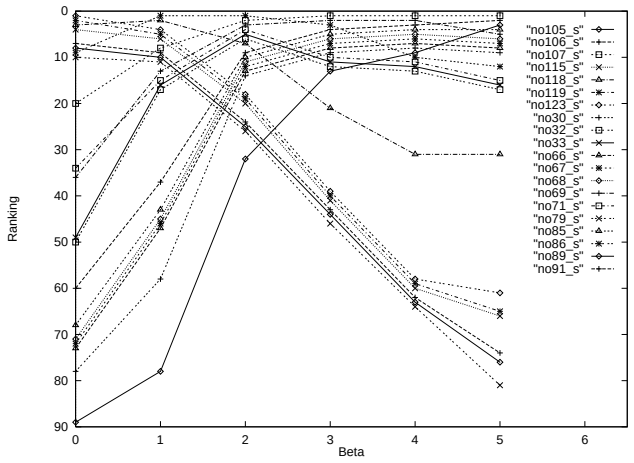

図 1: 関連度 $p(x;\ \theta(s))$ による順位の変遷

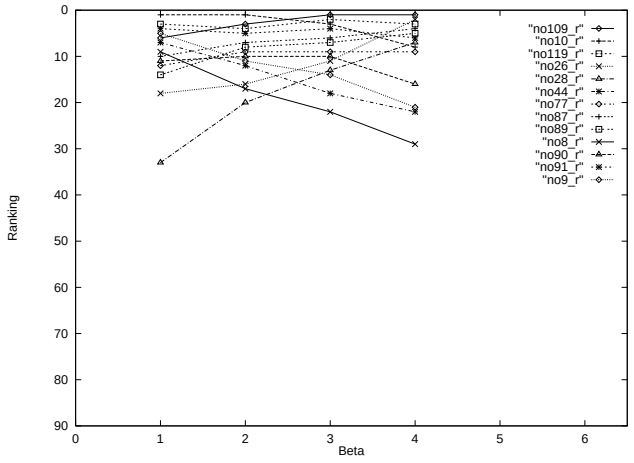

図 2: 関連度 $p(x;\ \theta(r))$ による順位の変遷

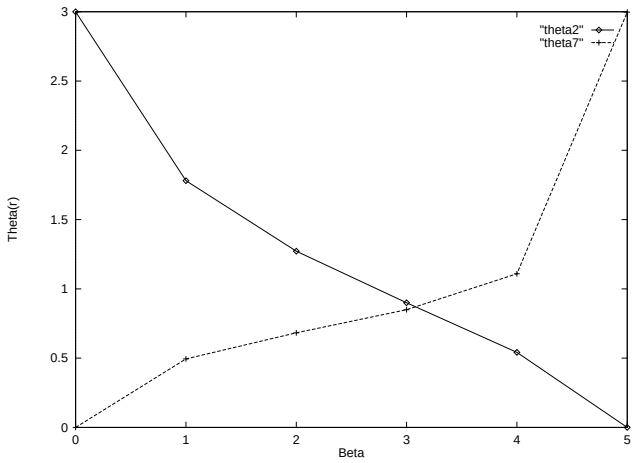

図 3: 意外性に関する操作前の m-測地線上のθ^2とθ^7

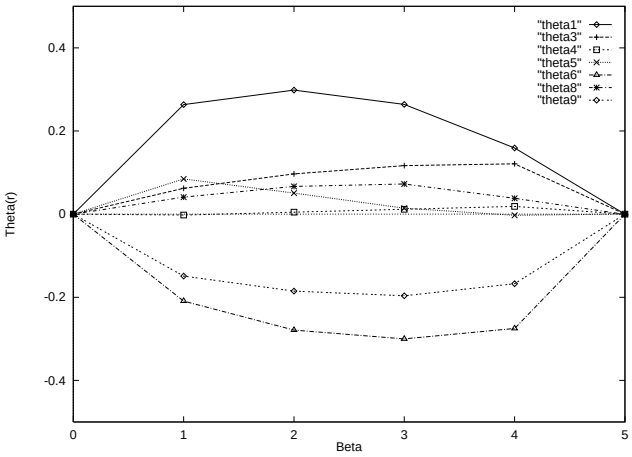

図 4: 意外性に関する操作による視点の重要度θ^k, $(k \neq 2, 7)$

参考文献

[1] 甘利 俊一, 長岡 浩司. 情報幾何の方法. 岩波講座 応用数学 [対象 12], 岩波書店, 1994.

[2] M. Mizumoto. Pictorial Representations of Fuzzy Connectives, Part I: Cases of t-norms, t-conorms and Averaging Operators. *Fuzzy Sets and Systems,* Vol. 31, pp. 217-242, 1989.

[3] 林 幸雄. 個々人の嗜好を重視した魅力的な情報空間の構築に向けて-情報幾何学と平均演算子-. 第 *2* 回 曖昧な気持ちに挑むワークショップ, pp. 69-72, 1997.

[4] 林 幸雄. データベースの数理モデルの視点パラメータの推定と関数空間におけるモデル拡張. 情報処理学会論文誌, Vol. 39, No. 6, pp. 1-11, 1998.

[5] Yukio Hayashi. Direct Calculation Methods for Parameter Estimation in Statistical Manifolds of Finite Discrete Distributions. *IEICE Trans. on Fundamental of Elect., Comm. and Comp. Sci.*, Vol. E81-A, No. 7, pp. 1486-1492, 1998.

[問合せ先]
林 幸雄
〒 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
Tel: 0761-51-1736, Fax: 0761-51-1149
E-mail: yhayashi@jaist.ac.jp
http://www.jaist.ac.jp/~yhayashi/

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	ゴミ処理の有料化	123	0.070080	5	5	2	1	1	1	1	1	2
2	紙を節約する方法	119	0.070080	5	5	3	1	1	1	1	5	1
3	家庭からの生活排水の汚れを減らす	118	0.070080	5	5	3	1	1	1	2	1	5
4	地球にやさしい移動の方法	115	0.070080	5	5	3	4	2	1	1	5	2
5	地球にやさしい発電	110	0.070080	3	5	4	4	3	1	1	5	2
6	身近な省エネによる二酸化炭素放出量の削減	109	0.070080	5	5	3	5	1	1	2	5	1
7	交通の隠れたコストを考える	91	0.070080	5	5	4	2	2	1	1	5	2
8	分散型発電技術がひらく可能性	89	0.070080	4	5	4	2	1	1	1	5	1
9	海を統治する法律	86	0.070080	3	5	4	1	1	1	4	3	4
10	紙消費量の変化と再利用の促進	79	0.070080	5	5	3	3	2	1	1	5	5

表 1: 関連度 $p(x;\theta(p))$ による情報 x のベスト 10 $(\beta = 0)$

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	海を統治する法律	86	0.213462	3	5	4	1	1	1	4	3	4
2	家庭からの生活排水の汚れを減らす	118	0.064293	5	5	3	1	1	1	2	1	5
3	身近な省エネによる二酸化炭素放出量の削減	109	0.064293	5	5	3	5	1	1	2	5	1
4	ゴミ処理の有料化	123	0.035285	5	5	2	1	1	1	1	1	2
5	紙を節約する方法	119	0.035285	5	5	3	1	1	1	1	5	1
6	地球にやさしい移動の方法	115	0.035285	5	5	3	4	2	1	1	5	2
7	地球にやさしい発電	110	0.035285	3	5	4	4	3	1	1	5	2
8	地球の人口を安定させる	107	0.035285	4	4	4	1	1	1	5	1	1
9	交通の隠れたコストを考える	91	0.035285	5	5	4	2	2	1	1	5	2
10	分散型発電技術がひらく可能性	89	0.035285	4	5	4	2	1	1	1	5	1

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	海を統治する法律	86	0.239031	3	5	4	1	1	1	4	3	4
2	地球の人口を安定させる	107	0.131183	4	4	4	1	1	1	5	1	1
3	食糧生産の維持/増大への技術のあり方の転換	69	0.131183	3	4	2	2	2	1	5	1	1
4	環境にとっての貿易の役割	71	0.039512	3	4	3	1	1	1	4	3	1
5	期待を裏切った技術進歩の例	33	0.039512	3	4	2	3	3	5	4	4	3
6	環境変化を克服する楽観的な可能性	32	0.039512	3	4	5	5	2	1	4	4	4
7	家庭からの生活排水の汚れを減らす	118	0.021684	5	5	3	1	1	1	2	1	5
8	身近な省エネによる二酸化炭素放出量の削減	109	0.021684	5	5	3	5	1	1	2	5	1
9	食糧消費を抑える	106	0.021684	4	3	3	1	1	1	5	1	1
10	乱獲される生物と減衰退する漁業	85	0.021684	4	3	3	1	1	1	5	1	1

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	地球の人口を安定させる	107	0.184170	4	4	4	1	1	1	5	1	1
2	食糧生産の維持/増大への技術のあり方の転換	69	0.184170	3	4	2	2	2	1	5	1	1
3	海を統治する法律	86	0.101074	3	5	4	1	1	1	4	3	4
4	食糧消費を抑える	106	0.055471	4	3	3	1	1	1	5	1	1
5	乱獲される生物と減衰退する漁業	85	0.055471	4	3	3	1	1	1	5	1	1
6	人間は陸上の純 1 次生産力 NPP の 40%近くを独占	68	0.055471	3	3	4	3	3	1	5	3	1
7	地球を圧迫する 3 つの要因	67	0.055471	3	3	3	3	3	1	5	4	2
8	地球の収容能力の限界	66	0.055471	2	3	4	3	2	3	5	2	1
9	温暖化に対処する為の農薬による 2 次的被害	30	0.055471	4	3	4	5	1	1	5	1	5
10	環境にとっての貿易の役割	71	0.030443	3	4	3	1	1	1	4	3	1

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	地球の人口を安定させる	107	0.157317	4	4	4	1	1	1	5	1	1
2	食糧生産の維持/増大への技術のあり方の転換	69	0.157317	3	4	2	2	2	1	5	1	1
3	食糧消費を抑える	106	0.086337	4	3	3	1	1	1	5	1	1
4	乱獲される生物と減衰退する漁業	85	0.086337	4	3	3	1	1	1	5	1	1
5	人間は陸上の純 1 次生産力 NPP の 40%近くを独占	68	0.086337	3	3	4	3	3	1	5	3	1
6	地球を圧迫する 3 つの要因	67	0.086337	3	3	3	3	3	1	5	4	2
7	地球の収容能力の限界	66	0.086337	2	3	4	3	2	3	5	2	1
8	温暖化に対処する為の農薬による 2 次的被害	30	0.086337	4	3	4	5	1	1	5	1	5
9	食糧と経済の動向	105	0.047383	3	2	5	1	1	1	5	1	1
10	海を統治する法律	86	0.026004	3	5	4	1	1	1	4	3	4

表 2: 関連度 $p(x;\theta(s))$ による情報 x のベスト 10 (上から$\beta = 1, 2, 3, 4$)

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	酸性雨の脇役は硝酸	10	0.022265	5	3	3	1	5	1	1	4	1
2	日本における酸性雨の被害	114	0.021372	5	3	3	1	5	1	3	3	1
3	紙を節約する方法	119	0.016528	5	5	3	1	1	1	1	5	1
4	交通の隠れたコストを考える	91	0.016460	5	5	4	2	2	1	1	5	2
5	煙突から酸性の降雨	9	0.016420	4	2	3	1	5	1	1	3	1
6	身近な省エネによる二酸化炭素放出量の削減	109	0.016391	5	5	3	5	1	1	2	5	1
7	燃やせないゴミの代表格プラスチック系の素材 5	44	0.015522	5	4	3	3	4	1	1	1	2
8	地球にやさしい移動の方法	115	0.015402	5	5	3	4	2	1	1	5	2
9	雨水中の酸から湖の荒廃	8	0.015129	4	3	3	1	5	1	1	1	1
10	巨大化した電力産業の改革の必要性	87	0.014648	4	4	4	3	2	1	1	5	1

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	酸性雨の脇役は硝酸	10	0.022937	5	3	3	1	5	1	1	4	1
2	日本における酸性雨の被害	114	0.021457	5	3	3	1	5	1	3	3	1
3	身近な省エネによる二酸化炭素放出量の削減	109	0.020408	5	5	3	5	1	1	2	5	1
4	紙を節約する方法	119	0.020015	5	5	3	1	1	1	1	5	1
5	交通の隠れたコストを考える	91	0.019370	5	5	4	2	2	1	1	5	2
6	地球にやさしい移動の方法	115	0.017755	5	5	3	4	2	1	1	5	2
7	巨大化した電力産業の改革の必要性	87	0.017379	4	4	4	3	2	1	1	5	1
8	分散型発電技術がひらく可能性	89	0.016438	4	5	4	2	1	1	1	5	1
9	エコロジーを計算に入れた価格付け	77	0.016100	4	5	4	1	2	1	1	4	1
10	何のための交通か	90	0.016097	5	4	4	2	2	1	1	5	3

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	身近な省エネによる二酸化炭素放出量の削減	109	0.021320	5	5	3	5	1	1	2	5	1
2	紙を節約する方法	119	0.020341	5	5	3	1	1	1	1	5	1
3	酸性雨の脇役は硝酸	10	0.020026	5	3	3	1	5	1	1	4	1
4	交通の隠れたコストを考える	91	0.019275	5	5	4	2	2	1	1	5	2
5	日本における酸性雨の被害	114	0.018627	5	3	3	1	5	1	3	3	1
6	巨大化した電力産業の改革の必要性	87	0.018232	4	4	4	3	2	1	1	5	1
7	分散型発電技術がひらく可能性	89	0.017765	4	5	4	2	1	1	1	5	1
8	地球にやさしい移動の方法	115	0.017560	5	5	3	4	2	1	1	5	2
9	エコロジーを計算に入れた価格付け	77	0.016564	4	5	4	1	2	1	1	4	1
10	何のための交通か	90	0.015837	5	4	4	2	2	1	1	5	3

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	身近な省エネによる二酸化炭素放出量の削減	109	0.017095	5	5	3	5	1	1	2	5	1
2	海に沈む土地	26	0.015948	4	3	5	5	1	1	4	1	1
3	紙を節約する方法	119	0.015871	5	5	3	1	1	1	1	5	1
4	巨大化した電力産業の改革の必要性	87	0.015810	4	4	4	3	2	1	1	5	1
5	分散型発電技術がひらく可能性	89	0.015563	4	5	4	2	1	1	1	5	1
6	交通の隠れたコストを考える	91	0.015394	5	5	4	2	2	1	1	5	2
7	二酸化炭素の増加予測	28	0.015247	3	3	5	5	1	1	1	4	1
8	酸性雨の脇役は硝酸	10	0.015108	5	3	3	1	5	1	1	4	1
9	エコロジーを計算に入れた価格付け	77	0.014664	4	5	4	1	2	1	1	4	1
10	日本における酸性雨の被害	114	0.014543	5	3	3	1	5	1	3	3	1

表 3: 関連度 $p(x;\theta(r))$ による情報 x のベスト 10 (上から$\beta = 1, 2, 3, 4$)

順位	タイトル	情報 x	関連度 $p(x;\ \theta)$	視点 i に対する関連性の 5 段階の評価値 1	2	3	4	5	6	7	8	9
1	地球の人口を安定させる	107	0.104943	4	4	4	1	1	1	5	1	1
2	食糧消費を抑える	106	0.104943	4	3	3	1	1	1	5	1	1
3	食糧と経済の動向	105	0.104943	3	2	5	1	1	1	5	1	1
4	乱獲される生物と減衰退する漁業	85	0.104943	4	3	3	1	1	1	5	1	1
5	食糧生産の維持/増大への技術のあり方の転換	69	0.104943	3	4	2	2	2	1	5	1	1
6	人間は陸上の純 1 次生産力 NPP の 40%近くを独占	68	0.104943	3	3	4	3	3	1	5	3	1
7	地球を圧迫する 3 つの要因	67	0.104943	3	3	3	3	3	1	5	4	2
8	地球の収容能力の限界	66	0.104943	2	3	4	3	2	3	5	2	1
9	温暖化に対処する為の農薬による 2 次的被害	30	0.104943	4	3	4	5	1	1	5	1	5
10	オゾン層破壊による被害	112	0.005225	5	3	4	1	1	5	4	1	1

表 4: 関連度 $p(x;\theta(q))$ による情報 x のベスト 10 ($\beta = 5$)