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

北陸先端科学技術大学院大学 知識科学研究科
林 幸雄 (yhayashi@jaist.ac.jp)

Contents

1. 主観的な情報とアクセス困難性
2. 動的な構造化に対する課題
3. 関係分布モデルの提案 (ファジィ論理との接点)
4. 視点の揺さぶりと意外性に関するデータ操作
5. 実験例における有効性
6. まとめ

==========

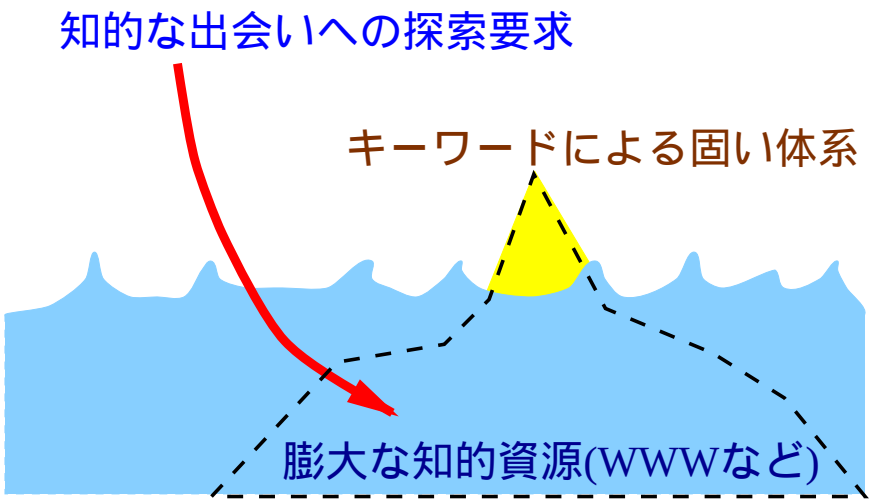
—

1.1. 魅力的な情報は ？

客観的なもの (ロゴス)	主観的なもの (パトス)
歴史的な事実	評価が未定なトピックス
辞書	小説, 記事
法律, 規則	経験などに基づく諺
観測データの羅列	さまざまな意見の集約
商業メディアを介した公式情報	口コミ情報
予め固定化された体系, スキーマ	視点に応じた関係構造
演繹や論理による関係	洞察や直観による関係
受動的な意味伝達	対話的な意味解釈

⇒ (他人の) さまざまな視点から吟味された情報をもとに, 自らの目的や意図に合った活用を見い出せること, すなわち, 現実世界に活かせる (生きた) 意味づけ/関係づけが要求される.

1.2. 情報過多によるアクセス困難性



==========

—

1.3. 曖昧な探索要求に対する技術動向

Dymanic Structure: 動的な関係付けの必要性 [Halasz88,91]

Casual User: アイデア着想などへの曖昧な意図 [Loeb92] や探索の要求 [Belkin92]

Social, Collaborative Filtering: 信頼できる人からの情報や解釈 [Malone87] を, 互いの得意対象で補間的に活用 [Goldberg92]

Long Term Goal: ユーザとの対話 (user profile [Belkin92]) により, 時間的に変化する興味や関心をシステムに反映

Customization: 個人に合った情報群の構造化/表示 [Bowman94][Loeb92]

Recommendation: インターネットにおける人的資源や推奨の活用 [Hill95][Shardanand95]

—

2. 動的な構造化に対する課題

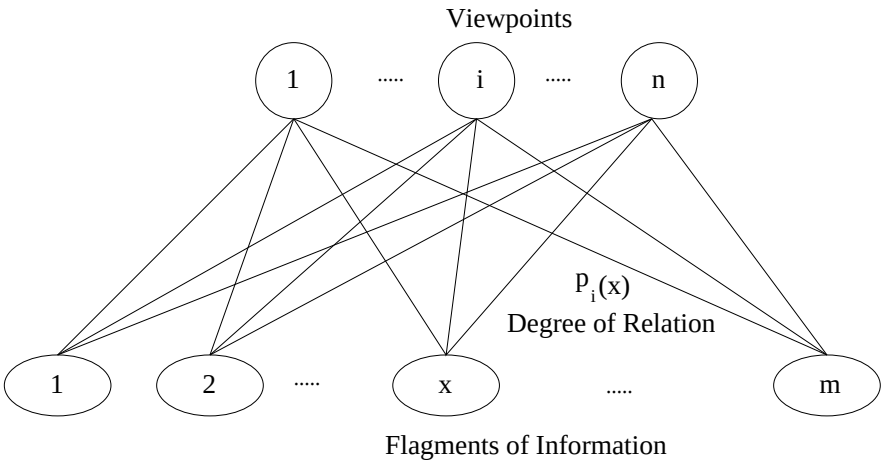
現状の技術手段では未解決 ：

1. 情報構造や知識体系の表現をどう定義するのか ？
2. 矛盾のない意味付けを求めると, 構造化/体系化のための労力が膨大になる.
3. さまざまな状況や目的に依存した関係をどう扱うのか ？

⇒ 個々の情報間を関係付けた厳密な構造化/体系化をあきらめて, 緩やかな関係構造の全体をパラメトリックに扱う方法論 (情報幾何学) に着目.

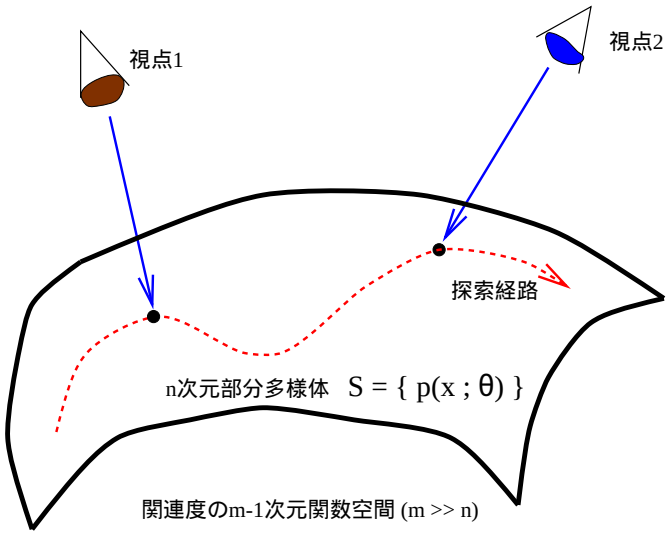
==========

3.1. 動的に情報空間を構築する関係分布モデル



⇒ 視点の重要度パラメータθで張られる関連度の関数空間 $S = \{p(x;\ \theta)\}$ 上で, 任意の視点に応じて関連度 (= 関係構造) を動的に生成.

3.2 視点に応じた関連度 $p(x;\ \theta)$ の生成



データ操作　:関連性の強弱が張る関数空間　(多様体) 上を探索

⇒ その関連度によるランキング表などから情報を選択.

3.3. 関係分布モデルの利点 (緩やかな関係構造)

- 利用者の質問に応じた動的で柔軟な構造化 (リンク時代の終焉 [Halasz88,91] に対処) ができる.
- 人間の主観に基づく曖昧な処理が扱える.
- 統計理論などに基づく理論解析の可能性がある.
- 構造を規定するための労力を大幅に削減できる (情報間の密な結合ではなく, 視点と情報間の疎な結合).
- 視点からの比較的明示的な全体構造/関係が把握できる.
- 情報間の関連性や解釈の自由度が大きい.

⇒ 構造破壊と再構造化を積極的に行なっている.

==========

3.4. ファジィ論理との接点

α-分布族 $S_\alpha = \{p_\alpha(x;\ \theta)\}$ よる関係分布モデルの具現化

$$p_\alpha(x;\ \theta) \stackrel{\text{def}}{=} \left[\frac{1-\alpha}{2}\sum_{i=1}^{n+1}\theta^i F_i(x)\right]^{\frac{2}{1-\alpha}},\ \ (\alpha \neq 1)$$

$\mathcal{X} = \{1, 2, \ldots, m\}$ 上の確率分布 $p_i(x) > 0,\ F_i(x) \stackrel{\text{def}}{=} [p_i(x)]^{\frac{1-\alpha}{2}}$.

$\Downarrow$ α-アファイン多様体$\tilde{S}$に拡張

$\tilde{S} = \{\tilde{p}(x;\ \xi) |\ \tilde{p}(x;\ \xi) > 0, \sum_{x\in\mathcal{X}} \tilde{p}(x;\ \xi) < \infty\}$,

$$\tilde{p}(x;\ \xi) \stackrel{\text{def}}{=} \left[\frac{1}{n}\sum_{i=1}^{n}\xi^i F_i(x)\right]^{\frac{2}{1-\alpha}} = \left[\frac{1}{n}\sum_{i=1}^{n} u_i(x)^\beta\right]^{\frac{1}{\beta}},$$

但し, $\beta \stackrel{\text{def}}{=} \frac{1-\alpha}{2}$, $\xi^i \stackrel{\text{def}}{=} n\beta\theta^i$, $u_i(x) \stackrel{\text{def}}{=} (\xi^i)^{\frac{1}{\beta}} p_i(x)$ とする.

4.1. 視点の揺さぶりと意外性に関するデータ操作

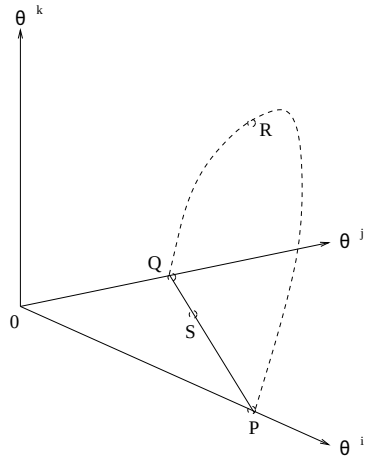
視点の揺さぶり操作: 利用者の (曖昧な) 意図の範囲を表すパラメータ空間の 2 点$\theta(p)$ と$\theta(q)$ を直線補間する各点$\theta(s)$ の経路上の操作. 通常は, 高々n'個 $(n' = |I| < n)$ の視点のみを非零の重要度の値$\theta^k(p) \neq 0$, $\theta^{k'}(q) \neq 0$ $(k, k' \in I)$ に選ぶ.

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.

⇒ 曖昧な意図に関する範囲を探索しながら, 奇抜な着想の触発をも同時に期待するための操作.

==========

4.2. 測地線上のデータ操作



探索範囲を示す 2 つの視点 PQ に対する揺さぶり操作 (実線 α-測地線) と発見的な操作 (点線 $-\alpha$-測地線を k 軸上に射影)
$\Rightarrow$ 特に, 指数型の m-測地線は, 感覚量 $\log p_i(x)$ の期待値 $\eta_i = \sum_{x\in\mathcal{X}} \log p_i(x) p(x;\ \theta)$ の補間に相当 (線形計算で実行可) .

—

5.1. 実験例

視点 i	視点 (属性) 名	θ^iの値が小 $\leftrightarrow$ 大
1	身近かさ	空間/心理的に遠い $\leftrightarrow$ 近い
2	問題へのアクション	事実を知る $\leftrightarrow$ 対策を考える
3	時間的な近さ	過去 $\leftrightarrow$ 未来
4	温暖化との関連性	弱い $\leftrightarrow$ 強い
5	酸性雨との関連性	弱い $\leftrightarrow$ 強い
6	オゾンホールとの関連性	弱い $\leftrightarrow$ 強い
7	食糧危機との関連性	弱い $\leftrightarrow$ 強い
8	エネルギー問題との関連性	弱い $\leftrightarrow$ 強い
9	化学汚染との関連性	弱い $\leftrightarrow$ 強い

地球環境問題を題材とした 125 個の各情報 x に対して, 9 個の各視点 i から関連度の感覚量 $\log p_i(x)$ を (主観的に) 5 段階評価.

—

5.2. 実験結果

例として,「対策案」と「食糧危機」に関心を持って情報探索する場合を想定.

- 視点の揺さぶり操作では,「対策案」から「食糧危機」へとランキングの変動が大きい.
- 一方, 意外性に関する操作では変動は小さく, 主に「身近さ」や「エネルギー問題」に関する情報がヒット.
- 新しいメッセージとして, “酸性雨による植物への育成被害” “発電や自動者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を気付かせる.

==========

—

6. まとめ

- 魅力的な主観情報の埋没と, それらを活用するための技術的課題 (現状の手段では未解決) を指摘し, 緩やかな関係構造の全体をパラメトリックに扱う方法論 (情報幾何学) に着目した.
- 情報幾何学の新しい応用として, 複数の視点に対する情報の関連度が張る関数空間から, 利用者の曖昧な意図に応じて動的に関係構造を生成するデータモデルを紹介した.
  この関係分布モデルをα-分布族によって具現化した.
- 利用者の曖昧な意図の範囲を補間しながら, 思いがけない意外な情報をも同時に探す方法を提案し, 実験例を通じてその有効性/可能性を示した.

==========

—

補足 1. 人々の推奨を活かした魅力的な情報空間

動的な関係構築の必要性: F. Halasz, Reflections on Note Cards: Seven Issues for the Next Generation of Hypermedia Systems, CACM'88

Personal/Collabo フィルタリング: D. Goldberg, et al., Using Collaborative Filtering to Weave an Information Tapestry, CACM'92.
S. Loeb, Architecting Personalized Delivery of Multimedia Information, CACM'92.

カスタム化の要望: C.M. Bowman, et al., Scalable Internet Resource Discovery: Research Problems and Approaches, CACM'94.

人的資源/推奨の活用: W. Hill, et al., Recommending and evaluating choices in a virtual community of use, CHI'95 Conf. Proc.'95.
U. Shardanand and P. Maes, Social Information Filtering: Algorithms for Automating "World of Mouth," CHI'95 Conf. Proc.'95.
M. Balabanovic and Y. Sholam, Content-Based, Collaborative Recommendation, CACM'97.

==========

—

補足 2.1. データモデルの比較

従来モデル: ハイパーテキスト等のように直接張られたリンクはもちろん, 属性値の条件一致などで関係付けられる (関係表などにおける) 間接的なリンクも含めた, 情報間の関係 (リンク) 操作から, 選択的に情報 (集合) を抽出.
情報とリンク集合中の個々の要素に対する代数的操作.

関係分布モデル: 視点に対する情報の関連性の強弱を表すリンクを, (重み付けて) 組み合わせた分布空間を考えて, リンク強度の分布空間に対する操作を介して, 選択的に情報 (集合) を抽出.
関連の強弱の付いたリンク強度の集まりである, 確率分布の関数空間に対する幾何学的操作.

==========

—

補足 2.2. データモデルにおける関係操作の変遷

< DB の関係操作の発展 >

RDB ： 関係代数による関係表の分割統合,

ER ： 関係属性の付与 (RDB 設計に便利),

SDM / OODB ： 関係の階層 / 抽象化,

HyperText DB: リンク編集,

関係分布モデル ： 関係の強弱の連続空間への操作.

==========

—

補足 3. 指数型分布族と混合型分布族 $(\alpha = \pm 1)$

Exp:

$$\begin{aligned} p(x;\ \theta) &= \exp\left[C(x) + \textstyle\sum_{i=1}^{n} \theta^i F_i(x) - \psi(\theta)\right], \\ &= p_0(x)\Pi_{i=1}^{n} \left[\tfrac{p_i(x)}{p_0(x)}\right]^{\theta^i} / \textstyle\sum_{x\in\mathcal{X}} \left\{ p_0(x)\Pi_{i=1}^{n} \left[\tfrac{p_i(x)}{p_0(x)}\right]^{\theta^i} \right\}, \end{aligned}$$

Mix: $p(x;\ \theta) = p_0(x) + \sum_{i=1}^{n} \theta^i (p_i(x) - p_0(x)),$

但し, $C(x) = \log p_0(x)$, $F_i(x) = \log \frac{p_i(x)}{p_0(x)}$とした.

$n = 2$ の例 ：m 次元の確率ベクトル $p, q \in M$ を結ぶ 2 つの方法

$$\begin{aligned} r(x) &= C_t \times p(x)^{1-t} q(x)^t, \\ s(x) &= (1-t)p(x) + tq(x). \end{aligned}$$

時間変数 $0 < t < 1$, 規格化定数 $C_t = \{\sum_{x\in\mathcal{X}} p(x)^{1-t} q(x)^t\}^{-1}$.

==========

—

補足 4. α-分布族の平均化作用

α-分布族と平均演算子の物理的解釈の対応

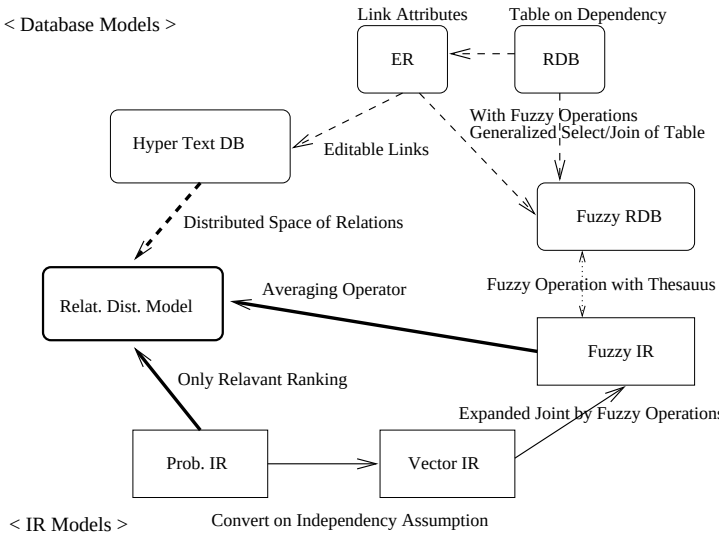
β	α	α-接続	平均の解釈	AND/OR 的作用
$-\infty$	∞	-	最小	AND
0	1	指数型	幾何平均	AND 的
1/2	0	リーマン	平方根平均	AND 的
1	-1	混合型	算術平均	中庸
2	-3	-	遠心力平均	OR 的
∞	$-\infty$	-	最大	OR

例 ：幾何平均 $z = [\ \Pi_i u_i\ ]^{\frac{1}{n}}$ は, 両辺の $\log$ をとると, 指数型 $\log z = \frac{1}{n}\sum_i \log u_i$ となる （$\beta \to 0$ の場合 ）.

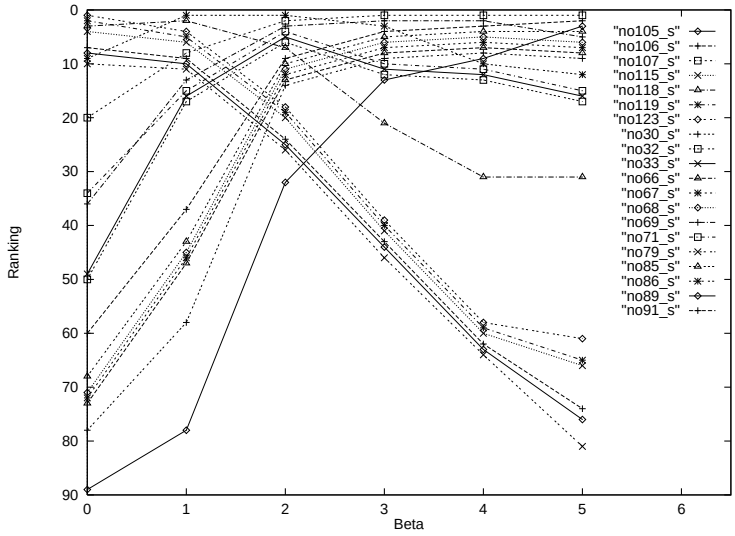
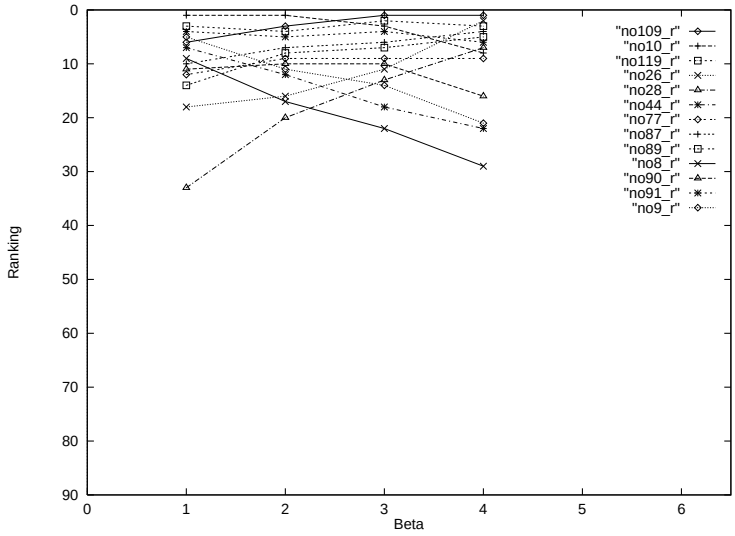
—

補足 5. 既存の DB, IR モデルに対する位置付け

関係分布モデルの拡散的な場へのデータ操作と平均化作用



補足 6.1. 視点の揺さぶり操作における順位の遷移



補足 6.2. 意外性に関する操作における順位の遷移