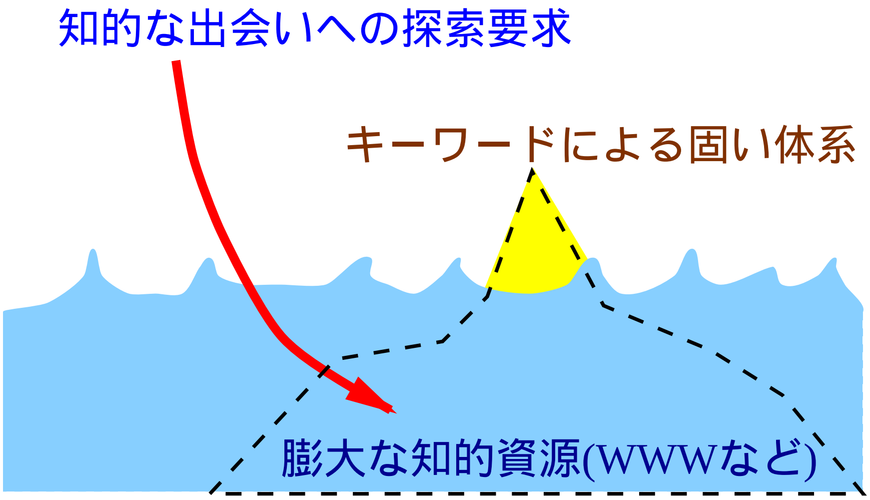
個人選好による情報アクセスに適したデータモデルについて

北陸先端科学技術大学院大学 林 幸雄

Contents

==

1. 情報過多によるアクセス困難性



知的な出会いへの探索要求

キーワードによる固い体系

膨大な知的資源(WWWなど)

==

2. 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]

3. 主観と客観の再考

3.1 人間的な主観性の回復 [菅野 90]

主観（人間知）: パスカルの直観的心情認識,
バシュラールの概念認識における非判明（多様化）,

- 個人的主観 → ファジィ理論が重視,
- 根源的な認識にかかわる普遍的主観,
  → 主観の内に認識の客観性を見出す観念論（カント）,
- 過去の価値観に対する反省の主観,

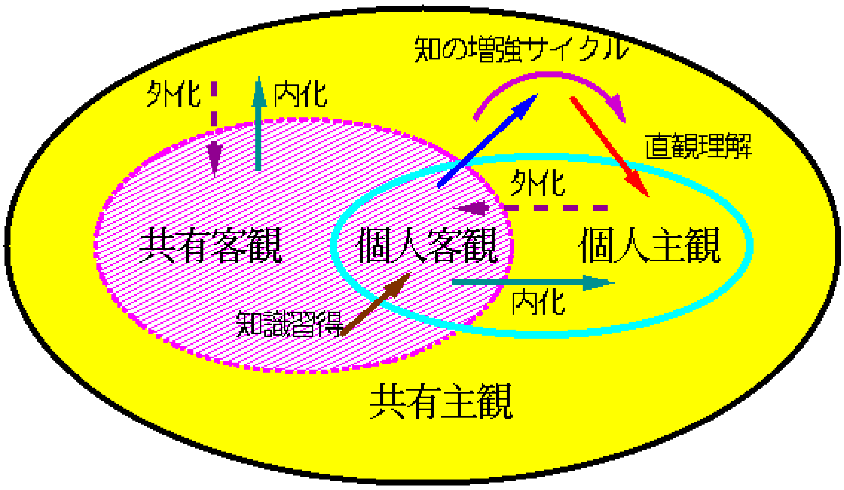
客観（機械知）: デカルトの方法的懐疑, 分析主義, 一意概念,
合理主義的主観.

我々は, 主客二原論よりも相補的とする立場（ボーア-ワイツェッカー）をとる.

==

3. 主観と客観の再考

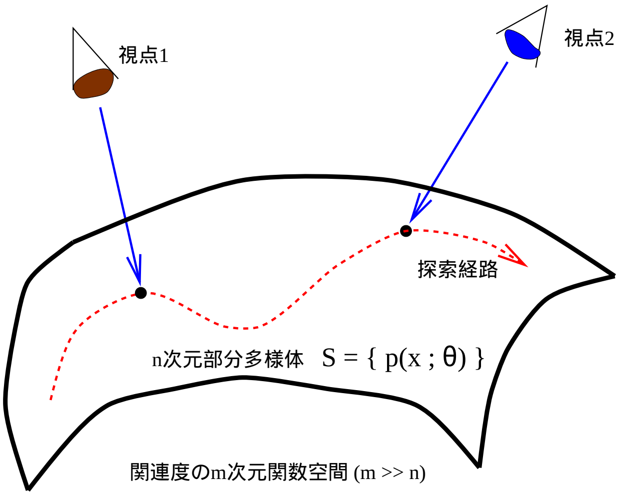
3.2 ネットワーク社会における知の流通, 活用, 創造



共有主観や共有客観からインスパイヤされた各個人の直観を原動力として, 新しい概念の形成や理解を促進 / 共有していく.

4. 動的に知の空間を構築するモデル

4.1 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.

4. 動的に知の空間を構築するモデル

4.2 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.

4. 動的に知の空間を構築するモデル

4.3 α-分布族による関係分布モデルの具現化

α-分布族 $S_\alpha = \{p_\alpha(x;\ \theta)\}$,

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

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

$\Downarrow$ α-アファイン多様体$\tilde{S}$に拡張

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

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

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

4. 動的に知の空間を構築するモデル

4.4 指数型分布族と混合型分布族 $(\alpha = \pm 1)$

Exp:

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

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

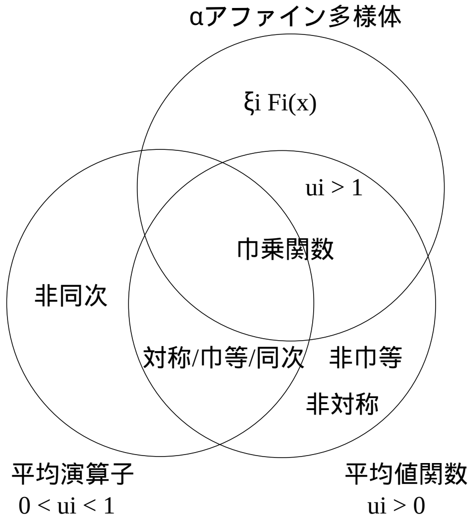
$$\begin{aligned} r(x) =&\ C_t \times p(x)^{1-t} q(x)^t, \\ s(x) =&\ (1-t)p(x) + tq(x). \end{aligned}$$

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

==

4. 動的に知の空間を構築するモデル

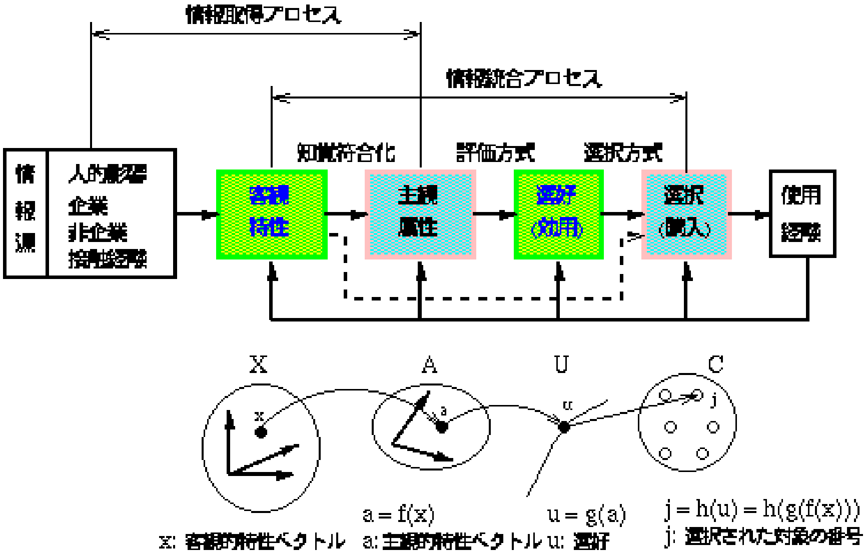
4.5. α-分布族, 平均値関数, 平均演算子の関数クラス



関係分布モデルは, ベクトル情報検索モデルの OR 結合をファジィ演算にしたファジィ情報検索モデルに相当 (+ 規格化作用).

5. 市場分析の個人選好モデルとの関連性

5.1. 多属性選択モデルの要約 [中西 84]



5. 市場分析の個人選好モデルとの関連性

5.2. 消費者需要モデル （特性→選好 ）

製品市場の特性 i の消費量 q_i に対する効用 $U(q)$ が,

Cobb-Douglas 型関数:

$$U(q) = \Pi_{i=1}^{n} q_i^{\theta^i}, \quad \Leftrightarrow \quad p(x;\ \theta) = \frac{\Pi_{i=1}^{n}[p_i(x)]^{\theta^i}}{\sum_{x \in \mathcal{X}} \left\{ \Pi_{i=1}^{n} p_i(x)]^{\theta^i} \right\}},$$

CES 型関数 ($\beta < 1$):

$$U(q) = \left[\sum_i \theta^i q_i^{\beta} \right]^{\frac{1}{\beta}}, \quad \Leftrightarrow \quad p_\alpha(x;\ \theta) = \left[\frac{1-\alpha}{2} \sum_{i=1}^{n+1} \theta^i \{p_i(x)\}^{\frac{1-\alpha}{2}} \right]^{\frac{2}{1-\alpha}}.$$

と規定され, それぞれ指数型分布族とα-分布族に対応する.

==

5. 市場分析の個人選好モデルとの関連性

5.3. 市場シェアモデル　(属性→選好→選択)

対象　(ブランド) $k \in \mathcal{X}$ の確率的効用 $Prob(k) \approx A_k$ が,

MNL モデル:

$$Prob(k) = \frac{\exp\left(\sum_{i=1}^{n} \theta^i X_{ki}\right)}{\sum_{j=1}^{m} \exp\left(\sum_{i=1}^{n} \theta^i X_{ji}\right)}, \quad \Leftrightarrow \quad p(x;\ \theta) = \sum_{i=1}^{n} \theta^i p_i(x),$$

MCI モデル:

$$Prob(k) = \frac{\Pi_{i=1}^{n} X_{ki}^{\theta^i}}{\sum_{j=1}^{m} \left\{\Pi_{i=1}^{n} X_{ji}^{\theta^i}\right\}}, \quad \Leftrightarrow \quad p(x;\ \theta) = \frac{\Pi_{i=1}^{n} [p_i(x)]^{\theta^i}}{\sum_{x \in \mathcal{X}} \left\{\Pi_{i=1}^{n} p_i(x)]^{\theta^i}\right\}},$$

と規定され, MNL モデルの $\log Prob(k)$ と MCI モデルの $Prob(k)$ はそれぞれ混合型分布族と指数型分布族に対応するが, α-分布族と一般化 MCI モデル $A_k = \Pi_{i=1}^{n} \Phi(X_{ki})^{\theta^i}$ との関数型は一致しない (標準形 $F_i(x) = [p_i(x)]^{\beta}$, $0 < p_i(x) < 1$ において).

==

5. 市場分析の個人選好モデルとの関連性

5.4. α-分布族, 平均演算子, 個人選好モデルの関数クラス

β	α	α-接続	平均（論理演算）	消費者需要	市場シェア
$-\infty$	∞	-	最小（AND）	CES 型	
0	1	指数型	幾何平均（AND 的）	Cobb-Douglas 型	MCI
1/2	0	リーマン	平方根平均（AND 的）	CES 型	
1	-1	混合型	算術平均（中庸）		MNL
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）.

5. 市場分析の個人選好モデルとの関連性

5.5. 関係分布モデルと個人選好モデルとの変数対応

関係分布モデル	市場シェア	消費者需要
θに対する関連度 (の対数) $\log p(x;\,\theta)$	ブランドの吸引力/魅力 $Prob(k) \approx A_k$	効用 $U(q)$
視点 i に対する 情報 x の関連度 $p_i(x)$	対象 k の i 番目の 属性 X_{ki}	需要/数量 (製品の物理特性) q_i
視点 i の重要度 パラメータ θ^i	個人の反応 パラメータ θ^i	個人の嗜好 パラメータ θ^i

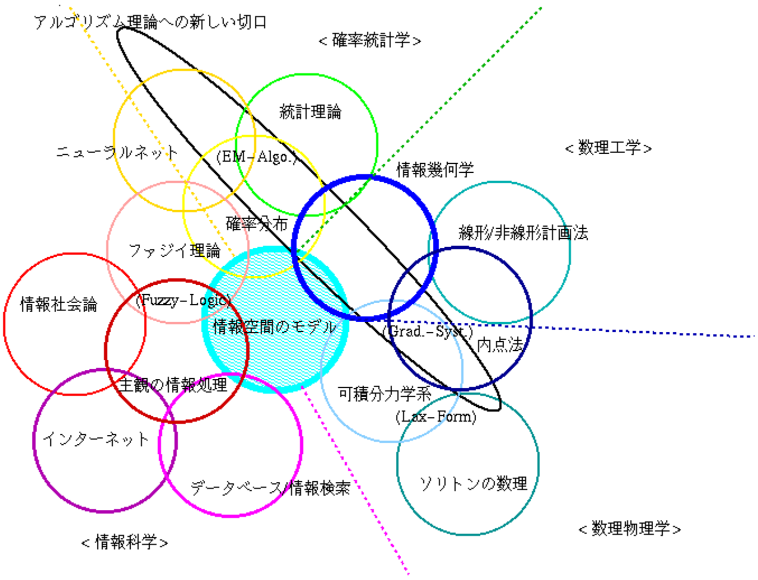
個人選好モデルの特性や属性との比較において, 視点に対する情報の関連度の主観性をよりいっそう強調した上で, 関係分布モデルはそれらと対応付け可能となる.

6. まとめ

- 人間の主観性を考慮した, 視点に対する情報の関連性の強弱に基づいて, 利用者の（複合的な）視点に応じて動的に関係構造を構築するデータモデルを紹介した.
  情報幾何学のα-分布族によってモデルを具現化した.
- 従来の関係表などにおける代数的なデータ操作と比較して, 本モデルにおける統計的多様体（関連度の関数空間）上の幾何学的なデータ操作や, そのファジィ平均化作用, 及び, 確率的情報検索モデルとの接点を示した.
- 市場分析の個人選好モデルとの, 関数クラスにおける理論上の密接な関連性や, 応用上の変数間の対応関係を指摘し, 非線形の選好空間を探索するための基本モデルとして位置付けた.

==

補足 0. 研究分野の広がり



補足 1.1. インタラクションの拡張

	物理世界	記号世界	知の世界
対話の相手	物理的実在	抽象的実在 （コード）	抽象的認知物
対話の結果	外界への動作	計算, 文書編集	概念理解, 状況把握
延長機能	身体	記号処理	直観的思考 ?
操作感覚	外的	内的	深的
道具の例	斧, 工作機械, 車	文書エディタ	次世代 DB ?

太古の人類から現代に至るまでの科学技術の発展によって, 物とのインタラクション → 記号とのインタラクション→ 情報や知識とのインタラクションへと拡張しつつある.

==

補足 1.2. 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) 意味づけが要求される.

==

補足 1.3. 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

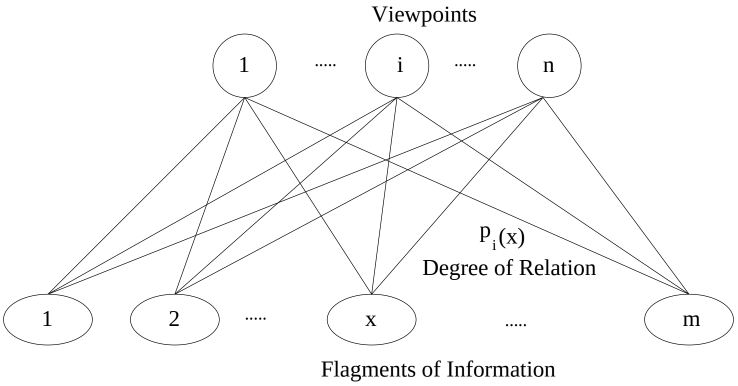
Personal/Collabo フィルタリング: D. Goldberg, et al., Using Collaborative Filtering to Weave an Information Tapestry, CACM'92.
S. Loeb, Architecting Personalized Delivery of Multimedia Information, CACM'92.

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. 関係 (確率) 分布モデル



⇒ 視点の重要度パラメータθで張られる情報の関連度の分布空間として多様体 $S = \{p(x;\ \theta)\}$ を考え，S に対する操作 (パラメータ変化による空間探索) を，探索経路上の視点に応じて情報の関係構造を動的に生成する DB 操作とする．

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 2.3. 関係分布モデルの利点

確率的な緩い結合:
- 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)
- 人間の主観 (確率) に基づく曖昧な処理,
- 統計理論などに基づく理論解析の可能性,

モデル構成上:
- データ規定労力の大幅削減 (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 / 関係の把握,
- 情報間の関連性や解釈の自由度が大.

→ 構造破壊と再度構造化を積極的に行なっている.

==

補足 2.4. Croft の IR モデルと指数型分布族

索引 X を持つ文書のある質問への関連度 $g(X)$ を

$$g(X) = \log \frac{P(X|\,Relevant)}{P(X|\,Nonrelevant)} = \sum_{i=1}^{n} x_i \log \frac{p_i(1-q_i)}{q_i(1-p_i)} + \sum_{i=1}^{n} \log \frac{1-p_i}{1-q_i},$$

とする (ベクトルIR :索引 × 質問). 但し, それぞれ独立な, 質問に関連する文書 d が項 x_iを含む確率 p_i, 質問に関連しない文書 d が項 x_iを含まない確率 q_iとする.

⇒ 関連性 p_iのみの IR を考えれば指数型

$$\log p(x;\ \theta) = \sum_i \theta^i \log p_i(x) - \psi(\theta),$$

に帰着.

==

補足 2.5. 既存の DB, IR モデルに対する位置付け

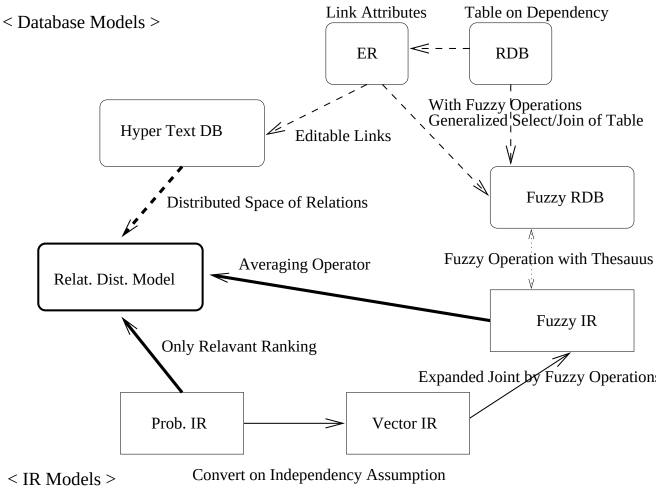


図 1: 関係分布モデルの関係場のデータ操作と平均化作用

補足 3.1. 平均値関数 (オペレーションズリサーチ)

定義 1 連続関数 $z = \phi(u_1, u_2)$ が次の *3* つの条件を満足するとき, これを平均値関数, その値を平均値という.

1. $u_1, u_2 > 0$ で, $min(u_1, u_2) \le \phi(u_1, u_2) \le max(u_1, u_2)$.
2. 関数 $\phi(u_1, u_2)$ は u_1, u_2 のそれぞれについて非減少関数であり, 停留点がない.
   $\phi(u_1, u_2)$ が微分可能な場合: $grad\ \phi \ge 0$, $div\ \phi > 0$.
3. 関数 $\phi(u_1, u_2)$ は *1* 次の同次関数である. すなわち, すべての $\gamma > 0$ について, $\phi(\gamma u_1, \gamma u_2) = \gamma\phi(u_1, u_2)$ が成立する (尺度変換による不変性の一種).

平均 : $\mathrm{min} \le$ 調和 $\le$ 幾何 $\le$ 算術 $\le$ 双対幾何 $\le$ 双対調和 $\le \mathrm{max}$

===

補足 3.2. 平均演算子（ファジィ論理演算）

定義 2 平均演算子 M は, M: $[0,1] \times [0,1] \rightarrow [0,1]$ なる関数であり, 次の条件を満たすものである.

1. $min(u_1, u_2) \leq M(u_1, u_2) \leq max(u_1, u_2)$.
2. $M(u_1, u_2) = M(u_2, u_1)$　（交換性）.
3. M は連続, 増加関数である.
4. $M(0,0) = 0$; $M(1,1) = 1$.
5. $M(u,u) = u$　（巾等性）.
6. $M \notin \{min, max\}$.

ファジィ演算：t-ノルム $\leq$ min $\leq$ 平均演算子 $\leq$ max $\leq$ t-コノルム

==

補足 3.3. 平均値関数と平均演算子

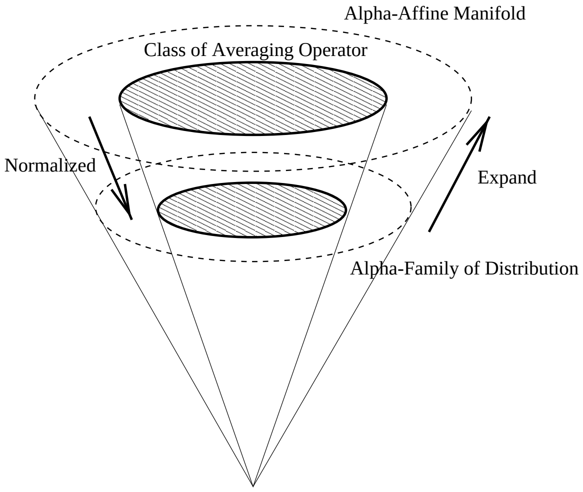


図 2: α-分布族と平均演算子のクラス （$0 \leq (\xi^i)^{\frac{1}{\beta}} p_i(x) \leq 1$ に制限）

補足 4. それぞれの関数クラス

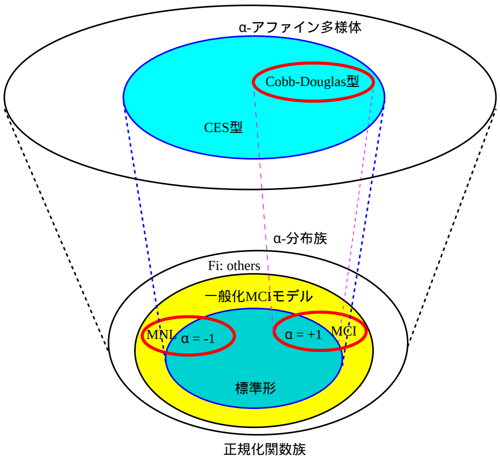


図 3: α-分布族, CES 型, 一般化 MCI モデルの包含関係

α-分布族: $p_\alpha^\beta = \beta \sum_i \theta^i F_i(x) \Leftrightarrow$ 一般化 MCI: $\log A_k = \sum_i \theta^i \log \Phi(X_{ki})$, …MNL: $\Phi(X_{ki}) = e^{X_{ki}}$, MCI: $\Phi(X_{ki}) = X_{ki}$