ラフ集合理論に基づくコンジョイントモデルに関する一考察

杉原　一臣*

A Study on Rough Conjoint Model Based on Rough Sets Theory

Kazutomi Sugihara

Conjoint analysis is a method for deriving the part worth value of each factor from the total evaluations. This can be also applied to multi-criteria ranking or choice problems. In our method which is called "Rough Conjoint Analysis", preference structure in conjoint analysis is represented by IF-Then rules. In this paper, the possibility of building a new model is considered in addition to the usefulness of the method.

Keywords: Conjoint Analysis, Rough Sets, Dominance Relation, Rule Extraction

１．はじめに

ラフ集合(Rough Sets)[1]は Z.Pawlak によって提案された情報近似のための数学的概念である．ラフ集合では，同値関係や類似関係，順序関係などによる集合を知識とみなして，与えられた集合をこの知識で表現するために，2 つの観点からの近似集合が提案されている．また，この近似に対して，多様な知識の表現が考えられることから，ラフ集合の拡張とその解釈に関する研究[1][2]が行われている．さらに，これらの近似により得られる集合から，IF-Then 形式のルール抽出も提案されていて，通常のデータ解析において定性的な解析結果を得ることを試みる研究も盛んである[3]．一方，コンジョイント分析(Conjoint Analysis)[4][5]は，対象データに関する諸要因とその組合せに対する全体評価から，各要因の部分効用値を求める手法である．また，各要因の部分効用値を求めることで，最も効果の高い要因の組合せを推定することができる．部分効用値を得る手法については，部分効用値から全体評価を構成するための結合法則により様々なコンジョイントモデルが提案されているが，現実問題において必ずしも仮定されたモデルが適切ではない場合が存在するため，モデルの妥当性に関する議論は尽きない．以上を踏まえて，諸要因と全体評価との間のデータ構造が未知である場合は，IF-Then ルールを用いる方がその構造を表現するのに適しているという立場から，先行研究として，ラフ集合の概念を用いたコンジョイント分析を検討し，数理モデルに関する制約を緩和することで，新たなコンジョイントモデルの構築を試みた[6][7][8].

本論文では，先行研究の分析手法をラフコンジョイント分析[9]と呼び，この分析手法を再度検討することによって，新たなモデル構築の可能性や分析手法の有用性を考察する．また，意思決定者によるデータ構造の理解を容易にするため，分析結果として得られる IF-Then ルールに着目し，その解釈や新たなルール表現についても併せて検討する．

本論文の構成は次の通りである．第 1 章では本論文の目的および背景を述べる．第 2 章ではラフ集合の概要を述べ，第 3 章ではラフコンジョイント分析について数値例を交えて説明する．第 4 章では，分析手法の有用性を高めることを狙いとして，得られた IF-Then ルールの解釈や新たなルール表現について検討する．第 5 章では，本論文のまとめと今後の課題について述べる．

＊ 経営情報学科

２．ラフ集合の基本概念

２．１．同値関係に基づくラフ集合

ラフ集合では，与えられた情報に対して 2 つの近似集合が求められる．通常，人間は与えられた情報を完全に理解することが難しいため，自身の持つ有限の知識を用いてその情報を近似的に理解することを試みる．ラフ集合における情報近似の概念は，決して特異なものではなく，人間の知識獲得の過程においてはごく自然のことである．

ある情報の集まりを情報システム S として定義する．

$$S = (U, Q, V, f) \tag{1}$$

ただし, U は対象の有限集合, Q は属性の有限集合, V は属性値の集合, f は $\forall q \in Q$ に関して $f: U \times Q \to V$ となる関数を表す．また $\forall x \in U, \forall q \in Q$ について， $f(x,q)$ は対象 x の属性 q に関する属性値を示している．

以上の情報システムを基に，任意の部分集合 $P(\subseteq Q)$ について，次の識別不能関係(Indiscernibility Relation)を定義する．

$$R_P = \left\{ (x,y) \in U^2 \mid f(x,q) = f(y,q), \forall q \in P \right\} \tag{2}$$

この関係は，部分集合 P で識別できない対象の対を表していて，$(x,y) \in R_P$ は「部分集合 P において，x と y は同じである」ことを意味している．この関係は同値関係であり，反射律(Reflexive)，対象律(Symmetry)，推移律(Transitive)を満たしている．また， x に関する識別不能関係から次の同値クラスを定義する．

$$R_P(x) = \left\{ y \in U \mid f(x,q) = f(y,q), \forall q \in P \right\} \tag{3}$$

このクラスを基に情報 $X(\subseteq U)$ を近似する． X を $R_P(x)$ で近似すると，次の下近似と上近似の集合が考えられる．

$$\text{下近似集合：}\ \underline{R_P}(X) = \left\{ x \in U \mid R_P(x) \subseteq X \right\} \tag{4}$$

$$\text{上近似集合：}\ \overline{R_P}(X) = \{ x \in U \mid R_P(x) \cap X \neq \phi \} \tag{5}$$

X と(４)，(５)式の間には，次の包含関係が成り立つ．

$$\underline{R_P}(X) \subseteq X \subseteq \overline{R_P}(X) \tag{6}$$

(７)式は「クラス $R_P(X)$ が確実に X に含まれている対象 x の集合」を表していて，「対象 x が X に含まれることが必然である」ことを意味している．また，(５)式は「クラス $R_P(x)$ が X に含まれている可能性のある対象 x の集合」を表していて，「対象 x が X に含まれる可能性がある」ことを意味している．以上のことから，X の 2 つの近似は，それぞれ X の存在する必然性と可能性を示している．このように，対象 X を把握することが困難である場合，下側と上側からそれに類似したクラスを用いて近似を与えるという考え方がラフ集合の情報近似の概念である．なお，下近似・上近似集合に関して，次の相補関係が存在する．

$$\underline{R_P}(X) = U - \overline{R_P}(X^C) \tag{7}$$

$$\overline{R_P}(X) = U - \underline{R_P}(X^C) \tag{8}$$

この下近似・上近似集合の含意関係から，ルールを抽出できることがわかっている．すなわち，得られた下近似と上近似を用いて，次の IF-Then ルールが得られる．

$$\text{IF} \quad x \in \left\{ R_P(y) \mid y \in \underline{R_P}(X) \right\}, \text{then exactly} \quad x \in X .$$
$$\text{IF} \quad x \in \left\{ R_P(y) \mid y \in \overline{R_P}(X) \right\}, \text{then possibly} \quad x \in X .$$

２．２．順序関係に基づくラフ集合

同値関係によるラフ集合では共通の値を持つデータ同士がクラスを形成し，そのクラスに基づき下近似・上近似が行われる．しかし，共通の値を持たないデータの場合，データが孤立するという特徴がある．そこで取扱うデータにより，同値関係より制約の緩い関係，すなわち類似関係などを用いたラフ集合への拡張がなされている．多基準意思決定問題を取扱うラフ集合の拡張としては，Greco らの順序関係によるラフ集合[10]がある．彼らのラフ集合は「より良い(悪い)データを持つものは，より高く(低く)評価される」という観点から定義されていて，ラフ集合で取扱える問題の範囲を大幅に広げているといえる．Greco らのラフ集合では，属性の値に大小関係を持つデータのみが取扱われている．以下に順序関係に基づくラフ集合の定義を示す．

定義 1．順序関係(dominance relation)

属性の集合 P について，「y が x を優越している」という関係を yD_px と表し，次のように定義する．

$$yD_Px \leftrightarrow \forall q \in P, f(y,q) \geq f(x,q) \tag{9}$$

ここで，$f(x,q)$ は対象 x の属性 q に関する大小比較可能な属性値を示している．D_P により，x に対する優越集合 $D_P^+(x)$ と被優越集合 $D_P^-(x)$ が次のように定義される．

$$D_P^+(x) = \{ y \in U \mid yD_Px \} \tag{10}$$

$$D_P^-(x) = \{ y \in U \mid xD_Py \} \tag{11}$$

定義 2．累積集合(cumulative set)

有限個のクラス Cl_r $(r = 1,2,\cdots,R)$ を用いて，累積集合を次のように定義する．

$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s \tag{12}$$

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s \tag{13}$$

ただし，Cl_r は次の性質を満たすものとする．

$$\exists r, \ x \in Cl_r \tag{14}$$

$$U = \bigcup_r Cl_r \tag{15}$$

$$\forall s,t (s \neq t) \quad Cl_s \cap Cl_t = \phi \tag{16}$$

$$Cl_R \succ \cdots \succ Cl_r \succ \cdots \succ Cl_1 \tag{17}$$

定義 3. 順序関係によるラフ集合

$D_P^+(x)$ による $Cl_t^{\geq}$ の下近似と上近似の集合を次のように定義する.

$$\underline{P}(Cl_t^{\geq}) = \{x \in U \mid D_P^+(x) \subseteq Cl_t^{\geq}\} \tag{18}$$

$$\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x) \tag{19}$$

また同様に, $D_P^-(x)$ による $Cl_t^{\leq}$ の下近似と上近似の集合を次のように定義する.

$$\underline{P}(Cl_t^{\leq}) = \{x \in U \mid D_P^-(x) \subseteq Cl_t^{\leq}\} \tag{20}$$

$$\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x) \tag{21}$$

これらの近似集合については, 次の相補関係が成り立っている.

$$U - \underline{P}(Cl_t^{\geq}) = \overline{P}(Cl_{t-1}^{\leq}) \tag{22}$$

$$U - \overline{P}(Cl_t^{\geq}) = \underline{P}(Cl_{t-1}^{\leq}) \tag{23}$$

(22)式の証明は以下の通りである.

$$\begin{aligned} U - \underline{P}(Cl_t^{\geq}) &= \{x \in U \mid D_P^+(x) \cap Cl_{t-1}^{\leq} \neq \phi\} \\ &= \{x \in D_P^-(y) \mid y \in Cl_{t-1}^{\leq}\} \\ &= \bigcup_{y \in Cl_{t-1}^{\leq}} D_P^-(y) \\ &= \overline{P}(Cl_{t-1}^{\leq}) \end{aligned} \tag{24}$$

(22)式とほぼ同じであるため, (23)式の証明については記述を省略する.

定義 3 の下近似集合については, 同値関係の場合と同様に IF-Then ルールが得られる. つまり, $x^* \in \underline{P}(Cl_t^{\geq})$ より,

IF $f(x,q_1) \geq f(x^*,q_1)$ and $f(x,q_2) \geq f(x^*,q_2)$ and $\cdots$ and $f(x,q_m) \geq f(x^*,q_m)$,
then exactly $x \in Cl_t^{\geq}$.

という形式の IF-Then ルールが得られる. また, $x^* \in \underline{P}(Cl_t^{\leq})$ より,

IF $f(x,q_1) \leq f(x^*,q_1)$ and $f(x,q_2) \leq f(x^*,q_2)$ and $\cdots$ and $f(x,q_m) \leq f(x^*,q_m)$,
then exactly $x \in Cl_t^{\leq}$.

という形式の IF-Then ルールが得られる.

3. ラフコンジョイント分析

コンジョイント分析は, 予め用意された諸要因の組合せに対する全体評価から各要因の効用を求める手法である. また, 求められた効用から, 最も効用の高い組合せを推定することができる. 通常, 全体評価は順序関係で得られていると仮定し, この順序関係から各要因の効用値を算出する. 効用値を算出する際には, 各要因の効用から全体評価を構成する結合法則が必要となる. コンジョイント分析で用いられる結合法則(モデル)としては, 加法モデルが一般的である. しかし, 現実の問題では加法モデルが必ずしも適切であるとは限らない.

また，この結合法則に関しては様々な議論がなされているが，現実の問題に遭遇したときに，各要因と全体評価との間の構造関係が明確であることは稀である．したがって，どのようなモデルが妥当であるかを議論することは難しい．そこで，先行研究として，データ構造を柔軟に表現できると共に，データからより詳細な知識が得られると考え，IF-Then 形式のルールを抽出する分析手法を提案した．

まず，サンプルデータの諸要因をカテゴリカルデータとして表したものを入力値，全体の評価を出力値とするデータセットを準備し，回帰分析(数量化 I 類)[11]を行う．回帰分析により推定された回帰係数を部分効用値とし，その値から，各属性における諸要因の効用に関する大小関係を得る．この操作により，サンプルデータにおいて定義 1 の順序関係を作成する．次に，意思決定者による全体の評価値を複数のクラスに分類する．これは，意思決定者が分析結果を理解するために IF-Then 形式のルールで知識を表現するのであれば，連続値より意思決定者の把握可能な範囲でクラス分けする方が望ましいという観点に立ったものである．

以上の手続きにより，コンジョイント分析で取り扱われるサンプルデータを変換し，順序関係によるラフ集合で分析することにより，IF-Then ルールの抽出が行われる．

マーケティング問題を例に分析の流れを示す．表-1 には 10 店舗のデータが表されていて，データは「立地」と「店舗形態」，そして評価の基準となる「年間売上」で構成されている．表-2 のカテゴリカルデータを入力値，表-1 の年間売上を出力値とする回帰分析を行った結果，各要因の部分効用値は表-3，表-4 のようになる．これらの結果から，立地に関しては住宅街，駅ビル，郊外，ビジネスの順で，店舗形態についてはディスカ，スーパー，コンビニの順に効用が高いことがわかる．また，部分効用値の結合法則として加法モデルを仮定した場合，各属性の中で最も効用の高い要因同士の組合せとなるので「住宅街」と「ディスカ」となるが，加法モデルの妥当性を検証していないため，この結論は適切とはいえない．

表-1：店舗に関するマーケティングデータ

店舗	立地	形態	年間売上 (百万円)
1	駅ビル	スーパー	31
2	駅ビル	ディスカ	35
3	郊外	コンビニ	24
4	郊外	スーパー	27
5	郊外	ディスカ	34
6	住宅街	スーパー	37
7	住宅街	コンビニ	30
8	ビジネス	スーパー	21
9	ビジネス	コンビニ	29
10	ビジネス	ディスカ	25

(注) ビジネス：ビジネス街，スーパー：スーパーマーケット，
コンビニ：コンビニエンスストア，ディスカ：ディスカウントストア

表-2：表-1 のカテゴリカルデータ

店舗	立地				形態		
	駅ビル	郊外	住宅街	ビジネス	スーパー	コンビニ	ディスカ
1	1	0	0	0	1	0	0
2	1	0	0	0	0	0	1
3	0	1	0	0	0	1	0
4	0	1	0	0	1	0	0
5	0	1	0	0	0	0	1
6	0	0	1	0	1	0	0
7	0	0	1	0	0	1	0
8	0	0	0	1	1	0	0
9	0	0	0	1	0	1	0
10	0	0	0	1	0	0	1

表-3：立地に関する部分効用値

	駅ビル	郊外	住宅街	ビジネス
部分効用値	0.00	-3.04	4.31	-5.14

表-4：形態に関する部分効用値

	スーパー	ディスカ	コンビニ
部分効用値	29.76	31.74	28.12

ここで，各属性における部分効用値の大小関係に着目し，順序関係によるラフ集合の分析を行う．分析を行うにあたり，全体評価を次の 4 つのクラスに分けている．

$$Cl_1 : [0,25),\ Cl_2 : [25,30),\ Cl_3 : [30,35),\ Cl_4 : [35,100]$$

「立地」と「形態」に関する順序関係に基づき，クラスの累積集合 $Cl_t^{\geq}, Cl_t^{\leq}$ を近似すると，次の下近似・上近似集合が得られる．

$Cl_t^{\geq}$ の下近似集合：$\underline{P}(Cl_2^{\geq}) = \{1,2,4,5,6,7,9,10\}$　$\underline{P}(Cl_3^{\geq}) = \{1,2,5,6,7\}$　$\underline{P}(Cl_4^{\geq}) = Cl_4 = \{2,6\}$

$Cl_t^{\geq}$ の上近似集合：$\overline{P}(Cl_2^{\geq}) = \{1,2,4,5,6,7,9,10\}$　$\overline{P}(Cl_3^{\geq}) = \{1,2,5,6,7\}$　$\overline{P}(Cl_4^{\geq}) = \{2,6\}$

$Cl_t^{\leq}$ の下近似集合：$\underline{P}(Cl_1^{\leq}) = Cl_1 = \{3,8\}$　$\underline{P}(Cl_2^{\leq}) = \{3,4,8,9,10\}$　$\underline{P}(Cl_3^{\leq}) = \{1,3,4,5,7,8,9,10\}$

$Cl_t^{\leq}$ の上近似集合：$\overline{P}(Cl_1^{\leq}) = \{3,8\}$　$\overline{P}(Cl_2^{\leq}) = \{3,4,8,9,10\}$　$\overline{P}(Cl_3^{\leq}) = \{1,3,4,5,7,8,9,10\}$

これらの近似集合から IF-Then ルールを得ることができる．例えば，店舗 1 は $Cl_3^{\geq}$ の要素であることから，「店舗 1 の組合せ(立地：駅ビル，店舗形態：スーパー)よりも(順序関係で)良い組合せであれば，年間売上はクラス 3 以上となる」というルールが得られる．また，従来のコンジョイント分析で着目されている最も有望な組

合せについては，下近似集合$\underline{P}(Cl_4^{\geq})$の要素が店舗2と6であることから，それぞれの立地と店舗に関する組合せと，その組合せよりも良い組合せが有望とみなされる．

このように，ラフコンジョイント分析は，諸要因と全体評価を関連付ける結合法則の制約を緩和し，IF-Thenルールによるデータ構造の柔軟な表現を可能にしている点で特徴的であるといえる．

4．ラフジョンジョイント分析で得られる IF-Then ルール

順序関係によるラフ集合の分析手法をコンジョイント分析に適用することにより，IF-Thenルールによる情報の獲得が可能となったが，順序関係に基づく近似集合に着目することで，これまで検討されてきたルールとは異なる形式のルールが得られる．例えば，順序関係に基づく下近似集合の内，

$$\exists s,t(s \leq t)\ \ x^* \in \left(\underline{P}(Cl_s^{\geq}) \cap \underline{P}(Cl_t^{\leq})\right)$$

となる対象x^*に関しては，次のIF-Thenルールが得られる．

$$x^* \in \left(\underline{P}(Cl_s^{\geq}) \cap \underline{P}(Cl_t^{\leq})\right)$$
$$\text{IF}\ \ f(x,q_1)=f(x^*,q_1)\ \text{and}\ f(x,q_2)=f(x^*,q_2)\ \text{and}\ \cdots\ \text{and}\ f(x,q_m)=f(x^*,q_m),$$
$$\text{then exactly}\ \ x \in \left(Cl_s^{\geq} \cap Cl_t^{\leq}\right).$$

このルールはx^*と同じものであれば，ある範囲のクラスに割り当てられることを意味している．従来考えられていたルールと比較すると，条件部の対象範囲が限定されているため，非常にコンパクトなルールになっている．前章で示したIF-Thenルールは個々の近似集合からルールが得られるため，少なくとも，与えられたデータの数だけルールを獲得できるが，$Cl_t^{\geq}, Cl_t^{\leq}$の下近似・上近似集合について，

$$\underline{P}(Cl_1^{\geq}) \supseteq \underline{P}(Cl_2^{\geq}) \supseteq \cdots \supseteq \underline{P}(Cl_{R-1}^{\geq}) \supseteq \underline{P}(Cl_R^{\geq})$$
$$\overline{P}(Cl_1^{\geq}) \supseteq \overline{P}(Cl_2^{\geq}) \supseteq \cdots \supseteq \overline{P}(Cl_{R-1}^{\geq}) \supseteq \overline{P}(Cl_R^{\geq})$$
$$\underline{P}(Cl_R^{\leq}) \supseteq \underline{P}(Cl_{R-1}^{\leq}) \supseteq \cdots \supseteq \underline{P}(Cl_2^{\leq}) \supseteq \underline{P}(Cl_1^{\leq})$$
$$\overline{P}(Cl_R^{\leq}) \supseteq \overline{P}(Cl_{R-1}^{\leq}) \supseteq \cdots \supseteq \overline{P}(Cl_2^{\leq}) \supseteq \overline{P}(Cl_1^{\leq})$$

となることから，結論部において冗長なルールが存在する．このことから，データの構造を詳細に検討するためには，意思決定者が容易に理解できるように冗長なルールを省く必要がある．

5．おわりに

ラフコンジョイント分析では，回帰分析等により属性の各要因に順序関係を与え，順序関係による情報近似を行うことで，コンジョイント分析の目的である最も効果の高い要因の組合せを推定すると共に，データ構造をより柔軟に表現できる．一方，情報近似により得られるルールの形式については，従来とは異なるものが考えられる．本論文では，獲得するIF-Thenルールの形式について注目し，近似表現を変えずに，異なるルールを得られることを示したが，他の形式も存在すると考えられる．今後は，IF-Thenルールの記述を詳細に検討することで，より柔軟なデータ構造の表現を模索したい．

本研究の遂行にあたり，平成21年度福井工業大学特別研究費(研究費B)の交付を受けた．

参考文献

[1] Z.Pawlak, Rough Classification, International Journal of Man-Machine Studies, Vol.20, pp.469-483, 1984.

[2] A.Skowron and C.M.Rauser, The Discerniblity Matrix and Functions in Information Systems, in R.Slowinski (ed) Intelligent Decision Support. Handbook of Application and Advances of the Rough Set Theory, Kluwer Academic Publishers, pp.331-362, 1992.

[3] L.Polkowski and A.Skowron：Rough Sets in Data Mining, I and II, Physica-Verlag , 1998.

[4] J.B.Kruskal : Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data, Journal of Royal Statist Society, Series B, Vol.27:2, pp.251-263, 1965.

[5] J.B.Kruskal and F.J.Carmone : MONANOVA, A Fortran-IV Program for Monotone Analysis of Variance (Non-Metric Analysis of Factorial Experiments), Behavior　Science, Vol.14, pp.165-166, 1969.

[6] Kazutomi Sugihara, Hiroaki Ishii and Hideo Tanaka, Conjoint Analysis Based on Rough Approximations by Dominance Relations Using Interval Regression Analysis, The IEEE 2002 World Congress on Computational Intelligence(CD-ROM), pp.763-766, 2002.

[7] 杉原一臣, 石井博昭, 田中英夫, ラフ集合による新しいコンジョイント分析の提案, 日本知能情報ファジィ学会誌, Vol.15, No.4, pp.421-428, 2003.

[8] Kazutomi Sugihara, Hiroaki Ishii and Hideo Tanaka, On Conjoint Analysis by Rough Approximations based on Dominance Relations, International Journal of Intelligent Systems, Vol.19, pp.671-679, 2004.

[9] 杉原一臣, ラフコンジョイント分析について, 日本オペレーションズ・リサーチ学会 2008 年春季研究発表会, 2008.

[10] S.Greco, B.Matarazzo and R.Slowinski, Rough Sets Theory for Multicriteria Decision Analysis, European Journal of Operational Research, Vol.129, pp.1-47, 2001.

[11] 河口至商：多変量解析入門, 森北出版, 1978.

（平成 22 年 3 月 31 日受理）