G-21

ラフ集合におけるコアと NN-GA におけるスキーマの比較研究

A Comparative Study between Cores in Rough Sets and Schemas in NN-GA

小幡 和樹 *　　中村鴻成 *　　原田 利宣 **
Kazuki Obata　Kosei Nakamura　Toshinobu Harada

1. はじめに

企画者が新たな製品やサービスを企画する際，アンケート調査結果などを用いて，企画対象の属性（原因）と評価（結果）の因果関係として得る試みがなされてきた．この行為は逆問題の解決に相当する．逆問題とは，一般的に「結果から原因を探る問題」を指し，その行為は逆推論と定義される．対して，「原因から結果を探る問題」を順問題と呼び，その行為は順推論と定義される．

一般的に，企画者による逆推論において，評価を満たす原因の組み合わせは一つではなく，複数個存在する．ここでいう，複数個の適合解のことを企画における「多様解」と呼ぶ．つまり，企画における多様解を求めて，はじめて逆問題を解いたと考えられる．ニューラルネットワークなどにより，逆問題解いた場合，1 つまたは数個の適合解を得ることができる．しかし，1 つまたは数個の適合解を得たとしても，それらが適合解の全てなのか，もっと良い解が存在するのではないかという疑問が残る［1］.

ここで様々な分野において，逆推論を行うデータマイニング手法として,ラフ集合と NN-GA（ニューラルネットワークを利用して知識ベースが作成され，遺伝的アルゴリズムによって知識ベースを参照しながら最適解を探索する）の二つの手法が挙げられる．両手法の共通点として，原因と結果の IF-THEN ルールの取得により，結果に大きく影響する原因を推論する点，多様解を得る点が挙げられる．

ラフ集合，NN-GA による逆推論の有効性を示した研究は多くある．井上らはラフ集合のデザインコンセプト策定，Web 検索システム，データセットの特徴抽出などへの応用し，ラフ集合による逆推論の有効性，実用性を示した［2］．齋藤らはラフ集合を用いて，ファザード（建物を正面から見た外観）の形態要素から伝統的であるか否かとう感性的な評価に対して逆推論を行った．これにより，ラフ集合が感性的な評価に対して，重要な原因の組み合わせを推論可能であること，ラフ集合が建築分野に応用可能であると示した［3］．榎本らはラフ集合と恒等写像モデル（NN）による最適解の比較を行った．これにより，両手法には逆推論である点，多様解を得る点という共通点があり，両手法から得られる多様解の特徴解明を行った［4］．田らはニューラルネットワークを利用して知識ベースが作成され，遺伝的アルゴリズムによって知識ベースを参照しながら最適解を探索する手法 NN-GA を実装し，自動車の形態要素の逆推論を行った．これにより, NN-GA による逆推論の有効性，得られる解の多様性を示した［5］.

しかし，ラフ集合と NN-GA の出力結果の特徴を比較し，得られる多様解の特徴を考察した研究はない．よって，両手法より得られる多様解を比較し，前述した数個の適合解を得たとしても，それらが適合解の全てなのか，もっと良い解が存在するのではないかという疑問について検証することは有意あると考えた．

そこで本研究では，ラフ集合により得られる IF-THEN ルールにおける頻出の IF（コア）と NN-GA により得られる IF-THEN ルールにおける頻出の IF（スキーマ）を比較し，両手法の出力結果の特徴考察を行った．

表 1　自動車の特徴における決定表

サンプル	カラー	造形	ドアタイプ	イメージ	選好
s1	色彩系	有機的	2 ドア	パーソナル	好き
s2	色彩系	曲線的	2 ドア	スポーティ	どちらでもない
s3	白黒系	曲線的	4 ドア	フォーマル	どちらでもない
s4	白黒系	有機的	4 ドア	パーソナル	好き
s5	白黒系	曲線的	4 ドア	パーソナル	どちらでもない
s6	色彩系	曲線的	2 ドア	スポーティ	好き

表 2　決定ルール

決定クラス	決定ルール	C.I.値
好き	IF[造型 : 有機的]THEN[選好 : 好き]	2/3
	IF[カラー : 色彩系]and[イメージ : パーソナル] THEN[選好 : 好き]	1/3
	IF[ドアタイプ : 2 ドア]and[イメージ : パーソナル] THEN[選好 : 好き]	1/3
どちらでもない	IF[カラー : 白黒系]and[造型 : 曲線的] THEN[選好 : どちらでもない]	2/3
	IF[造型 : 曲線的]and[ドアタイプ : 4 ドア] THEN[選好 : どちらでもない]	2/3
	IF[イメージ:フォーマル]THEN[選好:どちらでもない]	1/3
	IF[造形 : 曲線的]and[イメージ : パーソナル] THEN[選好 : どちらでもない]	1/3

*　和歌山大学大学院
**　和歌山大学

G-21

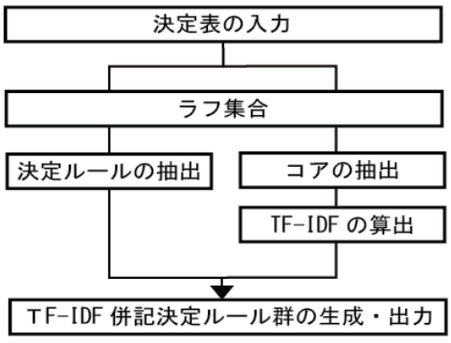

図 1　TF-IDF 併記決定ルール群表示のフローチャート

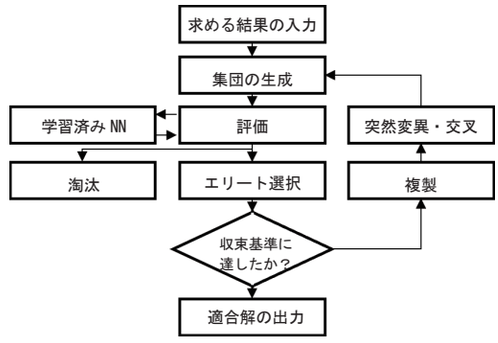

図 2　NN-GA の一連の流れ

2. ラフ集合概要

2.1　ラフ集合概説

ラフ集合理論は，1982 年にポーランドの Z.Pawlak によって提唱された理論［6］であり，対象の集合がもつ情報を粗く（ラフに）することで，対象の集合の丁度良い記述を求めることを目的としている．ラフ集合は，分析対象のデータに決定表を使用する．自動車の特徴についての決定表を示す［7］．表 1 の左端にある s1, ･･･，s6 はサンプルと呼び，分析対象の各自動車を指す．上端にある「カラー」などを属性，「色彩系」などの要素は属性値と呼び，各属性の実際の値を表す．右端の「選好」のような結論となる属性を決定属性，「好き」のような決定属性の属性値を決定クラスと呼ぶ．決定ルールは決定クラスとしたある結論を満たすための条件を IF-THEN 形式で記述したものである．例として表 1 における決定クラス「好き」の決定ルールをサンプル s1 にのみ着目して求めると，IF［カラー:色彩系］and［造形:有機的］and［ドアタイプ : 2 ドア］and［イメージ : パーソナル］THEN［選好:好き］となる．このような決定ルールを，決定クラス「好き」の下近似に属するすべての車（s1，s4）に対して求め，ブール演算を用いて条件部を再簡単系にしたものを，決定クラス「好き」における極小決定ルールと呼ぶ．ここでの下近似は，s2 や s6 のような決定クラスが異なるが属性値が全く同じサンプルを含まない，与えられた部分集合に確実に含まれるような集合を指す．また，求めた各決定ルールでそれぞれ C.I. 値（Covering Index）を算出できる．C.I. 値は決定ルール条件が当てはまるサンプルの数を，決定ルールの決定属性と同じ決定クラスに属するサンプルの数で割ることで求められ，決定ルール全体の中で特定の条件部がどれだけ重要度が高いかを判断する指標となる．本研究の比較対象となるコアとは以上の手順でもとめられた決定ルールのなかで頻出の属性値の組み合わせのことを指す．

2.2　使用するラフ集合システム

本研究では，中村らにより提案された，ラフ集合理論に TF-IDF を組み合わせた TF-IDF 併記決定ルール群表記法表示システムを使用する［8］．以下に TF-IDF 併記決定ルールについて説明する．TF-IDF は，自然言語処理でよく使用されており，文書中に含まれる単語の重要度を評価することを目的としている．TF(Term Frequency: 索引語頻度) は，ある文書におけるある単語の出現頻度を表わし，1 つの文書中に出現する回数が多い単語ほど大きくなる .IDF(Inverse Document Frequency: 逆文書頻度) は，ある単語のすべての文書中における出現率の逆数を表わし，すべての文書中において，出現 する回数が少ない単語ほど大きくなる .TF-IDF は，これら 2 つの指標を掛け合わせたものである . つまり，ほかと比較し て TF-IDF 値が大きい単語ほど，ある文書を特徴づける重要な単語だと解釈することができる．TF-IDF 併記決定ルール群表記法は TF-IDF をラフ集合に組み合わせることで，ある決定ルールを特徴づける重要な原因を発見することが可能となる手法である . 本研究では本ソフトで得られたコアをラフ集合により得られたコアとして扱う．

3. NN-GA 概要

ジェネティックアルゴリズム（Genetic Algorithm ; 以下 GA）とは，生物の進化過程を模倣し形式化した最適化問題や探索問題に対して利用される手法の一つである . GA では，ある結果に対して，最適な遺伝子（原因）の組み合わせを探索することが可能である．GA の特徴として，局所解に陥らない点，線形的な問題だけでなく，非線形的な問題にも探索が可能である点が挙げられる．

ニューラルネット - ジェネティックアルゴリズム（以下 NN-GA）とはニューラルネット（NN）と GA を組み合わせた逆推論の一手法である．NN-GA では求める結果を定め，その結果を推論するランダムに決められた遺伝子（原因）

G-21

をもつ染色体（個体）を N 個作成する．N 個の遺伝子を事前に学習済みの NN により適応度を評価する．ここでいう学習済み NN とは入力として，遺伝子を受け取り，求める結果か否かを分類し，確率として出力する NN である．NN により出力された，個体それぞれの求める結果である確率を適応度とする．次に，適応度の低い個体を淘汰し，高い個体を残す．次に，残った個体に複製，交叉，突然変異などの遺伝的処理を施し，再度個体を N 個作成する．この一連の作業を任意の回数繰り返すことで，求めたい結果に対して,大きな影響を及ぼす遺伝子をもつ個体の集団（適合解）が出力される．出力された適合解の中で頻出の遺伝子（スキーマ）を求める．以上が NN-GA の一連の動作である．

4. コアとスキーマの導出と比較

4.1　使用するデータセットの説明

本研究には，カリフォルニア大学アーバイン校（UCI : University of California Irvine）が管理，公開する心臓病のデータセット［9］から作成した，サンプル数 503．11 の原因（age: 年齢 ,sex: 性別 ,cp: 胸痛のタイプ ,trestbps: 安静時血圧 ,chol: 血中総コレステロール ,fbs: 空腹時血糖値 ,restecg: 安静時心電図結果 ,thalach: 最大心拍数 ,exang: 運動誘発性狭心症 ,oldpeak: 安静時に比べて運動により誘発される ST の低下 ,slope: 運動負荷時の ST セグメントの傾き）と 1 つの結果（心臓病の有無）から構成されるデータセットを使用した．行った処理として，欠損値の多いサンプル，原因の削除．また，原因の age，trestbps，chol，oldpeak は量的データであり，ラフ集合や NN-GA のスキーマ探索に不向きである．そのため,4 つの原因をそれぞれ四分位範囲に基づき，質的データに変換した．age は 50 歳未満，51 歳以上 55 歳未満，56 歳以上 60 歳未満，61 歳以上の 4 つのカテゴリからなる age_category とした．trestbps は 120 未満，120 以上 130 未満，130 以上 150 未満，150 以上の 4 つのカテゴリからなる trestbps_category とした．chol は 100 未満，100 以上 200 未満，200 以上 235 未満，235 以上 275 未満，275 以上の 5 つのカテゴリからなる chol_category とした．oldpeak は 0.2 未満，0.2 以上 1 未満，1 以上 2 未満，2 以上の 4 つのカテゴリからなる oldpeak_category とした．

表 4　各結果における正解データ

結果	正解データ
心臓病あり	cp : 0 ~ 2　restecg : 1,2　exang : 1 oldpeak : 2,3
心臓病なし	cp : 3　trestbps_category : 0,1 chol_category : 0,1　fbs : 0 restecg : 0　exang : 0 oldpeak : 0　slope : 0

表 3　心臓病データセットの原因、結果の詳細

原因	カテゴリ
年齢（age_category）（単位：歳）	0：50 未満　1：50 以上 55 未満　2：55 以上 60 未満 3：60 以上
性別（sex）	0：女性　1：男性
胸痛のタイプ（cp）	0：典型的な狭心症　1：非定型狭心症　2：非狭心症性疼痛 3：なし
安静時の血圧（trestbps_category） （単位：mmHg）	0：120 未満　1：120 以上 130 未満　2：130 以上 150 未満 3：150 以上
血中総コレステロール（chol_category） （単位：mg/dl）	0：100 未満　1：100 以上 200 未満　2：200 以上 235 未満 3：235 以上 275 未満　4：275 以上
空腹時血糖値（fbs）（単位：mg/dl）	0：120 以下　1：121 以上
安静時心電図結果（restecg）	0：正常　1：ST 波に異常あり 2：Estes の基準より、左室肥大の可能性大、または明らか
運動時最大心拍数（thalach）（単位：bpm）	0：120 未満　1：121 以上 141 未満　2：141 以上 160 未満 3：160 以上
運動時誘発性狭心症（exang）	0：なし　1：あり
運動により誘発される ST 低下（oldpeak）	0：0.2 未満　1：0.2 以上 1 未満　2：1 以上 2 未満　3：2 以上
運動負荷時の ST セグメントの傾き（slope）	0：上がり勾配　1：平坦　2：下り勾配
結果	カテゴリ
心臓病の有無	0：心臓病なし（血管の 50%未満の狭窄） 1：心臓病あり（血管の 50%以上の狭窄）

G-21

4.2　比較の基準となるデータについて

ラフ集合におけるコアと NN-GA におけるスキーマの精度の比較基準となる正解データを定義する．心臓病ありの結果に対する正解データは過去に本研究で使用するデータセットを分析し，重要となる原因を抜き出した既存研究の結果［10］［11］［12］と内科医の助言により作成した．心臓病なしの結果に対する正解データは厚生労働省が定める各原因毎の基準値［13］と内科医の助言により作成した．以下がその正解データである（表 4）．本研究では，正解データと一致するラフ集合におけるコアと NN-GA におけるスキーマの数を参考にラフ集合と NN-GA の精度を評価する．ここでいう一致とは，原因の組み合わせが，コアまたはスキーマとして得られた時，組み合わせを構成する原因すべてが正解データと一致していることを指す．精度の評価を行った後，TF-IDF 併記決定ルールと NN-GA それぞれの出力の特徴を考察する．

4.3　ラフ集合におけるコアの導出

本研究では，中村らの開発した TF-IDF 併記決定ルール群表記法表示システムを使用し，コアを得た． その結果，得られたコアは 2 種類に分類された．コアにいかなる条件が付加されても結果が変わらないコア，コアに条件が付加されることで結果が変化するコアの 2 種類のコアである．先行研究［8］より，前者を結論唯一型コア，後者を結論変化型コアとする．心臓病ありの結論唯一型コアが 154 個，心臓病なしの結論唯一型コアが 143 個，結論変化型コアが 33 個得られた．結論唯一型コアはその有無が結果に対して重要な意味を持つと考えられる．よって，精度の指標として，正解データと一致する結論唯一型コアを使用する．また結論変化型コアはその有無が結果に大きく影響することはないが，ある原因の組み合わせが付加されることにより，結果に対する影響が生まれると考えられる．これは，TF-IDF 併記決定ルール群表記法表示システム独自の特徴であると考えられるため，使用したソフトの出力結果の特徴として，結論変化型コアに着目する．

4.4　NN-GA におけるスキーマの導出

4.4.1　NN-GA における NN について

本研究では，全結合ニューラルネットワークを利用した．全結合型 NN は深層学習で用いられる手法の 1 つであり，中間層が全て全結合層になっている点が特徴である．具体的には，11 の原因から心臓病が存在するかしな

exang:1 ∧ slope:2	thalach_category:1★★☆ ∧ trestbps_category:2★★☆ ∧ sex:1★☆☆	→	target_01:1	0.077	{ 275, 291, 297, 309
	age_category:3★★★ ∧ restecg:0★★★	→	target_01:1	0.074	{ 262, 264, 271, 280
	thalach_category:0★★★ ∧ trestbps_category:3★★★	→	target_01:1	0.067	{ 262, 289, 302, 304
	trestbps_category:3★★★ ∧ oldpeak_category:3☆☆☆	→	target_01:1	0.064	{ 289, 299, 317, 319
	restecg:0★★☆ ∧ thalach_category:0★★☆ ∧ oldpeak_category:3☆☆☆	→	target_01:1	0.060	{ 285, 289, 307, 317
	age_category:3★★★ ∧ oldpeak_category:3☆☆☆	→	target_01:1	0.050	{ 349, 398, 401, 408
	chol_category:2★★☆ ∧ sex:1★☆☆ ∧ oldpeak_category:3☆☆☆	→	target_01:1	0.044	{ 207, 285, 289, 307
	age_category:1★★☆ ∧ trestbps_category:2★★☆ ∧ sex:1★☆☆	→	target_01:1	0.044	{ 207, 285, 307, 320
	age_category:0★★★ ∧ oldpeak_category:3☆☆☆	→	target_01:1	0.040	{ 273, 298, 299, 317

図 3　結論唯一型コア

slope:1	fbs:0★★★ ∧ age_category:0★★☆ ∧ sex:0☆☆☆	→	target_01:0	0.063	{ 54, 68, 75, 85, 95, 108, 120, 132,
	cp:3★★★ ∧ sex:0★★☆	→	target_01:0	0.054	{ 44, 85, 90, 97, 108, 120, 121, 152
	exang:0★★★ ∧ age_category:0★★☆ ∧ sex:0☆☆☆	→	target_01:0	0.054	{ 54, 68, 75, 85, 95, 132, 143, 147,
	restecg:0★★★ ∧ age_category:0★★☆ ∧ sex:0☆☆☆	→	target_01:0	0.044	{ 54, 75, 95, 132, 143, 147, 155, 18
	cp:2★★☆ ∧ sex:1★★☆ ∧ thalach_category:0☆☆☆	→	target_01:1	0.064	{ 211, 214, 220, 233, 238, 239, 249,
	trestbps_category:1★★★ ∧ exang:1★★☆ ∧ cp:2★★☆	→	target_01:1	0.050	{ 206, 214, 217, 220, 223, 227, 229,
	exang:1★★☆ ∧ cp:2★★☆ ∧ thalach_category:0☆☆☆	→	target_01:1	0.050	{ 211, 214, 220, 238, 239, 249, 362,
	restecg:2★★★ ∧ sex:1★★☆ ∧ thalach_category:0☆☆☆	→	target_01:1	0.047	{ 211, 214, 233, 239, 249, 284, 362,
	oldpeak_category:2★★★ ∧ sex:1★★☆ ∧ thalach_category:0☆☆☆	→	target_01:1	0.044	{ 220, 238, 249, 284, 362, 368, 386,

図 4　結論変化型コア

G-21

いかを二値分類する全結合 NN を構築した．NN の詳細について述べる．ネットワークの構造の作成および学習には Python の Keras ライブラリを使用した．実装した NN の構造 は Dence27-Dence60-Dropout(0.3)-Dense120-Dropout(0.5)-Dence20 -Dence1 である．バッチサイズは 10，エポック数は 30 とした．最適化アルゴリズムには Adam を用いた．全体データセットを 5 分割した交差検証を行い，得られた精度の平均が 82%のモデルを採用した．

4.4.2　NN-GA の詳細設定

NN-GA の全体フローとしては 3 節で説明したものと同様である．NN-GA の詳細の設定について述べる．

集団の個体数は 80, 各個体の遺伝子の数は 11（データセットの原因と同じ数）とした．各個体の評価の方法としては，NN より得られた値が 1 に近いほど，心臓病である可能性が高く，0 に近いほど心臓病でない可能性が高いと評価する．どの個体を次世代に残すかの選択手法としてはトーナメント方式を採用した．100 回の繰り返し後に収束するよう設定した．交叉は 8 割，突然変異は 2 割で発生するよう設定した．求められた適合解，80 個体の中で半数以上で見られた原因，原因の組み合わせをスキーマとする．

4.4.3　NN-GA におけるスキーマの導出

NN-GA を試行することで，適合解としてある結果となる個体が 80 体とスキーマが得られる．しかし，適合解は導出の過程に乱数が大きく関わるため，適合解，スキーマの内容が試行毎に変化する特徴がある．そのため，NN-GA を 100 回試行し，適合解とスキーマを 100 回求めた．100 回の試行より得られたスキーマ群の中で，頻出のスキーマを導出し，それを共通スキーマと定義する．この共通スキーマを後の比較に利用する．求められた共通スキーマの重要度を示す指標として，共通スキーマが適合解に現れた数を適合解全体で割った値を採用する．本論文ではこの値を共通スキーマ重要度と定義する．心臓病ありの共通スキーマの内，共通スキーマ重要度が 0.6 以上の共通スキーマ 66 個の内，21 個の共通スキーマが正解データと一致した．しかし，心臓病なしの共通スキーマは 12 個程度得ることができず，すべてが共通スキーマ重要度 0.05 以下であった．

4.4.4　コアとスキーマの比較

ラフ集合におけるコアと NN-GA におけるスキーマの精度の評価を行う．心臓病ありに対して，影響度の高い原因は両手法とも出力できていた．ラフ集合に関しては，得られたコア 159 個の内，22 個のコアが正解データと一致した

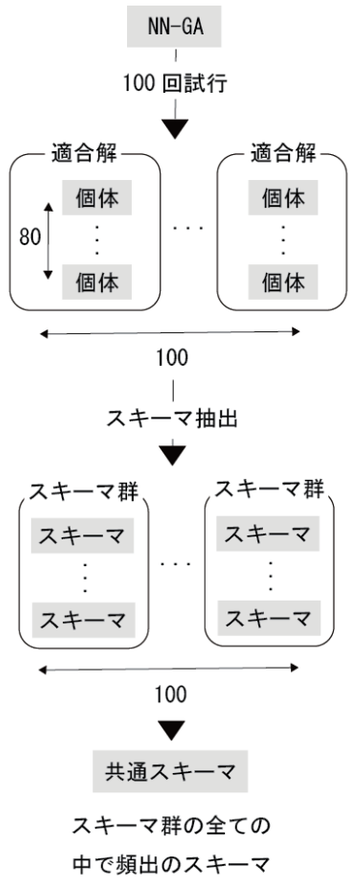

図 5　共通スキーマ導出のフローチャート

ため，コア全体の約 13%が有用なコアと判断した．対して，NN-GA では，得られた共通スキーマにおいて，共通スキーマ重要度が 0.6 以上の共通スキーマ 66 個の内，21 個が正解データと一致したため，共通スキーマ全体の約 31%が有用な共通スキーマと判断した．以上の割合の比較から，NN-GA より得られる共通スキーマの方が有用なデータを推論する精度が高いと考えた．心臓病なしに対して，ラフ集合は重要度の高い原因を出力できていたが，NN-GA では十分に出力することができなかった．ラフ集合は 143 個のコアの内 17 個が正解データと一致したため，コア全体の約 11%が有用なコアであると判断した．対して，NN-GA に関しては，共通スキーマ重要度が 0.05 以下のものが 12 個程度得られた．得られた共通スキーマの内，3 個は正解データと一致したが，共通スキーマ重要度がきわめて低いため，重要な原因をぬきだせてはいるが精度が悪いと判断した．

以上の比較から，ラフ集合は重要度の高い原因を安定して，出力することができるが，有用と考えられるコアの割合が少ない．そのため，有用なデータの選別に時間がかかるという特徴が考えられる．NN-GA は，すべての結果に対して，重要度の高い原因を安定して出力することはできないが，出力できた場合，有用と考えられる共通スキーマの

G-21

割合が大きく，有用なデータの選別が比較的容易であると考えられる．

4.4.5 多様性の評価

TF-IDF 併記決定ルール群表記法と NN-GA のそれぞれの特徴について考察する．ラフ集合については，結論変化型コアに着目する．結論変化型コアはコア自身の結果に対する重要度は高くないが，コアに対してある原因が付加されることにより結果に対する重要度が増すと考えられる．以上の記述について，本研究で導出した結論変化型コアに着目し，具体的に述べる．総コレステロール値が 275 以上と異常な値をとっているというコアに，運動時に狭心症の痛みが発生しない（exang：0）かつ安静時に心電図に異常が見られない（restecg：0）や運動時における心電図に異常が見られない（oldpeak_category：1，slope：0）という原因の組み合わせが付加されることにより心臓病なしという結果となる．また，年が若い（age_category：0）かつ運動時に狭心症の痛みが生じる（exang：1）や，運動時に心電図に異常が見られる（oldpeak_category：3，slope：2）などの原因の組み合わせが付加されることにより心臓病ありという結果となる．以上で挙げた付加部については，正解データとおおむね一致しており，心臓病の有無の結果に大きな影響を与える原因の組み合わせを抜き出しているといえる．よって，結論変化型コアにおけるコアはコア単体では結果に対して重要な意味をもつとはいえないが，付加部にある原因の組み合わせを読み解くことにより，結果にたいして重要な意味をもつ原因を考察できると解釈した．

NN-GA により出力された，心臓病ありに対する共通スキーマに着目する．この共通スキーマの中に，共通スキーマ重要度が低いが，興味深いデータが得られた．共通スキーマ 55 歳以上 60 歳未満かつ総コレステロール値が 275 以上が共通スキーマ重要度 0.05 として得られた．共通スキーマ 60 歳以上，総コレステロール値 100 未満が共通スキーマ重要度 0.7 以上であると出力されている．以上の結果から，60 歳以上であれば総コレステロール値が小さくても心臓病のリスクが高いと解釈できるが，55 歳以上 60 歳未満の年齢，総コレステロール値が 275 以上だと心臓病のリスクが高まると解釈できる．これは 55 歳以上 60 歳未満，総コレステロール値 275 以上の二つの要素が単体ではなく組み合わせであったからこそ，心臓病ありの結果への重要度が高まったと考えられる．つまり，NN-GA は組み合わせによる相乗効果を考慮できている手法であると考えられる．得られた年齢，性別に関する共通スキーマを基に，適合解のクラスタ分けを行ったところ以下のような結果が得られた．

年齢によるクラスタ毎の差は，前述したように総コレステロール値の大小であるとした．これについて内科医の意見は以下である．総コレステロール値は善玉コレステロールと悪玉コレステロールそれぞれの値の和であるため，総コレステロール値の大小によって心臓病のリスクが高まるとは断言できない．しかし，総コレステロール値 275 以上という値は善玉コレステロールを含めた値であっても異常な値であり，本研究で定義する心臓病のリスクが高まると考えられる．また，年齢は 60 歳以上というだけで心臓病のリスクが高いこと．年齢が 55 歳以上 60 歳未満であっても

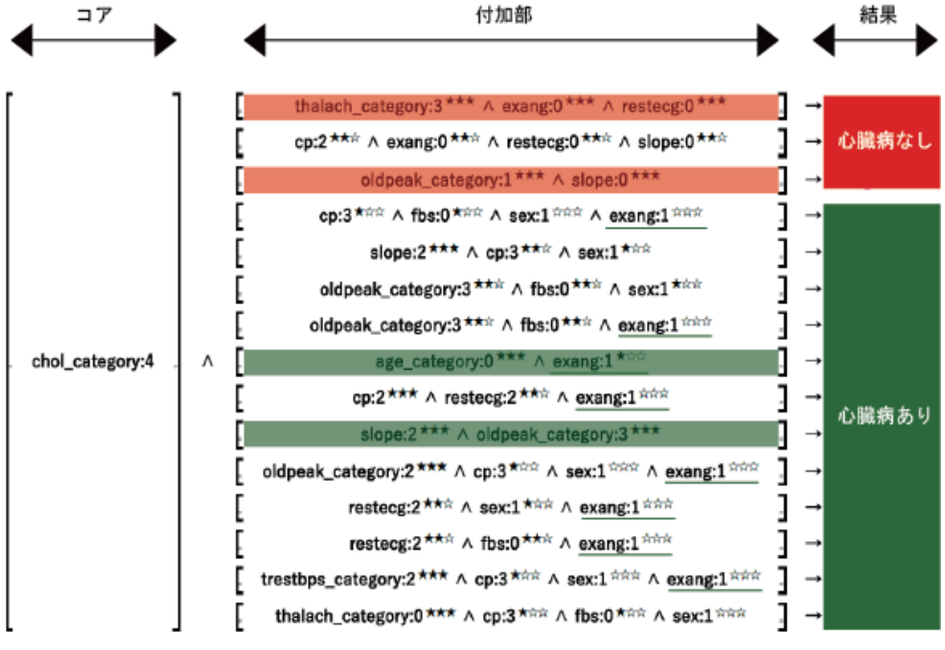

図 6　結論変化型コアの付加部による変化

G-21

心臓病のリスクが高いと考えられるが，総コレステロール値が 275 以上と異常な値をとった際，心臓病のリスクが上がると考えたため，私の解釈について同意が得られた．

性別による，女性の心臓病リスクを出力した適合会が 1 つしか得られなかったため，正確には判断できないが，男性の心臓病に対するリスクに血糖値はあまり関わりがないが，女性の心臓病に対するリスクは血糖値 121 以上であれば高まると解釈ができる．これについての内科医の意見は以下である．血糖値 121 以上という値は動脈硬化につながるため，本研究で定義する心臓病のリスクが高まると考えられる．しかし，それは男性，女性どちらにも言えることだと考えられるので，女性特有の問題とは断言できないと私の解釈について意見した．

本研究では．年齢と性別という 2 つの原因に着目したが，年齢に関しては多様解の存在が確認できた．よって NN-GA より得られる共通スキーマには多様性があると示唆された．また，0 歳以上であれば総コレステロール値が小さくても心臓病のリスクが高いと解釈できるが，55 歳以上 60 歳未満の年齢，総コレステロール値が 275 以上だと心臓病のリスクが高まるという解釈は，TF-IDF 併記決定ルール群表記法にて導出したコアから導くことは困難であった．そのため，NN-GA は TF-IDF 併記決定ルール群表記法により得られる結果から解釈することが難しい，結果に対して重要な原因を導出可能であると示唆された．

5　まとめ

本研究により以下の成果が得られた．

(1) UCI より引用した, 心臓病データセットを加工, 利用し，心臓病の有無を予測する NN を作成した．

(2) 作成した NN を遺伝的アルゴリズム（GA）に組み込み，結果に対して重要な原因，原因の組み合わせを探索する逆推論の一手法である NN-GA を実装した．

(3) 中村らの開発した，ラフ集合と TF-IDF を組み合わせた TF-IDF 併記決定ルール群表記法を導出するソフトにより求めたコア（結果に対して重要な意味をもつ原因）と NN-GA により求められた，スキーマ（結果に対して重要な意味をもつ原因）を比較し，両手法により得られる出力結果の特徴をまとめた．

(4) 重要である原因を抜き出す精度に関しては NN-GA の方が優れていた．

(5) TF-IDF 併記決定ルール群表記法，NN-GA の両手法とも組み合わせによる相乗効果を考慮している手法であった．

今後の課題としては以下が考えられる．

(1) NN-GA は心臓病なしに対する精度が非常に悪かった．改善案として，NN の精度向上，GA に対する設定を心臓病の有無それぞれによって使い分けるなどが考えられる．

(2) 今回使用した，TF-IDF 併記決定ルール群表記法は優れた手法ではあるが，使用するデータセットの数が膨大になるほど，結果出力の時間，得られる IF-THEN ルールが膨大となるという課題を抱えている．そのため，本研究では 500 程度のデータセットを使用した．しかし, 一般的に, データセットの数が増えるほど，NN の精度は向上する．つまり，NN-GA の精度も向上すると考えられる．以上のことから，10,000 以上のデータセットを NN に学習に用いた場合の NN-GA の出力の検証は今回の結果と異なるのか検証する必要があると考えられる．

(3) NN-GA の出力結果に多様性が示唆されたが，年齢，性別のみでクラスタ分けを行った結果からの推測であるため，使用した原因 11 個を加味したクラスタ分けを行い，多様性について検証する必要があると考えられる．

(4) NN-GA は，TF-IDF 併記決定ルール群表記法にて導出したコアから導くことは困難な結果に対して重要な原因を導出可能であると示唆されたが，現状，少数しかか確認できていない．そのため，以上の可能性について検証する必要があると考えられる．

謝辞

本研究における正解データ作成などに際して，本学キャンパスライフ・健康支援センター長　小河健一教授にご指導をいただきました．この場をお借りして心よりお礼申し上げます．

参 考 文 献

［1］ 原田利宣，森典彦 : 恒等写像モデルを応用した多様解・自動車コンフィギュレーション決定支援システムの構築 (2)，デザイン学研究，41(1)，pp51-58，1994.

［2］ 井上勝雄，原田利宣，椎塚久雄，工藤康生，関口彰 : ラフ集合の感性工学への応用，海文堂出版，2009.

［3］ 齋藤篤史，宗本順三，松下大輔 : 感性評価に基づく形態要素のラフ集合を用いた組合せ推論の研究，日本建築学会計画系論文集，594（ 8591），pp85-91，2005 .

［4］ 榎本雄介，原田利宣 : ラフ集合と恒等写像モデルによる最適解の比較研究，デザイン学研究，51（5），pp.1-8，2005.

［5］ 田慕玲，森典彦 : 目標イメージに適する自動車の形態を探索するデザイン支援システム - ジェネティックア

ルゴリズムによる製品形態の逆推論, デサイン学研究, 41（6），1995.

［6］ Pawlak，Z.：Rough sets，International Journal of Computer and Information Sciences，11（5），pp341-356，1982.

［7］ 森紀彦，田中英夫，井上勝男：ラフ集合と感性，海文堂出版，2004.

［8］ 中村鴻成，原田利宣：ラフ集合を用いた TF-IDF 併記決定ルール群表記法の提案と個人投資家の属性データに対する分析，和歌山大学院修士研究，2022 年度.

［9］ The UCI machine learning repository [online]. Available at: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

［10］Vikas Chaurasia，Saurabh Pal：Caribbean Journal of Science and Technology，Vol.1，pp208-217，2013.

［11］Umesh N. Khot, Monica B. Khot, Christopher T. Bajzer, Shelly K. Sapp, E. Magnus Ohman, Sorin J. Brener, Stephen G. Ellis, A. Michael Lincoff, Eric J. Topol：Caribbean Journal of Science and Technology1，JAMA，290(7)，pp898-904，2013.

［12］Robert Detrano, PhD, Andras Janosi, Walter Steinbrunn, Matthias Pfisterer, Johann-Jakob Schmid,, Sarbjit Sandhu, Kern H. Guppy, PhD, Stella Lee, and Victor Froelicher,：MDTHE AMERICAN JOURNAL OF CARDIOLOGY，Vol.64，pp304-310，1989.

［13］厚生労働省 生活習慣病予防のための健康情報サイト：https://www.e-healthnet.mhlw.go.jp/information/metabolic-summaries/m-05