

心理学的トポロジーに基づく心理状態の可視化
Visualization of Mental States Based on Psychological Topology

2018年9月

埼玉大学大学院理工学研究科（博士後期課程）
理工学専攻 人間支援・生産科学コース（主指導教員 平原 裕行）

青木 滉一郎

要旨

近年、可視化研究の適用領域は拡大し、自然現象やコンピュータ上で処理される情報にとどまらず、人々の行動や感情、関係性も、定量的かつ視覚的に表現するような試みがなされている。このような傾向は、人間を中心としたデータの測定および解析、可視化技術の発展に基づくものである。同時に、多様化が進む現代社会においては、個人や集団が直面する心理社会的な問題も様々であり、個々人の価値観やニーズ、心理的特性などを把握した上で、適切な対処へつなげる必要がある。すなわち、可視化技術の心理領域への応用に対して、社会的な要請が高まってきているといえよう。

心理学や社会科学の領域では、対象者の内面を明らかにするために、語りや記述といった言語、テキストを分析する質的なアプローチがなされてきた。このような研究アプローチの利点は、具体的な個人の内面を個別に、より詳細に把握できる点にある。その半面、質的研究におけるデータの分析と解釈には多くの時間を要するとともに、妥当性・信頼性の高い分析結果を得るためには、高度な専門性が求められる。そのため、研究実施にあたり分析者の負担が大きくなることや、データの分析および解釈内容について、専門外の人と理解を共有することが困難になることが、課題として挙げられる。

これに対し、近年、言語データないしはテキストデータの定量的・客観的な分析手法として、テキストマイニングの技術が急速な発展を遂げてきた。テキストマイニングでは、発話や文書に含まれる様々な語句の頻度に注目し、これらを発話や文書の特徴を定量データとして、多次元ベクトルの形式で表現することが多い。言語データを通じて、対象者の心理状態を分析、可視化する際にも、発話における語句の出現頻度を基にしたベクトルデータが、分析の単位となることが想定される。このとき、多次元ベクトルの形式をとるデータの特徴や、データ間の関連性を二次元、あるいは三次元空間上で表現することが、可視化を行う上での課題となる。

そこで、本研究では、言語データに基づいて個人・集団の心理状態とその変化を可視化するために、自己組織化マップ (SOM) とファジィクラスタ分析を組み合わせる可視化手法を提案した。SOM は、多次元空間上のデータを、個々のデータ間の

位置関係を維持しながら、低次元空間へ非線形的にマッピングする解析手法である。これにより、対象者の心理状態を反映し、多次元ベクトルの形式で表される言語データを二次元空間上にマッピングし、その特徴や時間変化を表現することが可能となる。さらに、SOMにより得られたマッピングの結果に対して、ファジィクラスタ分析を適用することで、マップ上の領域についてのクラスタリングを試みる。これにより、マップ上の各領域の性質が明らかになるとともに、縦、横といった空間軸に、心理状態の評価軸を設定できると考えた。以上の提案手法により、言語データが示す心理の多様な特質を、心理学的なトポロジーの観点から、二次元平面上で可視化することが本研究の目的である。

本研究では、以下の2段階の処理を通じて、研究目的へアプローチするための手法を提案した。第一に、SOMとファジィクラスタ分析という非線形解析の手法を組み合わせ、多次元の言語データを二次元空間上にマッピングした。また、個々のデータの座標から、言語の語り手の心理を理解するため、二次元空間上に心理状態の評価軸（心理的評価軸）を設定した。非線形解析により得られた心理的評価軸は、従来の線形解析では見出されなかった、言語データ間の心理学的なトポロジーを表現するものである。このような評価軸上で、個々人の心理状態とその変化を可視化する手法を提案した。第二に、個々人が示す評価軸上の遷移、すなわち心理変化の特徴をベクトルデータとして表現し、ファジィクラスタ分析によるクラスタリングを実施した。これにより、対象集団に共通してみられる心理変化のパターンを分類し、可視化することが可能となった。

本研究では、オリンピック・パラリンピック選手を含む、各競技のトップアスリートに対するインタビュー記事を Web 上から収集し、インタビュー中の選手の発話を言語データとして分析に用いた。心・技・体の各要素がパフォーマンスに影響を及ぼすスポーツにおいては、アスリートの心理面に焦点を当て、コンディショニングやトレーニングを行っていくことが重要である。提案手法では、言語データを分析の対象とすることで、質問紙による調査と比較しても、アスリートの内面をより詳細に理解することができる。さらに、分析結果を心理学的なトポロジーに基づいて効果的に可視化することにより、専門知識の有無に関わらず、アセスメントの結果をアスリート

本人，ないしは監督，コーチといった関係者との間で共有することがより容易に実現できると推察される。

上記の提案手法を用いた処理により，アスリートの発話に基づく言語データは，二次元空間上へマッピングされ，特徴的な布置を示した。すなわち，マッピングの結果から，二次元空間上に心理的評価軸を見出すことができ，個人の心理状態とその変化が，直観的な方法で可視化された。得られた可視化結果は，対象者の心理状態の多様な側面を示すとともに，その変遷を動的に表現するものであった。さらに，マップ上での座標遷移に基づくクラスタリングにより，集団レベルでの心理変化のパターンを抽出し，その特徴をとらえることが可能となった。

今後の課題は，対象者の幅を広げ，言語データから抽出される心理的評価軸や，心理変化のパターンについて，集団間で比較を行っていくことである。多様な言語データ，ならびに分析結果の蓄積を通じて，より普遍的な評価軸や，心理変化のパターンを抽出することが必要である。

目次	
要旨	i
目次	iv
図目次	viii
表目次	xii
第 1 章. 緒論	1
1.1. 緒言	1
1.2. 研究の背景	1
1.2.1. 可視化研究の領域	1
1.2.2. 言語データの量的分析	2
1.2.3. 言語データの質的分析	4
1.2.4. 言語データの分析における課題	5
1.3. 研究の目的	6
1.4. 論文の構成	8
1.5. 結言	9
第 2 章. 分析の基礎事項	10
2.1. 緒言	10
2.2. 情報可視化	10
2.2.1. 一般的可視化研究について	10
2.2.2. テキストデータの可視化	11
2.2.3. 心理状態の可視化	14
2.3. テキストマイニング	18
2.3.1. テキストマイニングとは	18
2.3.2. テキストマイニングの解析手法	18
2.3.3. テキストマイニングの研究事例	20
2.4. 心理データの収集	23
2.4.1. 量的調査と質的調査	23
2.4.2. 心理的アセスメント	23

2.5.	感情の定義とその関連理論.....	26
2.5.1.	感情の定義	26
2.5.2.	感情に関連する理論.....	28
2.6.	本研究における解析手法	32
2.6.1.	自己組織化マップ	32
2.6.2.	ファジィクラスタ分析	34
2.7.	結言.....	39
第3章.	解析の基盤：心理学的トポロジー.....	40
3.1.	緒言.....	40
3.2.	処理の概要.....	40
3.3.	言語データの収集	41
3.4.	言語データからの Modal word の選定.....	45
3.5.	TF-IDF 値および TF-DF 値の計算.....	49
3.6.	自己組織化マップの作成	51
1)	代表ベクトルの初期値の設定	53
2)	入力ベクトルの分類	53
3)	代表ベクトルの更新	54
3.7.	ファジィクラスタ分析（ニューロンのクラスタリング）	55
3.8.	トポロジー的評価	56
3.9.	結果および考察.....	62
3.9.1.	Modal Word Mi の選定結果	62
3.9.2.	ファジィクラスタ分析によるニューロンのクラスタリング結果.....	64
3.9.3.	各クラスタの特徴に関する考察.....	70
3.9.4.	心理的評価軸の抽出.....	78
3.9.5.	アスリートの心理状態および変化の可視化.....	83
3.9.6.	位相空間上での座標遷移の可視化	92
3.10.	結言	94
第4章.	解析：心理変化のパターン抽出・可視化.....	96

4.1. 緒言.....	96
4.1.1. 処理対象データ.....	96
4.1.2. 心理変化の解析手法.....	96
4.2. 遷移ベクトルの定義.....	97
4.3. ファジィクラスタ分析（アスリートのクラスタリング）.....	101
4.3.1. 可達行列による解析.....	101
4.3.2. クラスタリングの結果.....	110
4.3.3. クラスタの特徴.....	114
4.4. 心理変化パターンの可視化.....	125
4.4.1. TF-IDF 空間上の遷移を示す TF-IDF ベクトル $\mathbf{G}c$ に基づく可視化.....	126
4.4.2. TF-IDF 空間上の遷移を示す TF-IDF ベクトル $\mathbf{G}'c$ に基づく可視化.....	130
4.4.3. TF-IDF 空間上の遷移を示す TF-IDF ベクトル $\mathbf{I}c$ の値に基づく可視化.....	132
4.4.4. TF-IDF 空間上の遷移を示す TF-IDF ベクトル $\mathbf{I}'c$ に基づく可視化.....	137
4.5. 結言.....	140
第 5 章. 結論.....	141
5.1. 緒言.....	141
5.2. 総合結論.....	141
5.3. 今後の研究への提言.....	145
5.4. 結言.....	146
謝辞.....	147
参考文献.....	148
本研究に関する講演論文および講演発表.....	153
本研究に関する原著論文.....	153
付録 A. 対話的な可視化システム.....	154
A.1. 緒論.....	154
A.2. アプリケーションの概要.....	154
A.3. アプリケーションの入力データ.....	155
A.4. アプリケーションの実行画面.....	157

A.5. 結論..... 168

目次

図 1.1	Text Pager に送信されたメッセージに基づく感情の可視化.....	3
図 1.2	「源氏物語」の内容の可視化.....	4
図 2.1	Word clouds の例	12
図 2.2	Text visualization by the Galaxies	13
図 2.3	STREAMIT	13
図 2.4	2変数間の関係性の可視化	15
図 2.5	外向的・内向的な人物が形成する社会的ネットワーク	16
図 2.6	形態素ラティスの例	19
図 2.7	MeCab による解析例	20
図 2.8	次元説における感情の布置	31
図 2.9	SOM の入力層および出力層	32
図 2.10	「暑い」の概念の定量化.....	35
図 3.1	心理学的トポロジーに基づく分析および可視化のプロセス	41
図 3.2	Modal word Mi が示す感情空間上の速度ベクトル(a)および心理的評価 軸(b),(c)	46
図 3.3	Modal word Mi の選定プロセス	47
図 3.4	言語データのベクトル表現と二次元平面へのマッピング	58
図 3.5	心理的評価軸の抽出プロセス.....	59
図 3.6	マップ上でのベクトル遷移の可視化	61
図 3.7	メンバシップ行列 F (TF-IDF 値)	65
図 3.8	可達行列 F (TF-IDF 値)	66
図 3.9	メンバシップ行列 F (TF-DF 値)	67
図 3.10	可達行列 F (TF-DF 値)	68
図 3.11	類似度 R に応じたニューロンのクラスタリング結果 (TF-IDF 値) ..	69
図 3.12	類似度 R に応じたニューロンのクラスタリング結果 (TF-DF 値) ..	69
図 3.13	類似度 $R = 0.76$ において形成されるクラスタ (TF-IDF 値)	71
図 3.14	CL1-5 に属するニューロンの代表ベクトル成分 (TF-IDF 値)	71

図 3.15	CL6-10 に属するニューロンの代表ベクトル成分 (TF-IDF 値)	72
図 3.16	類似度 $R = 0.79$ において形成されるクラスタ (TF-DF 値)	72
図 3.17	CL1-5 に属するニューロンの代表ベクトル成分 (TF-DF 値)	73
図 3.18	CL6-10 に属するニューロンの代表ベクトル成分 (TF-DF 値)	73
図 3.19	各クラスタにおいて顕著な値を示した Modal Word (TF-IDF 値) ...	75
図 3.20	各クラスタにおいて顕著な値を示した Modal word (TF-DF 値)	76
図 3.21	マップ上のクラスタの双対グラフ (TF-IDF 値)	77
図 3.22	マップ上のクラスタの双対グラフ (TF-DF 値)	78
図 3.23	マップ上に設定された心理的評価軸 (TF-IDF 値)	80
図 3.24	マップ上に設定された心理的評価軸 (TF-DF 値)	83
図 3.25	アスリート $c(=33)$ の心理状態および変化の可視化事例	85
図 3.26	アスリート $c(=47)$ の心理状態および変化の可視化事例	86
図 3.27	アスリート $c(=54)$ の心理状態および変化の可視化事例	87
図 3.28	アスリート $c(=33)$ の心理状態および変化の可視化事例	89
図 3.29	アスリート $c(=47)$ の心理状態および変化の可視化事例	90
図 3.30	アスリート $c(=54)$ の心理状態および変化の可視化事例	91
図 3.31	入力ベクトル \mathbf{H}_j の座標遷移を加えた双対グラフ	92
図 3.32	入力ベクトル \mathbf{H}'_j の座標遷移を加えた双対グラフ	93
図 4.1	マップ上の領域 $\mathbf{CL}t$	97
図 4.2	マップ上の領域 $\mathbf{CL}'t$	98
図 4.3	ベクトル $\mathbf{G}c$ のメンバシップ行列 \mathbf{F}	102
図 4.4	ベクトル $\mathbf{G}c$ の可達行列 \mathbf{F}	103
図 4.5	ベクトル $\mathbf{G}'c$ のメンバシップ行列 \mathbf{F}'	104
図 4.6	ベクトル $\mathbf{G}'c$ の可達行列 \mathbf{F}'	105
図 4.7	ベクトル $\mathbf{I}c$ のメンバシップ行列 \mathbf{F}''	106
図 4.8	ベクトル $\mathbf{I}c$ の可達行列 \mathbf{F}''	107
図 4.9	ベクトル $\mathbf{I}'c$ のメンバシップ行列 \mathbf{F}'''	108
図 4.10	ベクトル $\mathbf{I}'c$ の可達行列 \mathbf{F}'''	109

図 4.11	各領域 CLt において顕著な値を示したキーワード	115
図 4.12	クラスタ C1 および C2 ($R = 0.70$) における平均ベクトル成分 wt	116
図 4.13	クラスタ B1 ($R = 0.81$) における平均ベクトル成分 wt	116
図 4.14	クラスタ A2, A5, A7, A8, A11 ($R = 0.90$) における平均ベクトル成分 wt	117
図 4.15	クラスタ A14, A15, A16, A18 ($R = 0.90$) における平均ベクトル成分 wt	117
図 4.16	クラスタ C1 ($R = 0.70$) における平均ベクトル成分 wt	118
図 4.17	クラスタ A1 ($R = 0.90$) における平均ベクトル成分 wt	119
図 4.18	各領域 $CL't$ において顕著な値を示したキーワード	120
図 4.19	クラスタ C1 ($R = 0.70$) における平均ベクトル成分 $w't$	121
図 4.20	クラスタ B1, B2, B7 ($R = 0.81$) における平均ベクトル成分 $w't$	122
図 4.21	クラスタ A1, A2, A4, A8, A9, A10 ($R = 0.90$) における平均ベクトル成 分 $w't$	123
図 4.22	クラスタ A13, A15, A19, A20, A22, A24 ($R = 0.90$) における平均ベク トル成分 $w't$	123
図 4.23	クラスタ C1 ($R = 0.70$) における平均ベクトル成分 $w't$	125
図 4.24	クラスタ A1 ($R = 0.90$) における平均ベクトル成分 $w't$	125
図 4.25	ベクトル Gc が形成するクラスタ C1 ($R = 0.70$) およびクラスタ B1 (R $= 0.81$)	127
図 4.26	ベクトル Gc が形成するクラスタ A2, A5, A7, A11, A14, A16, A28 ($R =$ 0.90)	128
図 4.27	ベクトル Gc が形成するクラスタ B2 ($R = 0.81$) およびクラスタ A8, A15 ($R = 0.90$)	129
図 4.28	ベクトル $G'c$ が形成するクラスタ C1 ($R = 0.70$) およびクラスタ A1 (R $= 0.90$)	131
図 4.29	ベクトル Ic が形成するクラスタ C1 ($R = 0.70$) およびクラスタ B1, B2, B7 ($R = 0.81$)	133

図 4.30	ベクトル \mathbf{Ic} が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A1, A4, A9 ($R = 0.90$)	134
図 4.31	ベクトル \mathbf{Ic} が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A2, A10, A13 ($R = 0.90$)	135
図 4.32	ベクトル \mathbf{Ic} が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A8, A15, A19, A20, A22, A24 ($R = 0.90$)	136
図 4.33	ベクトル \mathbf{Ic} が形成するクラスタ A1 ($R = 0.90$)	139
図 A.1	入力データの形式	156
図 A.2	コントロールパネル	158
図 A.3	マップ	158
図 A.4	データ読み込み後のマップ	159
図 A.5	マップ上での表示成分の選択	161
図 A.6	SOM の解析結果 (Display mode 1)	162
図 A.7	SOM の解析結果 (Display mode 2)	163
図 A.8	Element の選択的表示	164
図 A.9	SOM の解析結果 (Display mode 2, Element 1-3 を選択)	165
図 A.10	SOM の解析結果 (Display mode 3)	166
図 A.11	マップ上での座標遷移を示すアニメーションのイメージ	167
図 A.12	マップ上のニューロンの移動	168

表目次

表 2.1	先行研究における基本感情のカテゴリ	30
表 3.1	アスリート c(=3)の発話における各 Modal Word (M1 ~ M10) の出現頻度	50
表 3.2	アスリート c(=3)の発話における各 Modal Word (M11 ~ M20) の出現頻度	51
表 3.3	選定された Modal word Mi	62
表 3.4	感情の評価軸上における極性.....	63
表 4.1	Gc に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)	110
表 4.2	G'c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)	111
表 4.3	Ic に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)	111
表 4.4	I'c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)	112

第1章. 緒論

1.1. 緒言

本章では，研究背景として，関連領域における先行研究と研究課題を示した上で，新たに提案する手法と研究の目的，研究対象について述べる．また，本章末で本論文全体の構成に関して説明する．

1.2. 研究の背景

1.2.1. 可視化研究の領域

情報の視覚的な表示方法，あるいはその可視化のための自動化技術は，様々な領域において重要な役割を果たしており，これまで多くの研究者が可視化技術の研究開発に取り組んできた．可視化技術は，数値データの並びなどそのままでは直感的に把握しにくい，あるいはそのまま直接目視しただけでは，その属性を読み取ることができないような物理的対象の特性を視覚的に捉えることができるようにするために用いられる．

一般に可視化の研究は，科学的可視化と情報可視化という2つのアプローチに大別される．前者は，自然現象や生体現象，分子運動といった，物理的実体をもつデータの視覚的表現による可視化である．これに対し，後者は，数量データやテキストデータを対象とした可視化である．本論では，後者の情報可視化に焦点を当て，特に言語データの分析とその結果の可視化の手法を研究の対象とする．

可視化研究の適用領域は拡大し，自然現象やコンピュータ上で処理される情報にとどまらず，人々の行動や感情，関係性までもが，定量的かつ視覚的に表現されるようになった．このような傾向は，人間を中心としたデータの測定および解析，可視化技術の発展に基づくものである．同時に，多様化が進む現代社会においては，個人や集団が直面する心理社会的な問題も様々であり，個々人の価値観やニーズ，心理的特性などを把握した上で，適切な対処へつなげる必要がある．心理領域への可視化の応用は，こうした社会的要請にも応えながら，発展が期待される研究といえよう．

1.2.2. 言語データの量的分析

調査対象者の語りや記述を記録した言語データ、ないしはテキストデータは、複雑かつ多様な構造をもつため、これらの分析に特化した様々な解析手法が、テキストマイニングの領域で開発されてきた。例えば、新聞記事(八塚, 2007)や日記(Cavicchiolo et al., 2015)、ソーシャルメディアへの投稿(Cohn et al., 2004)など、様々な形式のテキストデータが、客観的かつ定量的な手法で分析されている。また、面接を通じて得られる言語データ(Seale et al., 2006)も、同様に分析の対象となりうる。これらのデータからは、そこに記載された客観的な事実に加えて、語り手や書き手の心情および態度、パーソナリティといった、個人の内面に関わる情報をも得ることができる。

言語データやテキストデータから、心理状態に関わる情報を抽出するための分析技術が発展するとともに、これらの情報を効果的に可視化する手法についても、注目が集まるようになった。例えば、Back et al. (2010) は、9.11 アメリカ同時多発テロ事件の発生時に、アメリカ国内で Text Pager (アメリカにおける無線呼び出しサービス) へ送信されたメッセージに注目し、人々の感情を示すキーワードの割合を算出した。

図 1.1 は、Anger (e.g., hate, annoyed), Sadness (e.g., crying, grief), Anxiety (e.g., worried, fearful) という 3 種類の感情を示すキーワードの増減を、折れ線グラフで表したものである。折れ線グラフは、事件の発生直後から 18 時間に渡って、怒りに関するキーワードが増加していく様子を示している。また、井波ら(2013) は、源氏物語のテキストを対象に、対立概念を示すキーワードの頻度分布に注目し、感情的な側面から物語の内容を可視化する手法を提案した。図 1.2 では、縦軸と横軸がいずれも作品中の時間軸に相当する。図 1.2(a)の左側では、「恋」や「愛」、「冷淡」や「恨めしい」といった相反する感情が分化しているのに対し、右側の中央部では、これらの感情が混在する形で、作品中に表現されていることがわかる。

先行研究の特徴として、対象となる集団の記述、あるいは文学作品を構成するテキストの群を対象としている点が挙げられる。すなわち、心理状態やその変遷の集団的・全体的な特徴が示されている一方で、個人の心理的特徴に焦点を当てた可視化はなされていない。個人に対する心理的支援の重要性を考慮すれば、個人単位での心理状態の分析および可視化の手法について、検討の余地があるといえる。

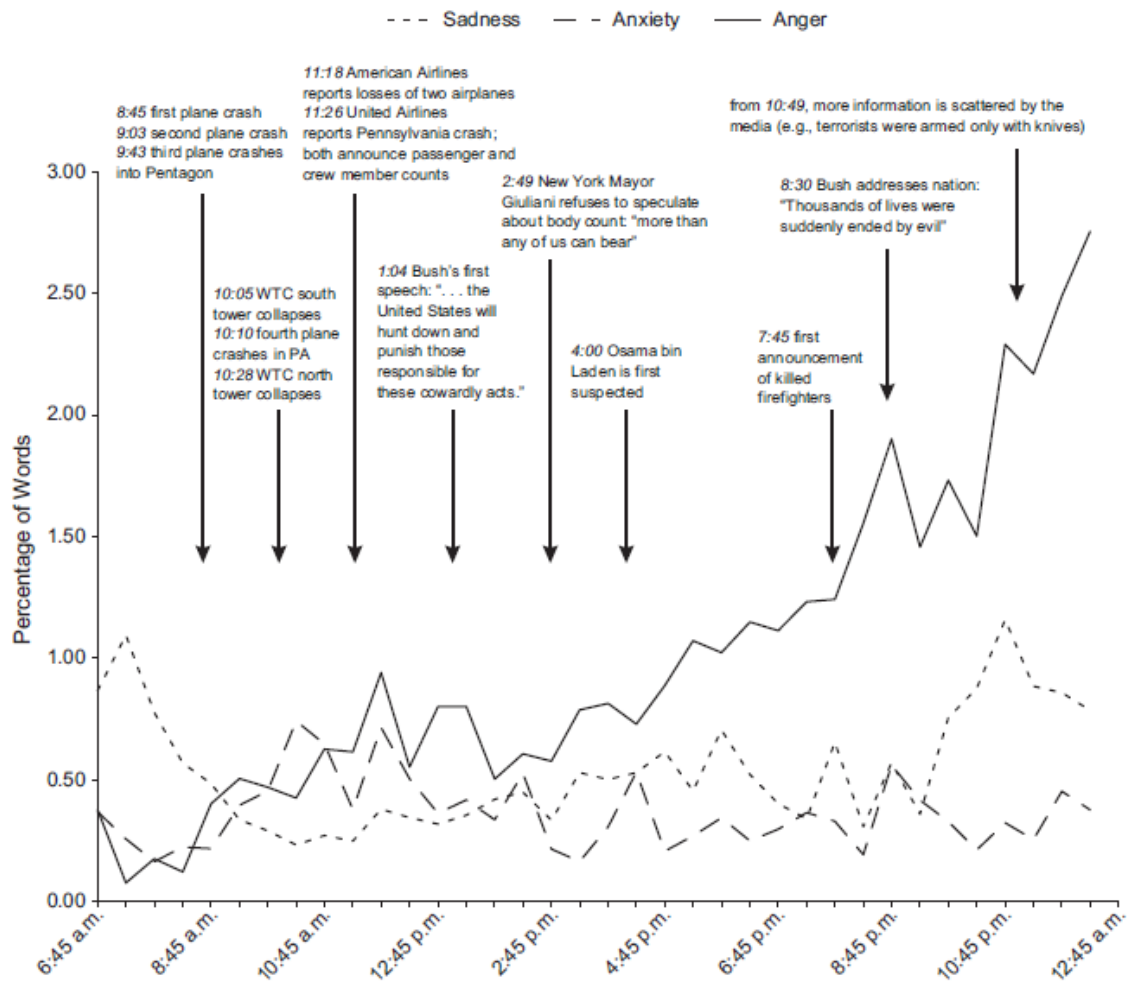


図 1.1 Text Pager に送信されたメッセージに基づく感情の可視化

(文献 (Back et al., 2010) より転載)

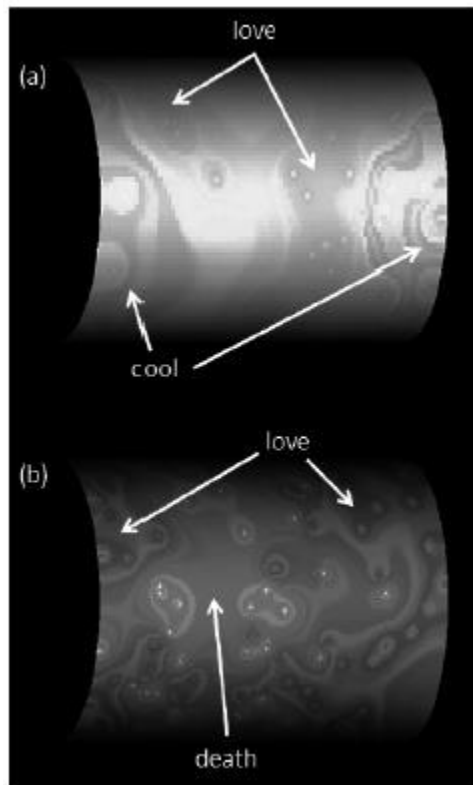


図 1.2 「源氏物語」の内容の可視化
 (文献 (井波ら, 2013) より転載)

1.2.3. 言語データの質的分析

これまで、心理学や社会科学の領域においても、語りや記述から対象者の内面を明らかにするアプローチがなされてきたが、これらは質的な研究手法に基づくものが中心であった。質的研究においては、理論的な枠組みに基づいて言語データの分析と解釈を行うことで、人手による分析結果から主観性を排除することが必要となる。分析手法としては、グラウンデッド・セオリー・アプローチや、KJ法などが用いられることが多い。例えば、専門家にヘルスケアプログラムを受講してもらい、体験についてのインタビューを行った研究 (Irving et al., 2014) や、消費者から寄せられたコメントをもとに、製品に対するニーズの抽出・分類を行った研究 (NI et al., 2017) において、これらの手法が採用されている。

言語データを対象にした場合、質的な分析を行うメリットは、具体的な個人の内面を個別に、より詳細に把握できる点にある。その半面、質的研究におけるデータの分

析と解釈には多くの時間を要するとともに、妥当性・信頼性の高い分析結果を得るためには、高度な専門性が求められる。そのため、研究実施にあたり分析者の負担が大きくなることや、データの分析および解釈内容について、専門外の人と理解を共有することが困難になることが、課題として挙げられる。

1.2.4. 言語データの分析における課題

言語データに対して、定量的な分析と可視化を行う手法を確立することは、上に挙げた質的研究の課題を補完するという点でも、意義があるといえるだろう。テキストマイニングは、質的研究と量的研究の特徴を兼ね備えた分析手法とされており（小平ら、2007）、言語データが有する多様な情報を効率的に抽出し、効果的に可視化するための補助的（前处理的）な手法として用いることが期待できる。価値観の多様化が進む現代では、クライアントや消費者など、対象となる個々人の心理的な特質をより深く理解し、適切な支援やサービス、製品開発へとつなげることが求められる。そのために、分析結果に対する解釈や洞察を促進するツール、あるいは、専門外の人と理解を共有するためのツールとして、言語データの分析・可視化手法の確立が望まれる。

言語データやテキストデータの特徴を定量的にとらえ、解析に用いるための手法としては、ベクトル空間モデル（Salton, et al., 1975）による手法がよく知られている。これは、分析対象となる個々のデータ（文書）を多次元空間上のベクトルデータとして表現する手法である。ベクトルデータとすることで、数式などを用いた数理的なモデルを適用することができる。このベクトル化においては、それぞれの文書に含まれる語句、ないしはキーワードの出現頻度を求め、これらをベクトルデータの成分とする。このような文書のベクトルデータ化によって、特定の文書を検索したり、類似度に応じて文書のクラスタリングを行ったりといったコンピュータによる処理が可能になる。このように、言語データの定量的な分析を行う上では、個々のデータをベクトルの形式で表現することが重要な鍵となる。

多次元データとして表される言語データを、二次元または三次元の平面上で可視化するためには、次元の削減が不可欠である。そのために、従来は因子分析や主成分分析、多次元尺度構成法といった、統計解析の手法が用いられてきた。これらの手法に

より、多次元データは二次元または三次元の直交座標系にマッピングされ、視覚的に表現することが可能となる。しかしながら、次元の削減にあたっては、データが有する情報の損失が避けられない。また、これらの手法は、データの線形分布を前提とした線形解析であり、非線形データの場合に、その特徴が解析結果に反映されない恐れがある。

1.3. 研究の目的

本研究で扱う言語データは、対象となる集団・個人の内面に関わる多くの情報を含んでおり、その定性的な性質から、従前では質的な研究やアセスメントの対象となることが多かった。近年は、言語データを含む、人々の語りや記述を取り扱うテキストマイニングの発展により、定量的な解析の手法も提案されている。言語データを定量的に取り扱う場合、特徴として含む情報の多さから、多次元データとしての性質をもつことが多い。言語データの解析を通じた、心理状態の可視化について検討する際には、データの多次元性を考慮する必要がある。

言語データに表される、心理状態の多様な性質を効果的に表現するためには、多次元の精度を損なわないような分析および可視化手法が不可欠と考えられる。この点を考慮し、本研究では、言語データに基づき個々人の心理状態とその変化を可視化するために、自己組織化マップ (Self-Organizing Map: SOM, Kohonen, 1995) とファジィクラスタ分析 (Shinkai, 2008) を組み合わせて用いることにした。SOM は、多次元空間上のデータを、個々のデータ間の位置関係を維持しながら、低次元空間へ非線形的にマッピングする解析手法である。これにより、言語データを含む多次元データの関係性を、二次元平面上で表現することが可能となる。さらに、SOM により得られた二次元平面上のマッピング結果に対し、ファジィクラスタ分析により、マップ上の領域についてクラスタリングを試みる。すなわち、多次元データの関係性を示す二次元のマップを、その特徴に応じていくつかの領域に分けることで、マップ上にデータの評価軸を設定することが可能となると考えた。

言語データが示す心理の多様な特質を、心理学的なトポロジーに従って、二次元平面上で可視化することが本研究の目的である。

具体的には、以下の2段階のプロセスにより、研究目的へのアプローチを試みる。第一に、SOMとファジィクラスタ分析という非線形解析の手法を組み合わせ、多次元の言語データをマッピングすることで、二次元平面上に表れる心理状態の評価軸（心理的評価軸）を抽出する。非線形解析により得られる心理的評価軸は、従来の線形解析では見出されなかった、言語データ間の心理学的なトポロジーを表現するものと想定される。第二に、集団の心理変化を心理的評価軸上の遷移として表現し、集団内に共通してみられる心理変化のパターンを明らかにする。

この2段階の取り組みを通じて、個人および集団の心理状態とその変遷を、視覚的にわかりやすく表現することが可能となる。また、特定の心理状態と可視化パターンとの関係性に基づき、対象者の心理的特徴を直観的に理解することも可能となり、心理的アセスメントなどへの活用が期待できる。

本研究において、その言語データを収集し、心理状態の可視化を行う対象となるのは、オリンピック・パラリンピック選手を含む、各競技のトップアスリートである。アスリートを対象とした分析、ならびに可視化を行う理由の一つとして、近年、スポーツ領域における心理的コンディショニングの重要性が増していることが挙げられる。スポーツにおけるパフォーマンスの発揮には、「心・技・体」という言葉に表されるように、身体能力や競技に特化した技術だけでなく、心理面を適切にコントロールする能力が不可欠である。トップアスリートの中には、集中やリラクゼーションの獲得、モチベーションの維持などをもたらす心理的スキルを獲得するために、メンタルトレーニングを導入する選手もみられる。メンタルトレーニングの実施においては、個人の有する心理的スキルを評価するためのアセスメントが不可欠であり、様々な評価手法が考案されてきた。著者らは、心理的スキルの量的なアセスメント手法として、質問紙を用いた調査を行うとともに、競技集団ごとに適切な評価指標を得るための分析手法を提案している（Aoki et al., 2017）。これに対し、本研究では言語データを分析の対象として採用することで、アスリートの内面をより詳細に理解することを想定している。さらに、分析結果を心理学トポロジーに基づいて効果的に可視化することにより、専門知識の有無に関わらず、アセスメントの結果をアスリート本人、ないしは監督、コーチといった関係者との間で共有することがより容易に実現できると推察

される。スポーツ領域においても、心理データの活用が注目を集めている現状を踏まえれば、本研究の提案手法をアスリートに適用し、心理状態の可視化に関する知見を得ることには、一定の意義があると思われる。

1.4. 論文の構成

本論文の構成は次の通りである。まず、第1章では、本研究の研究背景および提案手法、研究目的、研究対象を示し、最後に論文全体の構成について触れる。

第2章では、本研究の分析に関わる基礎事項として、情報可視化およびテキストマイニングの領域における知見と、心理学におけるデータ収集の方法、感情の定義について説明する。また、分析に用いた解析手法である、自己組織化マップおよびファジィクラスタ分析の基本概念と用語の定義を示す。

第3章および第4章においては、本研究における提案手法の内容と、その実行結果および考察、得られた結論について記す。提案手法は、心理学的トポロジーの導入(第3章)、集団における心理変化のパターン抽出・可視化(第4章)という二段階のプロセスにより、言語データの分析および可視化手法の確立と、可視化結果から見出される心理変化のパターン抽出を試みるためのものである。第3章(Aoki et al., 2018a)、第4章(Aoki et al., 2018b)にて報告された研究内容は、いずれも2018年にオンラインジャーナル“Psychology (<http://www.scirp.org/journal/psych/>)”に掲載された、著者らの論文(Aoki et al., 2018a, 2018b)に基づくものである。

まず、本研究における心理学的トポロジーとは、物理的実体をもたない心理状態間の関連性や、そこから見出される心理状態の評価軸など、個人や集団の特質を相対化して観測するための心理的空間における計量的枠組みを表す。座標を入れてベクトル空間化し、さらに距離を入れて位相空間化した心理的空間上へのデータのマッピングと、可視化の手法を提案するのが、本研究における第一の取り組みである。

また、心理学的トポロジーをベースにして表現された、個々人の心理状態とその変化に関する可視化結果から共通のパターンを抽出し、解釈を試みる。これにより、個人が示す可視化のパターンと、集団に典型的なパターンとを比較しながら解釈を行うといった分析方法へつなげることができる。すなわち、提案手法の重層的な活用を見

据えて、集団に対する可視化を行うのが、本研究における第二の取り組みである。

第8章では、上記の提案手法に着手する前に行った、心理状態の可視化のための対話的な可視化システムの開発の成果について報告する。具体的には、SOMによる解析結果を視覚的に表示するアプリケーションの開発と、その機能および実行に関わる詳細を示し、明らかになった課題について述べる。

以上の研究成果を踏まえ、明らかになった提案手法の有効性と課題、今後の発展の方向性などを、結言（第5章）として述べる。また、付録として、本研究への着手にあたり作成した、心理状態の可視化システムのプロトタイプとなるアプリケーションの詳細について記載する。

1.5. 結言

本章では、本研究に関わる研究領域として、情報可視化および言語データの量的・質的分析に焦点を当て、同領域における先行研究と研究課題について説明した。また、これらの研究背景を踏まえた本研究の提案手法と目的、研究対象について述べた上で、論文全体の構成を示した。

第2章. 分析の基礎事項

2.1. 緒言

本章では、研究の準備として、第3章・第4章における処理に関わる内容として、情報可視化、テキストマイニング、心理データの収集、感情の定義とその関連理論について説明する。また、第3章・第4章の処理で用いる解析手法として、自己組織化マップおよびファジィクラスタ分析について述べる。

2.2. 情報可視化

本節では、研究の目的となる「情報可視化」に焦点を当て、関連する先行研究について述べる。具体的には、テキストデータや、多次元データの可視化に加え、心理状態の可視化に関する研究を取り上げ、本研究との比較を行う。

2.2.1. 一般的可視化研究について

目に見えない現象や事象を見えるようにする、あるいは、一見すると特徴をとらえにくい情報を、視覚的にわかりやすく提示する、といった試みは総称して、可視化といわれている。いずれも、視覚から多くの情報を得て判断を行う、人間の認知的特性に即した情報の分析と提示の手法であるといえる。情報化社会の進展に伴い、昨今では、様々な領域で多くのデータを効率的に収集し、分析と評価を行い、効果的に伝達することが求められている。このような社会構造の変化を受けて、可視化の技術開発や活用に関わる研究は、今後も発展を遂げていくことが予想される。

現在、可視化研究は、科学的可視化と情報可視化とに大別され、それぞれが対象とするデータや、適用される領域も異なっている。科学的可視化は、自然現象や生態減少、分子運動など、物理的実体をもったデータを対象とする。これらのデータは、空間座標に対応付けられた値をもち、3次元空間上へのマッピングが可能である。一方、情報可視化は、数量データやテキストデータなど、物理的実体をもたないデータを対象とする。これらのデータは、空間座標と対応付けられる値をもたず、可視化を行う上でいずれの値に注目するかは、分析者の手に委ねられる。

本研究で行う可視化は、対象者の語りから得られる言語データを対象としており、後者の情報可視化に該当する。言語データの解析結果に対して、心理学的な視点に基づく評価軸を抽出し、可視化を試みる点に本研究の特色がある。

以下の節では、情報可視化の中でも、本研究と関連のある、テキストデータの可視化と、心理状態の可視化に関わる先行研究を概観する。

2.2.2. テキストデータの可視化

膨大な量の文書が電子化され、世界中でやりとりされるようになって以来、テキストデータの解析と可視化技術は、様々な領域で関心を集めている。データマイニングから派生したテキストマイニング技術の発展に伴い、テキストデータの集合から抽出された情報を効果的に表現するために、多くの可視化手法が確立されてきた。

テキストデータの可視化は、特定の時点における静的なデータを対象とするか、時系列的な変化を記録した動的なデータを対象とするかによって、用いられるアプローチや技術が異なる。前者の静的なテキストデータの可視化技術は、特徴ベースの可視化、あるいはトピックベースの可視化という観点から分類することができる (Liu et al., 2014)。

特徴ベースの可視化は、テキストに含まれる語句のまとまりや、文章構造の把握などを目的とした研究である。代表的な技術として、文書中に含まれる語を、出現頻度に応じて大きく表示する Word clouds がよく知られている (Wu et al., 2011)。図 2.1 は、カンファレンスの抄録に含まれるキーワードを Word clouds で表したもので、(a), (d) は 1999 年、(b), (e) は 2005 年、(c), (f) は 2010 年のデータを対象としている。

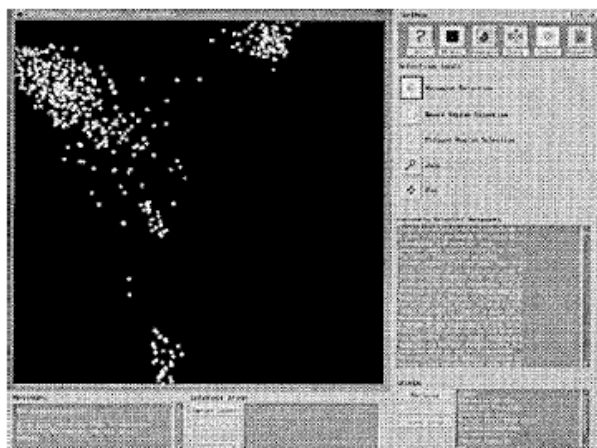


図 2.2 Text visualization by the Galaxies

(文献 (Wise, 1999) より転載)

時間的な属性をもつ動的なテキストデータの可視化は、文書内容の時間的な変遷をとらえる上で重要な試みである。例えば、Alsakran (2011) らが開発した、対話的な可視化システム“STREAMIT” (図 2.3) は、時間的な変化とともに変わるクラスタ構造は、画面上で近い位置に配置される。また、複数の文書がクラスタを形成する場合には、それらを同系色の色で同じクラスタであることを表示する。画面上には、時間の進展とともに新たな文書が追加されていき、その都度データ間の関連性やクラスタ構造が変化する。その様子をアニメーションで表現するのが、同システムの特徴である。

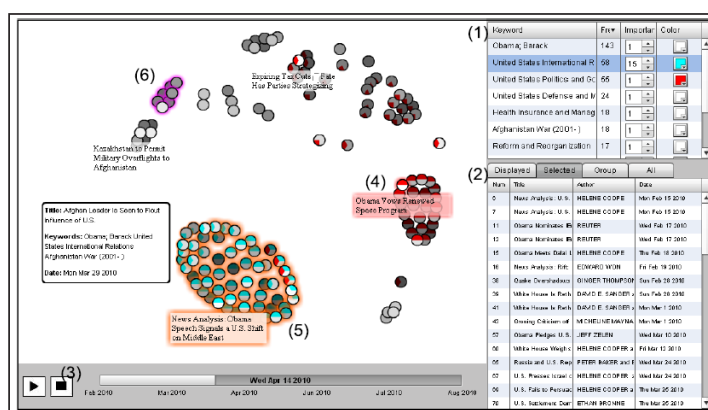


図 2.3 STREAMIT

(文献 (Alsakran, 2011) より転載)

2.2.3. 心理状態の可視化

心理学の研究者、あるいは心理臨床に携わる専門家は、被験者やクライアントから得たデータに基づき、個人ないしは集団に対する心理的な評価を行う。対象となる人々の内面について、深い理解や新たな洞察を促し、関係者との情報共有を円滑に進めるため、データを分かりやすく視覚的に表現することは重要である。また、心理学領域の研究では、統計解析の手法が数多く用いられてきた。そのため、統計データの可視化手法の確立にあたり、心理学が果たした役割は大きいと考えられる。

例えば、図 2.4 の上側の散布図は、横軸が個々人の情緒性、縦軸が犯罪に対する恐れの高さを表している。散布図上のデータの分布からは、両データが正の相関関係にあることが読み取れる。しかしながら、こうした従来のグラフ表現は、膨大かつ複雑なデータや、時系列データの可視化には不適切との指摘もある (Heer et al., 2012)。そこで、Web アプリケーションフレームワークを活用し、心理データを動的に可視化する試みがなされている (Ellis, 2015)。図 3.4 の下側の散布図は、Web サイト上に設置されたアプリケーションの実行画面である。サイトの訪問者は、グラフ上にプロットされたデータを男性と女性の層別に表示したり、グラフの縦軸・横軸をなす変数を自由に切り替えたりして、2 変数間の相関関係を確認することができる。このように、統計解析に基づく従来の可視化表現を、アプリケーションなどを通じて拡張していくことも、1 つの発展の方向性といえる。

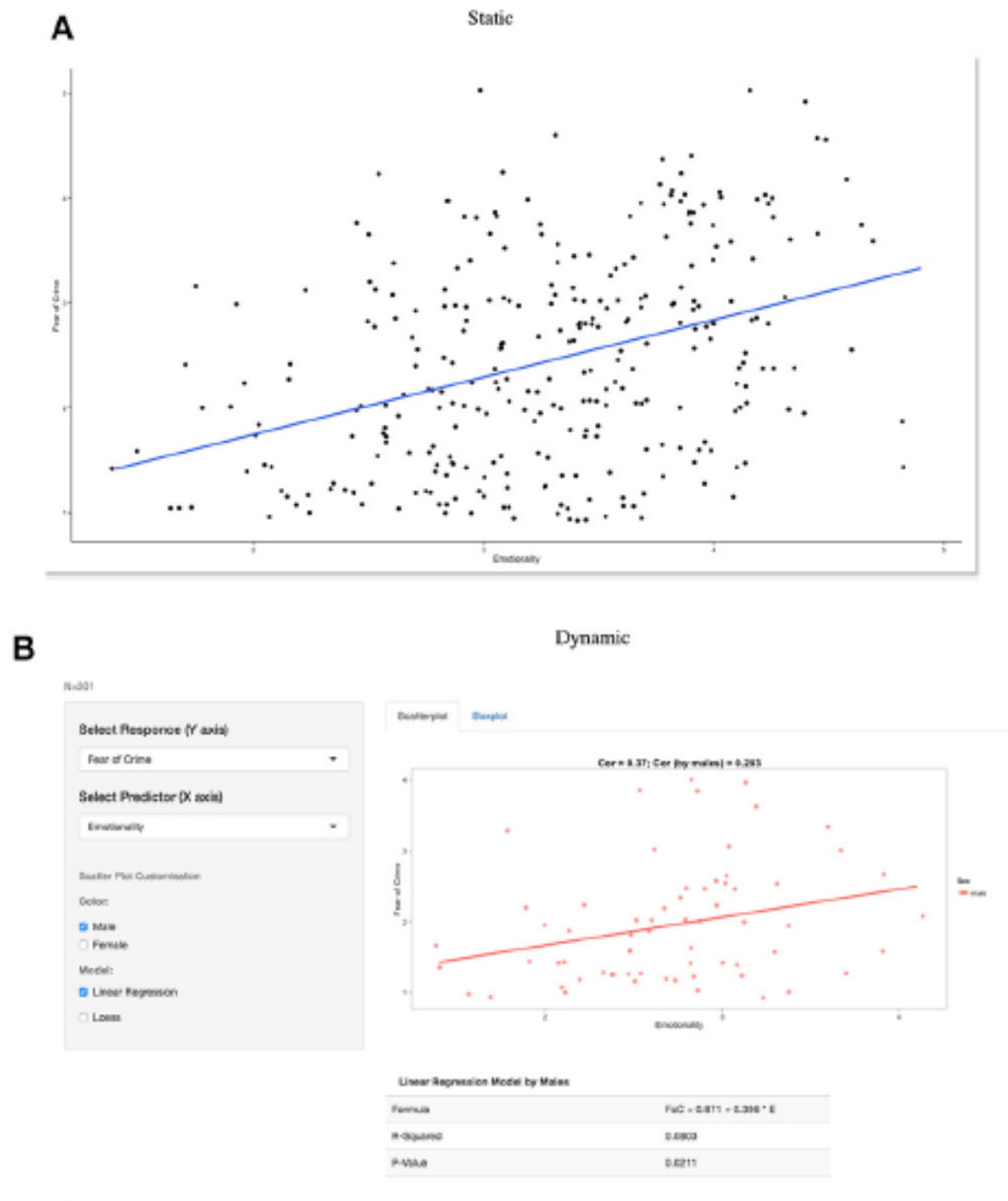


図 2.4 2変数間の関係性の可視化

(文献 (Ellis, 2015) より転載)

これに対し、個人および集団が示す心理社会的な性質を、統計解析とは異なる方法で解析し、可視化を行った研究も見受けられる。例えば、Friggeri (2012) らは、個々の性格特性と、SNS 上で形成される社会的ネットワークの構造との関係性に注目し、

データの分析および可視化を行った。可視化に用いられたデータは、SNS 上で実施が可能な性格検査の結果と、同 SNS における友人関係に関するデータである。図 2.5 の左側は、外向的な人物が形成したネットワーク構造を、右側は、内向的な人物が形成したネットワーク構造を可視化したものである。1つ1つの丸は SNS 上での友人を表し、同じコミュニティに所属する友人は、同じ色で表現される。下図より、外向的な人物は幅広いコミュニティに所属しており、いずれのコミュニティも所属する人員の数は少ない。これに対し、内向的な人物は少数のコミュニティに所属し、その多くが、多数の人員により構成されていることがわかる。

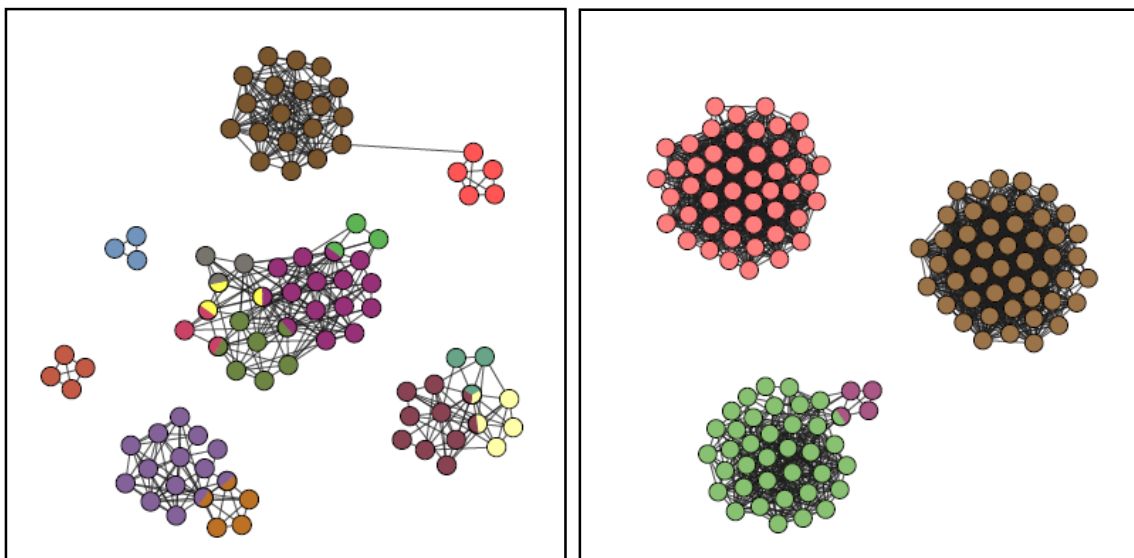


図 2.5 外向的・内向的な人物が形成する社会的ネットワーク

(文献 (Friggeri, 2012) より転載)

このように、心理学領域においても、データ分析を支援する技術として、新たな可視化手法の開発に着手した研究が行われつつある。今後は、分析対象や解析手法の多様化に伴い、従来のグラフとは異なる可視化表現の重要性も増していくであろう。

上で示した先行研究は、いずれも情報可視化に関わる研究であり、扱うデータは、空間的な位置に対応する値をもたない。それゆえに、個々のデータを位置づける座標軸を設定せず、データ間の関連性の読み取り易さに焦点を当てた研究が中心である。

‘図 3.4’に示すような二次元グラフを用いる場合、扱うことのできるデータは2変数に

限定され、テキストデータや言語データのような、多次元データの特徴を表現することは困難である。

本研究では、言語データに対する非線形解析を実施し、多次元データ間の関連性を二次元のマップ上で可視化する。このとき、平面上に評価軸を設定し、座標軸上で個々の心理状態の評価を行う点に、独自性があるといえる。提案手法は、個々人が二次元平面上で示す座標の遷移、すなわち心理変化を、物理的な運動になぞらえて表現するものである。こうした点で、科学的可視化が表す物理現象のように、より自然な形で心理の動きをとらえ、直観的に理解することが期待できる。

2.3. テキストマイニング

本節では、言語データの定量的な解析に用いられるテキストマイニングの手法と、先行研究の説明を行う。具体的には、本研究で用いた形態素解析の詳細について述べ、本研究に関連する研究事例として、テキストに示される感情に焦点を当てた先行研究を取り上げる。

2.3.1. テキストマイニングとは

テキストマイニングは、データマイニングから派生して用いられるようになった用語であり、構造化されていないテキストから、目的に応じて情報や知識を掘り出す方法と技術の総称である(金, 2012)。広義には、文書集合から必要な文書を検索する技術や、あらかじめ設定されたカテゴリへの分類、内容の類似性に基づくクラスタリングなどの技術も含まれる(那須川, 2006)。しかしながら、マイニングという言葉がもつ意味を踏まえれば、情報検索のような既知の情報獲得ではなく、膨大な文書データから新規の情報を発見する技術こそが、真のテキストマイニングであると、例えば Hearst (1999) は主張している。

2.3.2. テキストマイニングの解析手法

テキストマイニングの対象となるテキストデータは、自然言語で書かれた質的なデータであり、これらを定量的な解析に用いるための処理が必要となる。那須川(2006)は、文書の記述内容を統計的に分析可能な状態にするためのプロセスを自然言語処理、統計的な分析を行うプロセスをマイニング処理と呼び、この2つの処理を区別している。本節では、自然言語処理の代表的な手法であり、本研究においても実施した形態素解析について、具体的な解析事例を挙げながら解説する。

形態素解析とは、文章を構成する文単位を対象とした解析であり、それぞれの文を形態素(言語が意味をもつ最小単位)に分割した後に、その品詞を同定する処理のことである。コンピュータ上で形態素解析を行うために、様々なソフトウェアが開発されているが、本研究では、テキストマイニング用ソフトウェア「Tiny Text Miner: TTM」(松村ら, 2009)において、形態素解析エンジン「MeCab」を呼び出して実行し、解

析を行った。

MeCab では、コスト最小法と呼ばれる手法を用いて、形態素解析を実施する。最小コスト法とは、一つ一つの語や語の接続にコストを与えた辞書を作成し、総コスト数が最小となるように文を形態素に分割する手法である（村田ら、2012）。このとき、使われることが少ない単語や、接続することがない単語、品詞間には高いコストが設定される。そのため、総コスト数が低いほど、より適切な分割結果であると判断できる。

コスト最小法では、与えられた入力文字列と辞書に含まれる形態素を比較し、可能な分割のパターンを網羅した形態素ラティスを作成する（工藤、2005）。ここで、形態素ラティスの例を図 2.6 に示す。図中の四角で囲われた語が、形態素の分割パターンを示しており、語と語の間にはコストが設定されている。b を文の開始状態、e を終了状態とすると、その間には 6 通りの経路が存在する。それぞれの経路について、コストの合計値を算出し、最も値の小さい経路が形態素解析結果として出力される。

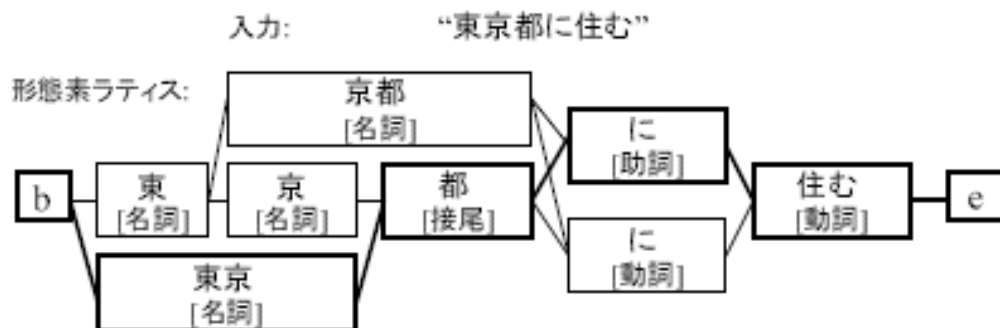


図 2.6 形態素ラティスの例

（文献（工藤，2005）より転載）

ここで、本研究の分析に用いたインタビュー記事の一文について、形態素解析を行った結果を図 2.7 に示す。画面は MeCab による解析結果の出力画面であり、入力文（“僕は中学時代から陸上部に所属して長距離走をやっていました”）を個々の形態素に分割した結果と、それぞれの品詞が表示されている。

テキストの分類を行う方法と、極性のラベルが付与されたデータを用いて機械学習を行い、学習結果に基づいて分類を行う方法などがある。前者の例として、Taboadaら(2011)は、単語の極性とその強度に関する辞書をもとに、語の強調(e.g., very good)および否定(e.g., not good)の表現を考慮に入れながら、様々なジャンルのユーザレビューが示す感情を分析するシステムを開発した。また、後者の例として、Sharmaら(2012)は様々な特徴選択の指標と機械学習の手法を組み合わせ、映画のレビュー文に対する感情分析を行った結果、増分比(Gain Ratio)を用いてサポートベクターマシンによる分類を実施した場合に、高い分類制度を得ることを示している。

その他にも、同様の研究として、テキストからの感情抽出に関する研究がある。感情分析と比較すると、テキストが示す感情をより細やかに分類したり、感情の強度を数値化したりする点で、心理状態に関するより詳細な情報を得るための技術であるといえる。感情抽出のアプローチは、テキストを文単位で分析するものと、語の単位で分析を行う方法とに分けることができる。前者の例として、感情を示すタグを付与した文章のデータベース(感情コーパス)を作成し、これを学習データとして機械学習を行い、対象となる文の感情を数値化する研究がある(McCallum, 1998)。また、後者の例として、菅原(2009)らは、感情に関係する語のデータベースを作成し、それぞれの語に感情のカテゴリと、その強度に関する情報を付与している。与えられたテキストに対して、上記のデータベースを参照することで、テキスト中に含まれる語を特定し、感情の強度を示すスコアを算出できる。

以上のように、テキストマイニングの分野においても、テキストから抽出可能な情報の1つとして、人々の感情が扱われていることが分かる。これらの研究と、本研究との違いは、個々のデータを心理学的な観点からとらえ、データ間のトポロジー的な性質を解明しようとする点にある。テキストマイニングを用いた先行研究では、感情を表すキーワードや文章を定義する際に、心理学において提唱された感情のカテゴリや、評価次元を参照している場合が多い(Plutchik, 1980; Russell, 1980)。しかしながら、テキストや言語データは、量的なデータと比較して、人々の内面に関するより多くの情報を含むものである。これらのデータに対して、定量的な解析技術が発展してきた現在では、人々の記述や語りから、新たな心理の特質が見出される可能性がある。

そのため、本研究では、分析の指標となる心理状態を規定せず、解析および可視化の結果から心理的評価軸を抽出する、探索的なアプローチを行う。これにより、言語データ収集の利点を最大限に活かした、心理状態の分析および可視化手法の開発につながると考えるに至った。

2.4. 心理データの収集

本章では、対象者の心理状態に関するデータの収集方法について、心理学における心理的アセスメントの観点から分類し、レビューする。その中でも、言語データの収集方法である面接法を中心に、それぞれの手法の特徴について説明する。

2.4.1. 量的調査と質的調査

本研究で用いる言語データは、収集された段階では、数量的に表現することができない質的なデータである。これに対し、統計データなど、数量的な表現や分析が可能な情報が量的データとなる。社会調査や心理的アセスメントなど、対象とする領域に応じてデータの収集方法は様々であるが、いずれも質的データをを得るための質的調査、あるいは、量的データをを得るための量的調査として分類することが可能である。得られるデータの性質や、データ収集のプロセスは調査方法に応じて異なるため、研究目的に応じて適切な方法を選択する必要がある。

2.4.2. 心理的アセスメント

心理学、特に心理臨床の領域では、クライアントの抱える心理的な課題や、本人の心理的特性、パーソナリティなどを理解するために、様々な技法を用いて評価を行う。このプロセスを心理的アセスメントという。主なアセスメントの技法は、面接法、観察法、質問紙法、投影法、作業検査法の5つである。これらの技法は、臨床場面で用いられるだけでなく、量的調査、あるいは質的調査の技法として用いられることもある。以下の節では、代表的なアセスメント技法として、面接法、観察法、質問紙法を取り上げ、データ収集、すなわち調査手法という観点から、これらのアセスメント技法の特徴について述べる。その中でも、言語データの収集に用いられる面接法に焦点を当て、他の技法との比較を行う。

1) 面接法

心理学における面接法は、クライアントの診断や治療を目的とした臨床的面接法と、データ収集を目的とした調査的面接法とに大別される。また、調査的面接は、調

査のプロセスにおける自由度に応じて、構造化面接、半構造化面接、非構造化面接という分類がなされる。あらかじめ設定された質問項目、質問順に従って、形式的な面接を行うのが構造化面接である。これに対し、質問項目を事前に設定せず、自由な形式で行われる面接が非構造化面接である。半構造化面接では、事前に質問項目を設定した上で、面接中の対象者の回答や反応、状況などを踏まえて、質問の内容や表現、順序などを柔軟に変更することができる。

観察法が対象の行動に焦点を当てた技法であるのに対し、面接法は、対象となる人々の語りを通じて、その人の心理状態や性格、価値観といった内面を明らかにするものである。すなわち、収集の対象とするデータそのものが異なるが、いずれもデータの分析・解釈に主観が入りやすいという点では共通している。特に、面接法では、調査者と対象者の相互作用が避けられないため、分析・解釈の際に客観的な視点を維持することが重要となる（保坂，2000）。

2) 観察法

観察法とは、日常的な場面や実験的に設定された状況下において、人間や動物の行動を観察し、記録を行い、その内容について分析を行うことで、行動の質的・量的な特徴を明らかにする技法である。観察においては、対象のありのままの行動を記述したり、特定の姿勢やジェスチャー、発話などの頻度を記録したり、対象の行動の程度（声の大きさ、活動への集中度など）や、行動から受ける印象について評価したりといった方法で、記録が行われる（中澤，1997）。

自然な状況で行われる観察は、対象のありのままの行動をデータとして幅広く収集し、分析に用いることができるという長所がある。しかしながら、観察法が扱うのは客観的な観察が可能な行動であり、対象の感情や思考、人格といった、内面に直結するデータを得ることは難しい。心理的側面を含めた分析を行う際には、観察法を他のアセスメント技法と組み合わせて用いるなど、分析が主観に偏らないための工夫が必要となる。

3) 質問紙法

心理学における質問紙法は、対象者の心理的特性の測定を目的としたものであり、妥当性・信頼性の検証された心理尺度を用いて行われることが多い。回答方法は、質問文に対して「はい」または「いいえ」で回答する2件法や、これに「どちらでもない」を加えた3件法、「程度」や「頻度」の測定を意図した5件法および7件法、選択肢の順位付けを求める順位法などがある（宮下，1998）。

面接法と比較すると、質問紙法は実施が比較的容易であり、実施者は高度なスキルを必要としない。また、集団を対象とした調査が可能であるため、一度に多数のデータを収集することができる。さらに、選択式の設問からは量的データを得ることができ、客観的な評価や統計解析を実施することが可能である。

このように、様々な利点が存在する一方で、質問紙法で得られるデータは、事前に定められた設問の範囲に限定されるという欠点もある。面接法のように、対象者の反応に応じて質問内容を変更・追加するなど、柔軟な対応を行うことは困難である。

2.5. 感情の定義とその関連理論

本研究では、言語データを分析し、対象者の心理状態を定量化して、その変化を視覚的に表現することを試みる。心理状態を説明するための心理学的な構成概念は様々あるが、その中で、感情へ焦点を当てた分析を行う。これにより、対象者の短時間で顕著な心理変化もとらえることができ、その様相を可視化することが可能となる。

本章では、本研究において言語データから対象者の心理状態を分析するために、その基礎となる感情の概念を明確化する。さらに、心理状態とその変化についての可視化を試みる上で、感情を重視することの意味について説明する。そのために、心理学における感情の定義と、それらの関連する理論について述べる。

2.5.1. 感情の定義

本研究の提案手法では、言語データをもとに対象者の心理状態を定量化するため、発話中に含まれる語のうち、心理状態を反映する語 (Modal word) の出現頻度に注目する。特に、一過性の明確な感情体験である情動へ焦点を当てることで、対象者の経験した心理状態とその変化を明確にとらえ、可視化することを想定している。そこで、本節では、Modal word の選定に関わる重要な概念として、感情および情動の特徴を示す。また、感情および情動に焦点を当てた可視化の有効性について述べる。

心理学においては、「感情」についての統一的な定義は存在せず、研究者ごとに様々な説明がなされている。例えば Ortony ら (1988) は、「感情とは、人が心的過程の中で行う様々な情報処理のうちで、人、物、出来事、環境についてする評価的な反応である」と述べている。ここでの「評価」とは、特定の対象について「好き—嫌い」、「良い—悪い」といった判断を下すことであり、これに伴って生理的、行動的、心理的な「反応」が生じるプロセスとして、感情を定義している。このような評価と反応のプロセスは、日常生活の多くの場面で経験するものであり、我々が対象について認識し、行動を選択する上で不可欠な精神活動であるといえる。

また、その特徴に基づく感情の定義として、Shiota ら (2012) は、1) 個体と種の生存に役立つこと (適応的側面)、2) 刺激に対する反応であること、3) 認知、主観的体験、生理的变化、行動の4要素からなること、という3点を挙げている。例えば、接

近する車や野生動物など、外的な脅威に対して生じる恐怖の感情は、逃避行動を引き起こし、個体の生存を有利にする働きをもつ。これは、感情の適応的側面を表す特徴である。また、恐怖の感情は、車、動物といった外的刺激によって生じた反応としてとらえることができる。感情を喚起する刺激は、必ずしも外的な要因であるとは限らず、過去の記憶などの内的刺激によって特定の感情が引き起こされることもある。上記の例では、車や動物を刺激として受け取るとともに、これらを脅威と認識するプロセス（認知）を経て、恐怖の感情（主観的体験）を経験するとともに、血圧の上昇や筋収縮といった生理反応、逃避という表出行動が生じている。こうした特徴からも、感情と認知、行動との関連性を見出すことができるので、感情が我々の精神活動全般において、中核的な役割を果たしているといえる。

上記のような特徴に加え、心理学領域における「感情 (affect)」は、様々な心理的反応、状態を総称する概念として用いられる。そのため、種々の感情は、その性質に応じて、「情動 (emotion)」、「気分 (mood)」、「選好 (preference)」といった用語で区別することが可能である。情動とは、原因となる対象が明らかで、生理的覚醒を伴うことが多く、喚起と消失のタイミングが明確な感情（大平，2010）を指す。感情の中でも、急激に起こる一過性の反応であり、行動の動機づけに関連した感情状態や体験が、情動に分類される（磯，2005）。これに対し、比較的長時間にわたって持続するものの、必ずしも原因が明確でなく、それほど強くない快または不快の感情状態を気分と呼ぶ（大平，2010）。また、対象についての「好き—嫌い」や、「快—不快」に関する直観的な判断を選好と呼び、感情に含めることがある（大平，2010）。

上記のような諸側面を考慮すると、我々が経験する感情は、その持続時間や強度、対象の明確さといった点で、分類が可能であることが分かる。その中でも、情動の喚起は比較的短時間で起こり、気分や選好と比較すると、急激な感情の変化として経験されるという特徴がある。すなわち、発話に表れる種々の心理状態の中でも、情動については、その変化をより明確にとらえることが期待できる。この点を踏まえ、本研究の提案手法においては、心理変化を分かりやすく可視化するために、情動を反映する語を中心に、対象者の心理状態を表す Modal word を選定する。

また、感情は、情動や気分、選好などの反応・状態を総称する用語であるが、情動

と同義に使用されることもある。以下、様々な心理反応・状態の総称として「感情」を用いる際には“affect”，情動と同義で「感情」を用いる際には“Emotion”ないしは“Emotional”を訳語にあてることで，本研究ではこれらを区別する。

次節では，感情を理解するための枠組みとして，基本感情説と次元説を取り上げ，それぞれの理論が対象とする感情（emotion）の具体的な要素（“喜び”，“怒り”，“悲しみ”など，個々の感情を表す用語）を示す。この用語に基づき，言語データに含まれる語の選別を行うことで，対象者の情動を示す Modal word を抽出できる。

2.5.2. 感情に関連する理論

前節で取り上げた感情の定義や分類を踏まえ，本研究では，発話に表れる情動を主な分析の対象とする。情動を反映する Modal word を選定する際には，個々の感情（emotion）をいくつかのカテゴリに分類したり，少数の評価軸上に位置づけたりといった，先行研究の枠組みを参照することが役立つと思われる。そこで，本節では，感情に関連する代表的な理論として「基本感情説」と「次元説」を取り上げ，本研究における Modal word の選定基準を明確にする。

心理学では，感情をとらえる枠組みとなる理論として，基本感情説と次元説が知られている。基本感情説は，喜びや怒り，悲しみといった，生得的かつ普遍的な感情表出，すなわち基本感情（basic emotions）のカテゴリを想定するものである。基本感情の定義および分類方法は研究者ごとに異なるが，ある感情が基本感情であるか否かを判断するための基準として，Ekman (1992) による9つの基準がよく知られている。Ekman によれば，1) 感情に対応する特徴的・普遍的な表出信号（喜び：笑顔，悲しみ：泣き顔，など）が存在すること，2) 動物にも類似した表出行動がみられること（怒り：歯をむき出す，など），3) 感情に対応する特異的な生理変化のパターンが存在すること，4) 感情を喚起する事象に普遍性が認められること（急激な環境変化に対して誰もが経験する驚き，など），5) 感情に対応する表出行動，生理反応の間に一貫性があること，6) 刺激事象に対して，即時的に感情反応が生じること，7) 感情の持続時間が短いこと，8) 必ずしも意識化されない，自動的な評価のプロセスであること（原因を自覚する前に感じる不安，など），9) ‘8)’のプロセスにおいて，感情反応が自発的

に生じること、の9点が、基本感情を同定するための判断基準になるという。このように、判断基準をどのように設定するかに応じて、基本感情の分類カテゴリやその内容にも違いが生じてくる。ここでは、様々な研究者が提唱する、代表的な基本感情のカテゴリを表 2.1 に示す。

第3章（‘心理学的トポロジーの導入’）では、以下の基本感情のカテゴリ（表 2.1）と、二次元平面上に布置された感情エピソード（図 2.8）を参照し、発話に含まれる語の中から、これらに該当するものを Modal word として選定した（‘3.4 言語データからの Modal word の選定’）。このとき、アスリートの心理測定や、記述データの分類、単語の感情極性の推定に関わる先行研究を併せて収集し、同様に Modal word の選定基準とした。

表 2.1 先行研究における基本感情のカテゴリ

研究者	基本感情	基本感情（訳）
Izard, 1971	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	怒り, 軽蔑, 嫌悪, 苦痛, 恐怖, 罪悪感, 興味, 喜び, 恥, 驚き
Plutchik, 1980	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	受容, 怒り, 予期, 嫌悪, 喜び, 恐怖, 悲しみ, 驚き
Ekman et al., 1982	Anger, disgust, fear, joy, sadness, surprise	怒り, 嫌悪, 恐怖, 喜び, 悲しみ, 驚き
Gray, 1982	Rage and terror, anxiety, joy	激怒と恐怖, 不安, 喜び
Tomkins, 1984	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	怒り, 興味, 軽蔑, 嫌悪, 苦痛, 恐怖, 喜び, 恥, 驚き
Oatley, et al., 1987	Anger, disgust, anxiety, happiness, sadness	怒り, 嫌悪, 不安, 幸福, 悲しみ

基本感情説では、感情の基本単位として複数のカテゴリを設定するのに対し、次元説では、個々の感情をより少数の評価次元に還元し、これを基本単位として取り扱う。例えば、Russell (1980) は、感情の評価次元として「快—不快」, 「活性—不活性 (覚醒—睡眠)」の二次元を想定し、これらの次元軸上に個々の感情が位置づけられるとした。図 2.8 は、様々な感情の二次元平面上での布置を示しており、次元説の立場に立つ Russell らの研究成果として、多くの文献で取り上げられてきたものである (中村, 2018)。

Russell ら (1999) は、快—不快, 活性—不活性の軸上で定義される原初的な感情状態を”Core affect”と呼んでいる。これは、非内省的に生じるが、意識化することの可

能な神経生理学的状態である。Core affect が意識されるとともに、その原因となる対象（人やもの、出来事に関する心的表象）が意識化される。対象はそれぞれ、Core affect に特定の変化を引き起こす性質（affective quality）をもつ。また、意識化された Core affect については、特定の対象への原因帰属が行われ、これを「帰属された affect」と呼ぶ。これらの感情反応（Core affect の意識化、対象の affective quality の知覚、対象への原因帰属など）が、喜びや怒りといった個々の感情に先立って生じるとみなすのが、Russell らの提唱する次元説（Core affect 説）の特徴である。様々な感情反応は、個々の感情に関する概念カテゴリ（感情プロトタイプ）と照合され、その類似性に基づいて、喜び、怒りなど特定の感情概念（感情エピソード、Emotional episode）が生じる。

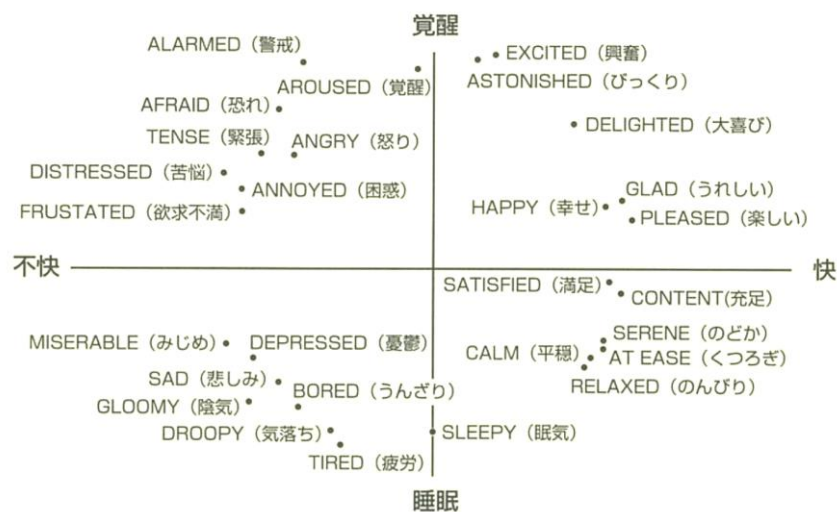


図 2.8 次元説における感情の布置

(文献 (中村, 2018) より引用)

2.6. 本研究における解析手法

本節では、提案手法で用いられる、自己組織化マップおよびファジィクラスタ分析の説明を行う。自己組織化マップに関しては、一般的に知られている解析手法として逐次型 SOM のアルゴリズムを紹介し、ファジィクラスタ分析については、分析の基礎となるファジィ理論について述べる。

2.6.1. 自己組織化マップ

自己組織化マップ (Self-Organizing Map: SOM) は、Kohonen (1995) により提唱されたニューラルネットワークアルゴリズムで、高次元データを低次元空間上へ非線形写像する解析手法である (金, 2017)。

SOM のニューラルネットワークは、入力層と出力層 (競合層, あるいはマップとも呼ぶ) という 2 層構造からなる (図 2.9)。入力層には、分析対象の個体を表す入力ベクトルが配置される。また、二次元平面への写像を行う場合、出力層は格子状のマップとして出力されることが多い。マップ上の各格子はニューロン (ユニット) と呼ばれ、それぞれが入力ベクトルの成分数に対応した代表ベクトルをもつ。

ここで、個体 $j(j = 1, 2, \dots, n)$ の入力ベクトルを $\mathbf{H}_j(h_{j,1}, h_{j,2}, \dots, h_{j,i})$ 、マップ上の $k(t = 1, 2, \dots, k)$ 個のニューロンがもつ代表ベクトルを $\mathbf{W}_t(w_{t,1}, w_{t,2}, \dots, w_{t,i})$ とおく。通常の SOM では、 \mathbf{W}_t の値を乱数により決定する。また、SOM のニューラルネットワークにおいて、 \mathbf{W}_t の成分は、入力ベクトルの各成分に対応する重みとして表される。

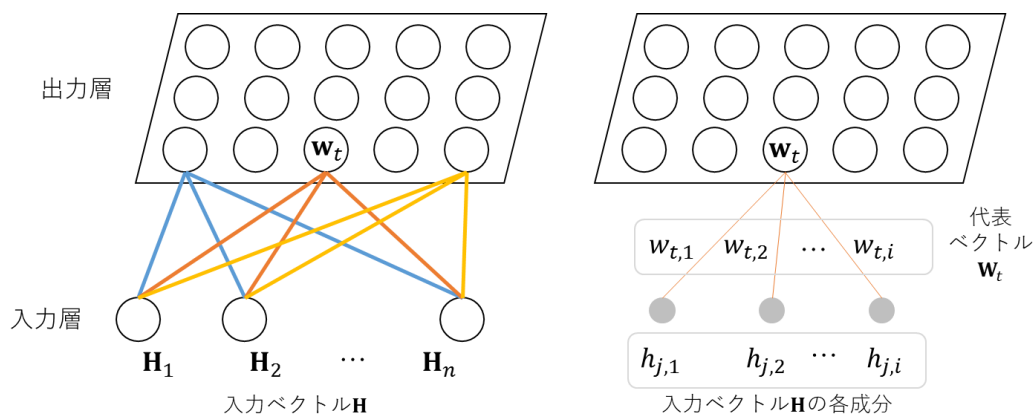


図 2.9 SOM の入力層および出力層

自己組織化マップのアルゴリズム

SOMにより、多次元データを2次元平面上に写像する際のアルゴリズムを、以下に示す。

- 1) 入力ベクトル \mathbf{H}_j と、マップ上のニューロンがもつすべての代表ベクトル \mathbf{W}_t とを比較し、下記の式に従って、ユークリッドノルムが最も小さい代表ベクトル \mathbf{W}_c をもつニューロン c を決定する。これを勝者ニューロンと呼ぶ。

$$\|\mathbf{H}_j - \mathbf{W}_c\| = \min_t \{\|\mathbf{H}_j - \mathbf{W}_t\|\} \quad (5.1.)$$

- 2) 勝者ニューロン c とその近傍ニューロン t がもつ代表ベクトル \mathbf{W}_t を、下記の式に従って更新する。

$$\mathbf{W}_t(r+1) = \begin{cases} \mathbf{W}_t(r) + h_{ci}(r)\{\mathbf{H}_j(r) - \mathbf{W}_t(r)\} & (t \in \mathbf{V}_c) \\ \mathbf{W}_t(r) & (t \notin \mathbf{V}_c) \end{cases} \quad (5.2.)$$

$$h_{ci}(r) = \alpha(t) \exp\left\{-\frac{\|\mathbf{LM}_c - \mathbf{LM}_t\|^2}{2\sigma^2(r)}\right\} \quad (5.3)$$

ここで、 $h_{ci}(r)$ は近傍関数と呼ばれ、勝者ニューロン c とその近傍ニューロン t との距離に応じて \mathbf{H}_j の更新への影響を調整する。 \mathbf{LM}_c は勝者ニューロン c の座標を、 \mathbf{LM}_t は近傍ニューロン t の座標を表す、2次元上の座標位置ベクトルである。 $\alpha(t)$ は学習率の係数を、 $\sigma^2(r)$ は勝者ニューロンの近傍領域 \mathbf{V}_c の半径を表す。すなわち、勝者ニューロン c との座標間の距離が $\|\mathbf{LM}_c - \mathbf{LM}_t\| \leq \sigma^2(r)$ となるような近傍ニューロン t は、近傍領域 \mathbf{V}_c に所属し、式(5.2.)による更新の対象となる。 $\alpha(t), \sigma^2(r)$ はいずれも、学習回数 r を変数とする単調減少関数である。最も簡単な単調減少関数は、 $1 - \frac{r}{T}$ として設定される。ただし、 T は事前に設定した学習の総回数である。

- 3) すべての入力ベクトル \mathbf{H}_j に対して、上記‘1)’と‘2)’を繰り返し実行する。

上記のアルゴリズムにより、個々の多次元データ（入力ベクトル）に対して、マップ上の勝者ニューロンの座標が決まる。ソフトウェアで SOM の解析結果を出力する際は、入力ベクトルを対応する勝者ニューロンの座標にマッピングし、入力ベクトル間の関係性が分かるように表示することが多い。このとき、類似した特徴をもつ入力ベクトルは同じ座標、あるいは近傍にマッピングされることとなる。また、それぞれのニューロンがもつ代表ベクトルの値を、星図や折れ線グラフ、色彩の濃淡で表すことにより、入力ベクトルのもつ特徴を読み取りやすくなる。SOM の出力結果は、格子状（正方形）のマップのほか、蜂の巣状（六角形）のマップ上に表す方法が提案されている。

本節で示した SOM のアルゴリズムは、1つの入力ベクトルの分類（アルゴリズム‘1’）が完了する度に、勝者ニューロンと近傍ニューロンの代表ベクトルの更新（アルゴリズム‘1’）を行う、逐次学習型の SOM に該当する。これは、一般的な SOM のアルゴリズムとして、よく知られているものである。

本研究で用いた SOM は、すべての入力ベクトルの分類が完了した後に、すべての勝者ニューロンと近傍ニューロンの代表ベクトルを更新する、一括学習型のアルゴリズムを採用している。このアルゴリズムについては、第6章で詳述する。

2.6.2. ファジィクラスタ分析

1) ファジィ理論

人間の思考や判断、行動は「曖昧さ (fuzziness)」を含むものであり、定性的な情報という側面をもつ。ファジィ理論は、このように曖昧な性質をもった情報を、定量的に扱うことを目的として提唱された理論である。

2) ファジィ集合

例えば、気温が 30 度以上の「真夏日」や、75 歳以上の「後期高齢者」などは、定義がはっきりとした数学的な集合であり、他の集合との明確な区別が可能である。これに対し、会話などで用いられる「暑い」、「若い」といった表現は、明確な基準が存在せず、個人によって定義や解釈が異なる「曖昧な集合」である。このように、曖昧

な表現によって言及され、明確な境界をもたない集合を「ファジィ集合」という。

曖昧な表現を数学的に扱うためには、「暑い」、「若い」といった概念のもつ曖昧さを、定量的に表現する数学的な枠組みが必要である。ここでは例として、「暑い」というファジィ集合の概念を、グラフを用いた表現で説明する（図 2.10）。図の横軸は気温を、縦軸は「暑い」と感じる度合いを表す。この度合いは、0 から 1 までの実数値をとり、ファジィ集合「暑い」に対する「帰属度」と呼ばれる。グラフ中の折れ線は、28°C以上では誰もが「暑い」と感じ、24°C以下では「暑い」と感じる人はいないことを示している。24°Cから 28°Cの間では、気温が高くなるにつれて、「暑い」という概念への帰属度が高まることがわかる。このように、ファジィ集合では、個々の要素（例：温度）が、その集合（例：“暑い”）に所属する度合いを、帰属度として設定する。また、個々の要素と帰属度との関係を関数として表現する。これを「メンバシップ関数」と呼ぶ。

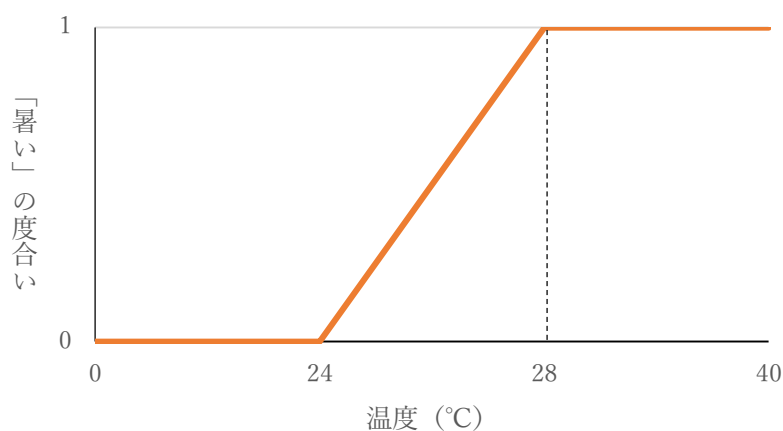


図 2.10 「暑い」の概念の定量化

全体集合 X を通常の集合とし、 X におけるファジィ集合を A とおく。このとき、ファジィ集合 A のメンバシップ関数 μ_A は、 $[0,1]$ に値をとる X 上の関数として、以下のように定義される。

$$\mu_A: X \rightarrow [0,1] \quad (5.4)$$

また、ファジィ集合 A のメンバシップ関数 $\mu_A(x)$ の値が正である x の集合の閉包を、ファジィ集合 A の「台集合 (support)」という。これを $\text{supp}\mu_A$ と表し、下記の式により定義する。ただし、集合 K の閉包 $C(K)$ とは、元の集合にその境界点を加えたものである。

$$\text{supp}\mu_A = C(\{x \in X \mid \mu_A(x) > 0\}) \quad (5.5)$$

ファジィ集合のメンバシップ関数の値が、0または1のいずれかの値しかとらない場合、これは通常の集合と等しくなり、「クリस्प集合」と呼ばれる。クリस्प集合への帰属度を表すメンバシップ関数を、特に「特性関数」という。

3) ファジィ集合の演算

ファジィ集合の基本的な演算は、通常の集合の演算に準ずるものである。それぞれの演算は、メンバシップ関数の関係式として、以下のように定義される。

- a. 相当 $A = B, \mu_A(x) = \mu_B(x)$
- b. 部分集合 $A \subseteq B, \mu_A(x) \leq \mu_B(x)$
- c. 和集合 $A \cup B, \mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x)$
- d. 共通集合 $A \cap B, \mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x)$
- e. 補集合 $\bar{A}, \mu_{\bar{A}}(x) = 1 - \mu_A(x)$
- f. 代数積 $AB, \mu_{AB}(x) = \mu_A(x)\mu_B(x)$
- g. 代数和 $A \dot{+} B, \mu_{A \dot{+} B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$
- h. 限界積 $A \odot B, \mu_{A \odot B}(x) = \max\{0, \mu_A(x) + \mu_B(x) - 1\}$
 $= 0 \vee (\mu_A(x) + \mu_B(x) - 1)$
- i. 限界和 $A \oplus B, \mu_{A \oplus B}(x) = \min\{1, \mu_A(x) + \mu_B(x)\}$
 $= 1 \wedge (\mu_A(x) + \mu_B(x))$
- j. 限界差 $A \ominus B, \mu_{A \ominus B}(x) = \min\{0, \mu_A(x) - \mu_B(x)\}$
 $= 0 \vee (\mu_A(x) - \mu_B(x))$

ファジィ集合の演算が、通常の集合の演算と異なるのは、矛盾律 ($A \cap \bar{A} = \emptyset$) と排中律 ($A \cup \bar{A} = X$) が成り立たないという点である。

4) ファジィ関係

あるものと他のものとの関わり合いを示す「関係」という一般概念も、数学的に定義できる関係から、曖昧で明確に規定できない関係に至るまで、様々な種類が存在する。例えば、「物質の分子量と密度」のような明確に定義された関係に対し、「クラス内における学生同士の仲の良さ」といった関係を、明確に規定することは困難である。このような曖昧な関係を、ファジィ集合の概念に基づいて定式化するために、「ファジィ関係」という概念が導入される。

ファジィ関係は2つの要素間の曖昧な関係を表す。これらの要素が属する全体集合を X, Y とすると、直積 $X \times Y$ におけるファジィ関係 R は、 $X \times Y$ 上のメンバシップ関数 $\mu_R(x, y)$ によって特徴づけられる集合 $R \subset X \times Y$ として定義される。メンバシップ関数 $\mu_R(x, y)$ は、要素 x と y との関係の度合いを示す。

$$\mu_R: X \times Y \rightarrow [0,1] \quad (5.6)$$

X と Y が有限集合である場合、 $X \times Y$ におけるファジィ関係 R は、行列として表すことができ、これを「ファジィ行列」という。このとき、ファジィ関係 R は、メンバシップ関数 $\mu_R(x_i, y_j)$ を要素としてもつファジィ行列 F によって、以下のように定義される。

$$F = (f_{i,j}), f_{i,j} = \mu_R(x_i, y_j), x_i, y_j \in X \times Y \quad (5.7)$$

ここで、 $0 \leq \mu_R(x_i, y_j) \leq 1, i = 1, 2, \dots, m, j = 1, 2, \dots, n$ であることから、ファジィ行列の要素は、すべて区間 $[0,1]$ の値をとる。

5) ファジィ関係の性質

直積 $X \times Y$ 上のファジィ関係については、4種類の基本的な性質が以下のように定義される。

- 1) 反射律： $\mu_R(x, x) = 1, \forall x \in X$
- 2) 対称律： $\mu_R(x, y) = \mu_R(y, x)$
- 3) 反対称律： $\mu_R(x, y) > 0, \mu_R(y, x) \Rightarrow x = y$
- 4) 推移律： $\mu_R(x, z) \geq \max_y \min\{\mu_R(x, y), \mu_R(y, z)\}$

特に、ファジィ関係が反射律と対称律をともに満たすとき、これをファジィ相似関係という。また、ファジィ関係が反射律、対称律、推移律をいずれも満たすとき、これをファジィ同値関係、またはファジィ類似関係という。

一般に、 $F^n = F \times F \times \dots \times F$ を、 $n \times n$ のファジィ行列の可達行列という。また、ファジィ行列 F によって表されるファジィ関係を R 、ファジィ行列 F^k によって表されるファジィ関係を R^k とする。このとき、 $n \times n$ のファジィ行列と、そのファジィ関係 R に対し、 $\hat{R} = \bigcup_{k=1}^n R^k$ を R の推移包という。

ファジィ行列 F においては、推移包 \hat{R} と可達行列 F^n が一致する。そのため、推移包 \hat{R} のファジィ行列の各要素 $\hat{f}_{i,j}$ は、他の要素を介した場合も含めて、要素 x_i と x_j の関連度の最大値を示す。

6) ファジィ関係の演算

ファジィ集合と同様に、ファジィ関係とそのメンバシップ関数においては、以下のような演算が成り立つ。

- 1) 部分集合 $R \subseteq S, \mu_R(x, y) \leq \mu_S(x, y), \forall (x, y) \in X \times Y$
- 2) 共通集合 $R \cap S, \mu_{R \cap S}(x, y) = \min\{\mu_R(x, y), \mu_S(x, y)\}$
- 3) 和集合 $R \cup S, \mu_{R \cup S}(x, y) = \max\{\mu_R(x, y), \mu_S(x, y)\}$
- 4) 補関係 $\bar{R}, \mu_{\bar{R}}(x, y) = 1 - \mu_R(x, y)$

- 5) 逆ファジィ集合 $R^{-1}, \mu_{R^{-1}}(x, y) = \mu_R(y, x)$
- 6) 代数積 $RS, \mu_{RS}(x, y) = \mu_R(x, y)\mu_S(x, y)$
- 7) 代数和 $R \dot{+} S, \mu_{R \dot{+} S}(x, y) = \mu_R(x, y) + \mu_S(x, y) - \mu_R(x, y)\mu_S(x, y)$
- 8) 限界積 $R \odot S, \mu_{R \odot S}(x, y) = \max\{0, \mu_R(x, y) + \mu_S(x, y) - 1\}$
- 9) 限界和 $R \oplus S, \mu_{R \oplus S}(x, y) = \min\{1, \mu_R(x, y) + \mu_S(x, y)\}$

7) ファジィ関係の合成

3つの集合 X, Y, Z において、 X と Y の間にはファジィ関係 R が、 Y と Z の間にはファジィ関係 S があるとする。このとき、 X と Z の関係に対して、マックスミニ合成が成立する。 R と S のマックスミニ合成は、 $X \times Z$ におけるファジィ関係 $R \circ S$ となり、以下の式によって定義される。

$$R \circ S, \mu_{R \circ S}(x, z) = \max_y \min\{\mu_R(x, y), \mu_S(y, z)\} \quad (5.8)$$

2.7. 結言

本章では、本研究の分析に関わる基礎事項として、関連領域である情報可視化、テキストマイニングの分野における知見について述べた。また、心理学におけるデータ収集の方法と、感情の定義および関連理論を示した。最後に、分析に用いた自己組織化マップとファジィクラスタ分析の基本概念と、それぞれの解析に関わる用語の定義について説明した。以上により、様々な研究領域における課題解決をもたらす試みとしての、本研究の位置づけが示された。

第3章. 解析の基盤：心理学的トポロジー

3.1. 緒言

本章では、提案手法の第一段階のプロセスとして、SOM とファジィクラスタ分析による言語データのマッピングと、マッピング結果からの心理的評価軸の抽出方法について説明する。まず、処理全体の概要を示した上で、処理の各段階における手続きに関して記載する。以降に、処理を実行した結果を示し、結果に対する考察と結言を述べる。

なお、本章に記載された内容は、2018 年にオンラインジャーナル“Psychology (<http://www.scirp.org/journal/psych/>)”に掲載された、著者らの論文 (Aoki et al., 2018a) に基づくものである。

3.2. 処理の概要

心理学的トポロジーに基づく、言語データの分析および可視化の手法を提案する。処理の手順は、図 3.1 に示す通りである。

まず、分析の対象として、オンライン上で公開されている、アスリートに対するインタビュー記事を収集した。調査であることを意識せず、自然に発せられた発言を対象とするようにするために、アスリート本人に対する直接のインタビュー調査は実施しなかった。次に、収集されたインタビュー記事から、Modal Word の選定を行った。本研究では、対象者の心理状態を反映すると思われる語句を、心理学の専門家によるアドバイスを受けながら、Modal Word として選定した。さらに、個々の発話における Modal Word の相対頻度 (TF: Term Frequency) と、特定の Modal Word を含む発話の割合 (IDF: Inverse Document Frequency) を求めた。これらをかけ合わせた TF-IDF 値を、心理状態の指標とし、トポロジー、すなわち空間上での距離が定義できるようにベクトルデータの形式にした。本研究において TF および IDF の指標を用いる意図は、多くのテキスト解析の先行研究で行われている、重要語句の抽出などとは異なり、可視化空間における発話間の距離を定めるための指標として用いる点にある。

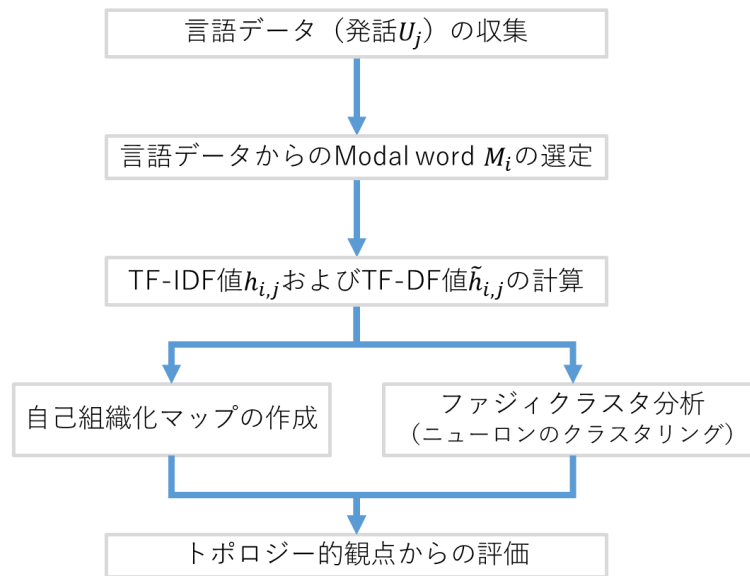


図 3.1 心理学的トポロジーに基づく分析および可視化のプロセス

算出された TF-IDF 値を用いて、自己組織化マップ (SOM: Self-Organizing Maps) の作成およびファジィクラスタ分析を実施した。ファジィクラスタ分析では、SOM の出力結果である、入力ベクトルの値に基づいて更新された、各ニューロンの代表ベクトルの値を分析に用いた。これにより、SOM の出力層 (マップ) を構成するニューロンのクラスタリングを行った。分析の結果、個々のアスリートの発話に対応する入力ベクトルがマップ上に分類され、それぞれの分類の位置を示す座標が得られた。また、マップ上のニューロンを、代表ベクトルの特徴に応じてクラスタリングすることで、マップを複数の領域 (いくつかのニューロンのまとまり毎) に分割した。最終的に、可視化された発話の座標についてトポロジー的な観点から考察を行い、さらに心理学的な解釈をすることにより、対象者の心理的な特徴を明らかにした。以下、これらの詳細について説明する。

3.3. 言語データの収集

オリンピック・パラリンピックの出場選手を含め、水泳や野球、ゴルフなどの競技に取り組む 60 名の日本人アスリート $c (= 1, 2, \dots, NC (= 60))$ を対象とした、日本語のインタビュー記事を Web 上の次の 5 種類のサイトから収集した。

- 1) あいおいニッセイ同和損保：アスリートインタビュー
https://www.challenge-support.com/athlete_interviews/
- 2) アスリート育成パスウェイ：アスリートライフスタイル（インタビュー）
<http://pathway.jpnsport.go.jp/lifestyle/interview.html>
- 3) Club Unisys+：プロゴルファーインタビュー
<http://www.unisys.co.jp/club/golf/golfers.html>
- 4) 東京オリンピック・パラリンピック競技大会組織委員会：東京 2020 アスリートビジット
<https://tokyo2020.org/jp/news/notice/20170706-01.html>
- 5) TOYO UNIVERSITY TOP ATHLETES 応援サイト 2016: 日本代表選手インタビュー
<http://www.toyo.ac.jp/site/toyo-topathletes/>

これらの記事のうち，アスリートの発言を示す一段落を「発話」 $U(Utterances)$ と定義し，これを分析対象の基本単位とした．個々のアスリートに対するインタビュー記事から，複数の発話を抽出し，最終的に 736 個の発話 $U_j(j = 1, 2, \dots, NU(= 736))$ が得られた．ここでは，収集したインタビュー記事の内容および発話の例として，アスリート $c(= 3)$ の発話 $U_j(j = 24, 25, \dots, 31)$ を以下に示す（個々の発話を鍵括弧でくくり，個人名については●●で表記した）．

(インタビュー記事の内容および発話の例：アスリート $c(=3)$)

U₂₄：「僕は中学時代から陸上部に所属して長距離走をやっていましたが，県大会に出られるかどうか程度レベルの選手でした．そんな中で横浜高校に入学して出会ったのが競歩でした．足が速くありませんでしたが，競歩の特徴的なクネクネとした動きが自分の中で面白いと思ったのが競技を始めるきっかけでしたね」

U₂₅：「実際に始めてみて，こんなにも色々な筋肉を使わなければいけないのか，という点が驚きでした．競歩は簡単に説明すると『足が垂直になるまで膝を伸ばす』，『どちらかの足が地面についていなければならない』というルールがあります．そのフォームを崩さず 20 kmもの距離を歩ききるためには，太もも裏から臀部，上半身の連動がとても大事になるんです．だからこそ，特徴的なフォームで歩く必要性に気づくことができました」

U₂₆：「僕自身，過去には何度か失格になったことがあります．特に今年（2016年）2月の日本選手権では残り 500m のところで失格となりました．非常に悔しかったですがフォームをもう一度見直して，翌3月の全日本競歩能美大会で優勝し，リオデジャネイロ五輪日本代表の切符をつかみ取ることができました」

U₂₇：「僕は理工学部の都市環境デザイン学科に在籍しています．そこで勉強していることは……環境のことを勉強しながら，未来に向けて構造物を建設することを学んでいます．授業では測量機械を使った実験などに取り組んでいます．ただ1年生の時は初めての寮生活や大学生活になれなくてうまく行かないこともありましたが，2年生からは競技も学業にもどんどん慣れてきて，今はそこまで苦ではなくなりました．競技と学業の両立に慣れるということが一番大切だったのかなって思います．授業に行くと，陸上部の寮では会えない友人もいますしね」

U₂₈：「また陸上部の寮で生活しているのですが，能美大会後に●●監督から寮長をやっしてほしいと指名されました．最初は『五輪候補に選ばれた分，国内合宿などで長期間

にわたって寮から離れることが多いので大丈夫かな?』と考えましたが、今は僕がいなくても周囲がしっかりとやってくれれば、という思いを持って寮長を務めています。

●●大学と言えば駅伝のイメージが強いかと思いますが、箱根駅伝などのおかげで長距離ブロックの中で競歩も少しずつ存在感が出てきていると感じます。自分自身も、寮でともに生活する同学年の●●たちから大きく刺激を受けています」

U₂₉:「僕にとってターニングポイントとなったのは大学1年生の12月にあった記録会です。当時の先輩であるロンドン五輪で活躍した●●(2015年卒業)さんの持っていたジュニア記録を1分ほど更新することができました。その記録が世界ジュニアランキングのトップだったからこそ、世界を意識し始めるようになりましたね」

U₃₀:「現在の日本競歩界は競技力がとても高くなっています。僕の出場する20kmでも世界ランク1~3位を日本人選手が独占しています。自分自身も国内合宿などに参加することで、レベルアップしていると感じます。現在僕は世界ランク3位ということもあってリオデジャネイロ五輪ではメダルは意識して臨むつもりですが、まずは自分の力を十分に発揮することを心がけています。その結果がいい順位につながるので、五輪へのプレッシャーを感じるか、ですか? もちろん日本代表なので……多少の不安はありますが、むしろ『早くオリンピックが来ないかな』とワクワクしています」

U₃₁:「そしてリオだけでなく長期的な目標も大事にされていて、4年後の東京五輪につながるようリオにしたいです。リオからメダルは狙いますが、東京で金を獲ること、世界ランク1位になることが自分にとっての最大の目標です。そして東京の次、2024年の五輪まで競技を続けていきたいですね。競技人生を終えたら、先輩たちが教えてくれたことを次世代に伝える立場になりたいと考えています。そのためにはまずリオデジャネイロ五輪を全力で臨むので、応援よろしくお願いします!」

3.4. 言語データからの Modal word の選定

インタビュー記事の発話中から、アスリートの感情を反映する語（以下、これらの語を“Modal word”と呼ぶ）を選定するにあたり、前提となる考え方と、選定方法について以下に述べる。

本研究では、可視化の対象とするアスリートの心理を、状態と変化という観点で区別する。前者の心理状態がスカラーとして定義されるのに対し、後者の心理変化はベクトルの形式で定義される。この心理変化の中核をなす概念として、感情の性質に注目し、Modal word の選定と出現頻度の分析を行うことで、感情の定量化を試みた。

‘2.5.2 感情に関連する理論’で述べたとおり、感情の中でも短時間かつ急激な心理変化としての性質をもつのが情動 (Emotion) である。このような感情 (情動) をとらえる枠組みとして、基本感情説や次元説といった理論が提唱されてきた。特に、次元説では、感情を三次元空間上に位置づけるアプローチ (Wundt, 1924; Matthews et al., 1990; 城, 2009) がなされている。感情のベクトルの性質を踏まえれば、個々の感情が三次元空間上に割り当てられる座標は、心理変化の速度に対応するものと解釈できる。これにより、心理変化としての感情の性質を、動的に表現することが可能となる。

そこで、前述の先行研究を参考に、感情を三次元空間上に位置づけるための評価軸 (u : 快—不快, v : 活性—不活性, w : 緊張—弛緩) を設定し、個々の Modal word が示す感情の性質を、3軸上の速度ベクトル $\mathbf{V}_{M_i} = \{u_{M_i}, v_{M_i}, w_{M_i}\}$ として表現した (図 3.2(a))。本研究では、 u, v, w を感情の評価軸とし、これらの3軸によって構成される三次元空間を感情空間と定義する。また、個々の感情の表出の程度は、前述の速度ベクトルから算出される速さ $|\mathbf{V}_{M_i}| = \sqrt{u_{M_i}^2 + v_{M_i}^2 + w_{M_i}^2}$ として、スカラー量で表される (図 3.2(b))。このとき、感情空間における Modal word の速度ベクトル \mathbf{V}_{M_i} および速さ $|\mathbf{V}_{M_i}|$ に基づき、特定の心理状態を表す心理的評価軸 $S_n (n \in \mathbf{N})$ を抽出することができる。

図 3.2(b)は、Modal word M_2 (勝つ) の速度ベクトル \mathbf{V}_{M_2} と Modal word M_4 (頑張る) の速度ベクトル \mathbf{V}_{M_4} が同方向で、速さについては、 $|\mathbf{V}_{M_4}| < |\mathbf{V}_{M_2}|$ であることを示している。そこで、感情の評価軸 v, w の間に心理的評価軸 S_1 を設定する。また、アスリートにとって、“頑張る”という語よりも、“勝つ”という語の方が勝利への意欲を強く反

映していると考えられるため、 S_1 を「闘争心」の心理的評価軸と名付ける。一方、図3.2(c)では、Modal word M_6 (嬉しい) の速度ベクトル \mathbf{V}_{M_6} と Modal word M_{14} (ミス) の速度ベクトル $\mathbf{V}_{M_{14}}$ がほぼ直交している。ベクトル \mathbf{V}_{M_6} は強い快感情を示しており、アスリートにとっては、競技への動機づけが高まっている状態であるとも推察される。これに対し、ベクトル $\mathbf{V}_{M_{14}}$ は不活性な感情状態を示し、動機づけの低下がうかがえる。そのため、これらのベクトルの終点を結ぶ心理的評価軸 S_2 を設定し、これを「自発性」と名付ける。

評価軸の抽出に際しては、様々な Modal word が示す感情の速度ベクトルを二次元空間上で可視化するために、自己組織化マップおよびファジィクラスタ分析の手法を用いた解析を行う必要がある。解析結果に基づく評価軸の抽出方法は、‘3.8 トポロジー的評価’において詳述する。

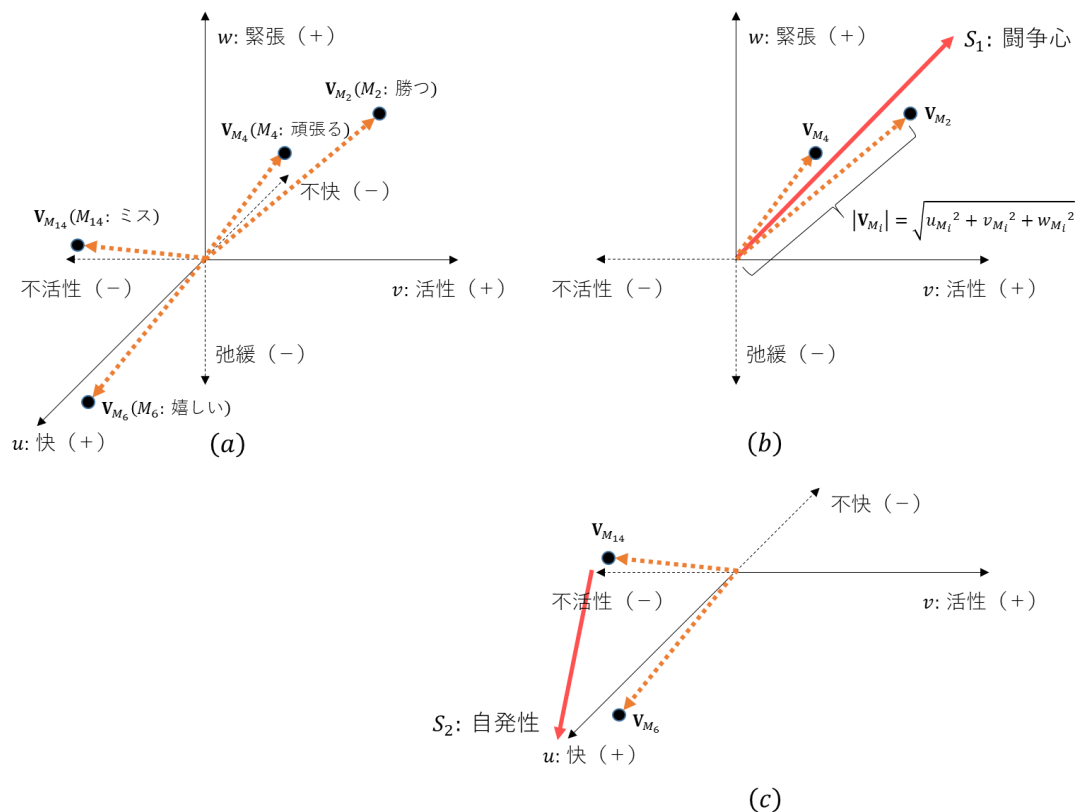


図 3.2 Modal word M_i が示す感情空間上の速度ベクトル(a)および心理的評価軸 (b),(c)

本研究では、対象者の発話に含まれる語のうち、上記の三次元空間上に位置づけられる語を、感情を反映する Modal word M_i として選定した。選定プロセスの詳細を、以下に示す。

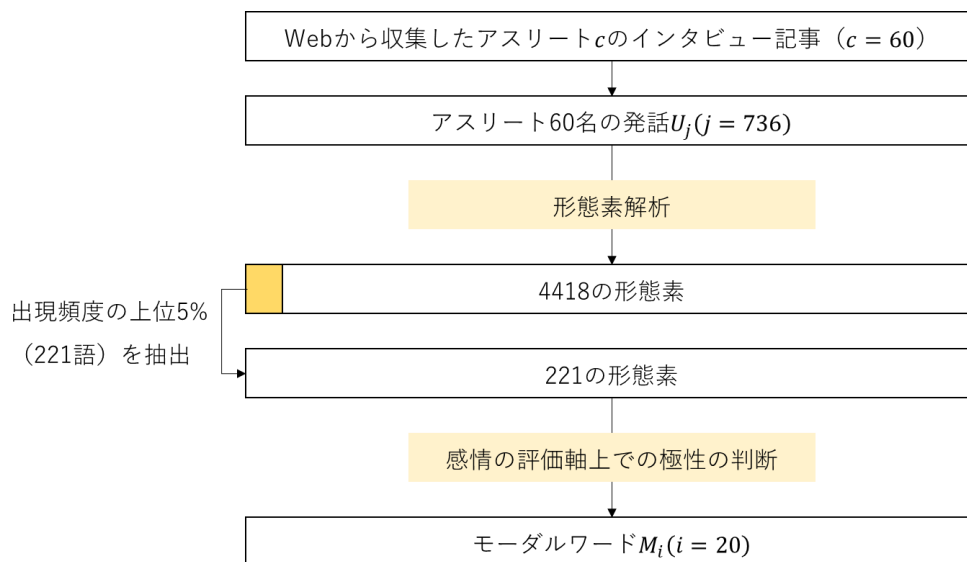


図 3.3 Modal word M_i の選定プロセス

まず、多くの先行研究でテキストデータの処理によく用いられている、テキストマイニング用ソフトウェア「Tiny Text Miner: TTM」(松村ら, 2009)を用いて、すべてのアスリートの発話 U に対する形態素解析を実施した。形態素解析とは、自然言語で書かれた文章を、個々の形態素(名詞、動詞、形容詞など、言語が意味をもつ最小単位)へと分割する手法である。これにより、アスリートの736個の発話 U は4,418種類で総数21,007個の形態素へと分割された。

次に、4,418種類の形態素のうち、出現頻度の上位5%に該当する221語を抽出した。これらの形態素から Modal word を選定するため、図 3.2 の評価軸上において、各形態素の示す極性(正負の方向性)を判断した。例えば、「目標」という形態素は、アスリートの動機づけと関連があり、心理的な活性や緊張の高まりを示すとともに、当事者にとってポジティブ(快)な意味をもつ語であると解釈できる。このことから、「目標」は、いずれの評価軸上でも正(+)の極性を示す Modal word と判断される。

一方で、「選手」のように評価軸上での極性の判断が困難な語や、「日本」、「スイング」といった固有名詞、スポーツ関連の用語などは、Modal word の候補から除外した。

以上のように、221 語の形態素について判断を行い、三種類の評価軸のいずれかにおいて、正 (+) または負 (-) の極性を示す語を選定した。選定に際しては、心理学の専門家である臨床心理士 2 名、スポーツの専門家であるスポーツトレーナー 1 名に意見を伺い、Modal word としての妥当性の検証を行った。最終的に、選定された語の中でも出現頻度の上位 20 語を、本研究における Modal Word $M_i (i = 1, 2, \dots, NM (= 20))$ とした。その上で、個々のアスリートの発話 U_j について、各 Modal word の出現頻度 $\tilde{f}_{ij} (i = 1, 2, \dots, NM, j = 1, 2, \dots, NU)$ を求めた。

3.5. TF-IDF 値および TF-DF 値の計算

本研究では、アスリートの発話中に含まれる語句に対して定量的な評価を行うため、TF-IDF 法を採用した。この手法では、文書に含まれる単語の出現頻度 (Term Frequency: TF, Luhn, 1957) および、特定の単語を含む文書の出現頻度である逆文書頻度 (Inverse Document Frequency: IDF) を求め、これらの値をかけ合わせた TF-IDF 値を算出する。TF-IDF 値は、ある文書における単語の重要度を示す指標であり、個々の単語について求めた TF-IDF 値が、その文書を表す特徴量となる。このとき、TF は単語の相対頻度、IDF は文書全体に対する単語ごとの重み付けの係数としての意味をもつ。本研究では、アスリートの発話 U_j を個々の文書としてとらえ、各 Modal Word M_i の TF-IDF 値を、下記の式に従って算出した。

$$h_{i,j} = f_{i,j} \cdot g_i \quad (3.1)$$

$$f_{i,j} = \frac{\tilde{f}_{i,j}}{\sum_k \tilde{f}_{k,j}} \quad (3.2)$$

$$g_i = \log_2 \frac{NU}{s_i} + 1 \quad (3.3)$$

ここで、 $\sum_k \tilde{f}_{k,j}$ は、発話 U_j に含まれるすべての Modal Word の出現頻度の合計を表しており、 $f_{i,j}$ は、 U_j における Modal Word M_i の出現頻度の比率を示す。また、 s_i は、 M_i を含む発話 U_j の総数である。上記の式より、 $h_{i,j}$ は $f_{i,j}$ と g_i の積であり、 U_j における Modal Word M_i の重要度を示している。

TF-IDF 法では、多くの文書に共通して含まれる語句よりも、特定の文書にのみ表れる語句の重要度を高く評価する。これに対し、文書集合からのキーワード抽出 (武田ら, 2001) や、文書要約 (小峰ら, 2002) に関する先行研究では、文書頻度 (Document Frequency: DF) を単語の重み付けに用いることがある。そこで、本研究では、TF と DF を組み合わせた TF-DF 値を新たに定義し、TF-IDF 値 $h_{i,j}$ と併せて分析に用いた。アスリートの発話 U_j に含まれる、各 Modal Word M_i の TF-DF 値 $\tilde{h}_{i,j}$ を本研究では、情報量 (エントロピー) で用いられる対数により次式のように定義する。

$$\tilde{h}_{i,j} = f_{i,j} \cdot (\log_2 s_i + 1) \quad (3.4)$$

クラスタ生成の観点から言えば、TF-IDF は各要素の違いを強調することで、多くのクラスタに分割するように働き、一つの大きなクラスタしか生成しないようなデータに対しては有効と思われる。一方、TF-DF は対象グループの中で共通に用いられる語句を強調して重み付けることで、それらをもつ要素同士の類似度が高く評価され、クラスタを構成するように働くので、各要素がバラバラでクラスタを構成し難いようなデータに対しては、有効と思われる。

ここで、‘3.3 言語データの収集’で取り上げた、インタビュー記事の内容および発話の例を用いて、TF-IDF 法および TF-DF 法の計算過程について述べる。

まず、アスリート $c(=3)$ の発話 $U_j (24 \leq j \leq 31)$ における、各 Modal Word M_i の頻度 $\tilde{f}_{i,j} (24 \leq j \leq 31)$ を表 3.2 および表 3.2 に示す。

表 3.1 アスリート $c(=3)$ の発話における各 Modal Word ($M_1 \sim M_{10}$) の出現頻度

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	$\sum_k \tilde{f}_{k,j}$
U_j	目標	勝つ	難しい	頑張る	楽しい	嬉しい	悪い	応援	自信	活躍	
24	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	1
27	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	1	1
30	0	0	0	0	0	0	0	0	0	0	0
31	2	0	0	0	0	0	0	1	0	0	3
s_i	65	54	52	38	36	35	36	44	26	27	
g_i	4.5012	4.7687	4.8231	5.2756	5.3536	5.3943	5.3536	5.0641	5.8231	5.7687	

表 3.2 アスリート $c(=3)$ の発話における各 Modal Word ($M_{11} \sim M_{20}$) の出現頻度

U_j	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	$\sum_k \tilde{f}_{k,j}$
	挑戦	勝てる	プレッシャー	ミス	不安	悔しい	負ける	集中	夢	成長	
24	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	1	0	0	0	0	1
27	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	1
30	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	3
s_i	24	44	25	27	22	22	24	23	18	22	
g_i	5.9386	5.0641	5.8797	5.7687	6.0641	6.0641	5.9386	6	6.3536	6.0641	

アスリートの発話のうち、 U_{31} に注目すると、Modal Word M_1 (“目標”) の出現頻度が2回、Modal Word M_8 (“応援”) の出現頻度が1回であり、Modal Word の頻度の合計値 $\sum_k \tilde{f}_{k,j}$ は3となる。すなわち、Modal Word M_1 のTF値 $f_{i,j} = \frac{2}{3} \doteq 0.6667$ となる。また、736($= NU$)の発話中で、Modal Word M_1 を含む発話 U_j は65個存在することから、 $S_i = 65$ であり、Modal Word M_1 のIDF値 $g_i = -\log_2 \frac{S_i}{NU} + 1 \doteq 4.5012$ と求めることができる。以上より、アスリート $c(=3)$ の発話 U_{31} において、Modal Word M_1 のTF-IDF値 $h_{i,j} = f_{i,j} \cdot g_i \doteq 3.0008$ となる。また、発話 U_{31} におけるModal Word M_1 のTF-DF値は、 $\tilde{h}_{i,j} = f_{i,j} \cdot (\log_2 S_i) \doteq 4.6816$ と算出される。

このように、それぞれのアスリートの発話ごとに、20語のModal Word M_i のTF-IDF値とTF-DF値をそれぞれ算出し、TF-IDF値を成分とする入力ベクトル $H_j = (h_{1,j}, h_{2,j}, \dots, h_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ および、TF-DF値を成分とする入力ベクトル $H'_j = (\tilde{h}_{1,j}, \tilde{h}_{2,j}, \dots, \tilde{h}_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ を得た。

3.6. 自己組織化マップの作成

自己組織化マップ (Self-Organizing Maps: SOM) は、Kohonen (1995) によって考案されたニューラルネットワークアルゴリズムであり、非線形写像により多次元データを低次元空間へマッピングする効果的な手法である。このアルゴリズムは、入力層

と出力層（マップ）という二層構造からなる。通常、マップは格子状の二次元平面として表される。マップ上の格子はニューロンと呼ばれ、それぞれにベクトル値（代表ベクトル）が割り当てられる。代表ベクトルは、入力データ（入力ベクトル）と等しい次元の成分をもつ。入力ベクトルは、個々の代表ベクトルとの比較を通じて、特定のニューロンへ分類され、マップ上に入力ベクトル間の関係性が可視化される。

通常の SOM のアルゴリズムでは、すべての入力ベクトルと代表ベクトルとが比較され、特定の入力ベクトルに対してユークリッドノルムが最小となる代表ベクトルをもつニューロン（勝者ニューロン）が決定する。その結果として、入力ベクトルはそれぞれの勝者ニューロンへ分類され、類似した特徴をもつ入力ベクトル同士が、マップ上の近い位置にマッピングされる。また、勝者ニューロンとその近傍のニューロンでは、分類された入力ベクトルに近づくよう、それぞれのもつ代表ベクトルの値を所与の式に従って更新する。入力ベクトルの分類と代表ベクトルの更新（学習）の繰り返しを通じて、多次元空間における入力ベクトルの関係性がマップ上に反映されていく。

SOM のアルゴリズムとしては、様々な手法が考案されており、クラスタ形成や特徴の抽出に特化した手法も存在する。一方、データの入力順に応じて解析結果が変化するようなアルゴリズム（逐次学習型 SOM）もあり、そのような SOM は、解析結果の再現性を保証できないという点から、試行毎に表示結果が変わる可能性があり、本研究の解析には不向きであるといえる。データの可視化を目的とした解析を行うためには、以下の4つの要件を満たす必要がある。

- 1) データの入力順に関わらず、常に同じ解析結果が得られる
- 2) マップ上の代表ベクトルの分布から、評価軸を見出すことができる
- 3) 隣接するニューロンがどのような領域を形成し、各領域がどのような性質を有するかを把握できる
- 4) 上記の‘3’について、客観的・自動的な処理が存在する

これらの要件を満たす SOM として、本研究では一括学習型の SOM (Batch-Learning

SOM: BLSOM) を採用し, Kanaya et al. (2001) により作成されたプログラム "BLSOM Viewer" による解析を実施した. この解析の詳細を以下に記す.

1) 代表ベクトルの初期値の設定

出力層の各ニューロンにおける代表ベクトルの初期値を設定するため, 入力ベクトル $\mathbf{H}_j = (h_{1,j}, h_{2,j}, \dots, h_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ および $\mathbf{H}'_j = (\tilde{h}_{1,j}, \tilde{h}_{2,j}, \dots, \tilde{h}_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ に対して主成分分析を実施した. これは, 各アスリートの発話 U_j における Modal Word M_i について求めた TF-IDF 値 (または TF-DF 値) を成分にもつ, 20 次元のベクトルデータである. 主成分分析の結果として得られた, 第一主成分, 第二主成分それぞれの固有ベクトルを $\mathbf{a}_1, \mathbf{a}_2$ とし, 標準偏差を σ_1, σ_2 とおく. 代表ベクトルの初期値 $\mathbf{W}_{l,m} = (W_{1,l,m}, W_{2,l,m}, \dots, W_{NM,l,m}) (l = 1, 2, \dots, L, m = 1, 2, \dots, M)$ は, これらの値を用いて, 下記の式により算出される (Kanaya et al., 2001).

$$\mathbf{W}_{l,m} = \mathbf{H}_{AV} + 5\sigma_1 \left[\mathbf{a}_1 \frac{(l-\frac{L}{2})}{L} + \mathbf{a}_2 \frac{(m-\frac{M}{2})}{M} \right] \quad (3.5)$$

$$L = \lceil 5\sigma_1 \rceil \quad (3.6)$$

$$M = \left\lfloor \frac{L \cdot \sigma_2}{\sigma_1} \right\rfloor \quad (3.7)$$

ここで, l は SOM のマップを構成するニューロンの列番号, m は行番号を表し, L, M はそれぞれ, マップの列数と行数である. \mathbf{H}_{AV} はすべての入力ベクトル \mathbf{H}_j (または \mathbf{H}'_j) の平均ベクトルを表す. ここで, $\lceil x \rceil$ は x の天井関数 (切り上げ), $\lfloor x \rfloor$ は x の床関数 (切り捨て) である.

2) 入力ベクトルの分類

個々の入力ベクトルは, 20 次元のユークリッド空間において, ユークリッドノルムが最小となる代表ベクトルをもつニューロン (勝者ニューロン) へと分類される. その分類の手順を以下に述べる.

まず, すべての入力ベクトルと代表ベクトル間のユークリッドノルムを算出し, j 番

目の入力ベクトル \mathbf{H}_j (または \mathbf{H}'_j) に対する勝者ニューロン $\mathbf{W}_{l,m}$ を決定する. このとき \mathbf{H}_j は, 下記の式を満たす勝者ニューロンと, その近傍ニューロンの集合 $\mathbf{V}_{l,\tilde{m}}$ に分類される.

$$\begin{cases} l - \beta(r) \leq \tilde{l} \leq l + \beta(r) \\ m - \beta(r) \leq \tilde{m} \leq m + \beta(r) \end{cases} \quad (3.8)$$

$$\beta(r) = \max\{0, \beta(1) - r\}, \beta(1) = \frac{l}{4} \quad (3.9)$$

ここで, r は学習の回数を示し, 入力ベクトルの分類および, 後述する代表ベクトルの更新が終了する度に 1 ずつ増加する. $\beta(r)$ は, 勝者ニューロンの近傍に含まれる範囲を示すパラメータであり, 学習回数 r の増加に伴って, その値は減少していく.

3) 代表ベクトルの更新

すべての入力ベクトルの分類が完了した後に, それぞれのニューロンがもつ代表ベクトルの更新が, 下記の式に従って行われる. 入力ベクトル \mathbf{H}_j を分析に用いた場合の, 代表ベクトルの更新式を以下に示す.

$$\mathbf{W}_{l,m}^{new} = \mathbf{W}_{l,m}^{old} + \alpha(r) \left[\sum_{\mathbf{H}_k \in \mathbf{V}_{l,\tilde{m}}} \frac{\mathbf{H}_k}{N_{l,\tilde{m}}} - \mathbf{W}_{l,m}^{old} \right] \quad (3.10)$$

$$\alpha(r) = \max\left\{0.01, \alpha(1) \left(1 - \frac{r}{T}\right)\right\} \quad (3.11)$$

$$\alpha(1) = 0.5 \text{ and } T = 100 \quad (3.12)$$

また, 入力ベクトル \mathbf{H}'_j を分析に用いた場合は, 代表ベクトルの更新式として, (3.10) の代わりに以下の(3.13)式を用いる.

$$\mathbf{W}_{l,m}^{new} = \mathbf{W}_{l,m}^{old} + \alpha(r) \left[\sum_{\mathbf{H}'_k \in \mathbf{V}_{l,\tilde{m}}} \frac{\mathbf{H}'_k}{N_{l,\tilde{m}}} - \mathbf{W}_{l,m}^{old} \right] \quad (3.13)$$

ここで、 $\alpha(r)$ は学習率係数と呼ばれるものであり、試行回数 r の増加に伴って減少する値である。また、 T は学習の総回数である。これらの式に従いながら、学習が進むにつれて、代表ベクトル $\mathbf{W}_{l,m}$ の更新の度合いは徐々に小さくなり、最適値へと収束していく。その収束性については、数学的に保証されている。

3.7. ファジィクラスタ分析（ニューロンのクラスタリング）

次に、SOMにより更新された代表ベクトル $\mathbf{W}_{l,m} = (W_{1,l,m}, W_{2,l,m}, \dots, W_{NM,l,m})$ を入力データとして、ファジィクラスタ分析を実施した。前述のSOMでは、Modal Word M_i のTF-IDF値を成分にもつ入力ベクトル \mathbf{H}_j および、TF-DF値を成分にもつ入力ベクトル \mathbf{H}'_j の分類結果に従い、 $\mathbf{W}_{l,m}$ の更新がなされた。すなわち、マップ上のニューロンがもつ $\mathbf{W}_{l,m}$ は、心理状態を示すModal Wordの特徴的な分布を反映したものであるといえる。このような特徴に応じて、マップ上の各領域を弁別するために、クラスタリング手法としてファジィクラスタ分析を採用した。

通常のクラスタ分析（ハードクラスタリング）における集合演算では、1つの要素が対象の集合に所属するか否かは真か偽の2値をベースとしている。これに対し、ファジィクラスタ分析における集合演算では、1つの要素が対象の集合に所属する度合いが0から1の値で表されることをベースとしている。このように、ファジィクラスタ分析では、対象のもつ多義性や、曖昧な性質を考慮した分析が可能であることから、自然言語で書かれた文書の分類などによく用いられており、本研究でも採用することにした。

分析にあたり、まず、すべての代表ベクトル $\mathbf{W}_{l,m}$ の間で相関係数を求め、この値を成分にもつ相関行列（メンバシップ行列 \mathbf{F} ）を以下のように定義した。

$$\mathbf{F} = \begin{pmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,T} \\ F_{2,1} & F_{2,2} & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ F_{T,1} & & & F_{T,T} \end{pmatrix} = \begin{pmatrix} \frac{\mathbf{W}_{1,1} \cdot \mathbf{W}_{1,1}}{|\mathbf{W}_{1,1}| \cdot |\mathbf{W}_{1,1}|} & \frac{\mathbf{W}_{1,1} \cdot \mathbf{W}_{1,2}}{|\mathbf{W}_{1,1}| \cdot |\mathbf{W}_{1,2}|} & \dots & \frac{\mathbf{W}_{1,1} \cdot \mathbf{W}_{L,M}}{|\mathbf{W}_{1,1}| \cdot |\mathbf{W}_{L,M}|} \\ \frac{\mathbf{W}_{1,2} \cdot \mathbf{W}_{1,1}}{|\mathbf{W}_{1,2}| \cdot |\mathbf{W}_{1,1}|} & \frac{\mathbf{W}_{1,2} \cdot \mathbf{W}_{1,2}}{|\mathbf{W}_{1,2}| \cdot |\mathbf{W}_{1,2}|} & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{W}_{L,M} \cdot \mathbf{W}_{1,1}}{|\mathbf{W}_{L,M}| \cdot |\mathbf{W}_{1,1}|} & \dots & \dots & \frac{\mathbf{W}_{L,M} \cdot \mathbf{W}_{L,M}}{|\mathbf{W}_{L,M}| \cdot |\mathbf{W}_{L,M}|} \end{pmatrix} \quad (3.14)$$

ここで、 $T = L \cdot M$ とおく。

次に、メンバシップ行列 \mathbf{F} は、下記の式に従って可達行列 $\tilde{\mathbf{F}}$ へと変換される。

$$\tilde{\mathbf{F}} = \underbrace{\mathbf{F} \otimes \mathbf{F} \otimes \dots \otimes \mathbf{F}}_T \quad (3.15)$$

$$\mathbf{F} \otimes \mathbf{F} = \begin{pmatrix} \text{Max}\{\text{Min}(F_{1,1}F_{1,1}), \text{Min}(F_{1,2}F_{2,1}), \dots, \text{Min}(F_{1,30}F_{30,1})\} & \dots & \text{Max}\{\text{Min}(F_{1,1}F_{1,30}), \text{Min}(F_{1,2}F_{2,30}), \dots, \text{Min}(F_{1,30}F_{30,30})\} \\ \vdots & \ddots & \vdots \\ \text{Max}\{\text{Min}(F_{30,1}F_{1,1}), \text{Min}(F_{30,2}F_{2,1}), \dots, \text{Min}(F_{30,30}F_{30,1})\} & \dots & \text{Max}\{\text{Min}(F_{30,1}F_{1,30}), \text{Min}(F_{30,2}F_{2,30}), \dots, \text{Min}(F_{30,30}F_{30,30})\} \end{pmatrix} \quad (3.16)$$

このとき、最大値関数 $\text{Max}(F_{1,1}F_{1,1}, F_{1,2}F_{2,1}, \dots, F_{T,T}F_{T,T})$ は、すべての要素積の中の最大値を示し、これは可達行列の計算における論理和に相当する。 $T(=L \cdot M)$ 回の試行の後に得られた可達行列 $\tilde{\mathbf{F}}$ の各成分は、代表ベクトル間の類似度 $R(0 \leq R \leq 1)$ を示す。ファジィクラスタ分析では、適切な類似度 R を基準として指定することで、類似度が指定された値以上となる要素同士が結合し、個々のクラスタを形成する。本研究においては、代表ベクトル $\mathbf{W}_{l,m}$ がクラスタの構成要素であるため、分析の結果を踏まえ、マップ上のニューロンを複数の領域に統合することが可能となる。

3.8. トポロジー的評価

SOMによる解析の結果、入力ベクトル \mathbf{H}_j および \mathbf{H}'_j を用いた100回の学習により、更新された代表ベクトル $\mathbf{W}_{l,m}$ が得られた。さらに、ファジィクラスタ分析の結果、代表ベクトル間の類似度 R に応じて形成されるクラスタ構造が明らかになり、マップ上のニューロンをいくつかの領域に統合することが可能となった。これらの処理を通じ

た、個々のアスリートの心理状態および変化の二次元平面上での可視化が、本研究で提案する手法である。

キーワードの分布に応じた領域分けがなされたマップ上で、各アスリートの入力ベクトル \mathbf{H}_j および \mathbf{H}'_j が割り当てられる座標と、その時間的な変化に注目することで、個々人の心理状態が変化する様子を視覚的にとらえることができると考えた。その具体的手法として、代表ベクトル $\mathbf{W}_{l,m}$ が示すキーワードの分布と、ニューロンのクラスタリング結果に基づき、マップ上に心理的評価軸を抽出することにした。この評価軸上で \mathbf{H}_j および \mathbf{H}'_j の座標が遷移する様子をアスリートごとに可視化し、その動きから読み取れる心理状態とその変化について、心理学的な視点から解釈を行うことが可能となる。

上記の提案手法では、分析対象の言語データ(発話 U_j)を 20 次元のベクトル ($\mathbf{H}_j, \mathbf{H}'_j$) として表現する。ベクトルの各成分は、発話中の各 Modal word M_i の TF-IDF 値 $h_{i,j}$, または TF-DF 値 $\tilde{h}_{i,j}$ であり、発話時点におけるアスリートの心理状態を示すものと想定する。これらのベクトルを SOM によりマッピングすることで、多次元ベクトル間の関連性が二次元空間上に可視化される。さらに、マッピングの結果に応じて、マップ上のニューロンがもつ代表ベクトル $\mathbf{W}_{l,m}$ の値が更新され、各入力ベクトル間の関連性を反映したマップを作成することができる(図 3.4)。下図は、'3.3 言語データの収集'で取り上げたアスリート $c(=3)$ の発話 U_{31} を例として、Modal word の TF-IDF 値をベクトル \mathbf{H}_{31} として表し、二次元空間上にマッピングする過程を示したものである。

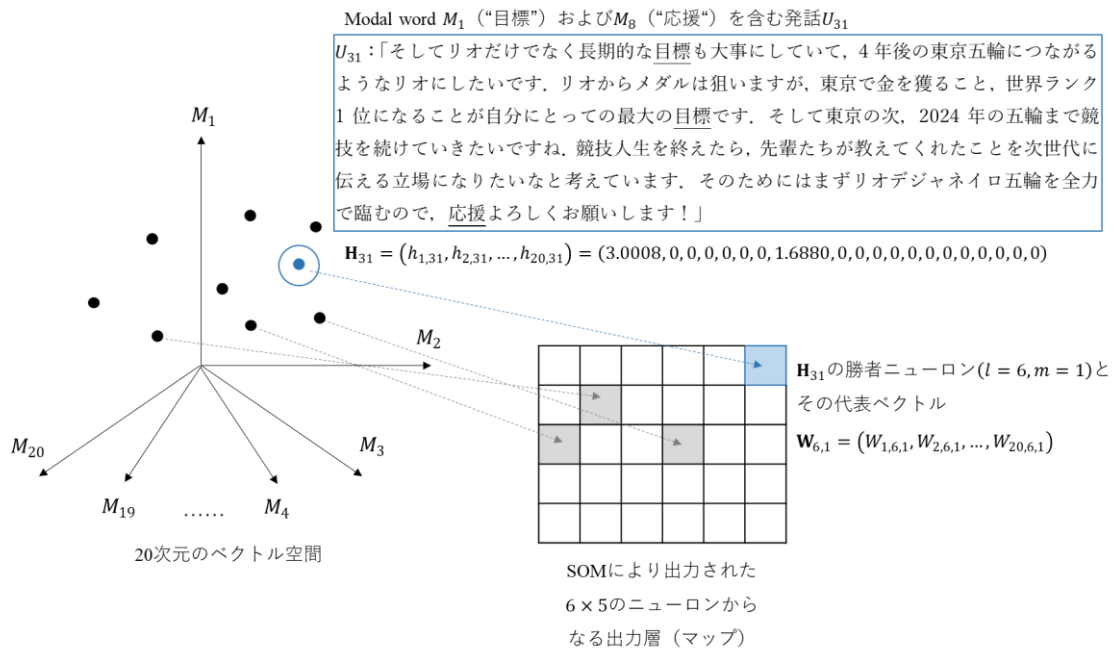


図 3.4 言語データのベクトル表現と二次元平面へのマッピング

代表ベクトル $W_{l,m}$ の値を用いたファジィクラスタ分析により、マップ上のニューロンはいくつかの領域(クラスタ)に統合される。所属するニューロンの代表ベクトルの成分によって、個々の領域は心理的な意味をもつと考える。これらの領域の位置関係から、マップ上に心理的評価軸を抽出することができる。

代表ベクトル $W_{l,m}$ の各成分 $W_{i,l,m}$ は、Modal word M_i の TF-IDF 値 $h_{i,j}$ または TF-DF 値 $\tilde{h}_{i,j}$ に相当し、ニューロン間での縦・横、あるいは対角線の軸に沿って、これらの値は増減する。ニューロン間に生じる代表ベクトル成分の増減から、Modal word が示す感情の分布を読み取ることができ、これらのニューロン間の距離が、感情の速さ(図 3.2(b)) に相当するものと考えられる。

本研究では、マップ上のニューロンを 10 のクラスタ (CL1-CL10) に分類し(図 3.5(b))、各クラスタにおける代表ベクトル成分の平均値を求めた。クラスタ間での平均値の増減に注目することで、個々の Modal word が示す感情の速さと、その心理変化の方向性が見出しやすくなった。以下では、各感情の速さと方向性を踏まえ、心理的評価軸を抽出するプロセスについて述べる。

例えば、マップ上の CL7 では、代表ベクトル $W_{l,m}$ の成分 $W_{i,l,m}$ の平均値が、いずれも

0に近い値を示している(図3.5(a)). これに対し, CL6では, $W_{4,l,m}, W_{8,l,m}, W_{15,l,m}$ の平均値が, 相対的に高い値を示している. これらは, Modal word M_4 (頑張る), M_8 (応援), M_{15} (不安)のTF-IDF値に相当する値である. すなわち, CL7からCL6にかけて代表ベクトル成分 $W_{4,l,m}, W_{8,l,m}, W_{15,l,m}$ の増加がみられることから, M_4, M_8, M_{15} が示す感情は, CL7からCL6への遷移として, 図3.5(b)の矢線のように表すことができる. このとき, 各Modal wordの感情の評価軸上における極性(図3.5(c))に注目すると, 緊張—弛緩の軸上においてのみ, すべてのModal wordが正(+)の極性を示していることがわかる. そこで, CL7からCL6を結ぶ矢線を, 「緊張」を表す心理的評価軸として抽出した.

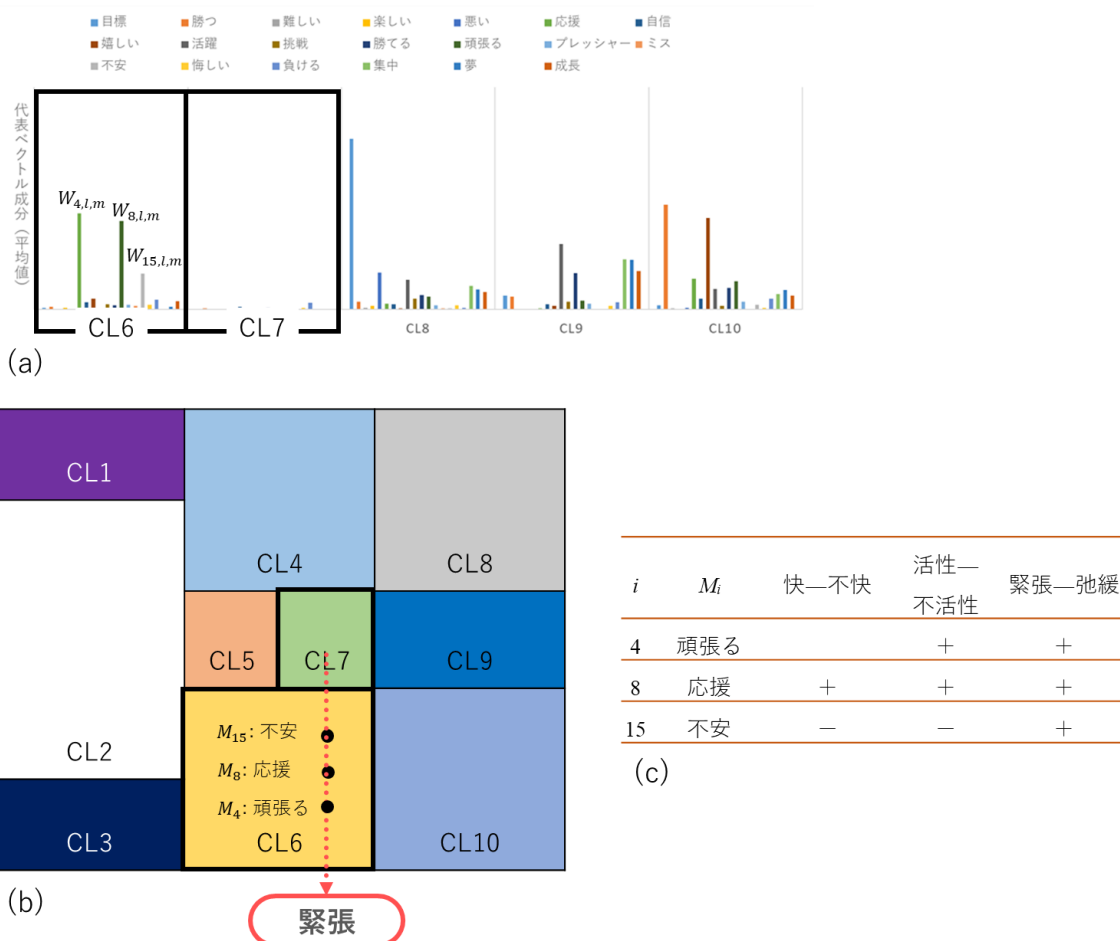


図 3.5 心理的評価軸の抽出プロセス

また、個々の発話に対応する入力ベクトルが、いずれの領域へマッピングされるのか（これを“座標”と呼ぶ）に注目することで、対象者の心理状態を解釈することが可能となる。さらに、時間の変化、すなわちインタビューの進展に伴って、入力ベクトルの座標が遷移する様子を矢線図で表すことにより、インタビュー中の心理変化がマップ上に可視化される（図 3.6）。このように、対象者の心理状態を示す入力ベクトルは二次元空間上にマッピングされ、心理変化に応じて領域間を遷移する。下図は提案手法により得られる可視化結果の概略図であり、マップ上の領域区分と、各領域を特徴づける Modal word を右上に示している。また、右下の図は、アスリート $c(=3)$ の発話 $U_{24}, U_{25}, \dots, U_{31}$ のマップ上での座標遷移を矢線図で表したものである。

SOM により決定される入力ベクトルの座標は、ベクトル間の位相関係を示すものであり、座標間の距離が意味をもたない。そのため、マップ上の座標遷移が示す心理変化を理解するためには、領域の隣接関係およびそれらの推移律に注目したトポロジー的な解釈が有効と判断した。そこで、本研究では、入力ベクトルが遷移するマップ上の各領域を要素とする集合に対して、要素間の新たな距離を定義することにより、位相空間としてとらえるようにする。この位相空間の導入により、領域の隣接関係からは、発話に対する心理的評価軸を抽出することができ、領域間の座標遷移からは、対象者の心理変化を読み取ることが可能となる。これらは、心理状態を表す多次元ベクトルのトポロジーを反映したものであり、本研究では、その性質を「心理学的トポロジー」と呼ぶ。

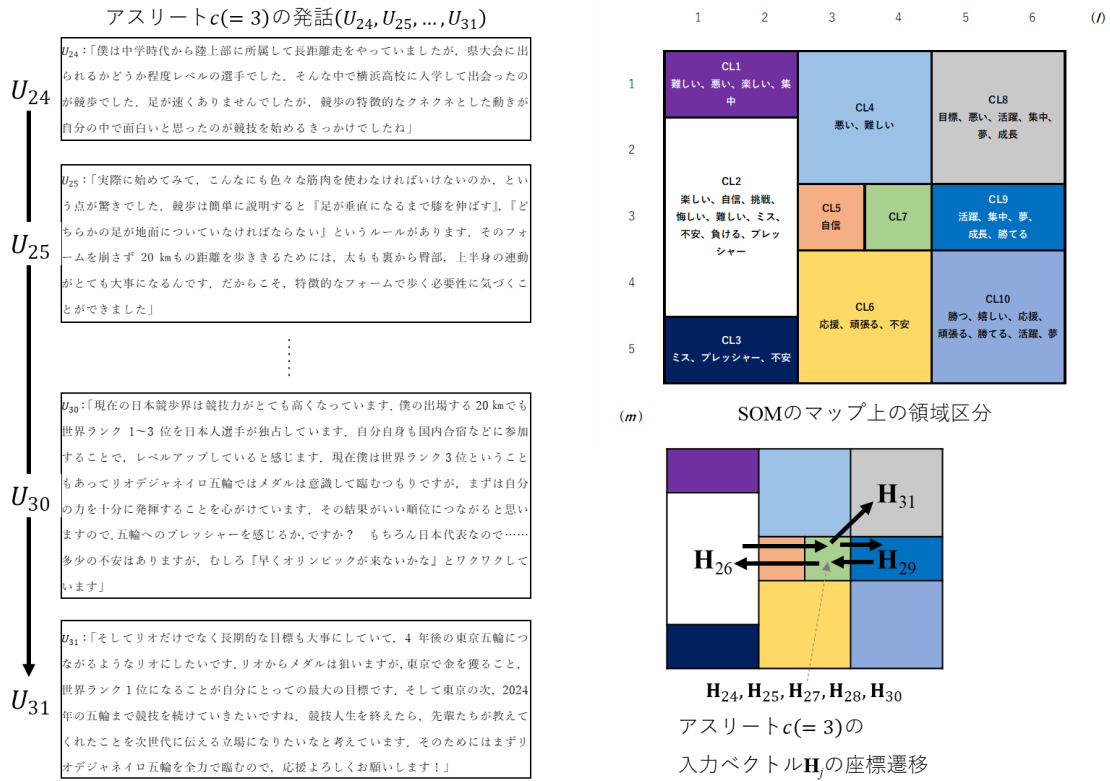


図 3.6 マップ上でのベクトル遷移の可視化

このとき、すべての発話 $U = \{U_1, U_2, \dots, U_{NU}\} (NU = 736)$ に対するファジィクラスタ分析の結果、生成されるクラスタを $CL_t (t = 1, 2, \dots, k, 1 \leq k \leq 736)$ とおく。そこで、集合 U とその冪集合 2^U で、 $(U, 2^U)$ として自然な位相を入れることができる。そして、 $CL_t \in 2^U$ であるので、この位相空間上でクラスタ CL_p, CL_q 間の距離 $d(CL_p, CL_q) (p = 1, 2, \dots, k, q = 1, 2, \dots, k)$ を導入することにより、クラスタの位置関係を把握することが可能となる。距離 d の定義については、‘3.9.3 各クラスタの特徴に関する考察’において詳述する。

3.9. 結果および考察

3.9.1. Modal Word M_i の選定結果

本研究では、60名のアスリート ($NC = 60$) に対するインタビュー記事を、1名につき1記事ずつ収集した。それぞれの記事には、5～20個の発話が含まれ、60名のアスリートから合計で736の発話 $U_j (j = 1, 2, \dots, NU (= 736))$ が得られた。また、前述したように、すべての発話 $U = \{U_1, U_2, \dots, U_{NU}\}$ は、形態素解析により総数 21,007 ($= NW$), 4,418種類 の形態素へ分割された。

これらの形態素から、アスリートの感情を反映する20語 ($NM = 20$) を Modal Word M_i として選定した。選定された20の Modal Word と、それぞれの出現確率 $\frac{e_i}{NW}$ を表 3.3 に示す。ここで、 $e_i = \sum_k \tilde{f}_{i,k}$ は、すべての発話 U における Modal Word M_i の出現頻度の合計を表す。また、表 3.4 は、感情の評価軸 (u : 快—不快, v : 活性—不活性, w : 緊張—弛緩) 上における, Modal word の極性 (正: +, 負: -) についての判断をまとめたものである。例えば、 M_5 (楽しい) は、快—不快および活性—不活性の評価軸上では正の極性を示すと判断し、これらを+の記号で示している。緊張—弛緩の評価軸上での極性については、判断が困難であるため、該当箇所を空欄とした。

表 3.3 選定された Modal word M_i

I	M_i	$\frac{e_i}{NW} \times 10^3$									
1	目標	4.61	6	嬉しい	2.19	11	挑戦	1.57	16	悔しい	1.29
2	勝つ	3.14	7	悪い	2.05	12	勝てる	1.52	17	負ける	1.24
3	難しい	2.71	8	応援	2.05	13	プレッシャー	1.43	18	集中	1.24
4	頑張る	2.48	9	自信	2.00	14	ミス	1.33	19	夢	1.19
5	楽しい	2.33	10	活躍	1.67	15	不安	1.29	20	成長	1.14

表 3.4 感情の評価軸上における極性

i	M_i	快—不快	活性—不活性	緊張—弛緩
1	目標	+	+	+
2	勝つ	+	+	+
3	難しい	-	-	+
4	頑張る		+	+
5	楽しい	+	+	
6	嬉しい	+	+	
7	悪い	-	-	
8	応援	+	+	+
9	自信	+	+	
10	活躍	+	+	+
I	M_i	快—不快	活性—不活性	緊張—弛緩
11	挑戦		+	+
12	勝てる	+	+	
13	プレッシャー	-	-	+
14	ミス	-	-	
15	不安	-	-	+
16	悔しい	-	+	
17	負ける	-	-	
18	集中	+		
19	夢	+	+	
20	成長	+	+	

アスリートの個々の発話 U_j について求めた, Modal Word M_i の出現頻度 $\tilde{f}_{i,j}$ をもとに, それぞれの TF-IDF 値 $h_{i,j}$ および, TF-DF 値 $\tilde{h}_{i,j}$ を算出した. 上記の手続きにより得られた入力ベクトル \mathbf{H}_j および \mathbf{H}'_j を, 自己組織化マップによる解析に用いた.

3.9.2. ファジィクラスタ分析によるニューロンのクラスタリング結果

入力ベクトル \mathbf{H}_j に対する主成分分析の結果，第一主成分および第二主成分の標準偏差は，それぞれ $\sigma_1 = 1.103, \sigma_2 = 1.082$ であった．この値から，SOM のマップの列数は $L = [5\sigma_1] = 6$ ，行数は $M = \left\lfloor L \cdot \frac{\sigma_2}{\sigma_1} \right\rfloor = 5$ となり，BLSOM の結果として， 6×5 のマップが出力された．また，マップを構成する 30 のニューロンに対する，入力ベクトル \mathbf{H}_j の分類結果を得た．さらに， \mathbf{H}_j の値に基づいて更新された代表ベクトル $\mathbf{W}_{l,m}$ の値に対して，ファジィクラスタ分析を実施した．分析に用いたメンバシップ行列 \mathbf{F} と，変換後の可達行列 $\tilde{\mathbf{F}}$ を図 3.7 および図 3.8 に示す．これらの図は，それぞれの行列の成分，すなわち代表ベクトル $\mathbf{W}_{l,m}$ 間の類似度を，値に応じて色分けしたコンター図である．

入力ベクトル \mathbf{H}'_j に対する主成分分析の結果，第一主成分および第二主成分の標準偏差は，それぞれ $\sigma_1 = 1.085, \sigma_2 = 1.080$ であった．この値から，SOM のマップの列数は $L = [5\sigma_1] = 6$ ，行数は $M = \left\lfloor L \cdot \frac{\sigma_2}{\sigma_1} \right\rfloor = 5$ となり，BLSOM の結果として， 6×5 のマップが出力された．また， \mathbf{H}'_j の値に基づいて更新された代表ベクトル $\mathbf{W}_{l,m}$ の値に対して，ファジィクラスタ分析を実施した．分析に用いたメンバシップ行列 \mathbf{F} と，変換後の可達行列 $\tilde{\mathbf{F}}$ を図 3.9 および図 3.10 に示す．

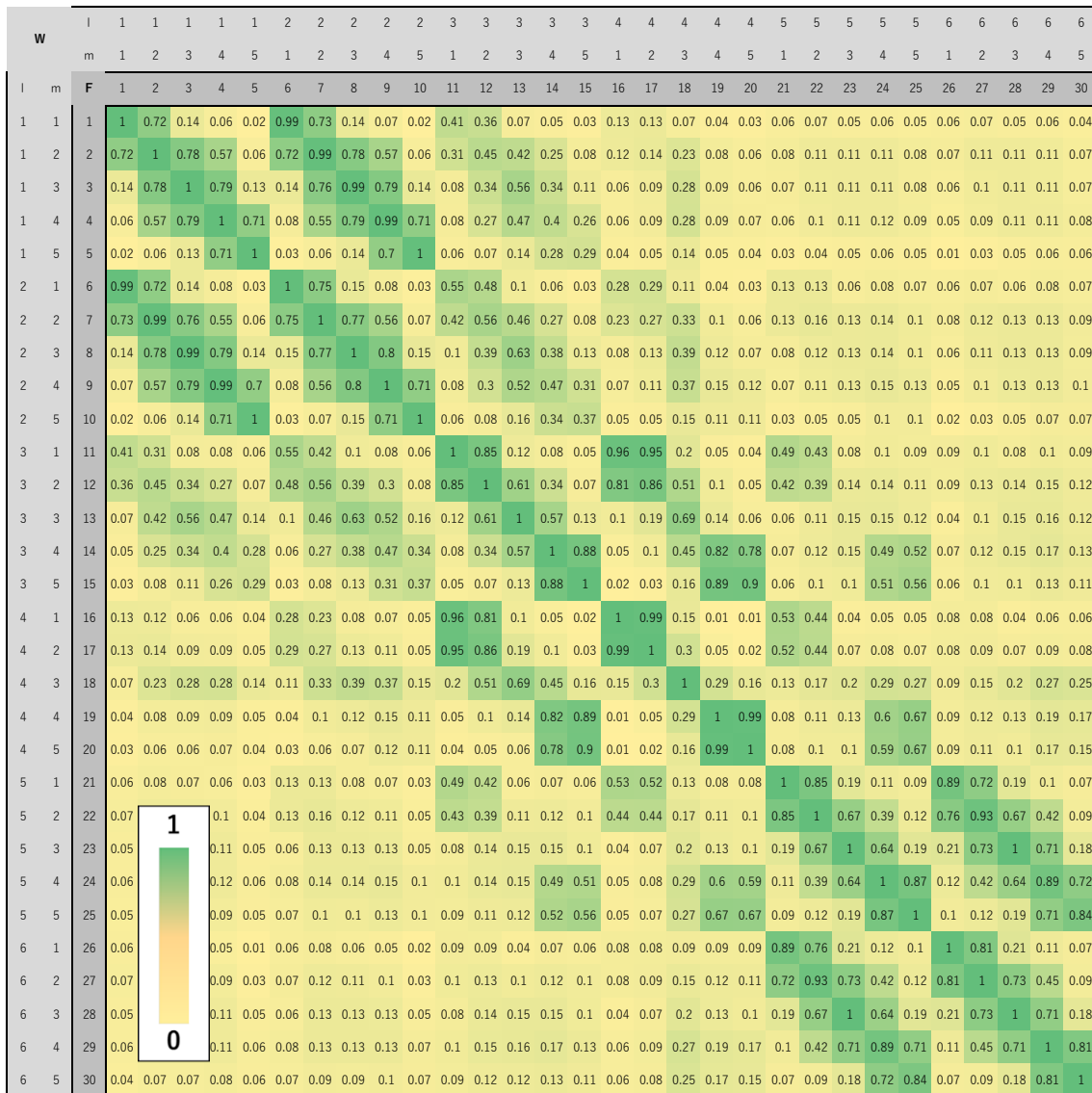


図 3.7 メンバシップ行列 F (TF-IDF 値)

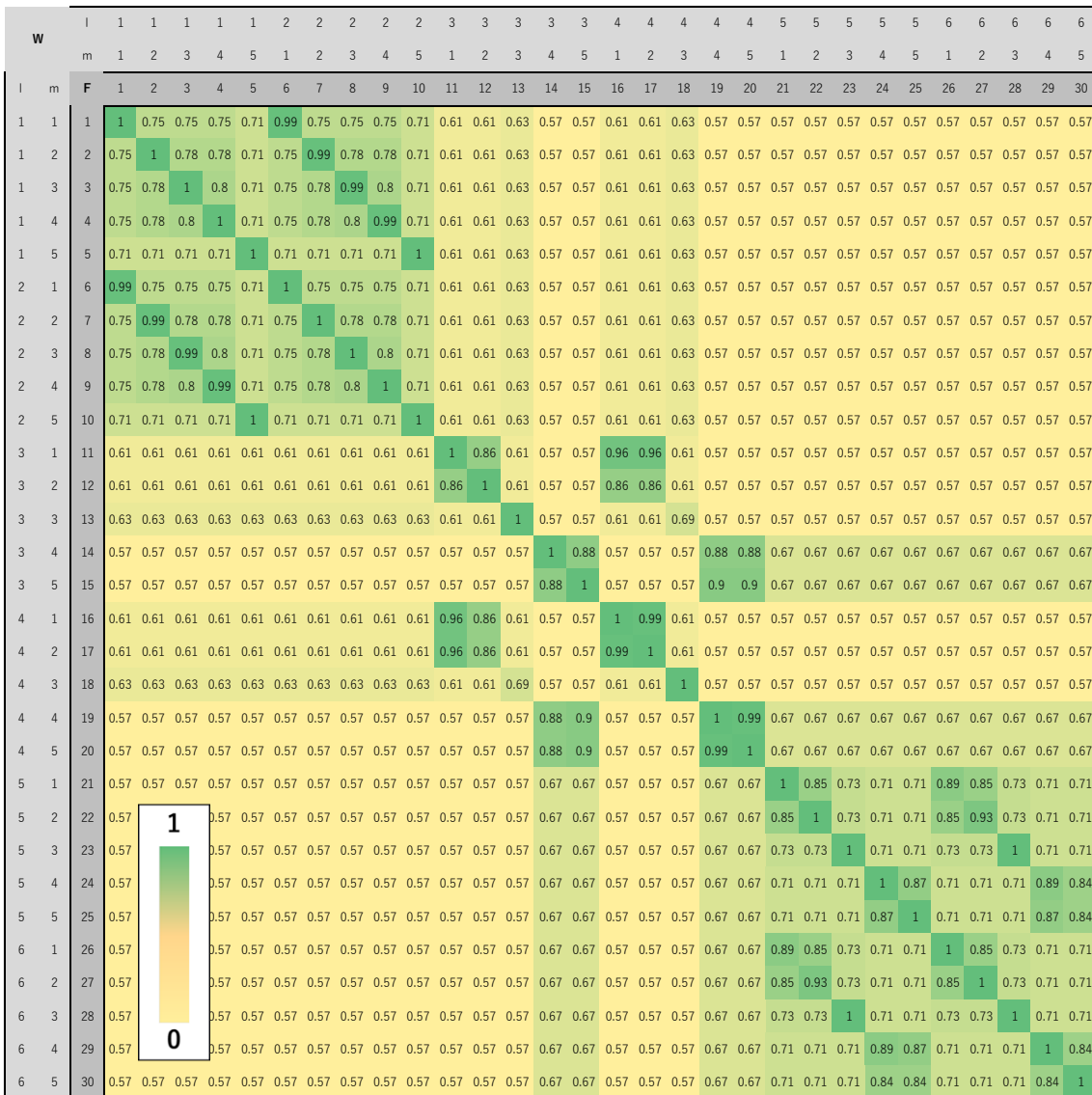


図 3.8 可達行列F (TF-IDF 値)

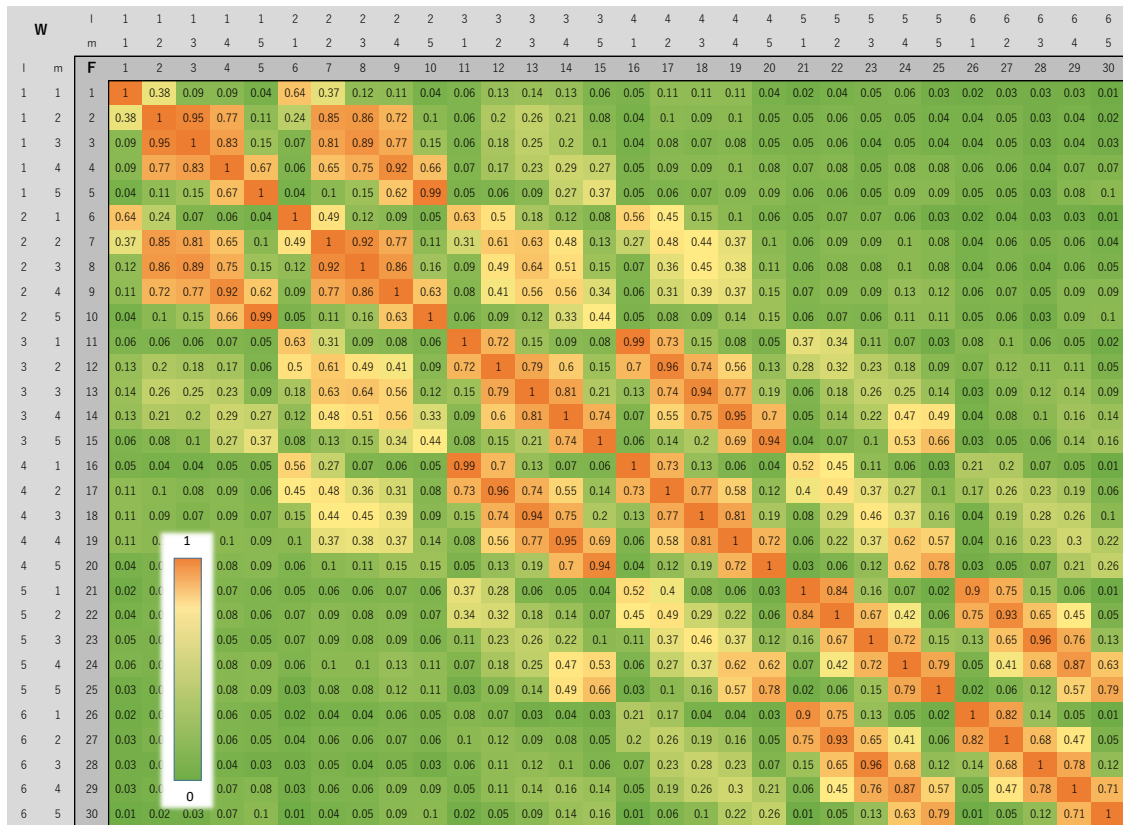


図 3.9 メンバシップ行列 F (TF-DF 値)

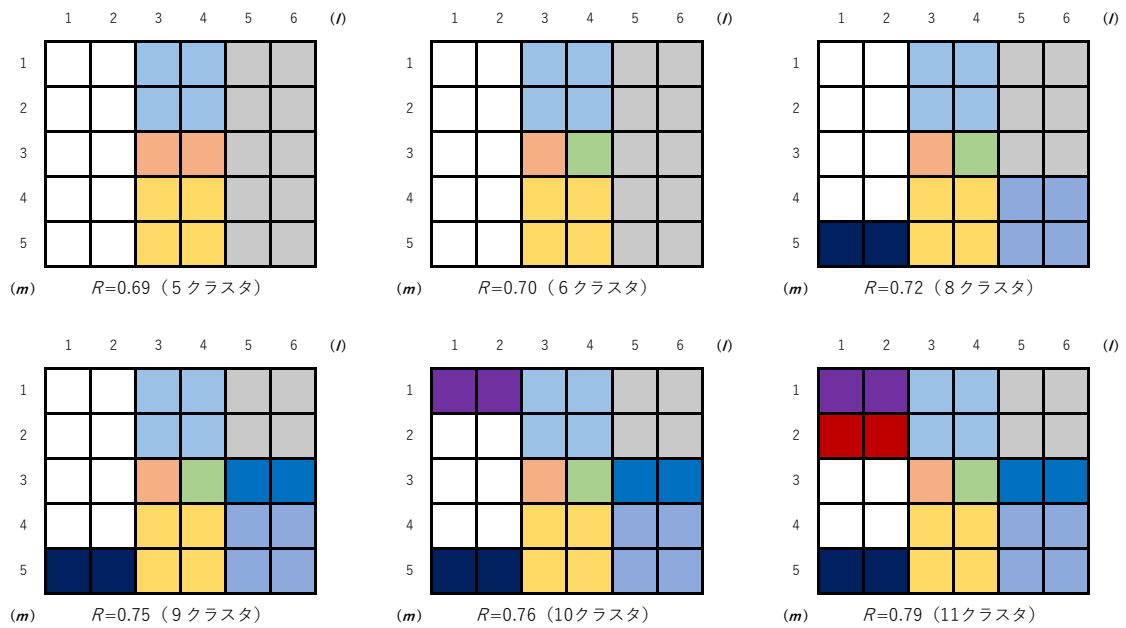


図 3.11 類似度 R に応じたニューロンのクラスタリング結果 (TF-IDF 値)

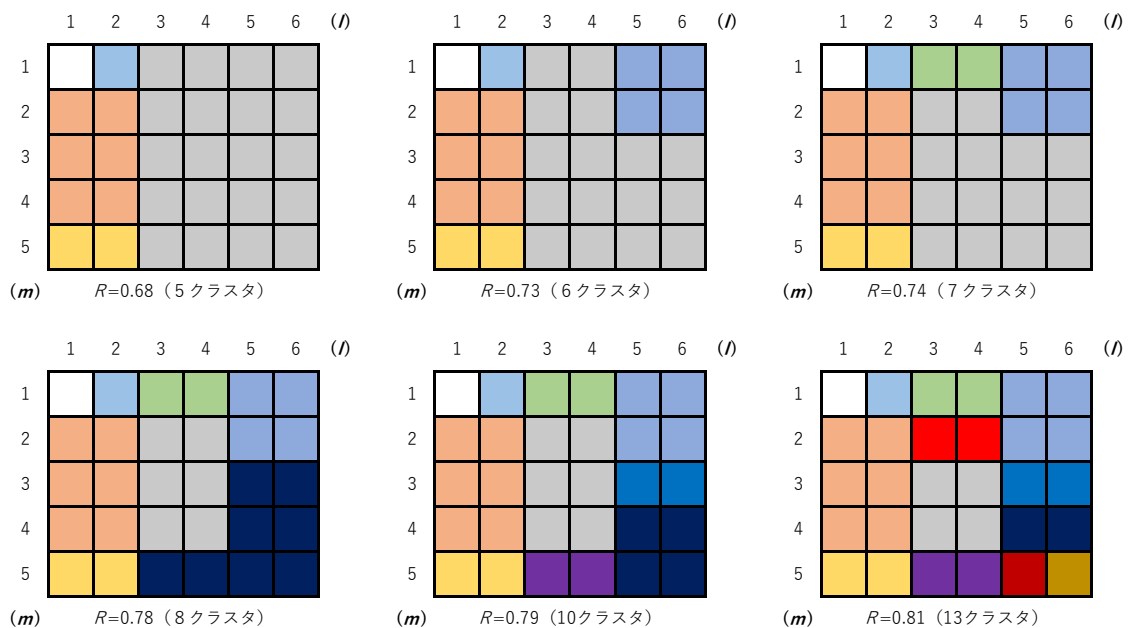


図 3.12 類似度 R に応じたニューロンのクラスタリング結果 (TF-DF 値)

上図は、SOM の結果として出力される 6×5 のマップであり、個々の格子がマップを構成するニューロンを表す。ファジィクラスタ分析において、類似度 R の値に応じて

異なるクラスタが形成される様子を表現するために、TF-IDF 値、TF-DF 値の分析結果のそれぞれについて、6 パターンのマップを作成した。例として、図 3.11 の左上のマップは、クラスタ形成の基準となる類似度 $R = 0.69$ と設定した場合に、マップ上の 30 のニューロンが 5 つのクラスタを形成することを示している。同様に、類似度 $R = 0.70$ と設定した場合には、6 つのクラスタが形成される。同じクラスタに所属するニューロンは同色で示され、クラスタリングの結果を視覚的にとらえることができる。

本研究では、SOM のマップを構成するニューロンをいくつかの領域に統合し、その配置に基づいてマップ上の評価軸を抽出するために、ファジィクラスタ分析を実施した。評価軸の抽出に際しては、マップの縦軸および横軸、対角線上に並ぶニューロンに注目し、その特徴を理解する必要がある。上に示されたマップのうち、図 3.11 の類似度 $R = 0.76$ および図 3.12 の類似度 $R = 0.79$ では、いずれも 10 のクラスタが形成され、縦軸および横軸、対角線上の領域が均等に分割されている。そのため、マップ上を区分する各クラスタが示す特徴から、評価軸を抽出できると考えられる。そこで、心理の専門家と協議・検討して、ここでは、最適なクラスタ数を 10 と設定した。そして、分類された 10 個の各クラスタの特徴について考察した上で、マップ上の評価軸について検討を行った。

3.9.3. 各クラスタの特徴に関する考察

TF-IDF 値を分析に用いた場合に、類似度 $R = 0.76$ において形成される 10 のクラスタを CL1, CL2, ..., CL10 と命名し (図 3.13), それぞれのクラスタに所属するニューロンがもつ代表ベクトル $\mathbf{W}_{l,m}$ の値に注目した。各クラスタに属するニューロンについて、クラスタ別に求めた代表ベクトル $\mathbf{W}_{l,m}$ の平均値を示す (図 3.14 および図 3.15)。

同様に、TF-DF 値を分析に用いた場合に、類似度 $R = 0.79$ において形成される 10 のクラスタを CL1, CL2, ..., CL10 と命名し (図 3.16), それぞれのクラスタに所属するニューロンがもつ代表ベクトル $\mathbf{W}_{l,m}$ の値に注目した。各クラスタに属するニューロンについて、クラスタ別に求めた代表ベクトル $\mathbf{W}_{l,m}$ の平均値を示す (図 3.17 および図 3.18)。

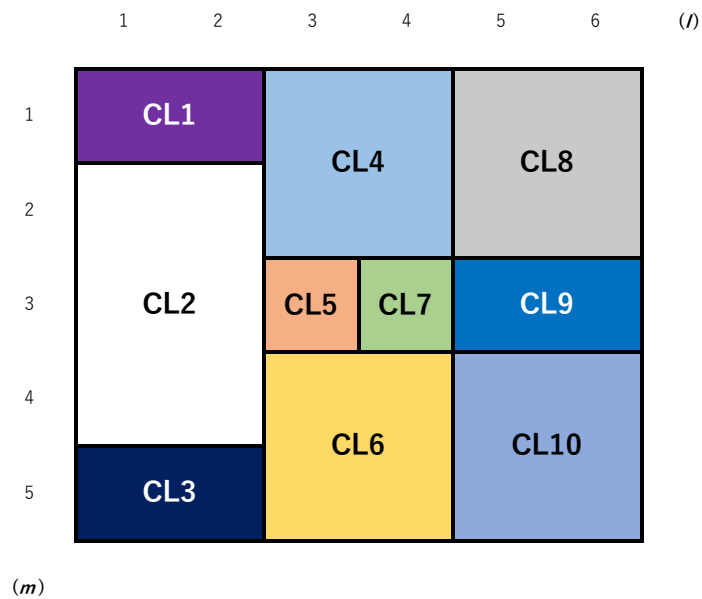


図 3.13 類似度 $R = 0.76$ において形成されるクラスター (TF-IDF 値)

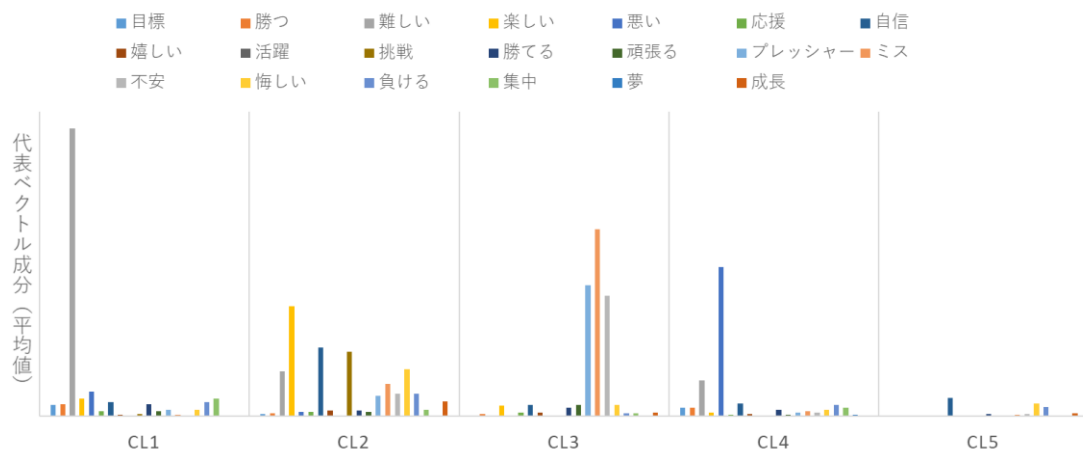


図 3.14 CL1-5 に属するニューロンの代表ベクトル成分 (TF-IDF 値)

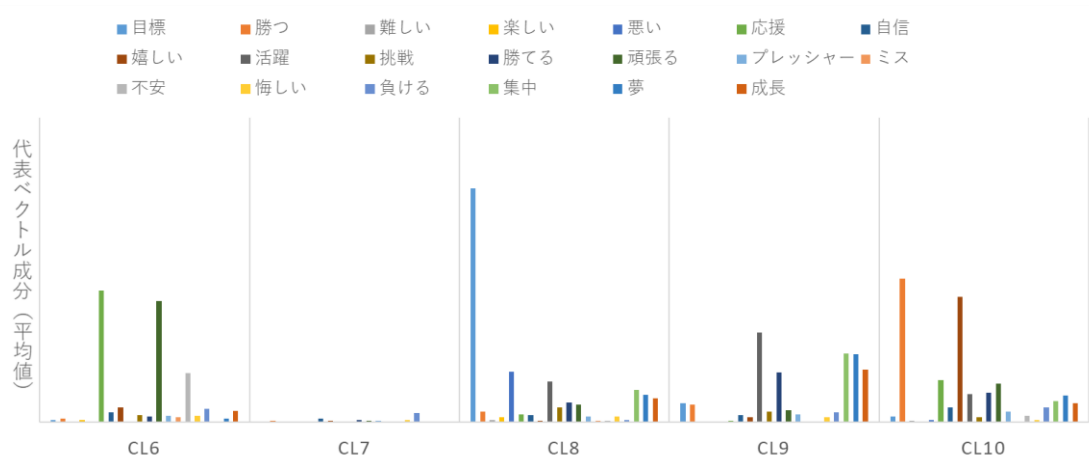


図 3.15 CL6-10 に属するニューロンの代表ベクトル成分 (TF-IDF 値)

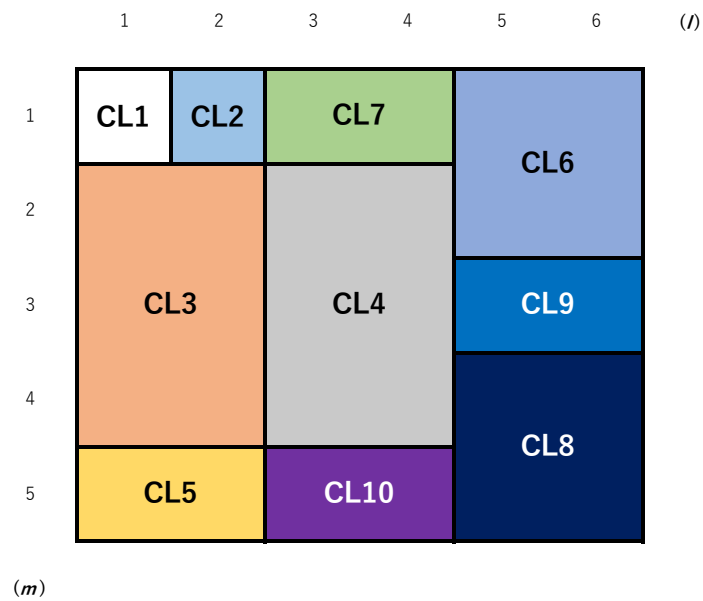


図 3.16 類似度 $R = 0.79$ において形成されるクラスター (TF-IDF 値)

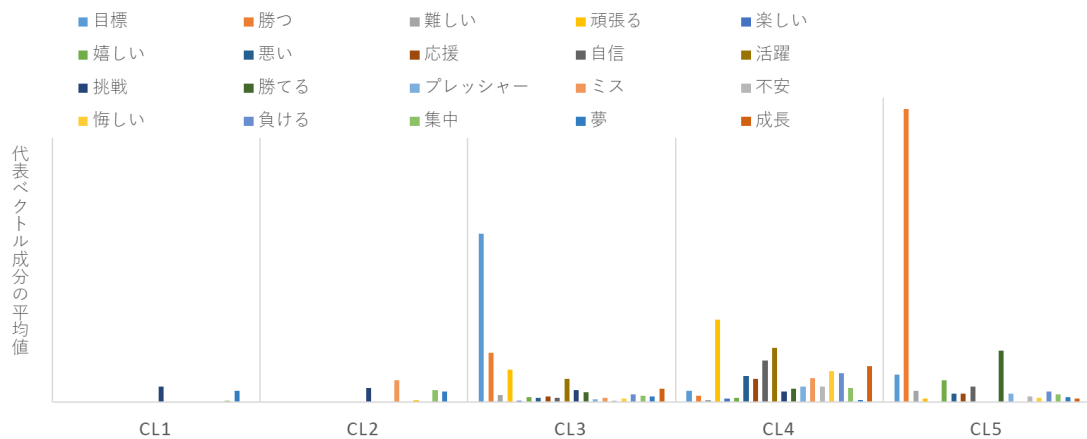


図 3.17 CL1-5 に属するニューロンの代表ベクトル成分 (TF-IDF 値)

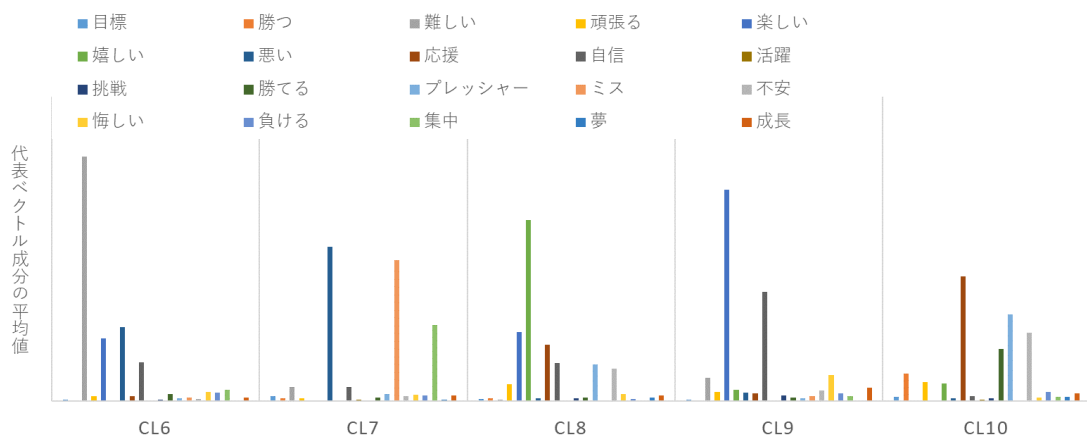


図 3.18 CL6-10 に属するニューロンの代表ベクトル成分 (TF-IDF 値)

これらの図から、各クラスタにおいてニューロンがもつ代表ベクトル $\mathbf{W}_{l,m}$ のうち、特に高い値を示す成分を特定することができる。代表ベクトルの成分は、入力ベクトル \mathbf{H}_j または \mathbf{H}'_j の成分、すなわち Modal Word M_i の TF-IDF 値 $h_{i,j}$ 、あるいは TF-IDF 値 $\tilde{h}_{i,j}$ に対応するものである。そのため、クラスタごとに顕著な値を示す代表ベクトルの成分を特定することで、マップ上の領域と、各 Modal Word の TF-IDF 値との関係が明らかになるといえる。

例えば、図 3.17 における CL1 の代表ベクトルに注目すると、いずれの成分も 0、あるいはそれに近い値を示している。このクラスタに属するニューロンには、入力ベク

トル \mathbf{H}'_j の中でも、ゼロベクトルのみが分類された。すなわち、Modal Word M_i を含まない発話が、CL1 の領域に分類されていることがわかる。これに対し、CL5 の代表ベクトルに関しては、Modal Word M_2 (= "勝つ") に関する成分が顕著に高い値を示している。このことから、マップの左下を占める CL5 の領域には、「勝つ」という Modal Word を数多く含む発話が分類されているといえる。以上のように、クラスタごとの代表ベクトル成分の平均値に注目し、特に高い値を示した Modal Word を、対応する領域に記載した(図 3.19 および図 3.20)。図中に記載された Modal Word はいずれも、TF-IDF 値または TF-DF 値に対応する代表ベクトル成分の平均値が、すべてのベクトル成分の平均値 (TF-IDF: 0.177, TF-DF: 0.245) を上回ったものである。

個々のクラスタを特徴づける Modal Word を記載した図 3.19 および図 3.20 は、マップ上における Modal Word の分布を示したものととらえることができる。この図に基づき、マップの縦軸および横軸、対角線に対応する心理的評価軸を抽出することが可能となる。

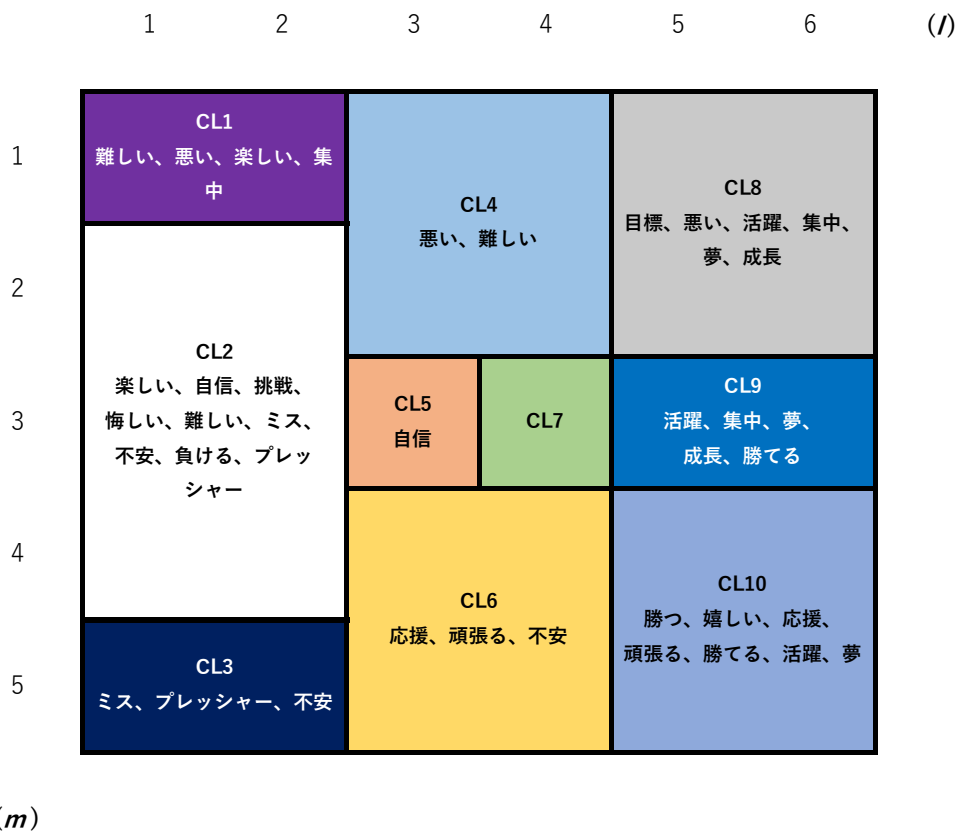


図 3.19 各クラスタにおいて顕著な値を示した Modal Word (TF-IDF 値)

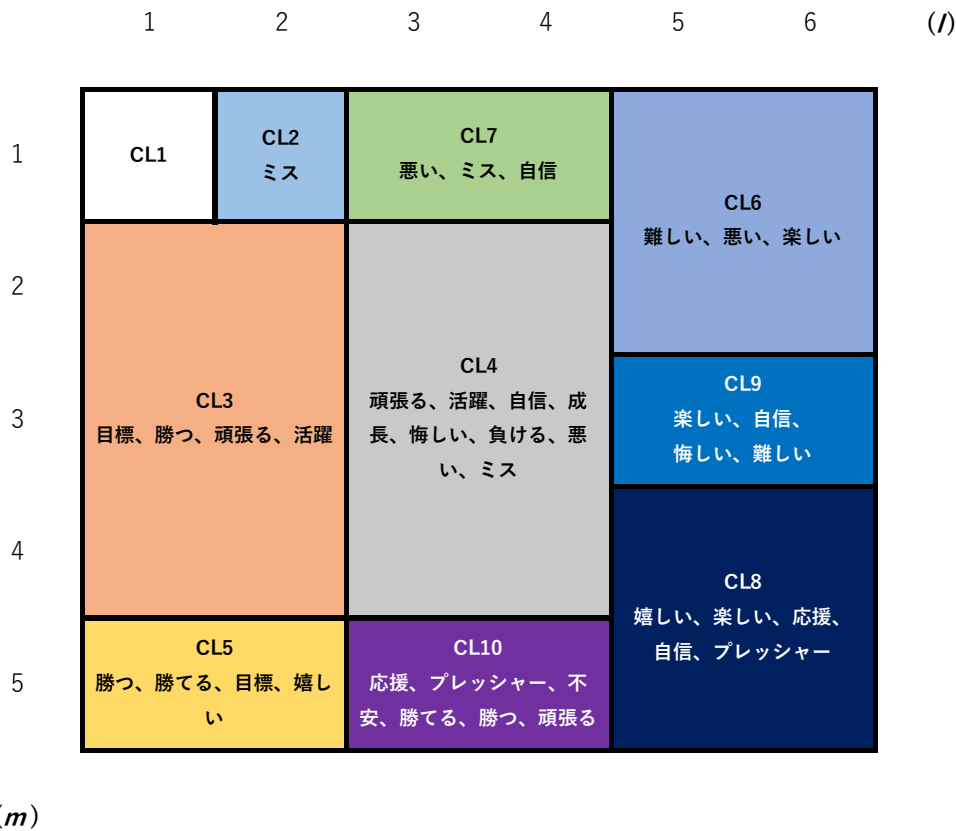


図 3.20 各クラスタにおいて顕著な値を示した Modal word (TF-DF 値)

また、図 3.13 および図 3.16 で示されたマップ上のクラスタについて、そのトポロジ的性質をとらえるために、矩形分割の双対グラフを作成した（図 3.21 および図 3.22）。これは、各クラスタの中心を頂点とし、隣り合うクラスタの頂点同士を線分で結ぶことで、クラスタの隣接関係を表したグラフである。

ここで、図 3.21 のグラフ上の頂点（各クラスタ）を $CL_t (t = 1, 2, \dots, 10)$ 、2つの頂点間の距離を $d(CL_p, CL_q) (p = 1, 2, \dots, 10, q = 1, 2, \dots, 10)$ とおく。ただし、距離 $d(CL_p, CL_q)$ は、双対グラフにおいて頂点 p から頂点 q へと至る最短経路長（ p から q に至るパスでそこに含まれる辺の数が最小のもので、その辺の個数）と定義する。例えば、 CL_1 と CL_7 を結ぶ最短経路は、 CL_4 を介した 2 つの辺を含むパスであることから、 $d(CL_1, CL_7) = d(CL_7, CL_1) = 2$ となる。このとき、本研究で導入した距離 d は以下の距離の公理を満たすことは明らかで、距離空間であるので、マップ上の領域の集合 T とその集合上の距離

d は、位相空間を構築する.

$$1) \text{ 任意の頂点 } CL_p, CL_q, CL_r \text{ について, } d(CL_p, CL_q) \leq d(CL_p, CL_r) + d(CL_r, CL_q) \quad (3.17)$$

$$2) d(CL_p, CL_q) = 0 \Leftrightarrow p = q \quad (3.18)$$

$$3) d(CL_p, CL_q) = d(CL_q, CL_p) \quad (3.19)$$

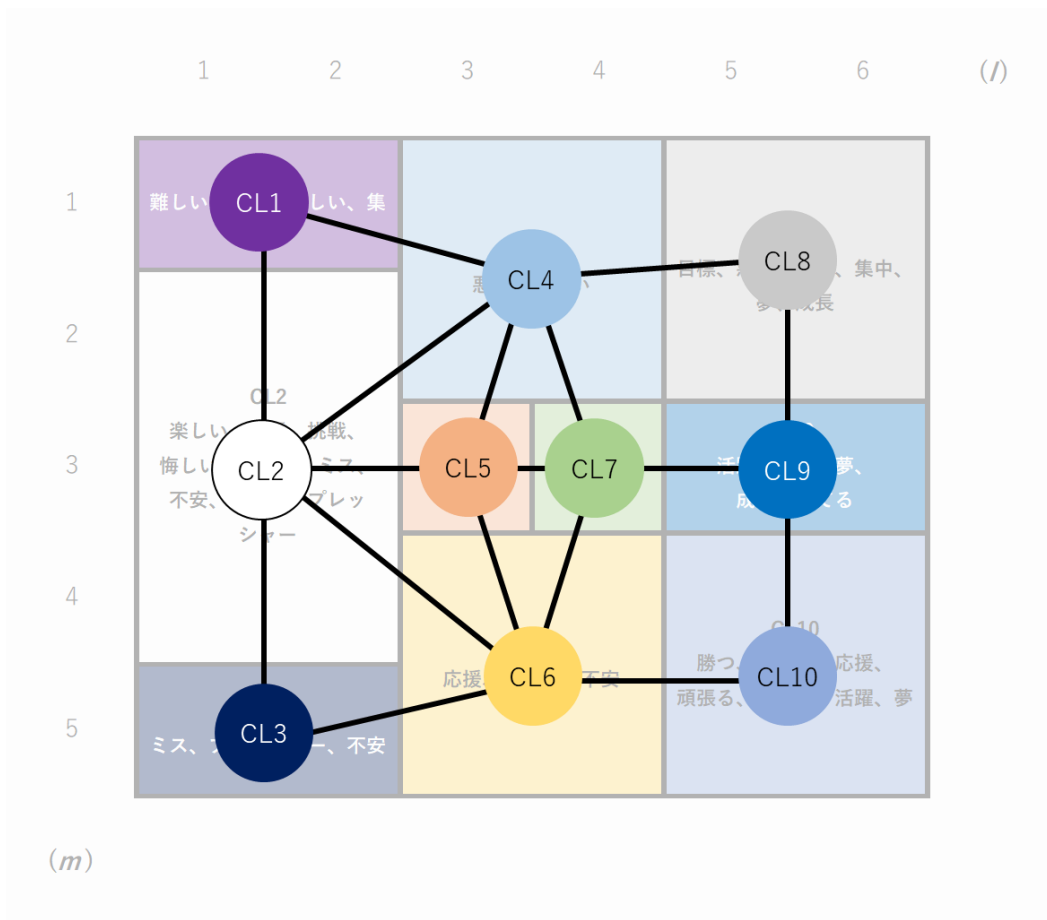


図 3.21 マップ上のクラスタの双対グラフ (TF-IDF 値)

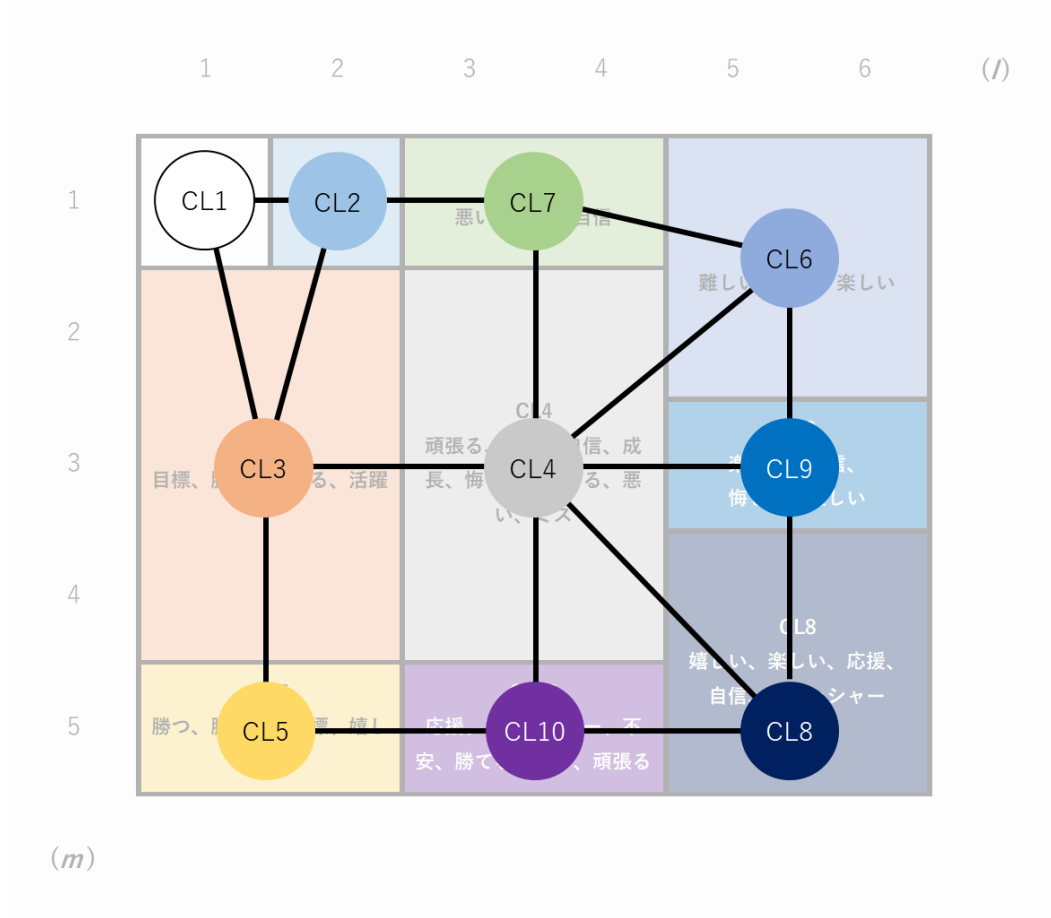


図 3.22 マップ上のクラスタの双対グラフ (TF-DF 値)

3.9.4. 心理的評価軸の抽出

SOM によるマッピングの結果について、心理学的な観点から解釈を行うために、心理状態の動力的な類似性について考察した。動力学は、座標の速度および加速度を表現するために用いられる学問であるが、この原則に従えば、我々の心理は状態と変化に分かれ、動的に表現される必要がある。つまり、心理の状態はスカラー、変化はベクトルとして定義されることになる。ここでは、SOM のマップ上における座標とその変化を表すベクトルに基づき、心理的変遷の表現を試みた。

‘3.8 トポロジー的評価’で述べた心理的評価軸の抽出プロセスと、Modal word の分布 (図 3.19, 3.20) に基づき、SOM のマップ上に抽出された心理的評価軸を図 3.23 および図 3.24 に示す。ここでは、赤色の実線による矢印がスカラー座標、点線による矢印がベクトル座標を示す軸となる。

1) TF-IDF 値を用いた場合の心理的評価軸

マップの中心に位置する CL7 は、代表ベクトル成分 $W_{i,l,m}$ の平均値がいずれも 0 に近い。すなわち、CL7 とマップ上の各領域との間でベクトル成分の増加がみられるため、これらの領域間に心理的評価軸を設定できると考えられる。

例えば、マップの上部に位置する CL4 では、「悪い」、「難しい」といったネガティブな Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「快—不快」および「活性—不活性」の評価軸において、これらの Modal word の極性がいずれも負であることが示されている。そのため、マップ中央から上部へと伸びる軸を設定し、これを「不快」を表す心理的評価軸とした。これに対し、マップの右側では、CL9 において、「活躍」や「集中」、「夢」といった Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「快—不快」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。そこで、中央から右側にかけて軸を設定し、これを「快」を表す心理的評価軸とした。

マップの左側に位置する CL2 では、「楽しい」、「自信」、「挑戦」といった Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「活性—不活性」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。これらの点を踏まえ、マップの中央から左側へと伸びる軸を「活性」、左側から中央へと伸びる軸を「不活性」を表す心理的評価軸とした。

マップの下部に位置する CL6 では、「応援」、「頑張る」、「不安」といった Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「緊張—弛緩」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。そこで、中央から下部へ伸びる軸を設定し、これを「緊張」を表す心理的評価軸とした。

抽出された 5 つの心理的評価軸は、いずれも図 3.1(a) に示した感情の評価軸 u, v, w に一致しており、心理変化としての感情表出の程度を表すものと解釈できる。

‘3) 各クラスタの特徴に関する考察’で示した双対グラフを参照すると、「緊張」の距離 $d(CL_7, CL_6)$ および、「不快」の距離 $d(CL_7, CL_4)$ 、「快」の距離 $d(CL_7, CL_9)$ の値は、いずれも 1 である。「活性」と「不活性」の両端の距離 $d(CL_7, CL_2) = d(CL_2, CL_7)$ のみ、値が 2 となる。図 3.22 では、活性と不活性の評価軸に沿って、「楽しい」、「自信」、「挑戦」

という Modal word によって特徴づけられる CL2 と、「自信」という Modal word によって特徴づけられる CL5 に、マップ上の領域が分割されている。TF-IDF 値を用いた分析では、自信という語の重要性がより強調され、活性・不活性の感情が細分化したことで、評価軸の両端の距離が延長したと考えられる。

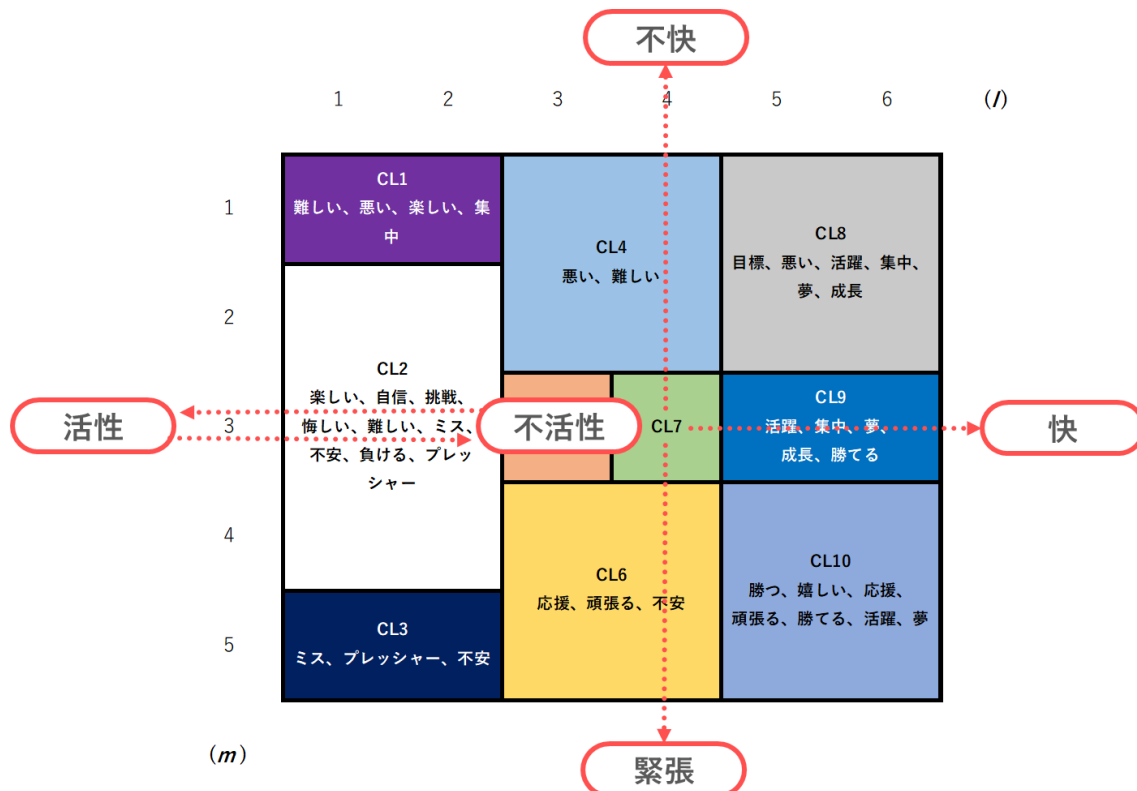


図 3.23 マップ上に設定された心理的評価軸 (TF-IDF 値)

2) TF-IDF 値を用いた場合の心理的評価軸

マップの左上に位置する CL1 は、代表ベクトル成分 $W_{i,l,m}$ の平均値がいずれも 0 に近い。すなわち、CL1 とマップ上の各領域との間でベクトル成分の増加がみられるため、これらの領域間に心理的評価軸を設定できると考えられる。

CL1 から左下の CL5 にかけては、「勝つ」、「目標」といった Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「快—不快」および「活性—不活性」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。これらの Modal word は、競技に対する能動的・積極的な姿勢を反映し、アス

リートの心理的な活性の高まりを示すものと考えられる。そこで、マップの左上から左下に至る縦軸を「活性」、反対方向へ向かう縦軸を「不活性」を表す心理的評価軸とした。

マップ上部の CL2 から CL6 にかけては、「難しい」、「悪い」といったネガティブな Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「快—不快」および「活性—不活性」の評価軸において、これらの Modal word の極性がいずれも負であることが示されている。そのため、マップ上部の横軸を「不快」を表す心理的評価軸とした。また、右上の CL6 から右下の CL8 にかけては、「嬉しい」、「楽しい」といったポジティブな Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「快—不快」および「活性—不活性」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。そのため、マップ右側の縦軸を「快」を表す心理的評価軸とした。

上記の心理的評価軸を踏まえると、マップ右下の CL8 は快感情を、左下の CL5 は活性の高さを示す領域であると解釈できる。加えて、CL5 から CL8 にかけては、「勝つ」、「目標」、「応援」、「プレッシャー」といった Modal word に対応する $W_{i,l,m}$ の平均値が高くなっている。表 3.4 では、「緊張—弛緩」の評価軸において、これらの Modal word の極性がいずれも正であることが示されている。そのため、マップの右下から左下に至る横軸を、「緊張」の高まりを示す軸とした。

マップの対角線における Modal word の分布に注目すると、中央の CL4 から左下の CL5 にかけては、「勝つ」、「勝てる」といった、競技に勝つことを意識した Modal word の値が高くなっている。一方、CL4 から右下の CL8 にかけては、「嬉しい」、「楽しい」といったポジティブな感情に関する Modal word の値が増加しており、自発的な競技参加の姿勢を反映したものと考えられる。そこで、マップの右上から左下に至る対角線を「闘争心」、左上から右下に至る対角線を「自発性」の評価軸とした。

TF-IDF を用いた場合と同様に抽出された、5つの心理的評価軸（快、不快、活性、不活性、緊張）は、いずれも図 3.1(a)に示した感情の評価軸 u, v, w に一致している。すなわち、心理変化としての感情表出の程度を表すものと解釈できる。一方、「闘争心」および「自発性」の心理的評価軸は、マップ上の Modal word の分布が示す意味を解

積することで、新たに抽出されたものである。これらは、感情空間上で異なる速度ベクトルを示す、多様な感情によって構成された、心理状態を表す評価軸であると推察される。

‘4) 各クラスタの特徴に関する考察’で示した双対グラフを参照すると、上記の7つの評価軸のうち、「活性」と「不活性」の両端の距離 $d(CL_1, CL_5) = d(CL_5, CL_1)$ および、「快」の距離 $d(CL_6, CL_8)$ 、「緊張」の距離 $d(CL_8, CL_5)$ の値は、いずれも2である。これに対し、「不快」の両端の距離 $d(CL_1, CL_6)$ および、「自発性」の距離 $d(CL_1, CL_8)$ 、「闘争心」の距離 $d(CL_6, CL_5)$ の値は、いずれも3となる。後者の不快、自発性、闘争心といった心理状態を示す程度は、アスリート間での個人差、あるいは個人間での変動が大きく、このことが位相空間上の距離に反映された可能性がある。可視化結果を読み取る上では、上記の3種類の評価軸上での心理変化は、他の評価軸と比べて、変化の程度が相対的に大きいものと理解する必要があるだろう。

以上のように、マップ上での座標とその変化が示す心理状態と心理変化をとらえるための心理的評価軸を設定した。以下の結果および考察では、TF-IDF 値および TF-IDF 値による分析に用いた個々人の入力ベクトル \mathbf{H}_j （または \mathbf{H}'_j ）とその分類結果、ならびにマップ上へ設定された心理的評価軸に基づく可視化を行い、個々人の心理状態とその変化について解釈を行った。

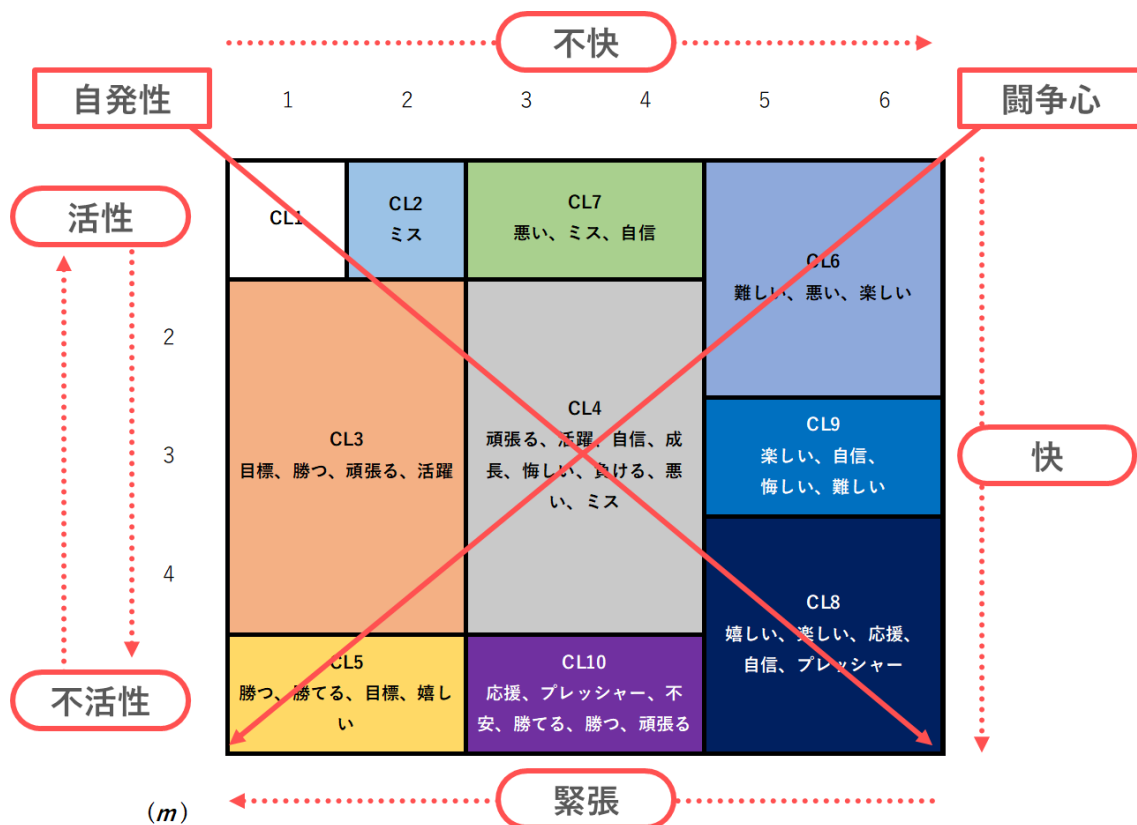


図 3.24 マップ上に設定された心理的評価軸 (TF-IDF 値)

3.9.5. アスリートの心理状態および変化の可視化

ここでは、TF-IDF 値および TF-IDF 値のそれぞれを用いた場合に、マップ上で特徴的な心理変化を示した 3 名のアスリートに関する可視化の結果 (図 3.25-3.30) を示す。可視化の結果について心理学的な解釈を加える上で、取り組む競技やカテゴリなど、アスリートの属性については言及しない。

1) TF-IDF 値を用いた場合の可視化結果

図 3.25 は、アスリート $c (= 33)$ の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、11 の発話 U_j が得られ、これらに対応する入力ベクトル \mathbf{H}_j の座標遷移がマップ上に表現されている。丸で囲われた数字は、SOM により個々人の入力ベクトル \mathbf{H}_j が割り当てられたニューロンの座標と、インタビュー中の発話 U_j に対応する順番を示している。これらの数字を繋ぐ矢線図は、マップ上を入

力ベクトル \mathbf{H}_j が遷移する軌跡を表しており、心理的評価軸上での座標遷移の様子に注目することで、アスリートの心理状態および変化についての解釈が可能となる。

図 3.25 からは、主に「緊張」の評価軸上での遷移のほか、「快」の評価軸上での遷移や、マップの右上 (CL8)、右下 (CL10) の領域への遷移が読み取れる。右上・右下への遷移は、2 種類の心理状態や心理変化が混在した状態であると考えられる。マップの右上を占める CL8 への遷移は、インタビュー中の発話で「目標」や「悪い」、「活躍」、「集中」、「夢」、「成長」などの Modal word を用いた場合に生じる。このとき、アスリートは目標の達成や試合での活躍、選手としての成長など、競技を通じた自己実現について言及していると推察される。同時に、「悪い」というネガティブな Modal word の使用は、競技生活の中で経験する不快な感情と関係があるといえよう。また、右下の CL10 への遷移は、「勝つ」、「嬉しい」といった Modal word の使用を示しており、活性度が高く肯定的な心理状態であることが読み取れる。

図 3.26 は、アスリート $c(= 47)$ の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、12 の発話 \mathbf{U}_j が得られ、これらに対応する入力ベクトル \mathbf{H}_j の座標遷移がマップ上に表現されている。特徴として、「不快」の評価軸上での座標遷移を中心に、「活性／不活性」および「快」、「緊張」といったすべての軸上での遷移を示している点が挙げられる。また、マップの左上 (CL1) の領域への遷移も顕著にみられる特徴であり、これは、「難しい」や、「悪い」、「楽しい」、「集中」といった Modal word の使用と関係がある。不快な感情を示しながらも、活性の高い状態 (CL2) へと遷移する可能性を含んだ状態であると考えられる。

図 3.27 は、アスリート $c(= 54)$ の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、17 の発話 \mathbf{U}_j が得られ、これらに対応する入力ベクトル \mathbf{H}_j の座標遷移がマップ上に表現されている。「活性／不活性」の評価軸上および、マップの右下の領域 (CL10) への遷移が中心であり、「楽しい」や「自信」、「勝つ」、「嬉しい」のような、肯定的な Modal word が発話中にみられる点が特徴的である。また、「緊張」の評価軸上や、マップの右上 (CL8) の領域への遷移もみられる。時には緊張を示しながらも、一貫して肯定的かつ積極的な心理状態を表明する様子が見えてくる。

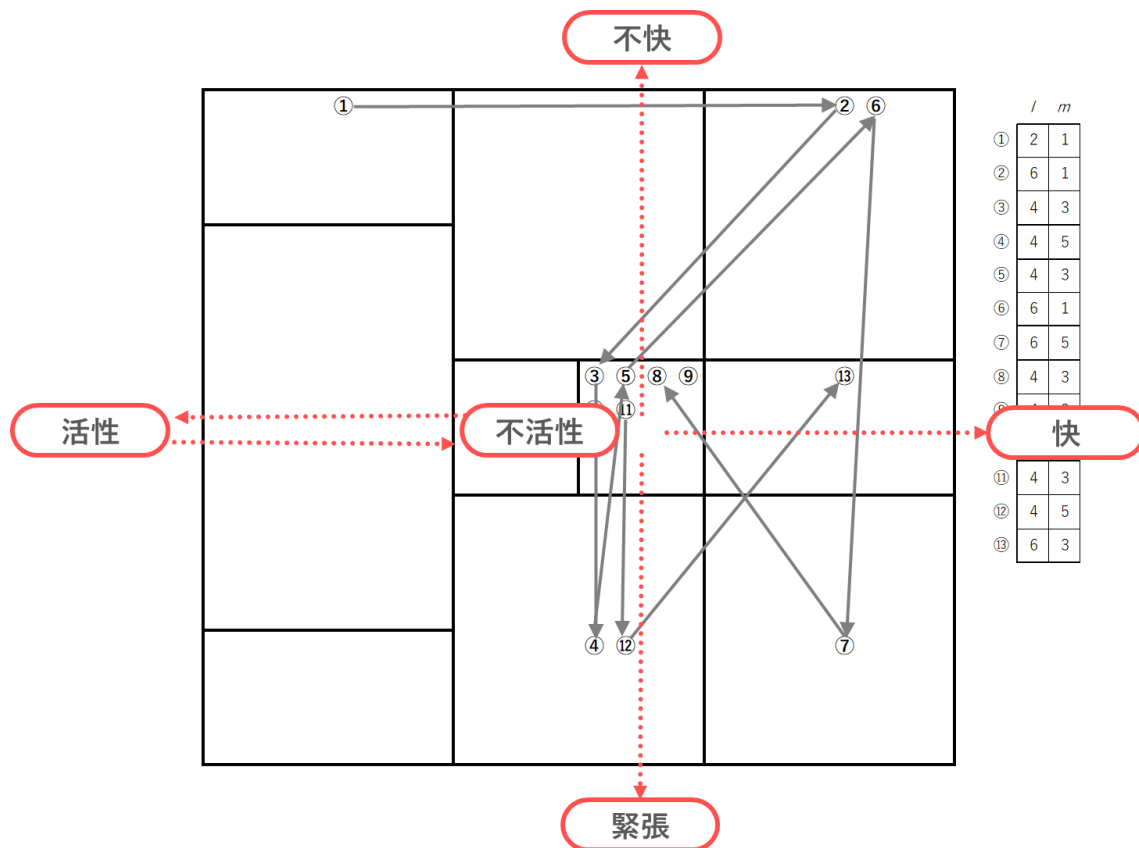


図 3.25 アスリート c(=33)の心理状態および変化の可視化事例

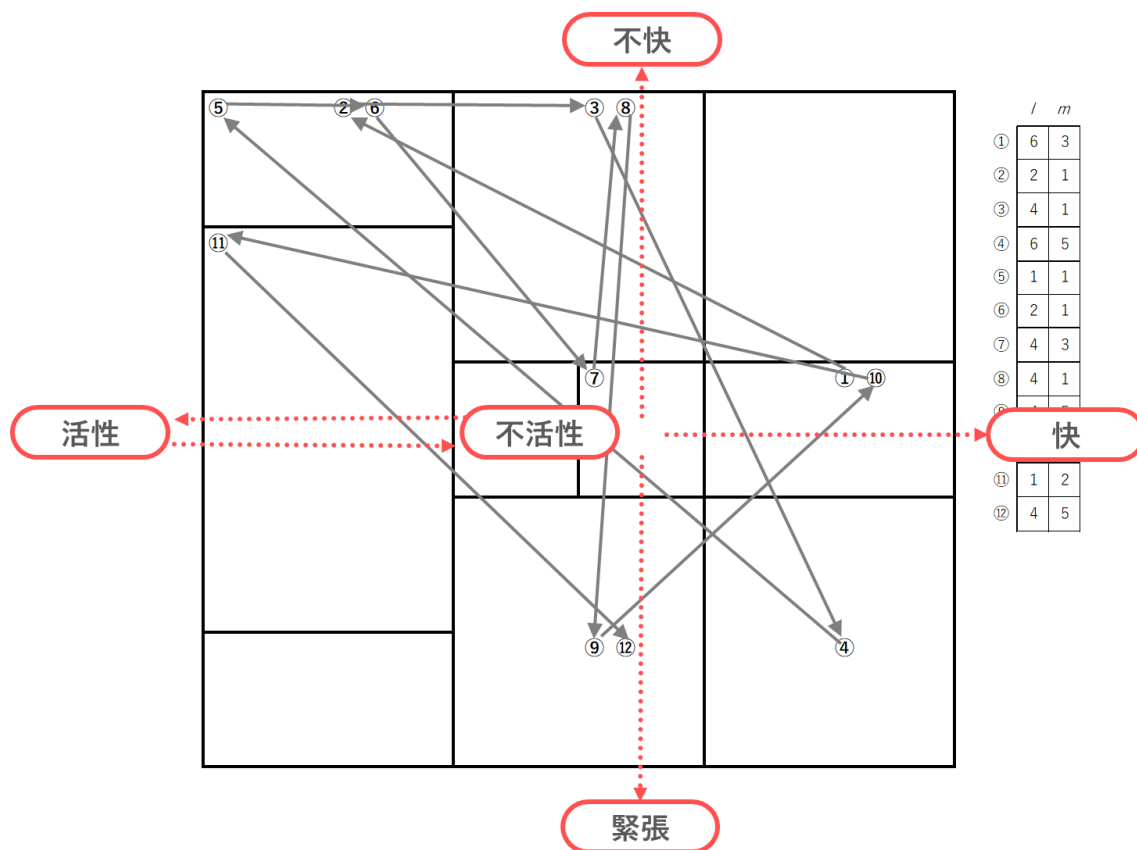


図 3.26 アスリート c(=47)の心理状態および変化の可視化事例

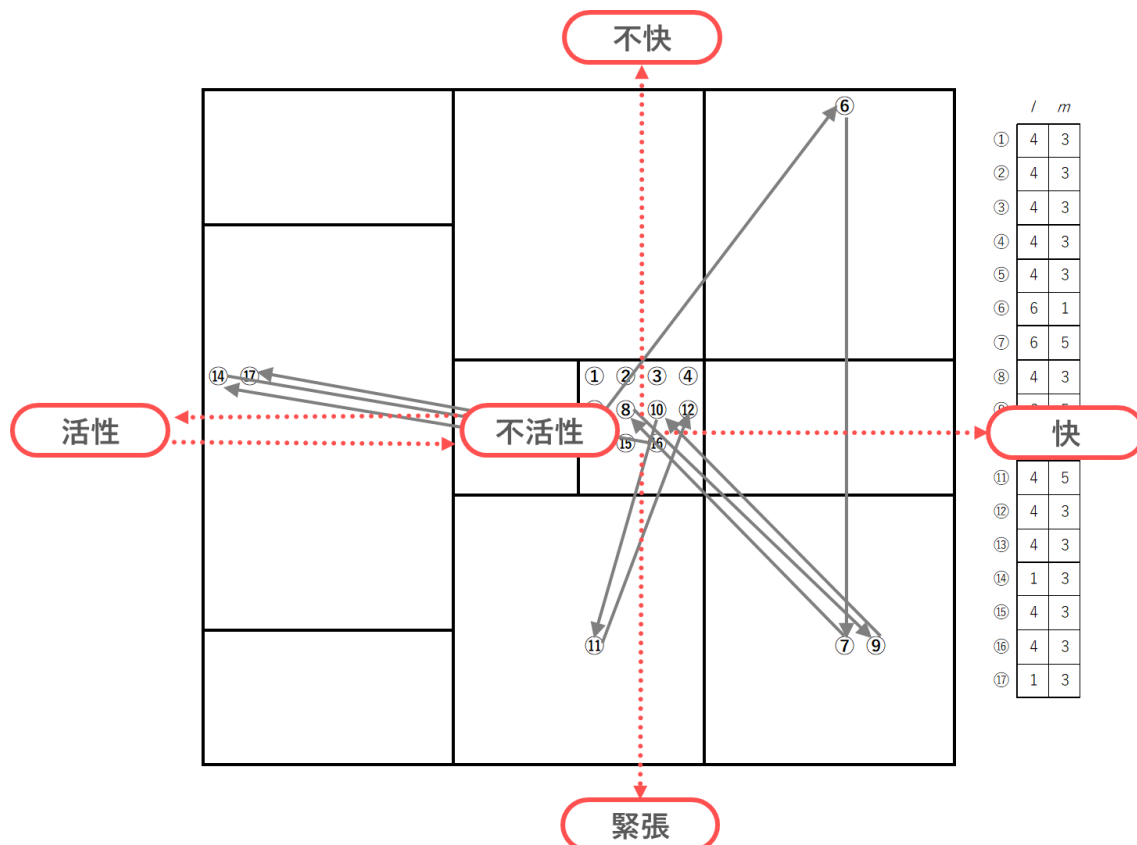


図 3.27 アスリート c(=54)の心理状態および変化の可視化事例

2) TF-DF 値を用いた場合の可視化結果

図 3.28 は、アスリート c(= 33)の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、13 の発話 U_j が得られ、これらに対応する入力ベクトル H'_j の座標遷移がマップ上に表現されている。図より、このアスリートは、マップ上の「活性／不活性」の評価軸上を中心とした遷移を示していることがわかる。これは、インタビュー中の発話において、「勝つ」、「目標」といったキーワードを多用した場合にみられる、入力ベクトル H'_j の座標遷移である。競技やスポーツ参加そのものに喜びを見出すというよりも、勝利や競技成績に焦点を当てているという点で、「活性／不活性」のベクトル座標軸上での変化は、アスリートの外的な動機づけ (Deci et al., 1985) に関係があると推察される。アスリート c(= 33)は、マップ上でほぼ一定の座標遷移を示しており、これは外的な動機づけによって生じる、競技に対しての一貫した態度であるといえよう。

図 3.29 は、アスリート $c(= 47)$ の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、12 の発話 U_j が得られ、これらに対応する入力ベクトル H'_j の座標遷移がマップ上に表現されている。マップ上の座標遷移は、主に「不快」の評価軸に沿った移動を示している。12 の入力ベクトル H'_j は、そのうちの 7 つが CL6 および CL7 の領域に分類されており、これは、アスリートの発話に「悪い」、「難しい」といったネガティブなキーワードが含まれていることを意味する。その他の入力ベクトルは、CL4 および CL5 の領域に分類されており、「活性」に関わる心理変化、ないしは「闘争心」の高まりが読み取れる。そのため、このアスリートにとっては、闘争心の高まりが、後に続く不快な感情の原因となっている可能性がある。

図 3.30 は、アスリート $c(= 54)$ の心理状態とその変化に関する可視化の結果である。同アスリートのインタビュー記事からは、17 の発話 U_j が得られ、これらに対応する入力ベクトル H'_j の座標遷移がマップ上に表現されている。主に、CL8 および CL9 の領域への遷移がみられることから、発話中に「嬉しい」、「楽しい」といったキーワードが含まれていたことがわかる。こうした傾向は、アスリートの快感情と競技に対する自発性を示すものであり、スポーツそのものを楽しむことがモチベーションの増加につながるような、内発的動機づけ (Deci et al., 1985) との関連性が示唆される。

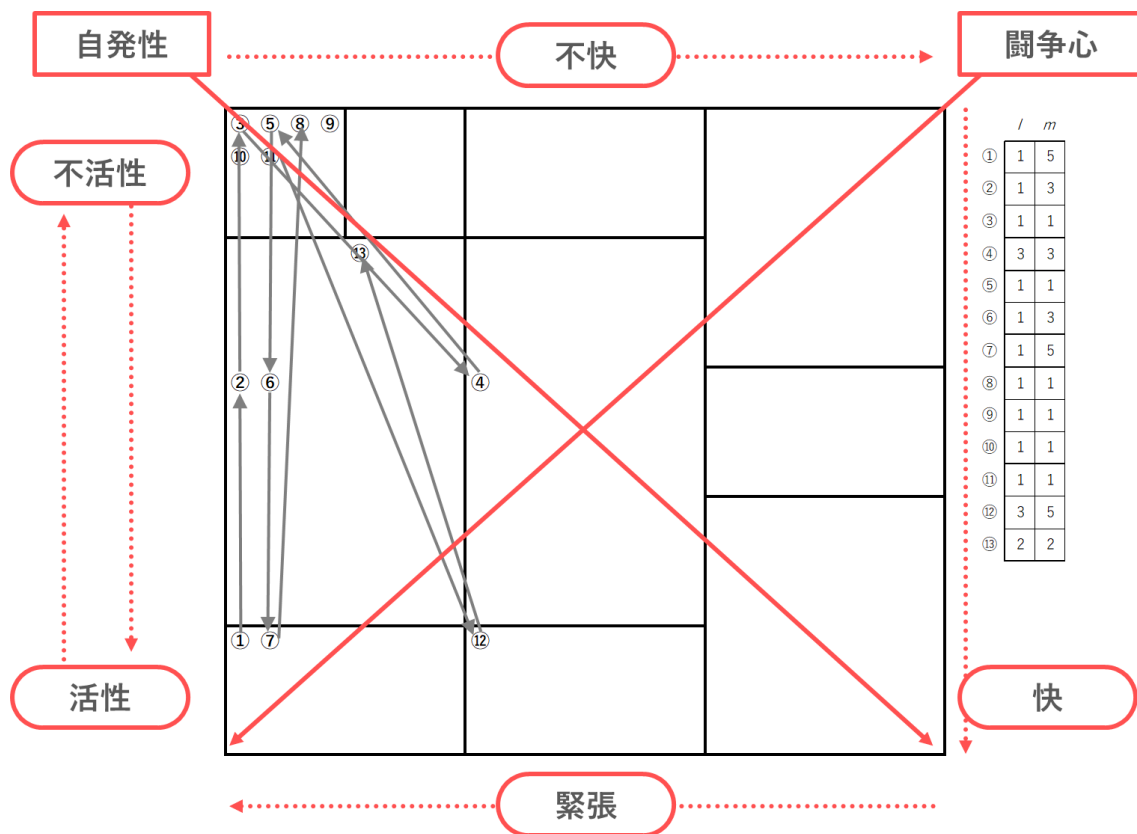


図 3.28 アスリート c(=33)の心理状態および変化の可視化事例

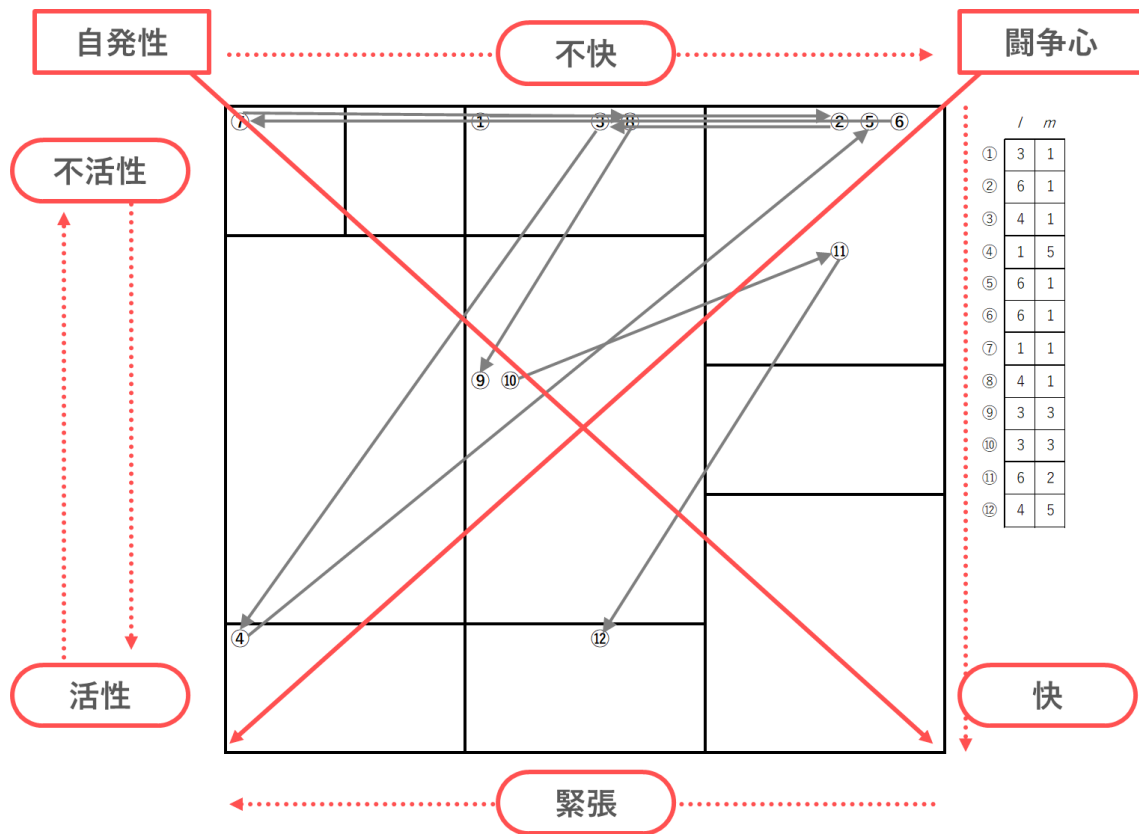


図 3.29 アスリート c(=47)の心理状態および変化の可視化事例

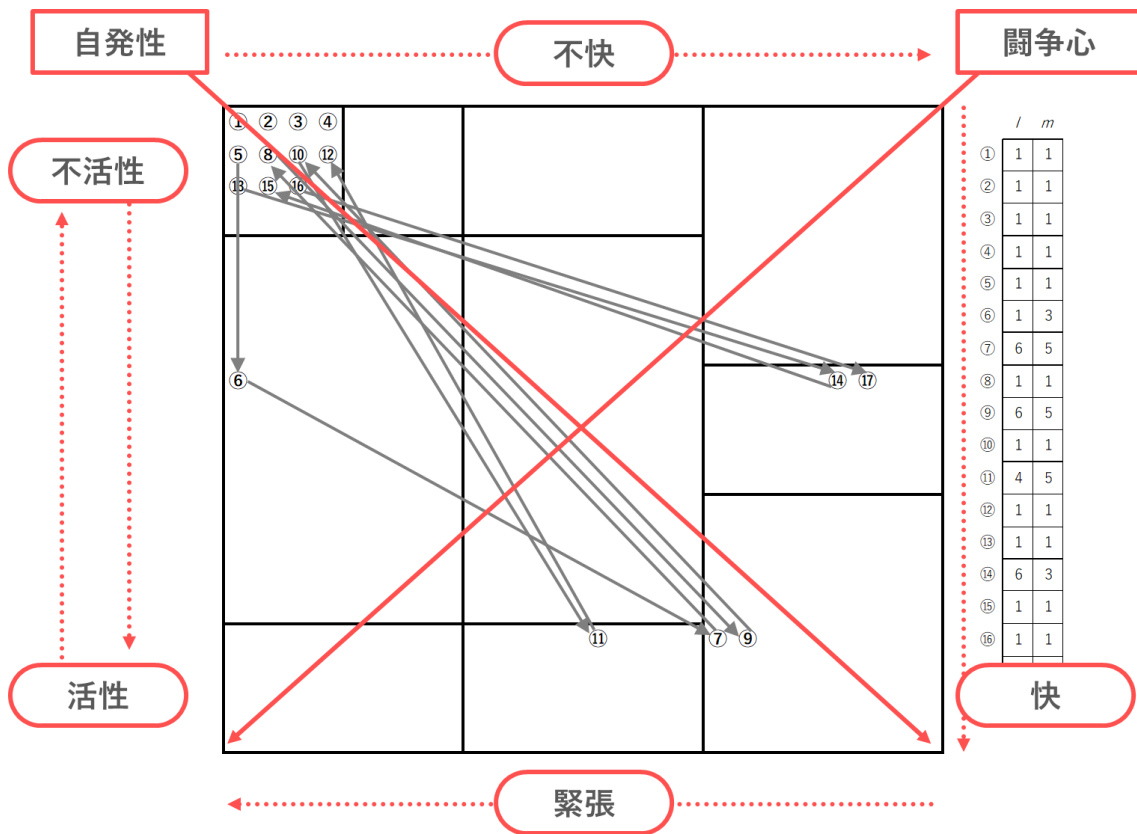


図 3.30 アスリート c(=54)の心理状態および変化の可視化事例

3.9.6. 位相空間上での座標遷移の可視化

下図は、図 3.21, 図 3.22 の双対グラフに、すべてのアスリートの入力ベクトル \mathbf{H}_j (図 3.31) および \mathbf{H}'_j (図 3.32) がグラフ上で示す座標遷移を加えて可視化したものである。赤色の点線は、ある頂点 CL_p から異なる頂点 CL_q へと、隣接する頂点を飛び越えて生じる座標遷移を示している。例えば、図 3.31 における CL_1 から、隣接する頂点 CL_2, CL_4 を飛び越えて、 CL_3 および $CL_6, CL_7, CL_8, CL_9, CL_{10}$ へと向かう座標遷移がみられる。このとき、赤の点線で示される CL_1 と CL_3 の頂点間の距離 $d(CL_1, CL_3)$ は、隣接する頂点間の遷移（黒の線分）を経た最短距離に等しいものと定義する。この場合も、前述の双対グラフと同様、任意の頂点 CL_p, CL_q, CL_r について、 $d(CL_p, CL_q) \leq d(CL_p, CL_r) + d(CL_r, CL_q)$ および $d(CL_p, CL_q) = 0 \Leftrightarrow p = q$, $d(CL_p, CL_q) = d(CL_q, CL_p)$ が成り立つことから、クラスター CL_t の集合 T と、クラスター間の距離 d は、位相空間を構築するといえる。

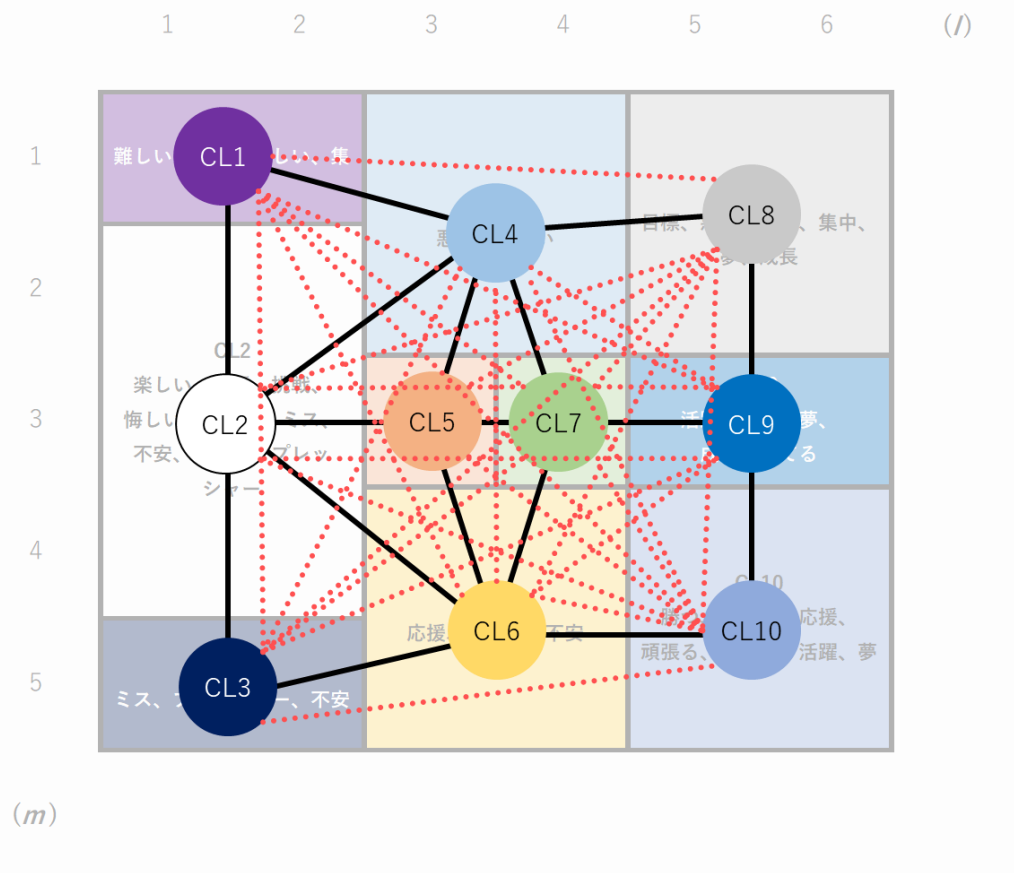


図 3.31 入力ベクトル \mathbf{H}_j の座標遷移を加えた双対グラフ

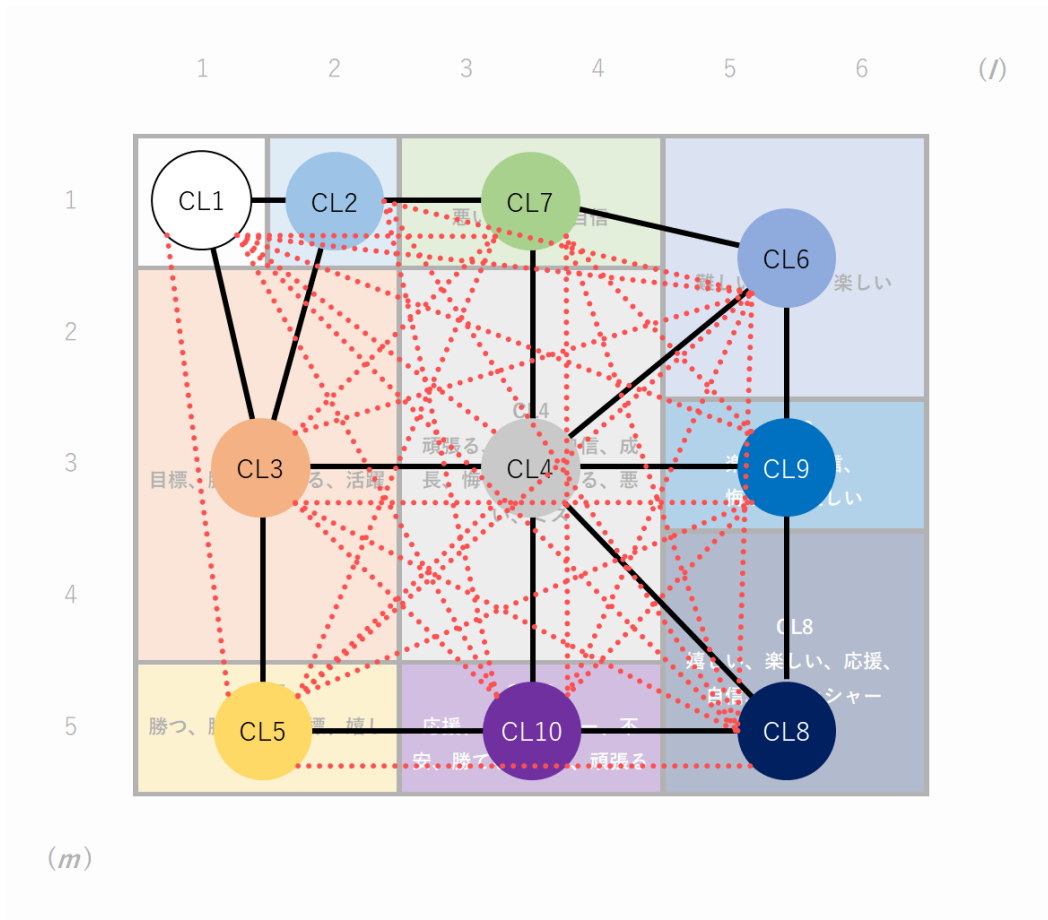


図 3.32 入力ベクトル H'_j の座標遷移を加えた双対グラフ

3.10. 結言

対象者の心理状態とその変化についての理解を促進するために、アスリートに対するインタビュー記事を対象として、言語データの分析および可視化手法を提案した。まず、形態素解析により言語データを形態素へと分割し、アスリートの心理状態を反映する語句をキーワードとして選定した。また、発話ごとのキーワードの出現頻度を求めた上で、TF-IDF法またはTF-DF法による重み付けを行い、この値を入力ベクトル \mathbf{H}_j または \mathbf{H}'_j の成分として分析に用いた。それぞれの発話の特徴を示す多次元ベクトルを、SOMおよびファジィクラスタ分析により解析し、二次元平面上へのマッピングを行うことで、個々人の心理状態および変化を可視化した。本章における結言は、以下の通りである。

- 1) 個人の心理状態とその変化を二次元平面上で可視化するために、SOMによる入力ベクトルのマッピングを行うことが有効であった。さらに、ファジィクラスタ分析により代表ベクトルのクラスタリングを行うことで、心理的評価軸が抽出され、可視化結果に対して心理学的な解釈を加えるための基準が得られた。
- 2) 本章で用いたファジィクラスタ分析により、心理の状態および変化を示す評価軸がそれぞれ抽出された。これにより、対象者の発話から読み取れる心理状態と変化について、トポロジー的な解釈を加えることが可能となった。本章では、入力ベクトル \mathbf{H}_j または \mathbf{H}'_j に対するSOMにより得られたマップが、いずれも10の領域に分割された。また、入力ベクトル \mathbf{H}_j を用いて作成したマップからは、心理の変化（感情）を示す5つの軸が抽出され、入力ベクトル \mathbf{H}'_j を用いて作成したマップからは、心理の変化を示す5つの軸と、心理状態を示す2つの軸が抽出された。
- 3) 本章では、Web上から収集したインタビュー記事に対して提案手法を適用し、心理状態と変化についての可視化を行った。その結果、対象のアスリート集団において、代表的な座標遷移のパターンを抽出し、その心理学的な意味について解釈を行うことができた。このことから、提案手法の有効性が裏付けられたと言える。

- 4) 提案手法は、言語データ処理を必要とする、クライアントへの心理アセスメントに活用することができる。今後、分析データが蓄積されていくことで、より普遍的な心理的評価軸の設定や、心理状態の時系列変化についての可視化も可能となると思われる。スポーツ領域のみならず、様々な領域における心理アセスメントにおいて、提案手法の活用が期待される。

第4章. 解析：心理変化のパターン抽出・可視化

4.1. 緒言

本章では，提案手法の第二段階のプロセスとして，アスリート集団の示す心理変化のパターン抽出と可視化の方法について説明する．

本章における処理では，第3章の分析結果として得られた2種類のデータを対象として，以下に示す2通りの方法で解析を行った．

なお，本章に記載された内容は，2018年にオンラインジャーナル“Psychology (<http://www.scirp.org/journal/psych/>)”に掲載された，著者らの論文 (Aoki et al., 2018b) に基づくものである．

4.1.1. 処理対象データ

1) TF-IDF 値に基づく分析結果

- ・ SOM による，入力ベクトル $\mathbf{H}_j = (h_{1,j}, h_{2,j}, \dots, h_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ の分類結果と，代表ベクトル $\mathbf{W}_{l,m} = (W_{1,l,m}, W_{2,l,m}, \dots, W_{NM,l,m}) (l = 1, 2, \dots, L, m = 1, 2, \dots, M)$ の更新された値
- ・ ファジィクラスタ分析によるニューロンのクラスタリング結果

2) TF-DF 値に基づく分析結果

- ・ SOM による，入力ベクトル $\mathbf{H}'_j = (\tilde{h}_{1,j}, \tilde{h}_{2,j}, \dots, \tilde{h}_{NM,j}) (j = 1, 2, \dots, 736, NM = 20)$ の分類結果と，代表ベクトル $\mathbf{W}_{l,m} = (W_{1,l,m}, W_{2,l,m}, \dots, W_{NM,l,m}) (l = 1, 2, \dots, L, m = 1, 2, \dots, M)$ の更新された値
- ・ ファジィクラスタ分析によるニューロンのクラスタリングの結果

4.1.2. 心理変化の解析手法

上記の‘1)’および‘2)’の分析結果のデータをもとに，各アスリートの入力ベクトル \mathbf{H}_j または \mathbf{H}'_j が，マップ上の各領域 CL_t に分類された個数 $N_{c,t}$ を求めた．さらに，これらの値に対して，(1)TF-IDF 法に準ずる手法，(2)TF-DF 法に準ずる手法で重み付けを行い，遷移ベクトル \mathbf{I}_c および \mathbf{I}'_c を得た．すなわち，4通りの方法でデータを作成し，分

析および可視化に用いた。

以下の節では、処理の各段階における手続きと、その結果および考察を処理の遂行順に記す。すなわち、4種類のデータを用いた、それぞれの分析および可視化の結果を順に述べていく。最後に、処理全体の結果を踏まえた結言を述べる。

4.2. 遷移ベクトルの定義

SOM のマップ上のニューロンに分類された、各アスリートの入力ベクトル \mathbf{H}_j の座標遷移に注目し、その集団的な傾向を可視化することが、本章における目的である。そこで、60名のアスリート c を対象に、個々人の座標遷移の履歴を表すベクトルデータを算出し、これらのデータを用いてファジィクラスタ分析を実施した。

まず、前章において10のクラスタに分割された、マップ上の各領域を $CL_t (t = 1, 2, \dots, 10)$ または、 $CL'_t (t = 1, 2, \dots, 10)$ と定義した。前者は TF-IDF 値に基づく分析結果、後者は TF-DF 値に基づく分析結果に相当する。クラスタ CL_t および CL'_t と、マップ上の領域との対応関係は、下記の通りである。

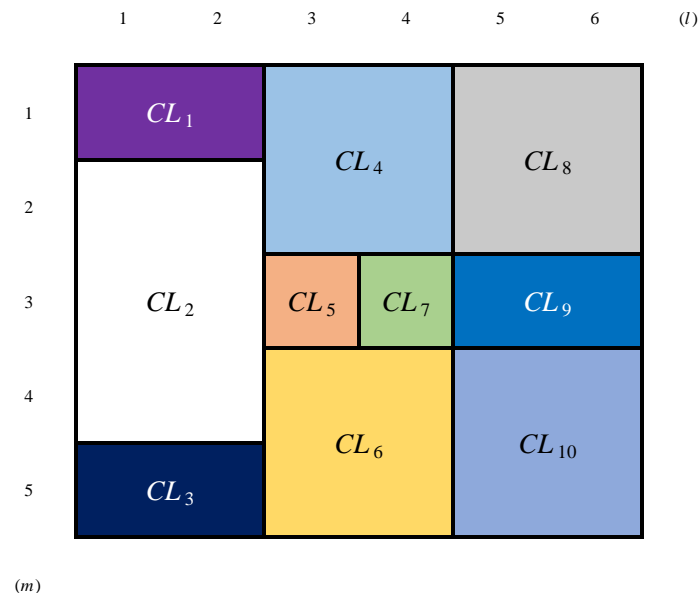


図 4.1 マップ上の領域 CL_t

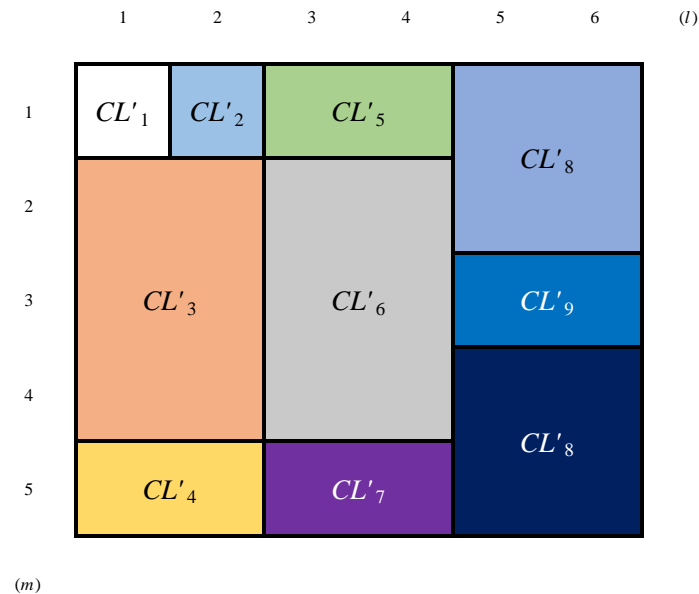


図 4.2 マップ上の領域 CL'_t

次に、SOMによる入力ベクトルの分類結果をもとに、領域 CL_t のそれぞれに分類された入力ベクトル \mathbf{H}_j (または H'_j) の個数をアスリート別に求め、それぞれ $N_{c,t}$ ($c = 1, 2, \dots, NC (= 60)$), $N'_{c,t}$ ($c = 1, 2, \dots, NC (= 60)$)とおいた。

本研究において、マップ上での各アスリートの個別の入力ベクトルの遷移パターンを、1つのベクトルとして定義し、グループ全体の遷移パターンや、個別の共通する遷移パターンをもつサブグループを特定するための計算の被演算子として扱う。このベクトルを、遷移ベクトルと名付ける。

本研究で導入する遷移ベクトルは、以下の3点を留意して定義した。

- 1) 遷移の軌跡に関して、その経路における途中の各地点 (領域 CL_t) への訪問順序を考慮すると、順列の個数の自由度が発生し、クラスタは形成しづらくなり、共通の遷移パターンを抽出することは困難になると見込まれる。よって、訪問の順序は区別しないこととした。
- 2) 遷移の軌跡に関して、その道のりの長さを考慮すると、軌跡のパターンが似ていても、ある特定地点 (領域 CL_t) への訪問回数が異なると別のクラスタとして分類

される。そこで、軌跡のパターンが似ているものは、なるべく同じクラスタに分類されるよう、(各領域に対する訪問回数) / (全訪問回数) で正規化するようにした。

- 3) 上記の‘2)’を導入しても、全員がある特定地点に共通して多く訪問した場合は、軌跡のパターンが類似したものと見なされ、区別がつかなくなり、1つのクラスタを形成してしまう可能性がある。これを回避するために、テキストマイニングで使われている IDF (逆文書頻度) の概念を、軌跡のパターンにおける訪問領域の回数の算定に導入した。すなわち、他の多くの対象者が共通に訪問している領域については、寄与率を下げ、他の対象者があまり訪問していない領域を訪問する場合は寄与率を上げて、軌跡のパターンを分類できるようにした。

具体的には、テキストマイニングの領域で用いられる TF-IDF 法と同様の基本概念に基づき、 $N_{c,t}$ ($N'_{c,t}$) の値について重み付けを行うことで、入力ベクトル \mathbf{H}_j (\mathbf{H}'_j) の分類個数に関する指標 $w_{c,t}$ ($w'_{c,t}$) を得た。 $w_{c,t}$ および $w'_{c,t}$ の計算式を、以下に示す。ここで、 NC_t は、入力ベクトル \mathbf{H}_j が領域 CL_t に分類されたアスリートの人数、 NC'_t は、入力ベクトル \mathbf{H}'_j が領域 CL'_t に分類されたアスリートの人数である。

$$w_{c,t} = u_{c,t} \cdot v_{c,t} \quad (4.1)$$

$$u_{c,t} = \frac{N_{c,t}}{\sum_{t=1}^{10} N_{c,t}} \quad (4.2)$$

$$v_{c,t} = \log_2 \frac{NC}{NC_t} \quad (4.3)$$

$$w'_{c,t} = u'_{c,t} \cdot v'_{c,t} \quad (4.4)$$

$$u'_{c,t} = \frac{N'_{c,t}}{\sum_{t=1}^{10} N'_{c,t}} \quad (4.5)$$

$$v'_{c,t} = \log_2 \frac{NC}{NC'_t} \quad (4.6)$$

また、第3章で用いた TF-DF 法についても、同様に $N_{c,t}$ ($N'_{c,t}$) の重み付けに適用し、TF-IDF 法に基づく重み付け手法との間で、分析結果を比較する。TF-DF 法を適用した場合の指標 $\tilde{w}_{c,t}$ ($\tilde{w}'_{c,t}$) は、下記の式に従って算出する。

$$\tilde{w}_{c,t} = u_{c,t} \cdot (\log_2 NC_t) \quad (4.7)$$

$$\tilde{w}'_{c,t} = u'_{c,t} \cdot (\log_2 NC'_t) \quad (4.8)$$

上述の式により、マップ上のすべての領域 CL_t について $w_{c,t}$, $\tilde{w}_{c,t}$ および $w'_{c,t}$, $\tilde{w}'_{c,t}$ の値を求め、これらの値を成分としてもつベクトルを以下のように定義した。

$$\mathbf{G}_c = (w_{c,1}, w_{c,2}, \dots, w_{c,10}) \quad (4.9)$$

$$\mathbf{G}'_c = (\tilde{w}_{c,1}, \tilde{w}_{c,2}, \dots, \tilde{w}_{c,10}) \quad (4.10)$$

$$\mathbf{I}_c = (w'_{c,1}, w'_{c,2}, \dots, w'_{c,10}) \quad (4.11)$$

$$\mathbf{I}'_c = (\tilde{w}'_{c,1}, \tilde{w}'_{c,2}, \dots, \tilde{w}'_{c,10}) \quad (4.12)$$

ベクトル \mathbf{G}_c および \mathbf{G}'_c は、いずれも第6章におけるTF-IDF値の分析に基づくデータであり、本章では前者をTF-IDF法、後者をTF-DF法により算出する。すなわち、ベクトル \mathbf{G}_c は「TF-IDF空間上の遷移を示すTF-IDFベクトル」、ベクトル \mathbf{G}'_c は「TF-IDF空間上の遷移を示すTF-DFベクトル」である。また、ベクトル \mathbf{I}_c および \mathbf{I}'_c は、いずれも第3章におけるTF-DF値の分析に基づくデータであり、本章では前者をTF-IDF法、後者をTF-DF法により算出する。すなわち、ベクトル \mathbf{I}_c は「TF-DF空間上の遷移を示すTF-IDFベクトル」、ベクトル \mathbf{I}'_c は「TF-DF空間上の遷移を示すTF-DFベクトル」である。

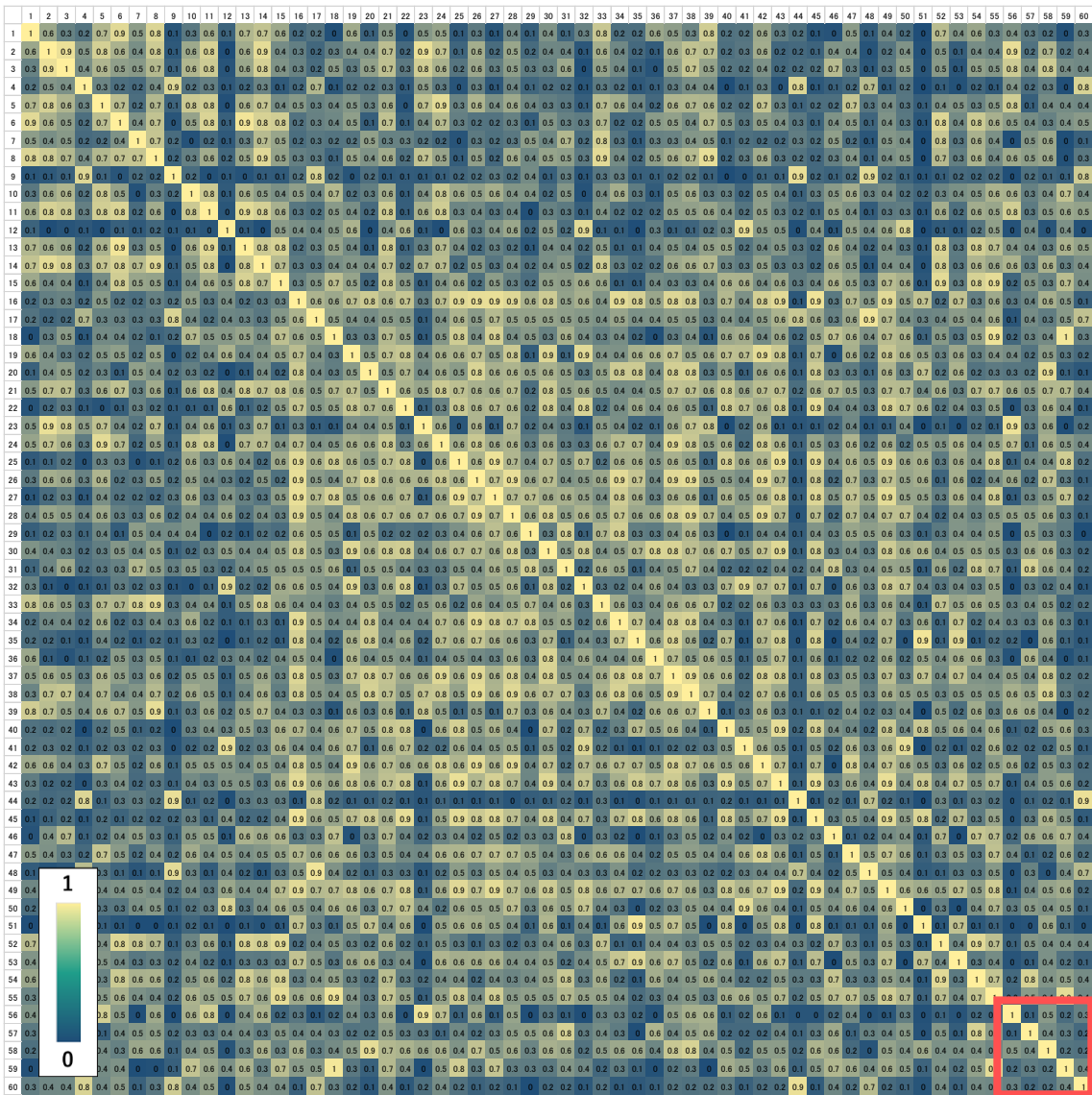
これらのベクトルデータは、アスリート c の入力ベクトルが分類されたマップ上の領域と、マップ上の遷移においてその領域を訪問した回数を示すものである。すなわち、入力ベクトル \mathbf{H}_j (\mathbf{H}'_j)のマップ上での座標遷移を表現するベクトルデータであり、個々の心理変化の特徴についての情報を含むと考えられる。そこで、すべてのアスリートについて求めた \mathbf{G}_c および $\mathbf{G}'_c, \mathbf{I}_c, \mathbf{I}'_c$ を用いて、60名のアスリートに対するファジィクラスタ分析を実施した。

4.3. ファジィクラスタ分析（アスリートのクラスタリング）

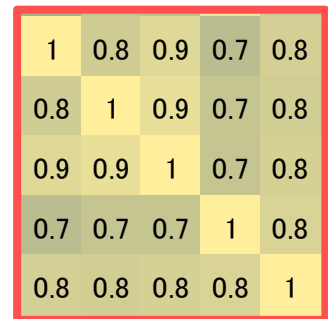
4.3.1. 可達行列による解析

すべてのアスリートのベクトル \mathbf{G}_c 間（または \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c の各ベクトル間）で相関行列を求め、これらの値を成分にもつメンバシップ行列 \mathbf{F} を作成した後に、可達行列 $\tilde{\mathbf{F}}$ への変換を行った。可達行列 $\tilde{\mathbf{F}}$ の成分である、ベクトル \mathbf{G}_c 間（または \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c の各ベクトル間）の類似度 R が、クラスタ形成の指標となる。類似度 R の基準値に応じて形成されるクラスタ数は変化するため、集団における心理変化パターンの抽出および可視化を踏まえて、最適なクラスタ数について検討した。

まず、分析に用いた各ベクトルのメンバシップ行列 \mathbf{F} および \mathbf{F}' , \mathbf{F}'' , \mathbf{F}''' と、変換後の可達行列 $\tilde{\mathbf{F}}$ および $\tilde{\mathbf{F}}'$, $\tilde{\mathbf{F}}''$, $\tilde{\mathbf{F}}'''$ を図 4.3-4.10 に示す。下図は、それぞれの行列の成分、すなわちベクトル \mathbf{G}_c 間（または \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c の各ベクトル間）の類似度を、値に応じて色分けしたコンター図である。

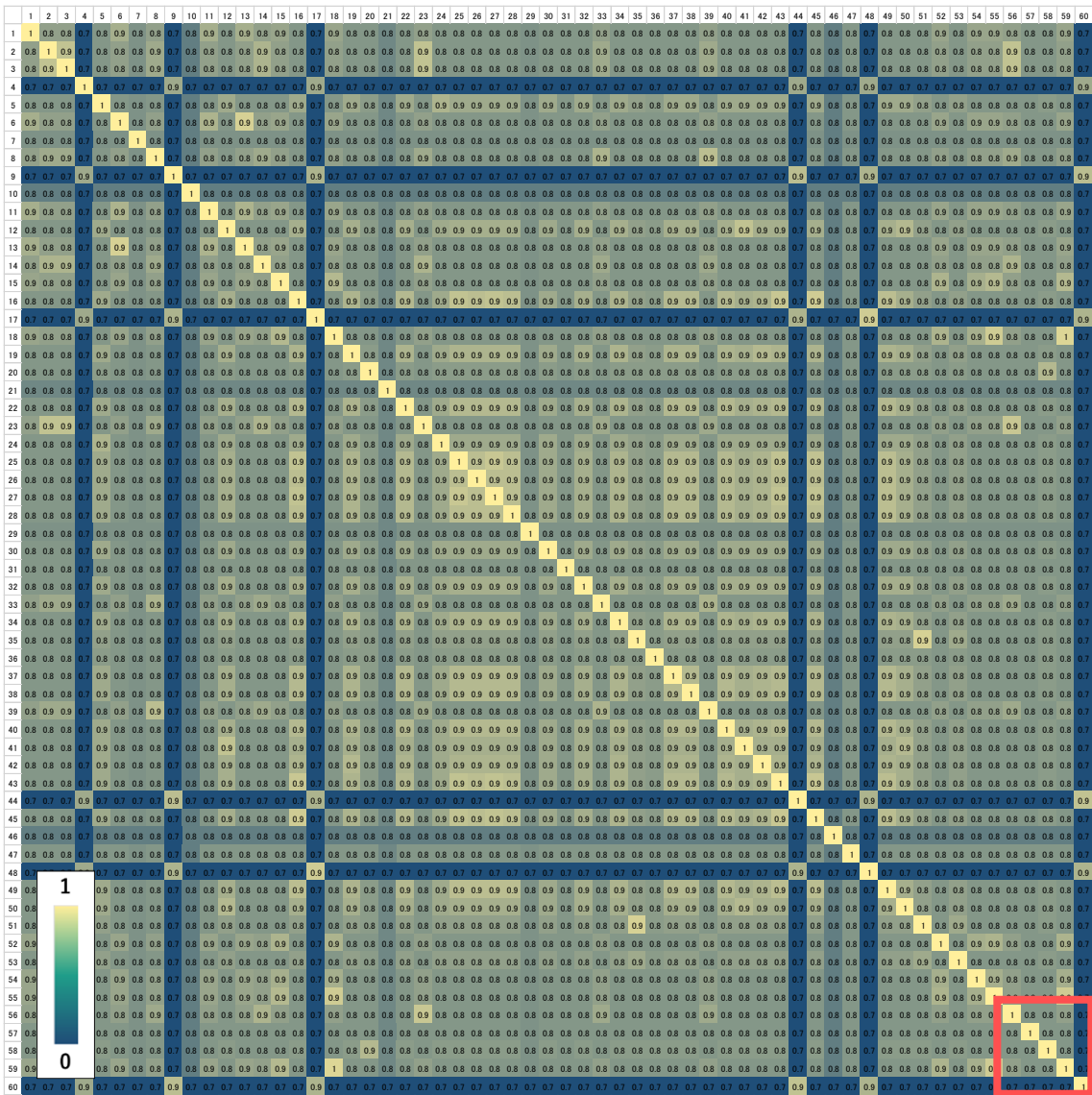


(a)

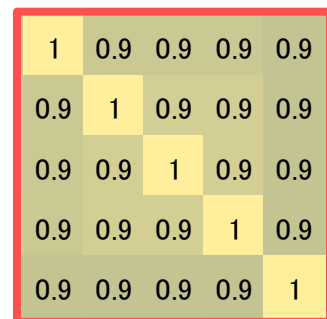


(b)

図 4.3 ベクトル G_c のメンバシップ行列 F

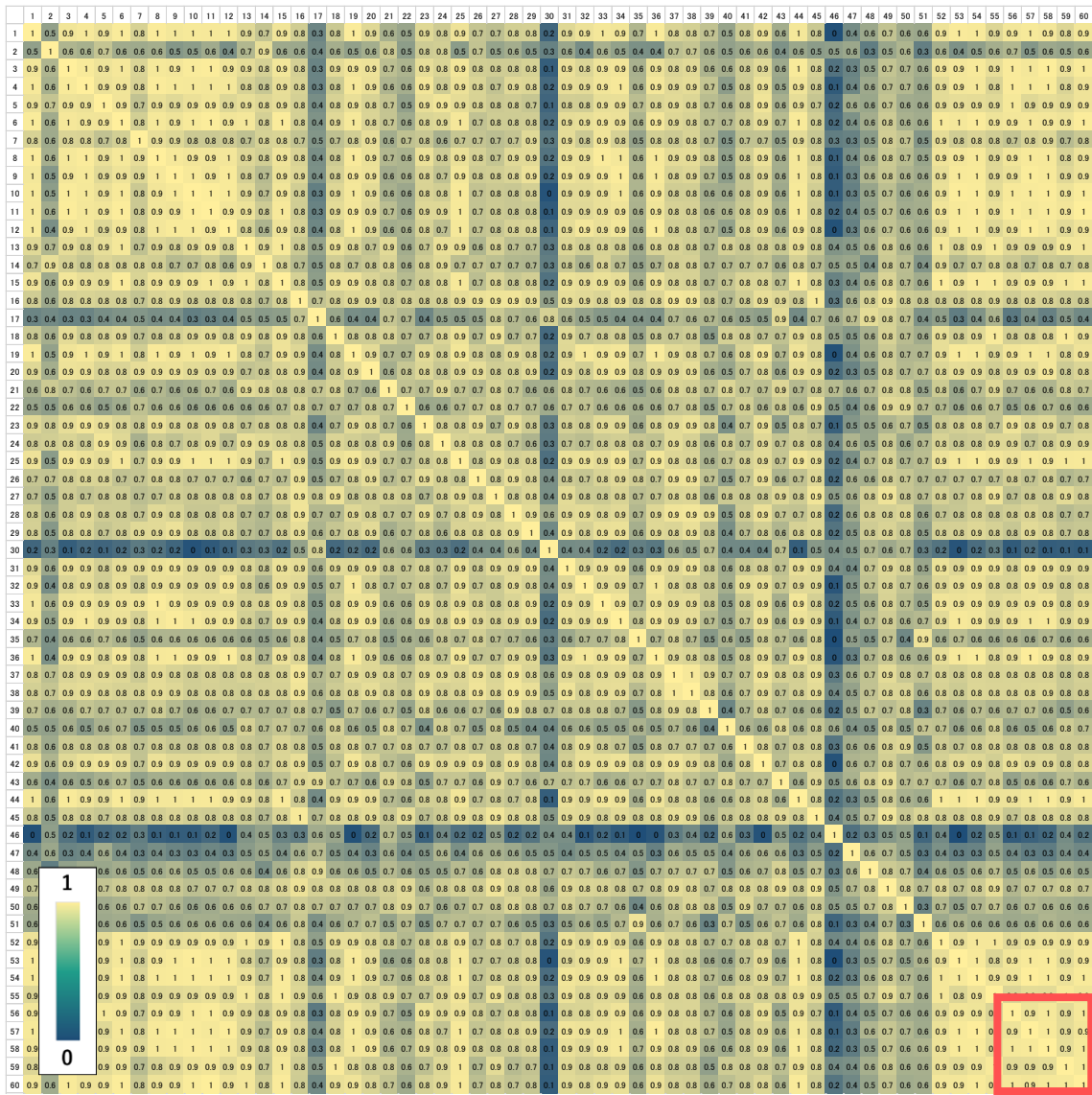


(a)

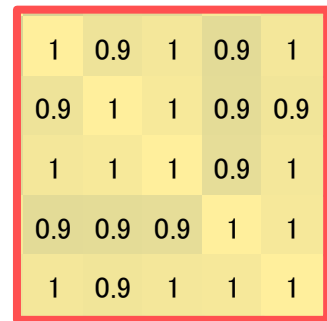


(b)

図 4.4 ベクトル G_c の可達行列 \bar{F}

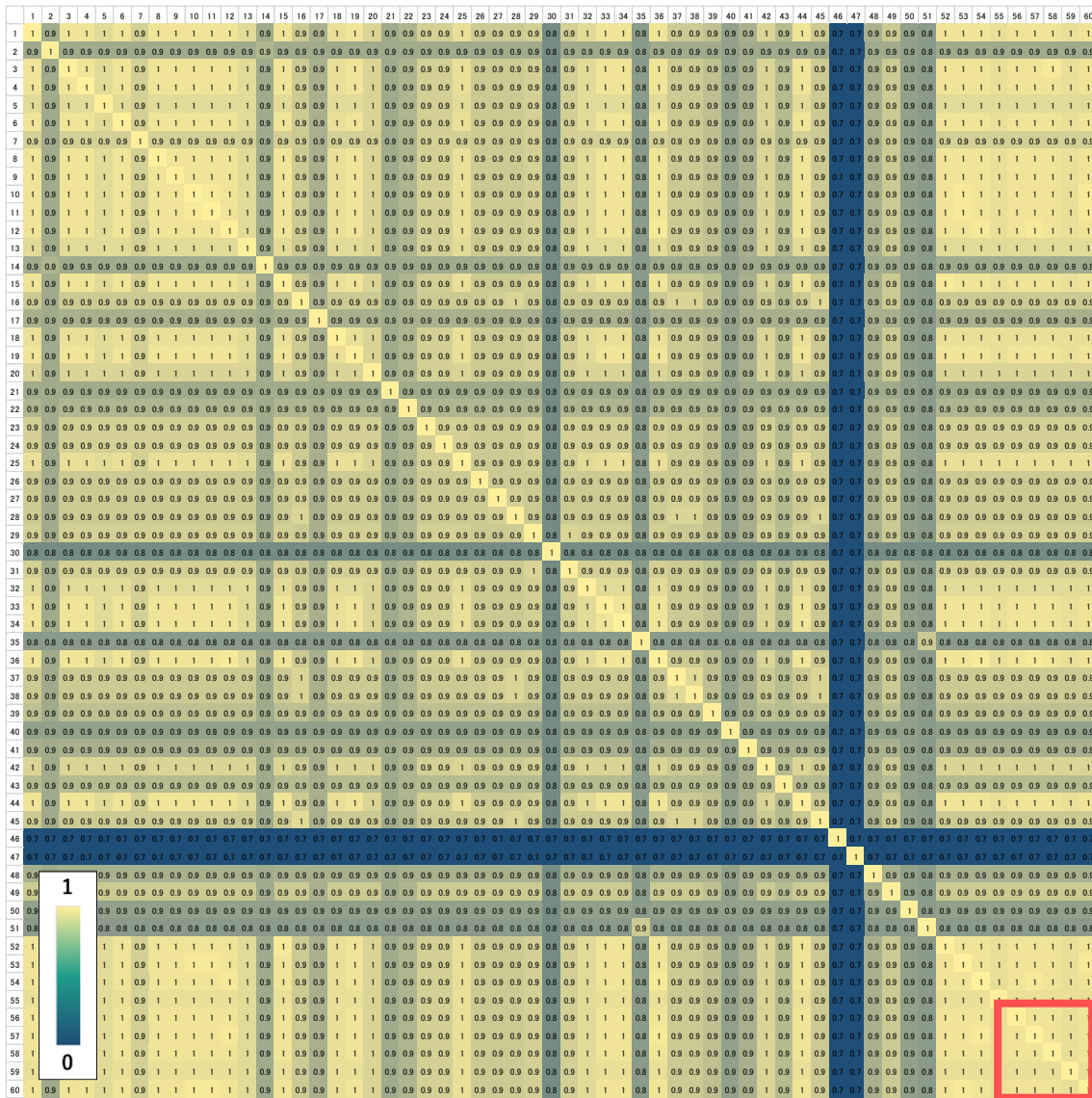


(a)

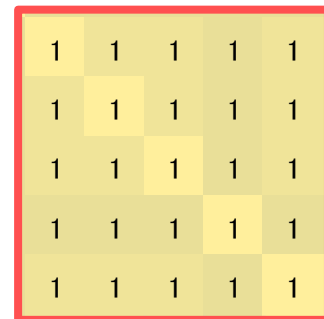


(b)

図 4.5 ベクトル G'_c のメンバシップ行列 F'



(a)



(b)

図 4.6 ベクトル G'_c の可達行列 F'

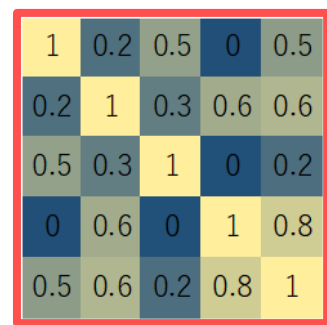
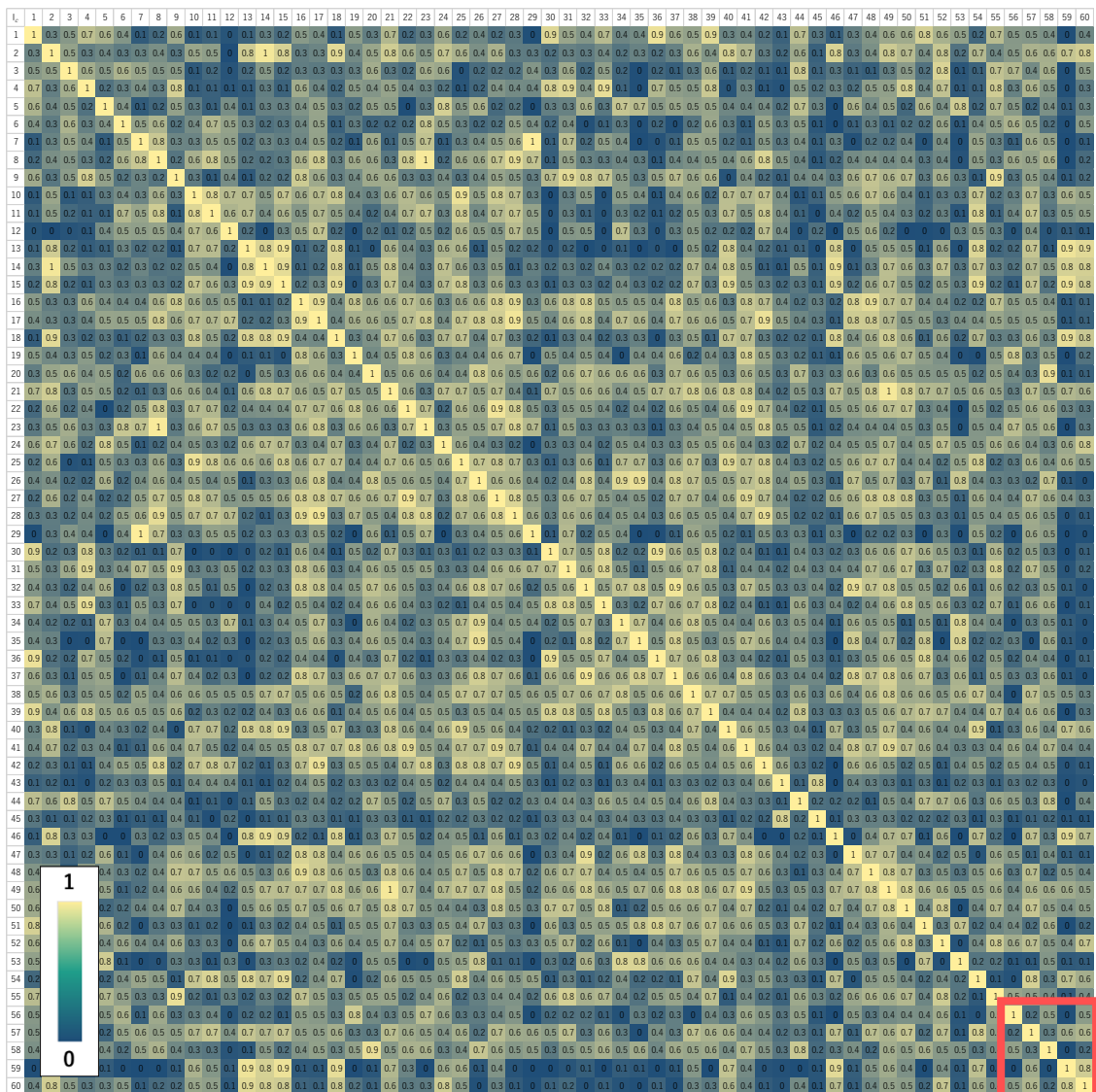
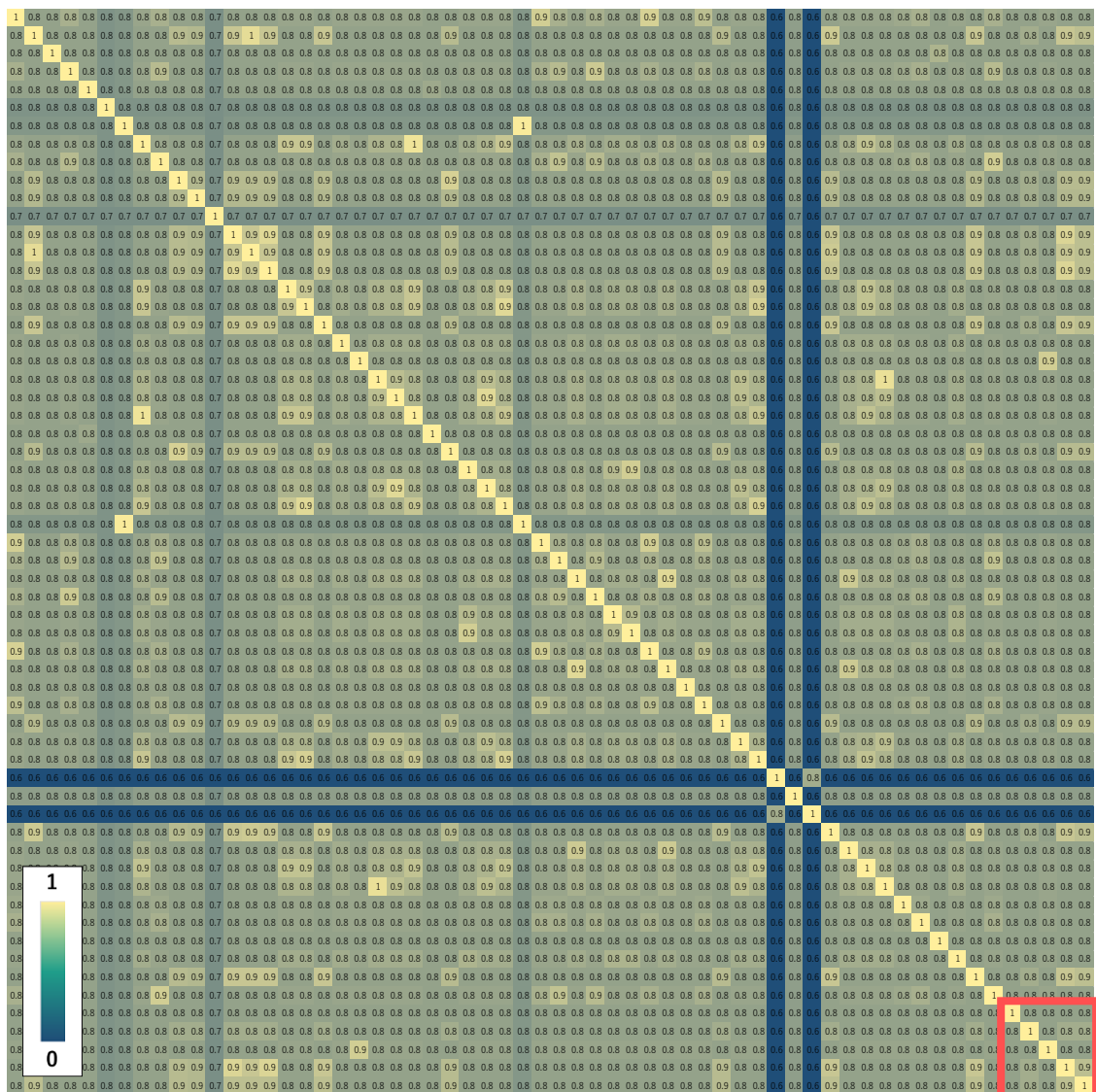


図 4.7 ベクトル I_c のメンバシップ行列 F''

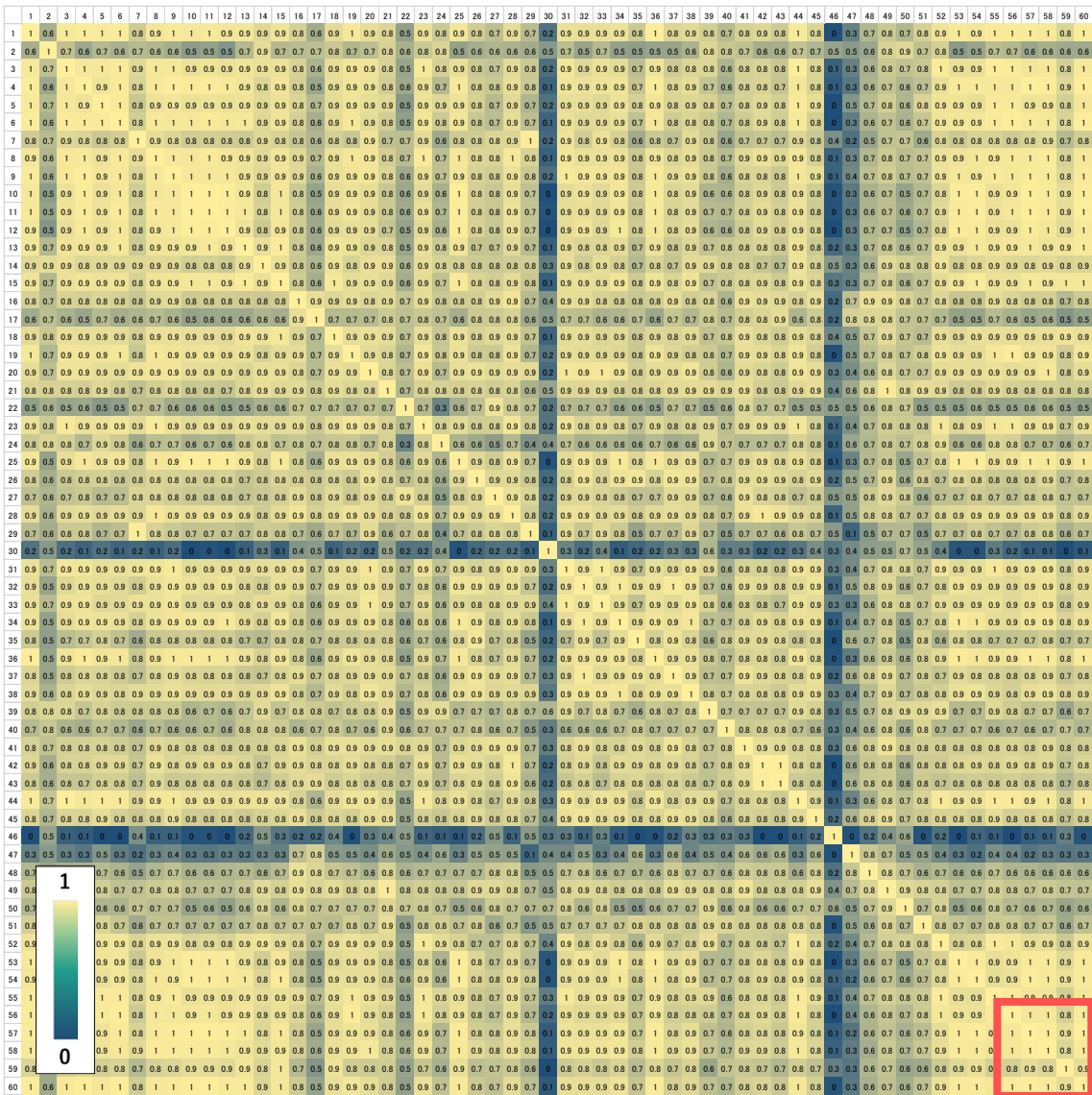


(a)

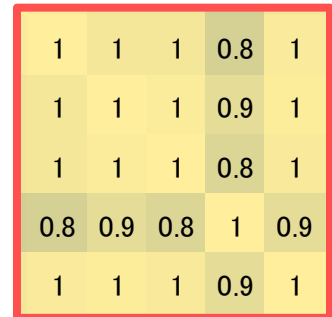
1	0.8	0.8	0.8	0.8
0.8	1	0.8	0.8	0.8
0.8	0.8	1	0.8	0.8
0.8	0.8	0.8	1	0.9
0.8	0.8	0.8	0.9	1

(b)

図 4.8 ベクトル \mathbf{l}_c の可達行列 \tilde{F}''

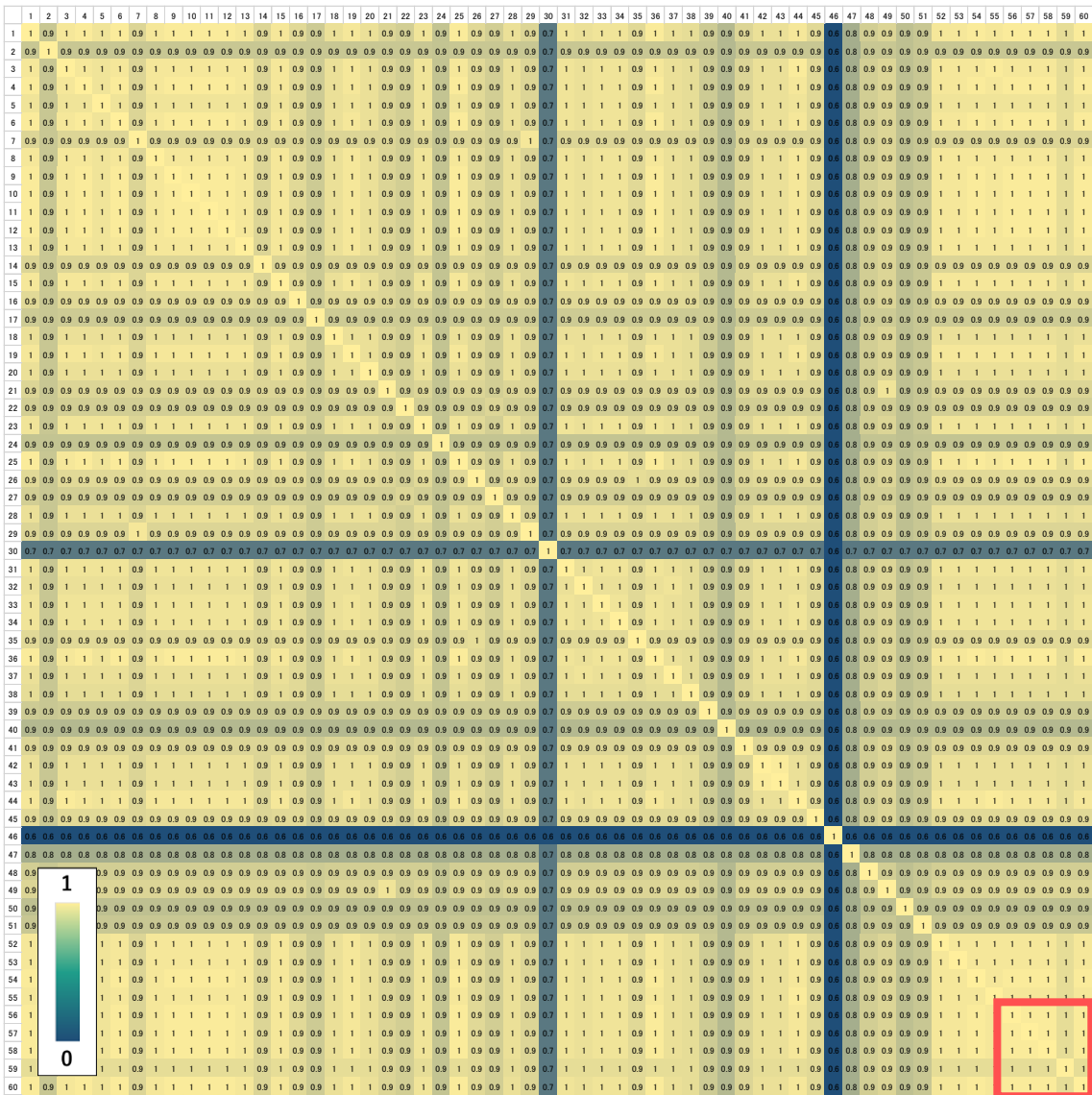


(a)

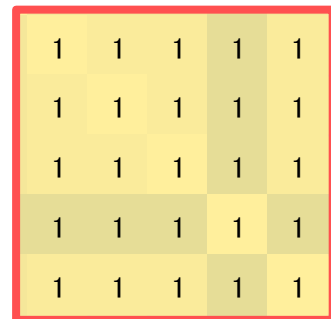


(b)

図 4.9 ベクトル I'_c のメンバシップ行列 F'''



(a)



(b)

図 4.10 ベクトル I_c の可達行列 F^m

図 4.3-4.10 の(b)は、いずれもコンター図の右下の領域を拡大したものである。可達行列への変換により、各ベクトル間の類似関係がより明確になり、近い性質をもつアスリート同士がクラスタを形成しやすくなったことがわかる。

4.3.2. クラスタリングの結果

次に、基準となる類似度 R の値に応じてベクトル \mathbf{G}_c および \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c が形成するクラスタ数と、それぞれのクラスタに属するアスリート c の一覧を表 4.1 および表 4.2 に示す。ここでは、形成されるクラスタ数に顕著な違いがみられた点に着目し、類似度 R が 0.70, 0.81 および 0.90 の場合の分析結果のみを記載する。また、 \mathbf{G}_c および \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c に対するクラスタ分析の結果を踏まえ、それぞれのデータに対する、後の分析および可視化の方針について述べる。

表 4.1 \mathbf{G}_c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)

R	cluster	athlete (c)	R	cluster	c	R	cluster	c	R	cluster	c	R	cluster	c
0.70	C1	1,6,8,14,33,39,5,11,13,15,21,52,54 .2,7,23,3,10,24,42,47,56,55,18,19, 25,30,37,38,40,43,31,46,57,58,59, 26,16,28,34,49,27,22,32,41,45,36, 20,35,51,53,29,50,12	A1	1	A12	14	A23	30	A34	46				
		A2									2,3,23,5 6	A13	15	A24
	C2	4,9,44,60,17,48	A3	4	A14	16,26,27, 45,28,25, 43,49	A25	32	A36	50				
	B1	1,6,8,11,13,14,52,33,39,2,15,3,54, 7,5,23,56,55,57,24,18,59,31,37,27, 29,16,26,28,30,35,38,42,25,43,45, 49,34,20,19,36,51,53,58,47,40,22, 32,12,41,50	0.90	0.90	0.90	0.90								
0.81	B2	4,9,44,60,17,48	A4	5	A15	17,48	A26	33	A37	52				
	B3	10	A5	6,13	A16	18,55,59	A27	34	A38	53				
	B4	21	A6	7	A17	19	A28	35,51	A39	54				
	B5	46	A7	8,39	A18	20	A29	36	A40	57				
			A8	9,44,60	A19	21	A30	37	A41	58				
	A9	10	A20	22	A31	38								
	A10	11	A21	24	A32	40								
	A11	12,41	A22	29	A33	42								

表 4.2 G'_c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)

R	cluster	athlete (c)	R	cluster	c	R	cluster	c
0.70	C1	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 36, 37, 38, 39, 41, 42, 44, 45, 49, 52, 53, 54, 55, 56, 57, 58, 59, 60, 21, 2, 22, 50, 40, 43, 17, 35, 48, 51, 30		A1	1, 3, 4, 5, 6, 8, 9, 10, 11, 12, 15, 19, 25, 32, 33, 34, 36, 42, 44, 52, 53, 54, 56, 57, 58, 60, 20, 31, 55, 59, 23, 24, 13, 29, 18, 7, 16, 26, 28, 37, 38, 27, 45, 49, 39, 43		A9	40
	C2	46		A2	2		A10	41
	C3	47		A3	14		A11	46
0.81	B1	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 23, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 36, 37, 38, 39, 41, 42, 44, 45, 49, 52, 53, 54, 55, 56, 57, 58, 59, 60, 21, 2, 22, 50, 40, 43, 17, 35, 48, 51, 30	0.90	A4	17		A12	47
	B2	46		A5	21	0.90	A13	48
	B3	47		A6	22		A14	50
				A7	30			
				A8	35,51			

表 4.3 I_c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)

R	cluster	athlete (c)	R	cluster	c	R	cluster	c	R	cluster	c	
0.70	C1	1,4,30,36,39,44,51,9,31,33,50,52,55,38,3,5,58,35,37,53,16,32,48,7,57,21,22,27,41,49,2,14,24,60,15,25,26,34,40,54,20,47,17,19,28,10,8,23,29,11,13,46,18,59,42,12,56,6		A1	1,36,30		A16	18		A31	44	
	C2	43,45		A2	2,14,46		A17	19		A32	45	
	B1	1,30,36,39,51,4,33,9,31,50,55		A3	3		A18	20,58		A33	48	
	B2	2,14,15,18,40,46,13,59,54,25,60,11,57,10		A4	4,33		A19	21,49		A34	50	
	B3	3,52		A5	5		A20	22,27		A35	51	
	B4	5,24		A6	6		A21	24		A36	52	
	B5	6		A7	7,29	0.90	A22	26,35	0.90	A37	53	
	B6	7,29		A8	8,23		A23	31		A38	54	
	0.81	B7	8,23,28,42,16,17,27,32,37,47,48,22,41,49,35,19,21,26,53,34		A9	9,55		A24	32,37,47		A39	56
		B8	12		A10	10,25		A25	34		A40	57
		B9	20,58		A11	11		A26	38		A41	60
		B10	38		A12	12		A27	39			
		B11	43,45		A13	13,15,59		A28	40			
B12		44		A14	16		A29	41				
B13		56		A15	17,28,42		A30	43				

表 4.4 I'_c に対するファジィクラスタ分析の結果 ($R = 0.70, 0.81, 0.90$)

R	cluster	athlete (c)	R	cluster	c	R	cluster	c
0.70	C1	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 2, 29, 50, 40, 22, 48, 17, 47	A1	3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 19, 23, 25, 31, 32, 33, 34, 36, 44, 52, 53, 54, 55, 56, 57, 58, 60, 14, 18, 20, 28, 42, 38, 45, 59, 41, 7, 26, 35, 37, 39, 21, 16, 27, 43, 49, 29, 17, 48, 22				
	C2	30	A2	2		A7	47	
	C3	46	A3	24		A8	50	
0.81	B1	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 2, 29, 50, 40, 22, 48, 17, 47	A4	30	0.90	A9	51	
	B2	30	A5	40				
	B3	46	A6	46				

1) ベクトル G_c が形成するクラスタ

表 4.1 より、類似度 $R = 0.70$ をクラスタ形成の基準としたとき、60 名のアスリートは 2 つのクラスタ (クラスタ C1, C2) を形成する。類似度 $R = 0.81$ を基準とした場合には、55 名のアスリートが 5 つのクラスタ (クラスタ B1-5) を形成するが、クラスタ B2 に所属するアスリートは C1 と同様であり、クラスタ B3 および B4, B5 はいずれも 1 名のアスリートが所属するのみである。類似度 $R = 0.90$ を基準とした場合、60 名のアスリートは 41 のクラスタに分割される。このうち、2 名以上のアスリートが所属するクラスタは、クラスタ A2, A5, A7, A8, A11, A14, A15, A16, A28 のみである。

類似度 $R = 0.81$ を基準とした場合に形成されるクラスタ B1, B2 には、全体の 9 割以上 (57 名) のアスリートが所属しており、類似度 $R = 0.90$ を基準とすると、クラスタ B1 は A2, A5, A7, A11, A14, A16, A28 といったクラスタに、クラスタ B2 は A8, A15 といったクラスタに分割される。上記の B1 と B2 の下位の 9 つのクラスタは、いずれも複数名のアスリートが所属するクラスタである。そこで、クラスタ B1 および B2 に所属する 57 名のアスリートを中心に、より詳細な分析および可視化を行うこととした。

2) ベクトル \mathbf{G}'_c が形成するクラスタ

表 4.2 より、類似度 $R = 0.70, 0.81$ をクラスタ形成の基準としたとき、60 名のアスリートは3つのクラスタ（クラスタ C1, C2, C3 およびクラスタ B1, B2, B3）を形成し、60 名中 58 名がクラスタ C1 と B1 に所属することがわかる。それ以外の2名は、個別のクラスタを単独で形成する（C2, C3, B2, B3）。また、類似度 $R = 0.90$ とした場合も、60 名中 46 名が同じクラスタ（A1）に所属し、その他のアスリートは、単独でクラスタを形成することがわかる（A2-A14）。

このように、ベクトル \mathbf{G}'_c を分析に用いた場合には、ほとんどのアスリートが1つのクラスタ（C1, B1, A1）に所属する。このことから、クラスタ C1（=B1）および A1 に所属するアスリートの \mathbf{G}'_c は、集団内に共通の座標遷移を示すものととらえ、これらのデータを分析および可視化の対象とした。

3) ベクトル \mathbf{I}_c が形成するクラスタ

表 4.3 より、類似度 $R = 0.70$ をクラスタ形成の基準としたとき、60 名のアスリートは2つのクラスタ（クラスタ C1, C2）を形成し、60 名中 58 名がクラスタ C1 に所属することがわかる。類似度 $R = 0.81$ を基準とした場合には、45 名のアスリートが主要な3つのクラスタ（クラスタ B1, B2, B7）を形成し、残りの15名は10のクラスタ（クラスタ B3-B6 および B8-B13）を形成するが、いずれのクラスタも1名、または2名のアスリートが所属するのみである。類似度 $R = 0.90$ を基準とした場合、60 名のアスリートは41のクラスタを形成し、いずれのクラスタも、所属するアスリートは1名から3名のみとなる。

類似度 $R = 0.81$ を基準とした場合に形成される、3つの主要なクラスタには、全体の75%に相当するアスリートが所属しており、各クラスタに所属するアスリート数は、いずれも10名以上である。そこで、これらの3つのクラスタ（クラスタ B1, B2, B7, $R = 0.81$) に所属するアスリートの \mathbf{I}_c は、集団内に典型的な座標遷移を示すものと判断した。以上を踏まえ、上記の3つのクラスタに所属する45名のアスリートを中心に、より詳細な分析および可視化を行うこととした。

4) ベクトル \mathbf{I}'_c が形成するクラスタ

表 4.4 より, 類似度 $R = 0.70, 0.81$ をクラスタ形成の基準としたとき, いずれの場合も, 60 名中 58 名のアスリートが同じクラスタ (C1, B1) に所属することがわかる. それ以外の 2 名は, 個別のクラスタを単独で形成する (C2, C3, B2, B3). また, 類似度 $R = 0.90$ とした場合も, 60 名中 52 名が同じクラスタ (A1) に所属し, その他のアスリートは, 単独でクラスタを形成することがわかる (A2-A9).

このように, ベクトル \mathbf{I}'_c を分析に用いた場合には, ほとんどのアスリートが 1 つのクラスタ (C1, B1, A1) に所属する. このことから, クラスタ C1 (=B1) および A1 に所属するアスリートの \mathbf{I}'_c は, 集団内に共通の座標遷移を示すものととらえ, これらのデータを分析および可視化の対象とした.

4.3.3. クラスタの特徴

以下では, ベクトル \mathbf{G}_c および \mathbf{G}'_c , \mathbf{I}_c , \mathbf{I}'_c が形成するクラスタの特徴について, それぞれ分析を行った結果を記載する.

1) TF-IDF 空間上の遷移を示す TF-IDF ベクトル \mathbf{G}_c が形成するクラスタの特徴

まず, 類似度 $R = 0.70, 0.81, 0.90$ を基準とした場合に形成される, 各クラスタの特徴を把握するため, それぞれのクラスタに所属するアスリートに対して, ベクトル \mathbf{G}_c の平均ベクトル \mathbf{G} を求めた. これにより, アスリートの入力ベクトル \mathbf{H}_j が分類されるマップ上の領域 CL_t に関して, クラスタ別の特徴を把握することができる. ここでは, 前章で明らかにされた, 領域 CL_t を特徴づけるキーワードの分布 (図 4.11) と併せて, 類似度 $R = 0.70, 0.81$ の場合のクラスタ C1 および C2 (図 4.12), 類似度 $R = 0.81$ を基準とした場合のクラスタ B1 (図 4.13), 類似度 $R = 0.90$ を基準とした場合のクラスタ A2, A5, A7, A8, A11 (図 4.14) および A14, A15, A16, A18 (図 4.15) について, 各クラスタに所属するアスリートのベクトル \mathbf{G}_c の平均ベクトル \mathbf{G} を算出し, 平均ベクトル \mathbf{G} の成分を $w_t (t = 1, 2, \dots, 10)$ とした.

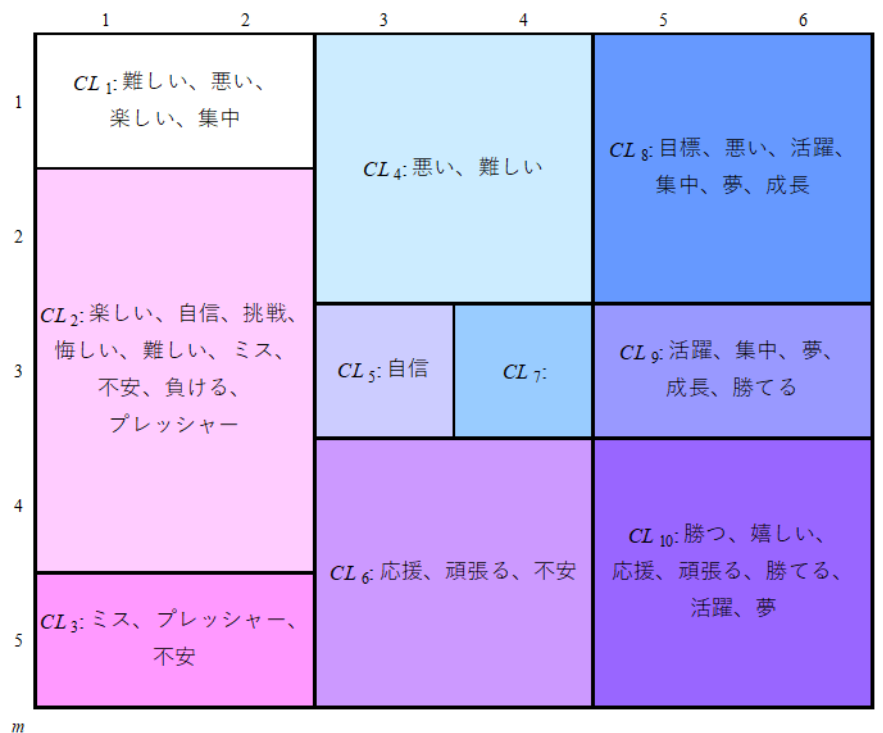


図 4.11 各領域 CL_t において顕著な値を示したキーワード

類似度 $R = 0.70, 0.81$ をクラス形成の基準とし、クラス C1, C2 および B1 に所属するアスリートの平均ベクトル \mathbf{G} を求めたところ、クラス C1 と B1 については、領域 CL_5 に対応する w_5 の値が0であったが、それ以外の成分間では明確な違いがみられなかった。一方、クラス C2 では、 w_5 の値が高い値を示しており、これは Modal word の中でも「自信 (M_9)」の TF-IDF 値に対応する。

また、類似度 $R = 90$ を基準とし、クラス B1 の下位クラスである A2, A5, A7, A11, A14, A16, A28 と、クラス B2 の下位クラスである A8, A15 に所属するアスリートの平均ベクトル \mathbf{G} を求めた。その結果、上位のクラスと同様に、B1 の下位クラスでは w_5 の値が0であるのに対し、B2 の下位クラスでは、この値が高くなる傾向が示された。領域 CL_5 に入力ベクトルが分類されたアスリートは、60 名中 6 名のみであったことから、発話中に「自信 (M_9)」という Modal word を含むか否かが、所属するクラスを左右する要因になったと考えられる。

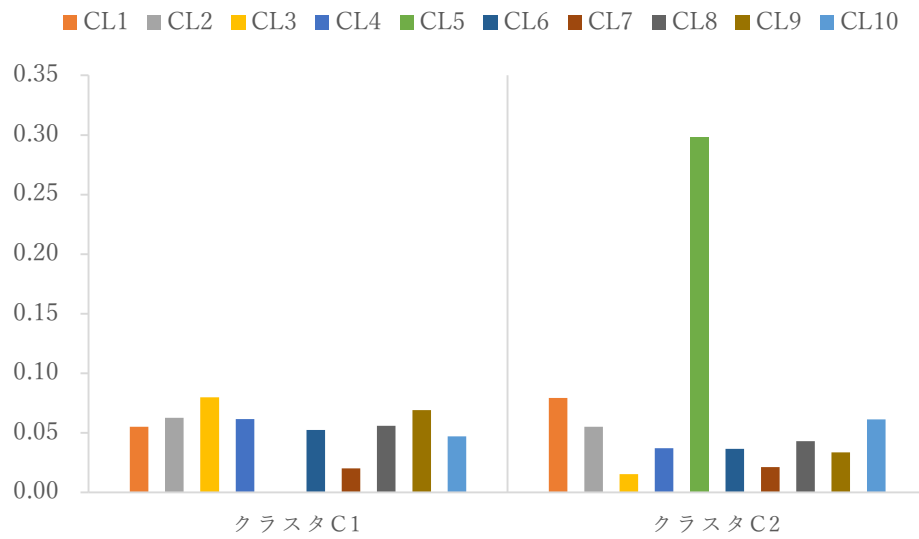


図 4.12 クラスタ C1 および C2 ($R = 0.70$) における平均ベクトル成分 w_t

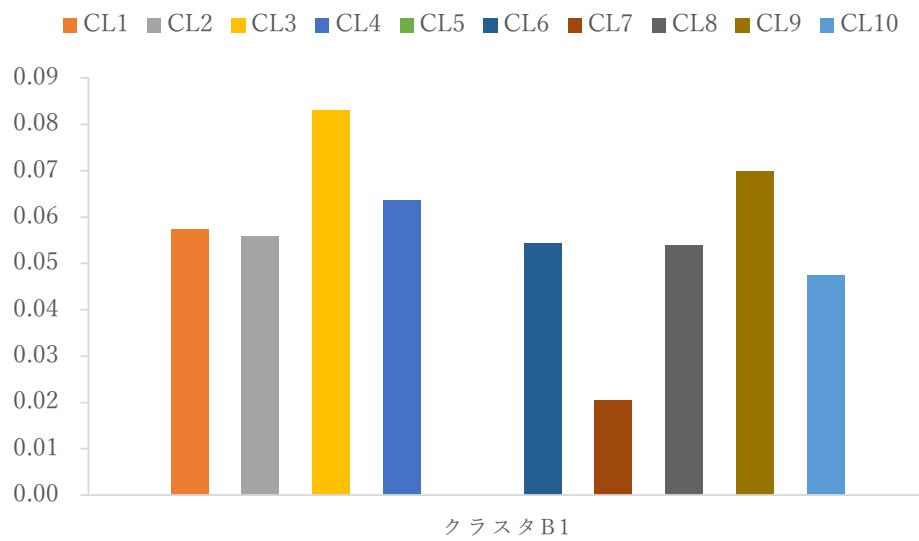


図 4.13 クラスタ B1 ($R = 0.81$) における平均ベクトル成分 w_t

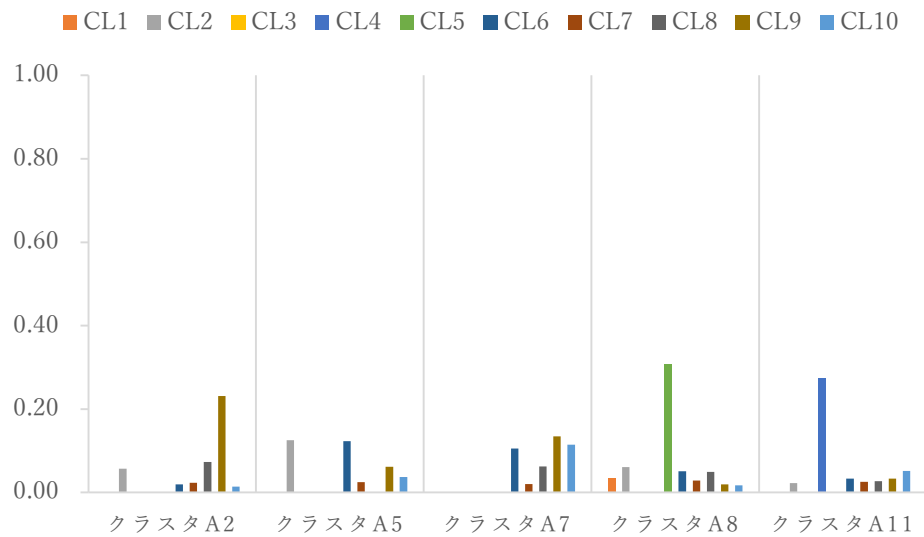


図 4.14 クラスタ A2, A5, A7, A8, A11 ($R = 0.90$) における平均ベクトル成分 w_t

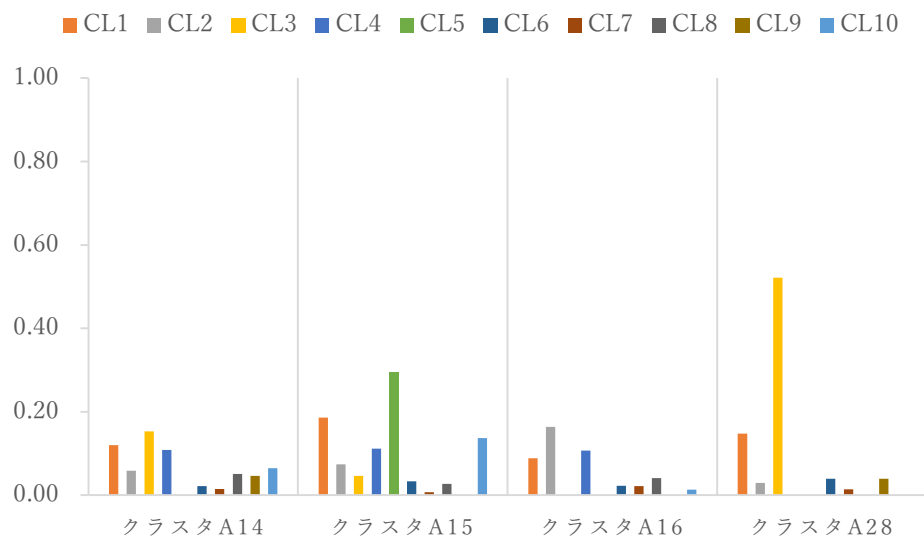


図 4.15 クラスタ A14, A15, A16, A18 ($R = 0.90$) における平均ベクトル成分 w_t

2) TF-IDF 空間上の遷移を示す TF-DF ベクトル \mathbf{G}'_c が形成するクラスタの特徴

ベクトル \mathbf{G}_c と同様に、類似度 $R = 0.70, 0.81, 0.90$ を基準とした場合の、各クラスタの特徴を把握するため、各クラスタに所属するアスリート間で、ベクトル \mathbf{G}'_c の平均ベクトル \mathbf{G}' を求めた。類似度 $R = 0.70, 0.81$ において形成されるクラスタは同数であるため、

ここでは、類似度 $R = 0.70$ の場合のクラスタ C1 (図 4.16) と、類似度 $R = 0.90$ の場合のクラスタ A1 (図 4.17) のみを分析の対象とした。このとき、平均ベクトル \mathbf{G}' の成分を $\tilde{w}_t (t = 1, 2, \dots, 10)$ とおく。

類似度 $R = 0.70, 0.90$ をクラスタ形成の基準とし、クラスタ C1 および A1 に所属するアスリートの平均ベクトル \mathbf{G}' を求めたところ、平均ベクトル成分 \tilde{w}_t の値について、両クラスタは同様の傾向を示した。いずれのクラスタにおいても、平均ベクトル成分 \tilde{w}_7 が顕著に高い値を示している。これは、ベクトル \mathbf{G}_c の分析と同様に、領域 CL_7 に多くの入力ベクトルが分類され、 $N_{c,7}$ が増加した結果、平均ベクトル成分 w_7 が高い値を示した結果と考えられる。

図 4.12 および図 4.13 と比較すると、図 4.16 と図 4.17 では、突出した値をもつベクトル成分と、他の成分との値の差がより顕著に示されている。発話 U_j の多くは Modal Word を含まないため、大半のアスリートの入力ベクトル \mathbf{H}_j には、ゼロベクトルが含まれる。これらのベクトルは、いずれもマップ上の同じ領域へ分類される。このようなアスリートの座標遷移に関わる共通性が、TF-DF 法により強調（ベクトル成分の計算で、高い評価値が付与）された結果、遷移ベクトル間の差異を見だしにくくなったことがうかがえる。

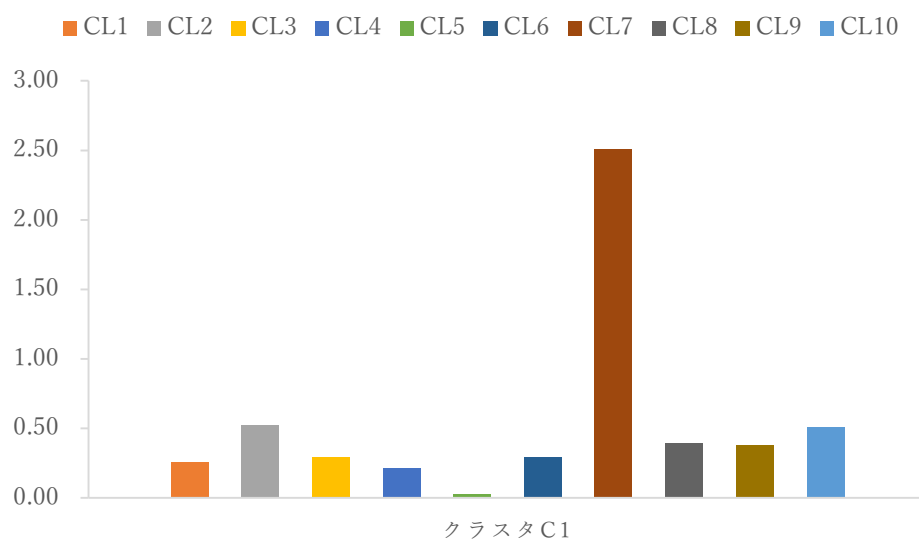


図 4.16 クラスタ C1 ($R = 0.70$) における平均ベクトル成分 \tilde{w}_t

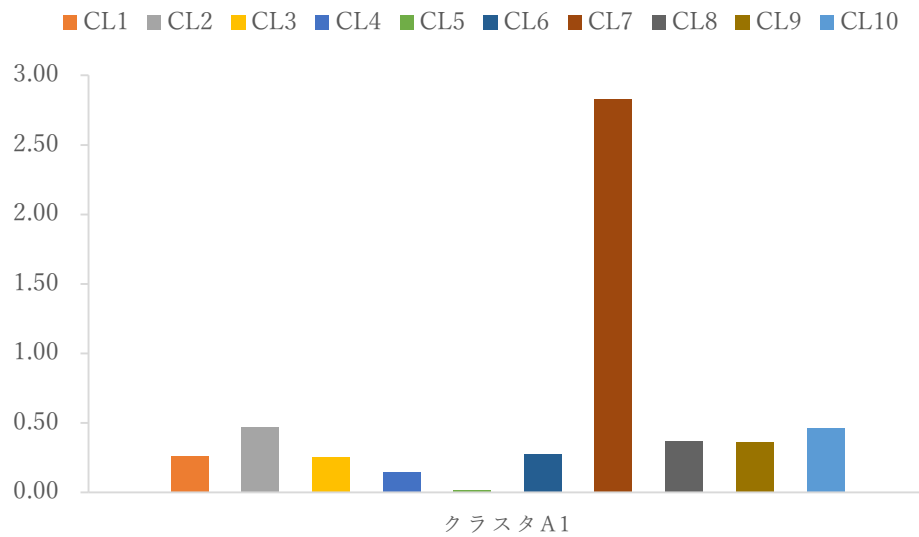


図 4.17 クラスタ A1 ($R = 0.90$) における平均ベクトル成分 \bar{w}_t

3) TF-DF 空間上の遷移を示す TF-IDF ベクトル \mathbf{I}_c が形成するクラスタの特徴

まず、類似度 $R = 0.70, 0.81, 0.90$ を基準とした場合に形成される、各クラスタの特徴を把握するため、それぞれのクラスタに所属するアスリート間で、ベクトル \mathbf{I}_c の平均ベクトル \mathbf{I} を求めた。これにより、アスリートの入力ベクトル \mathbf{H}'_j が分類されるマップ上の領域 CL'_t に関して、クラスタ別の特徴を把握することができる。ここでは、前章で明らかにされた、領域 CL'_t を特徴づけるキーワードの分布 (図 4.18) と併せて、類似度 $R = 0.70$ の場合のクラスタ C1 (図 4.19)、類似度 $R = 0.81$ を基準とした場合のクラスタ B1, B2, B7 (図 4.20) について、各クラスタに所属するアスリートのベクトル \mathbf{I}_c の平均ベクトル \mathbf{I} を算出した。このとき、平均ベクトル \mathbf{I} の成分を $w'_t (t = 1, 2, \dots, 10)$ とおく。

	1	2	3	4	5	6	l
1	CL_1 :	CL_2 : ミス	CL_5 : 悪い、ミス、集中		CL_8 : 難しい、悪い、楽しい		
2	CL_3 : 目標、勝つ、 頑張る、活躍		CL_6 : 頑張る、活躍、自信、 成長、悔しい、負ける、 悪い、ミス		CL_9 : 楽しい、自信、 悔しい、難しい		
3					CL_9 : 楽しい、自信、 悔しい、難しい		
4	CL_4 : 勝つ、勝てる、 目標、嬉しい		CL_7 : 応援、プレッシャー、 不安、勝てる、勝つ、 頑張る		CL_{10} : 嬉しい、楽しい、 応援、自信、プレッシャー		
5					CL_{10} : 嬉しい、楽しい、 応援、自信、プレッシャー		
m							

図 4.18 各領域 CL'_t において顕著な値を示したキーワード

類似度 $R = 0.70$ をクラスタ形成の基準とし、クラスタ C1 に所属するアスリートの平均ベクトル \mathbf{I} を求めたところ、平均ベクトル成分 w'_t はいずれも 0.10 以下の値を示し、成分間で値の顕著な違いはみられなかった。

また、類似度 $R = 0.81$ をクラスタ形成の基準とし、クラスタ B1, B2, B7 に所属するアスリートの平均ベクトル \mathbf{I} を求めたところ、これらの平均ベクトル成分 w'_t は、それぞれ異なる傾向を示した。クラスタ B1 では、 CL'_4 への入力ベクトル \mathbf{H}'_j の分類個数を示す w'_4 の値が、他の成分と比べて相対的に高い値であった。領域 CL'_4 では、代表ベクトル $\mathbf{W}_{l,m}$ の成分のうち、 $W_{1,l,m}$, $W_{2,l,m}$, $W_{6,l,m}$ および $W_{12,l,m}$ が高い値を示しており、これらは「目標 (M_1)」、「勝つ (M_2)」、「嬉しい (M_6)」、「勝てる (M_{12})」、といったキーワードの TF-IDF 値に対応する。そのため、クラスタ B1 に所属するアスリートは、競技における勝利などの具体的な目標について、より多く言及していると考えられる。

クラスタ B2 では CL'_9 への入力ベクトル \mathbf{H}'_j の分類個数を示す w'_9 の値が、他の成分と比べて相対的に高い値であった。領域 CL'_9 では、代表ベクトル $\mathbf{W}_{l,m}$ の成分のうち、 $W_{3,l,m}$, $W_{5,l,m}$, $W_{9,l,m}$ および $W_{16,l,m}$ が高い値を示しており、これらは「難しい (M_3)」、「楽

しい (M_5)」, 「自信 (M_9)」, 「悔しい (M_{16})」, といったキーワードの TF-IDF 値に対応する。そのため, クラスタ B2 に所属するアスリートは, 主にスポーツそのものの楽しさといった, 感情的な側面に焦点を当てていると考えられる。

クラスタ B7 では, CL'_5 および CL'_8 への入力ベクトル \mathbf{H}'_j の分類個数を示す w'_5, w'_8 の値が, 他の成分と比べて相対的に高い値であった。領域 CL'_5 では, 代表ベクトル $\mathbf{W}_{l,m}$ の成分のうち, $W_{7,l,m}, W_{14,l,m}$ および $W_{18,l,m}$ が高い値を示しており, これらは「悪い (M_7)」, 「ミス (M_{14})」, 「集中 (M_{18})」, といったキーワードの TF-IDF 値に対応する。同様に, 領域 CL'_8 では, 代表ベクトル $\mathbf{W}_{l,m}$ の成分のうち, $W_{3,l,m}, W_{5,l,m}$ および $W_{7,l,m}$ が高い値を示しており, これらは「難しい (M_3)」, 「楽しい (M_5)」, 「悪い (M_7)」, といったキーワードの TF-IDF 値に対応する。そのため, クラスタ B7 に所属するアスリートは, 競技における失敗や困難さといった側面に意識を向けていると考えられる。

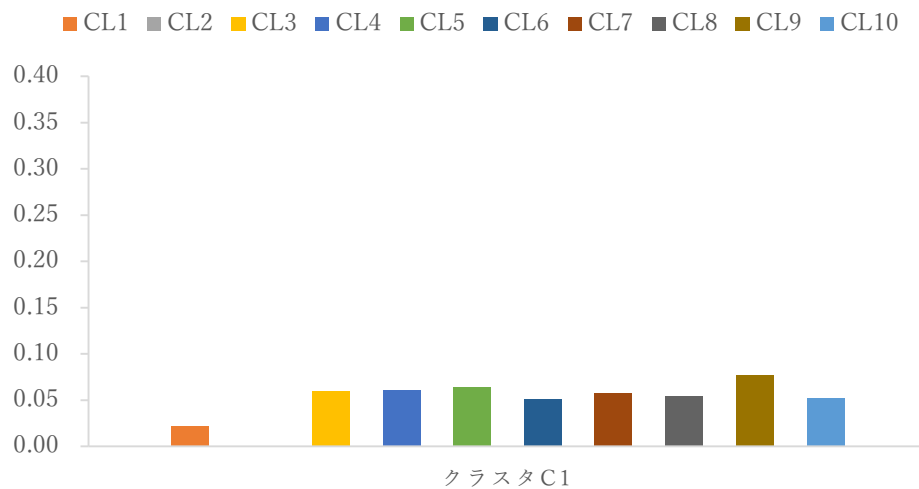


図 4.19 クラスタ C1 ($R = 0.70$) における平均ベクトル成分 w'_t

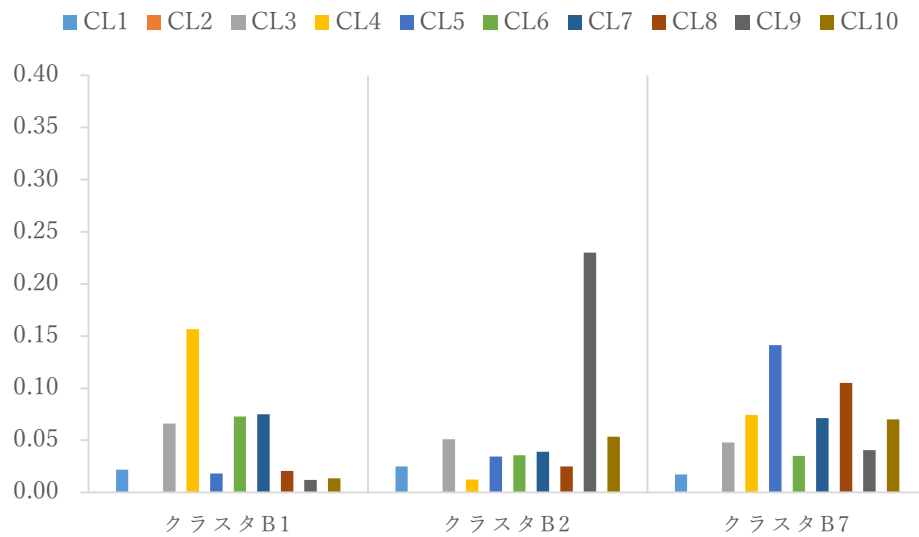


図 4.20 クラスタ B1, B2, B7 ($R = 0.81$) における平均ベクトル成分 w'_t

さらに、類似度 $R = 0.90$ を基準とした場合に形成される、12 のクラスタ（クラスタ A1, A2, A4, A8, A9, A10, A13, A15, A19, A20, A22, A24）に所属するアスリートについても、平均ベクトル \mathbf{I} を算出した（図 4.21, 図 4.22）。これらのクラスタは、類似度 $R = 0.81$ におけるクラスタ B1, B2, B7 の下位クラスタである。類似度 $R = 0.90$ におけるクラスタ A1, A4, A9 は、類似度 $R = 0.81$ において結合し、クラスタ B1 を形成する。同様に、下位のクラスタ A2, A10, A13 が結合して上位のクラスタ B2 を、下位のクラスタ A8, A15, A19, A20, A22, A24 が結合して上位のクラスタ B7 を形成する。

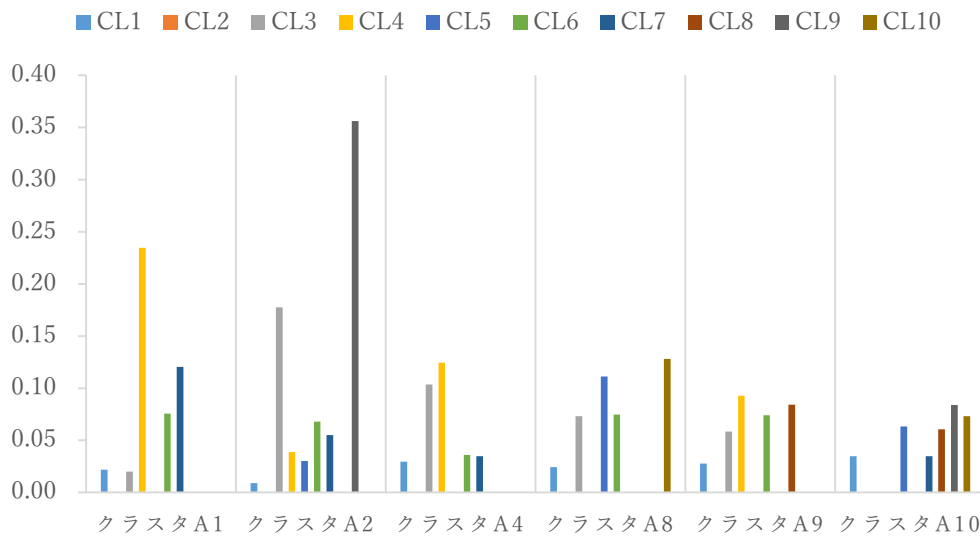


図 4.21 クラスタ A1, A2, A4, A8, A9, A10 ($R = 0.90$) における平均ベクトル成分 w'_t

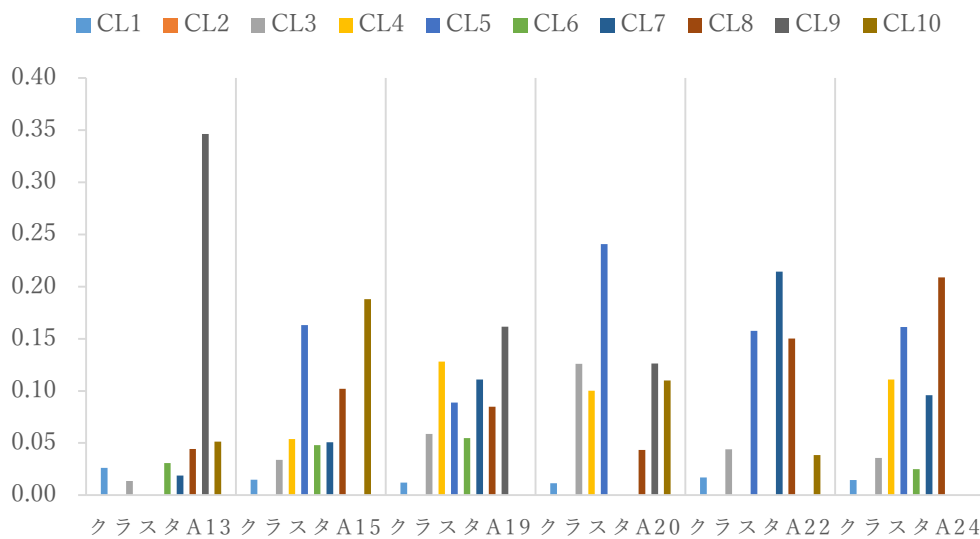


図 4.22 クラスタ A13, A15, A19, A20, A22, A24 ($R = 0.90$) における平均ベクトル成分 w'_t

図 4.21 におけるクラスタ A1, A4, A9 に注目すると、いずれのクラスタも、 CL_4 への入力ベクトル \mathbf{H}'_j の分類個数を示す、 w'_4 が最も高い値を示した。同様に、クラスタ A2, A10, A13 は、いずれも CL_9 への入力ベクトル \mathbf{H}'_j の分類個数を示す、 w'_9 が最も高い値を

示した。これらのクラスタから算出される平均ベクトル \mathbf{I} は、それぞれの上位クラスタであるクラスタ B1, B2 ($R = 0.81$) と同様の傾向を示していることがわかる。

図 4.21 および図 4.22 におけるクラスタ A8, A15, A19, A20, A22, A24 は、 CL'_5, CL'_8 への入力ベクトルの分類個数を示す w'_5, w'_8 だけでなく、 w'_4, w'_7 も相対的に高い値を示した。そのため、これらのクラスタに所属するアスリートは、インタビュー中に様々なキーワードを使用しており、その心理状態と変化は、多様な要因によって特徴づけられるといえよう。

4) TF-DF 空間上の遷移を示す TF-DF ベクトル \mathbf{I}'_c が形成するクラスタの特徴

ベクトル \mathbf{I}_c と同様に、類似度 $R = 0.70, 0.81, 0.90$ を基準とした場合の、各クラスタの特徴を把握するため、各クラスタに所属するアスリート間で、ベクトル \mathbf{I}'_c の平均ベクトル \mathbf{I}' を求めた。類似度 $R = 0.70, 0.81$ において形成されるクラスタは同数であるため、ここでは、類似度 $R = 0.70$ の場合のクラスタ C1 (図 4.23) と、類似度 $R = 0.90$ の場合のクラスタ A1 (図 4.24) のみを分析の対象とした。このとき、平均ベクトル \mathbf{I}' の成分を $\tilde{w}'_t (t = 1, 2, \dots, 10)$ とおく。

類似度 $R = 0.70, 0.90$ をクラスタ形成の基準とし、クラスタ C1 および A1 に所属するアスリートの平均ベクトル \mathbf{I}' を求めたところ、平均ベクトル成分 \tilde{w}'_t の値について、両クラスタは同様の傾向を示した。いずれのクラスタにおいても、平均ベクトル成分 \tilde{w}'_1 が顕著に高い値を示している。これは、ベクトル \mathbf{G}_c および \mathbf{G}_c を用いた分析と同様に、領域 CL'_1 への入力ベクトル \mathbf{H}'_j の分類個数 $N'_{c,1}$ が大きくなる傾向にあり、その値が TF-DF 法によってさらに重み付けられた結果と推察される。

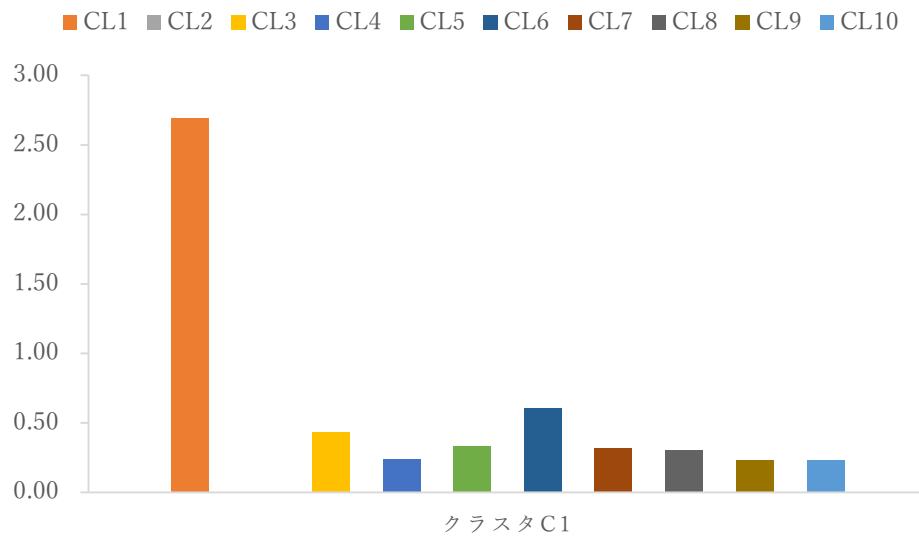


図 4.23 クラスタ C1 ($R = 0.70$) における平均ベクトル成分 w'_t

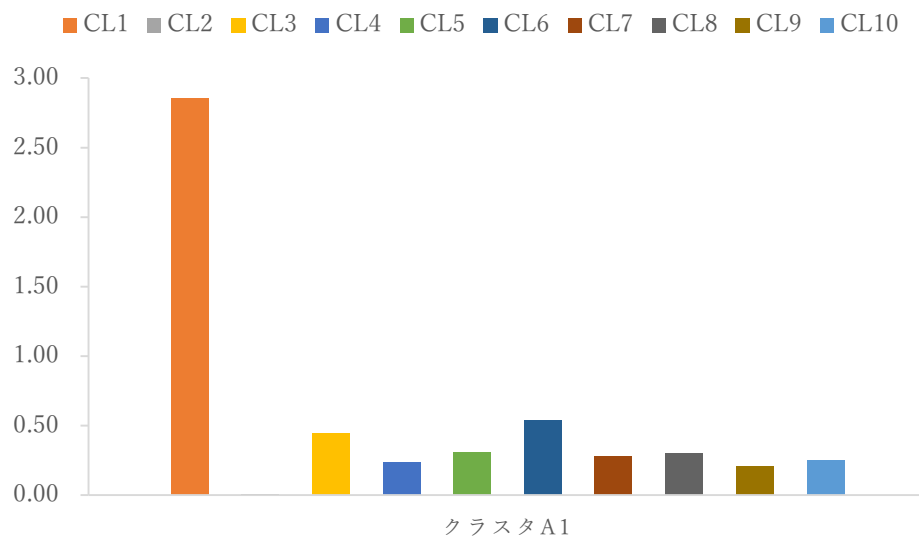


図 4.24 クラスタ A1 ($R = 0.90$) における平均ベクトル成分 w'_t

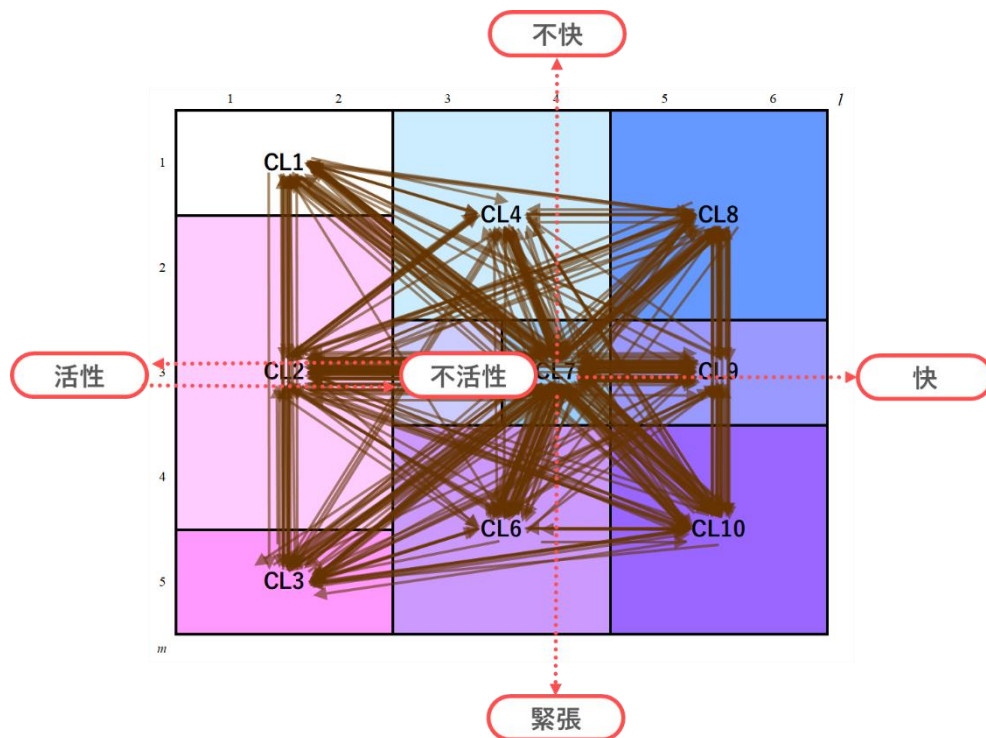
4.4. 心理変化パターンの可視化

本節では、第3章・第4章を通じて得られた複数の遷移ベクトルと、そのクラスタリング結果に基づき、各クラスタが示す心理変化のパターンについて可視化を行った。可視化の結果および、それぞれのパターンに対する心理学的な解釈を以下に示す。

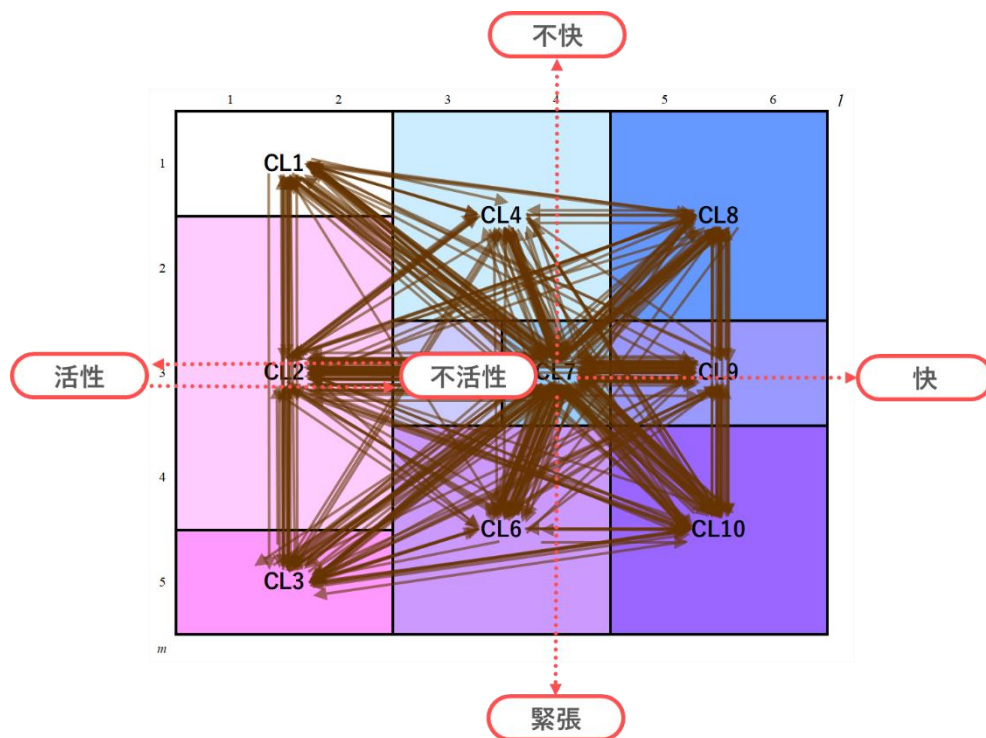
4.4.1. TF-IDF 空間上の遷移を示す TF-IDF ベクトル \mathbf{G}_c に基づく可視化

以下の図 4.25-図 4.33 は、マップ上に割り当てられた各アスリートの入力ベクトル \mathbf{H}_j の座標が遷移する軌跡を、矢線図によって可視化したものである。それぞれの可視化の図は、表 4.1-4.4 において同じクラスタに所属する選手同士の軌跡を重ね合わせることで、個々のクラスタに共通してみられる座標遷移のパターンを表現している。ここでは、類似度 $R = 0.70, 0.81$ を基準とした場合のクラスタ C1 および B1 (図 4.25), B2 (図 4.27), 類似度 $R = 0.90$ を基準とした場合のクラスタ A2, A5, A7, A11, A14, A16, A28 (図 4.26) および A8, A15 (図 4.27) について、可視化を行った。

これらの可視化の図により、各クラスタに所属するアスリートが示す座標遷移の傾向、すなわち集団における心理変化の共通パターンを認識することが可能となる。さらに、各クラスタの可視化の図と併せて、下位クラスタの可視化の図を示すことで、集団にみられる心理変化のパターンを細分化し、より詳細に検討できる。



クラスター C1 ($R = 0.70$)



クラスター B1 ($R = 0.81$)

図 4.25 ベクトル G_c が形成するクラスター C1 ($R = 0.70$) およびクラスター B1 ($R = 0.81$)

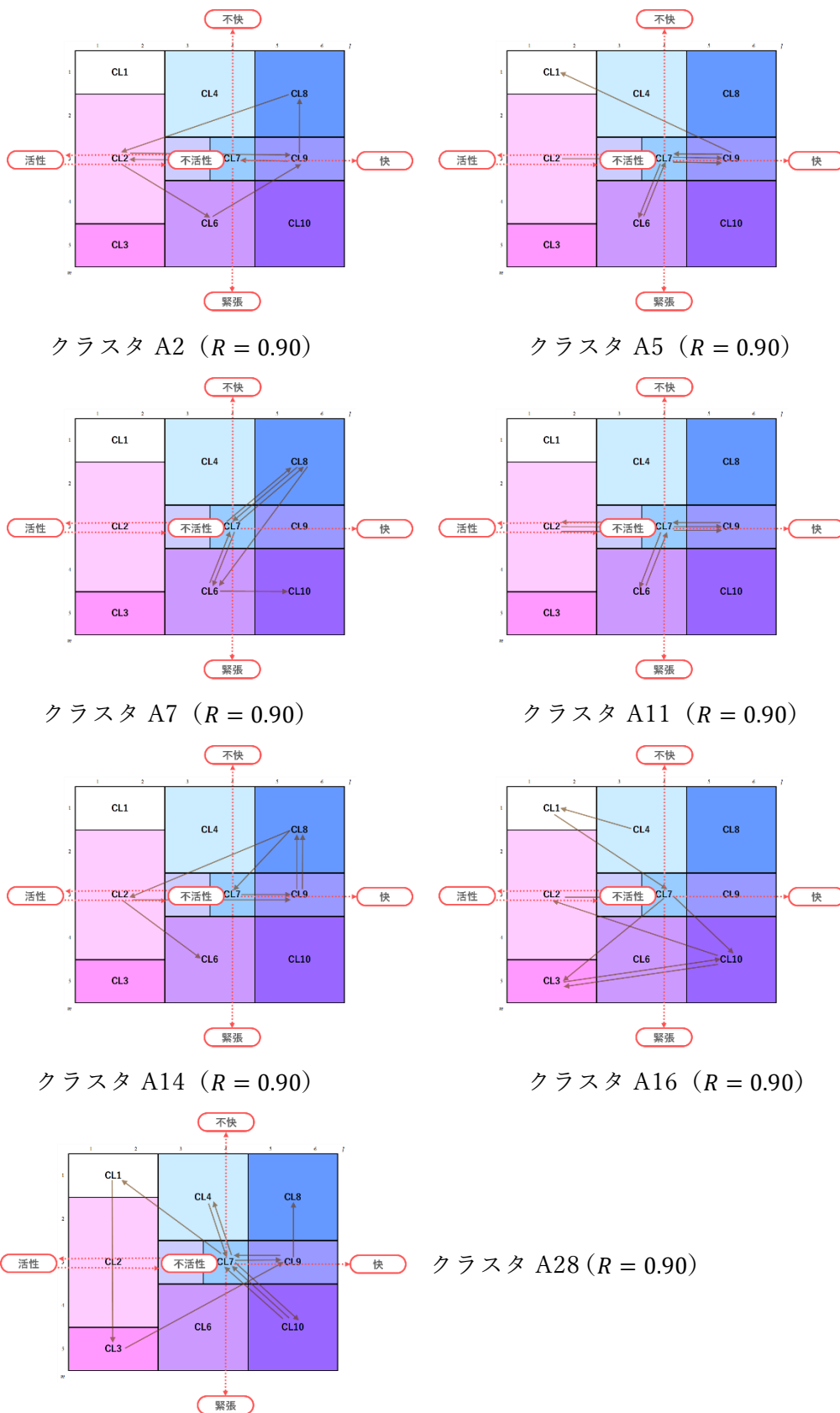
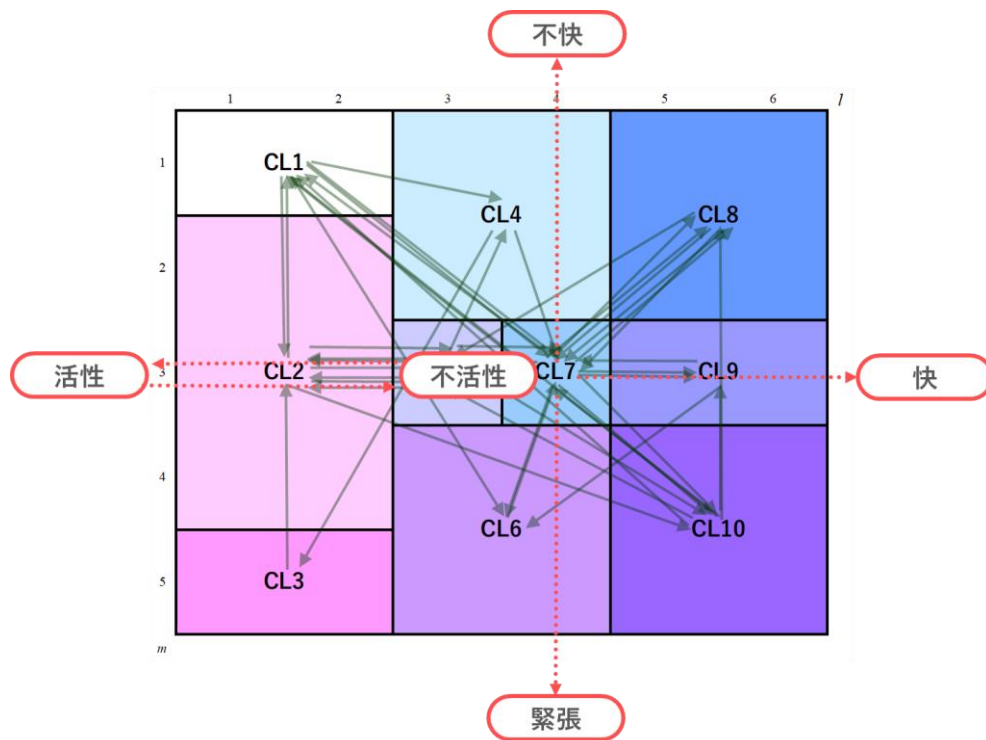
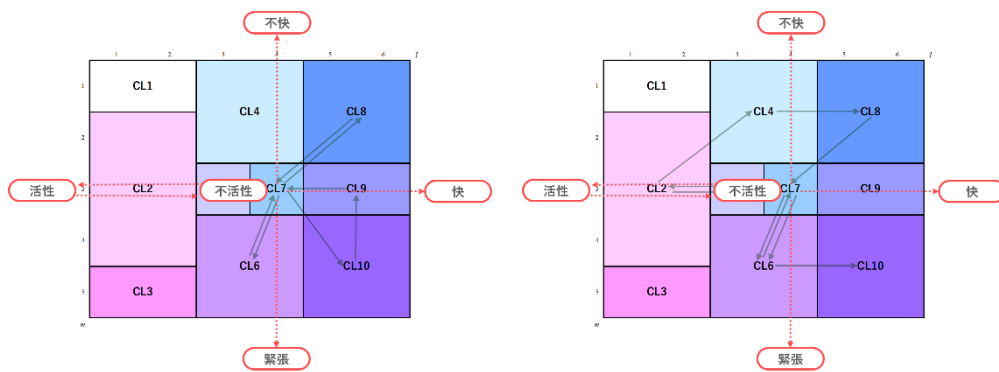


図 4.26 ベクトル G_c が形成するクラスタ A2, A5, A7, A11, A14, A16, A28 ($R = 0.90$)



クラスター B2 ($R = 0.81$)



クラスター A8 ($R = 0.90$)

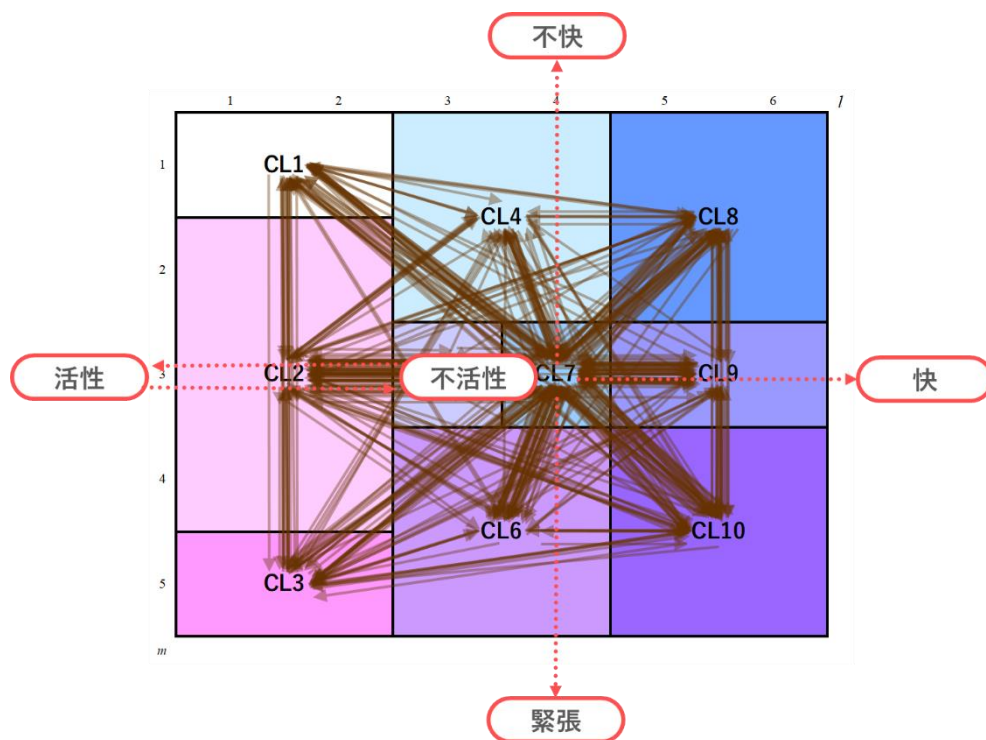
クラスター A15 ($R = 0.90$)

図 4.27 ベクトル G_c が形成するクラスター B2 ($R = 0.81$) およびクラスター A8, A15 ($R = 0.90$)

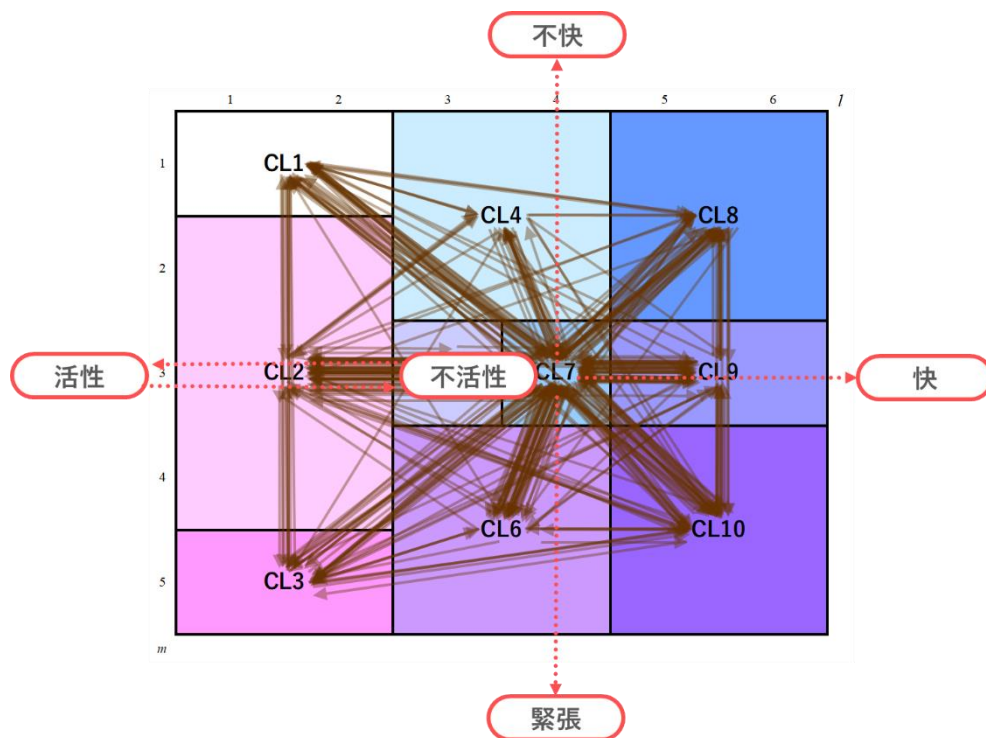
4.4.2. TF-IDF 空間上の遷移を示す TF-IDF ベクトル \mathbf{G}'_c に基づく可視化

ここでは、類似度 $R = 0.70, 0.90$ を基準とした場合の、クラスター C1 と A1 について行った可視化について説明する。作成された可視化の図を以下の図 4.28 に示す。

類似度 $R = 0.70, 81$ を基準とした場合に形成されるクラスター C1 および B1 には、58 名のアスリートが所属する。また、類似度 $R = 0.90$ を基準として形成されるクラスター A1 には、46 名のアスリートが所属しており、いずれも対象者集団の大半が属するクラスターとなっている。図 4.28 からは、クラスター C1 と A1 の間に明確な差異を見出すことは難しく、ベクトル \mathbf{G}'_c の値に基づくクラスタリングでは、特徴的な座標遷移を抽出することは困難といえよう。



クラスタ C1 ($R = 0.70$)



クラスタ A1 ($R = 0.90$)

図 4.28 ベクトル G'_c が形成するクラスタ C1 ($R = 0.70$) およびクラスタ A1 ($R = 0.90$)

4.4.3. TF-DF 空間上の遷移を示す TF-IDF ベクトル \mathbf{I}_c の値に基づく可視化

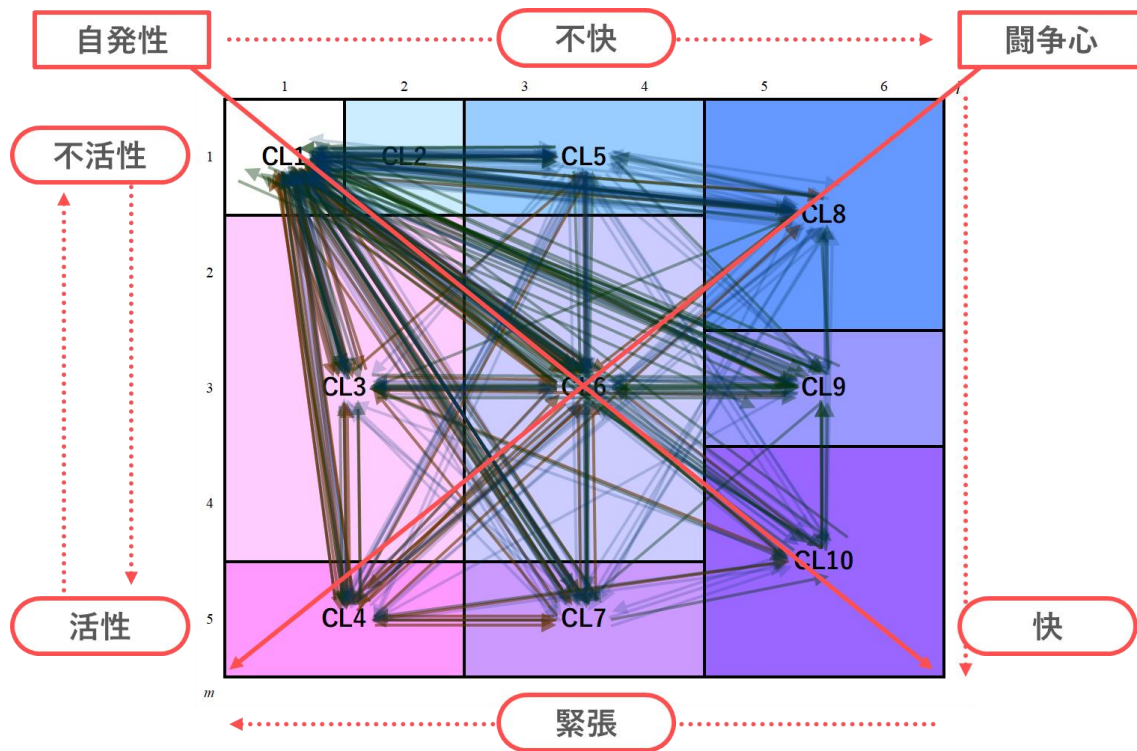
ここでは、類似度 $R = 0.70$ を基準とした場合のクラスタ C1 (図 4.29) と、類似度 $R = 0.81$ を基準とした場合のクラスタ B1, B2, B7 (図 4.30-7.32) のそれぞれについて、可視化を行った。作成された可視化の図を、以下に示す。

類似度 $R = 0.70$ を基準とした場合に形成されるクラスタ C1 には、58 名のアスリートが所属する。これらのアスリートは、類似度 $R = 0.81$ を基準とした場合、主に 3 種類の下位クラスタへと分割される (クラスタ B1, B2, B7)。クラスタ C1 の可視化の図では、マップ全体への座標遷移が示されているのに対し、下位クラスタでは、それぞれ異なる遷移の傾向がみられる。

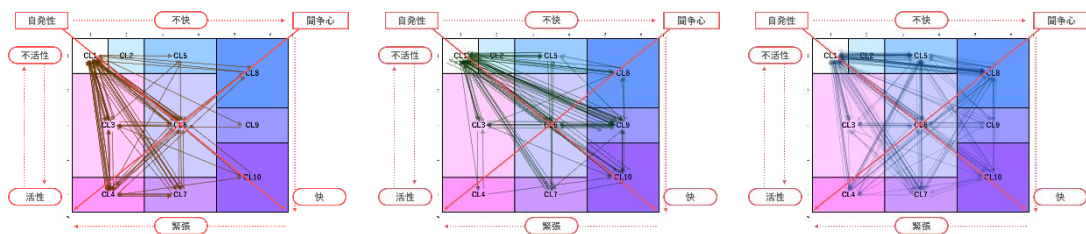
3 種類の下位クラスタのうち、クラスタ B1 では、主に「活性—不活性」の評価軸上での座標遷移がみられた。このような傾向は、インタビュー中の発話における、「目標」や「勝つ」といったキーワードの使用と関係がある。クラスタ B1 に所属するアスリートは、競技における目標達成に意識を向けており、これは、外的な事象および、自らの内面の意識的な側面への関心を示していると考えられる。

クラスタ B2 では、主に領域 CL_1 から CL_9 への座標遷移がみられ、アスリートの自発性と、快感情の増加を示すものと考えられる。こうした傾向は、インタビューにおける「楽しい」、「自信」といったキーワードの使用と関係がある。クラスタ B2 に所属するアスリートは、競技生活の中で経験するポジティブな感情に注目しており、これは特に、自らの内面の無意識的な側面への関心を反映していると思われる。

クラスタ B7 では、主に「不快」の評価軸上での座標遷移がみられた。こうした傾向は、インタビューにおける「難しい」、「悪い」といったキーワードの使用と関係がある。クラスタ B7 に所属するアスリートは、クラスタ B2 に所属するアスリートと同様、自らの内面の無意識的な側面に関心を向けていると考えられる。



クラスタ C1 ($R = 0.70$)



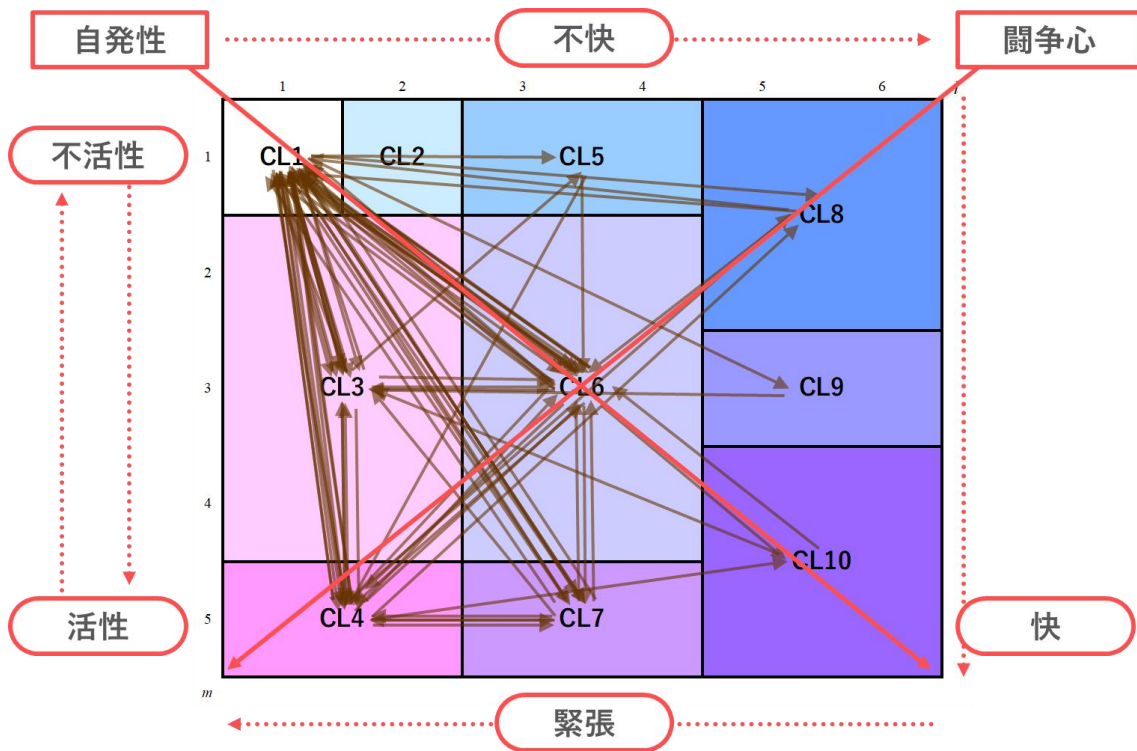
クラスタ B1

クラスタ B2

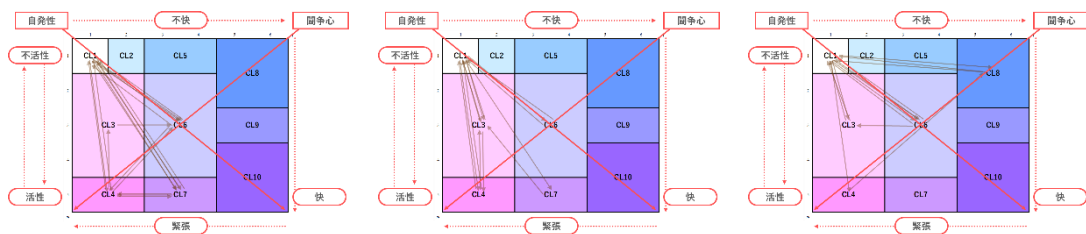
クラスタ B7

($R = 0.81$)

図 4.29 ベクトル I_c が形成するクラスタ C1 ($R = 0.70$) およびクラスタ B1, B2, B7
($R = 0.81$)



クラスタ B1 ($R = 0.81$)



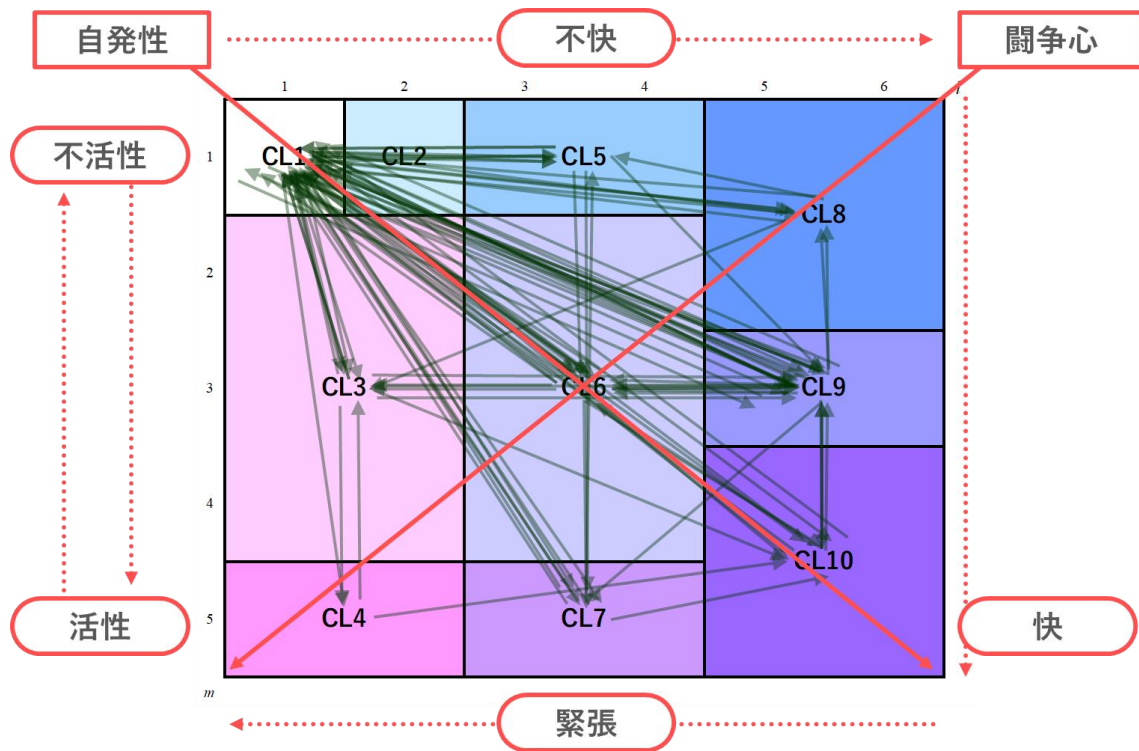
クラスタ A1

クラスタ A4

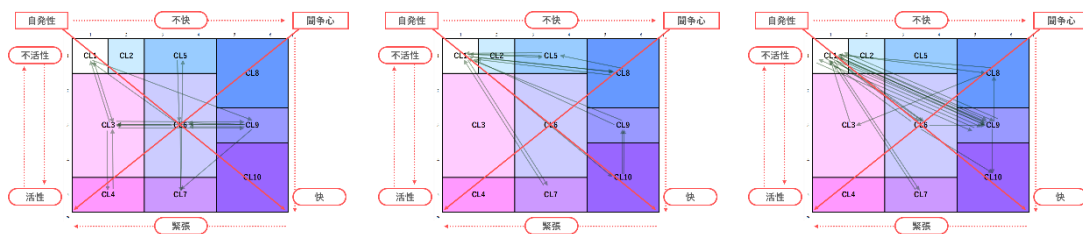
クラスタ A9

($R = 0.90$)

図 4.30 ベクトル I_c が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A1, A4, A9 ($R = 0.90$)



クラスタ B2 ($R = 0.81$)



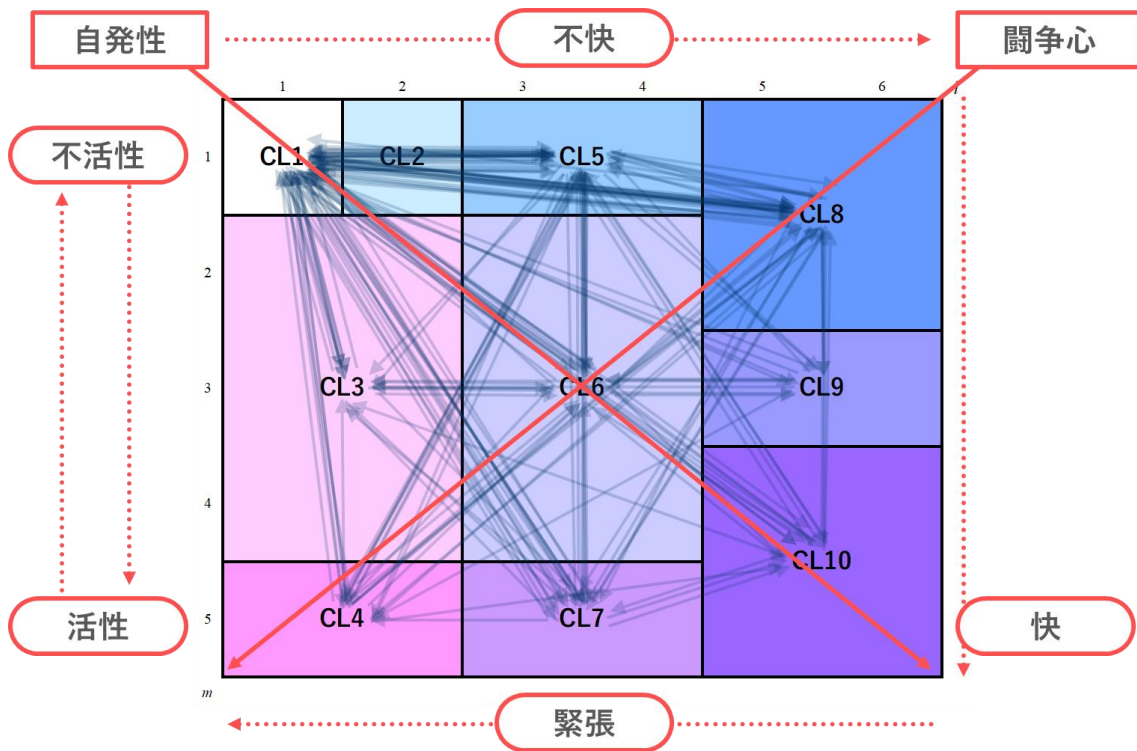
クラスタ A2

クラスタ A10

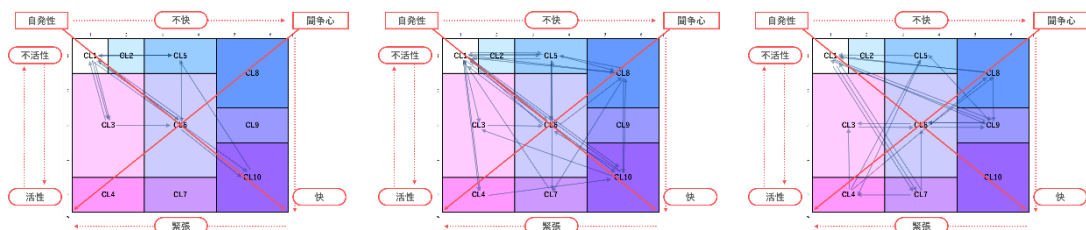
クラスタ A13

($R = 0.90$)

図 4.31 ベクトル I_c が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A2, A10, A13 ($R = 0.90$)



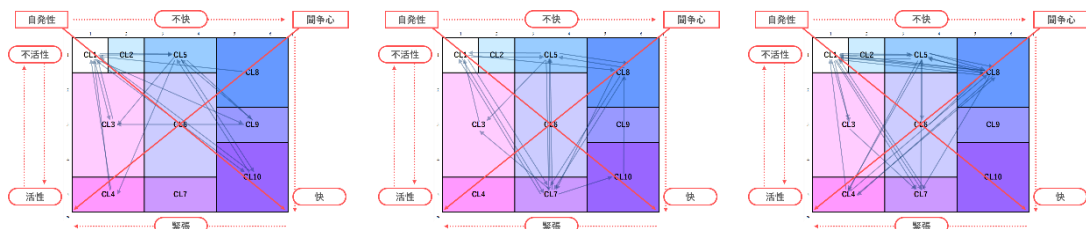
クラスタ B7 ($R = 0.81$)



クラスタ A8

クラスタ A15

クラスタ A19



クラスタ A20

クラスタ A22

クラスタ A24

($R = 0.90$)

図 4.32 ベクトル I_c が形成するクラスタ B1 ($R = 0.81$) およびクラスタ A8, A15, A19, A20, A22, A24 ($R = 0.90$)

図 4.30-図 4.32 を比較すると、類似度 $R = 0.81$ を基準とした場合のクラスタ B1 と B2 では、座標遷移の方向性が明確に示されている。クラスタ B1 では「活性—不活性」の評価軸上、クラスタ B2 では領域 CL_1 から CL_9 へかけての座標遷移が顕著であり、それぞれの下位クラスタにおいても、同様の遷移がみられる。これに対し、クラスタ B7 とその下位クラスタは、「不快」の評価軸上での遷移がみられる点は共通しているものの、その他にも様々な方向への座標遷移を示している。クラスタ B7 にみられる遷移パターンの多様性は、アスリートの不快な感情に付随した、精神の不安定性を反映していると考えられる。

個人の内面を象徴する 7 つの評価軸のうち、「快」および「不快」の軸は、「自発性」の軸上での遷移とも関連している。自発性が、個々人の心理状態における意識的な側面であるのに対し、快および不快は、本能的な情動に近く、無意識的な側面を表現するものである。また、「活性」および「不活性」の軸は、フロイトの提唱した前意識における、心的エネルギーの水準を表していると考えられる。「活気」の評価軸は、ユングの提唱した集合的無意識における、トリックスターの元型を象徴するものととらえられる。このように、言語データの解析から抽出された評価軸と、個々人が示す軸上での座標遷移に注目することで、心理状態の意識的側面のみならず、無意識的側面の理解につながる可能性が示唆された。

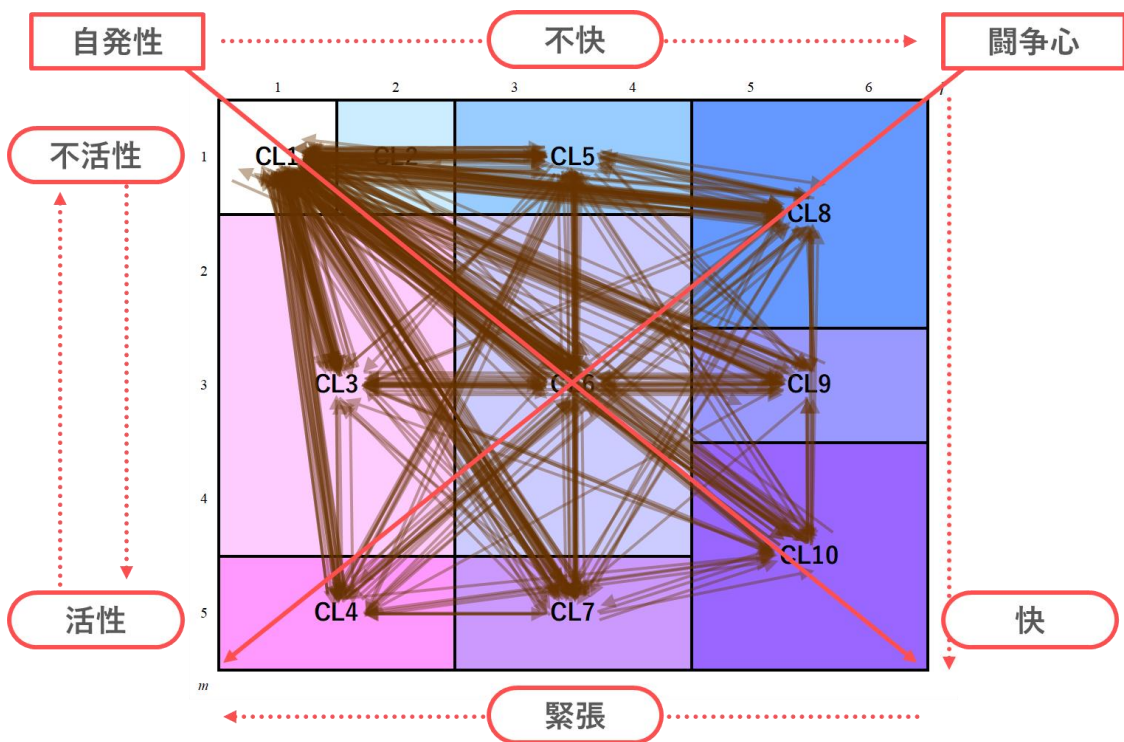
4.4.4. TF-DF 空間上の遷移を示す TF-DF ベクトル \mathbf{I}'_c に基づく可視化

ここでは、類似度 $R = 0.70, 0.90$ を基準とした場合の、クラスタ C1 と A1 について可視化を行った（図 4.33）。

類似度 $R = 0.70$ を基準とした場合に形成されるクラスタ C1 には、58 名のアスリートが所属する。このとき形成されるクラスタ C2、クラスタ C3 には、いずれも 1 名のアスリートが所属するのみである。また、類似度 $R = 0.90$ を基準とした場合に形成されるクラスタ A1 には、52 名のアスリートが所属する。このとき形成されるクラスタ A2-A9 も、それぞれ 1 名のアスリートが所属するのみである。すなわち、クラスタ C1 および A1 には、対象のアスリートの大半が所属する形となり、個々人がマップ上で示す遷移のパターンは様々であるために、可視化結果から特定のパターンを抽出する

ことは困難であった。

第3章で用いた TF-DF 値は、データ間の共通性を強調した重み付けの指標である。SOM の代表ベクトルの値から、マップ上の各領域の性質について考察する際には、TF-DF による重み付けにより、Modal word をバランスよく重み付けることができた。結果として、適切な心理的評価軸の抽出につながったといえよう。これに対し、本章におけるアスリートのクラスタリングでは、ニューロンへの入力ベクトルの分類個数を示す $N_{c,t}$ に対し、TF-DF 法による重み付けを行った。その結果、ベクトル \mathbf{I}'_c の成分 w'_t の偏りを強調することとなり、ベクトル間の差異を見出しにくくなったと考えられる。データ間の差異を強調する TF-IDF 法は、まとまりのあるデータ群を分割する際に、共通性に焦点を当てる TF-DF 法は、多様な特徴をもつデータをクラスタリングする際に用いるなど、分析の目的やデータの性質に応じて、重み付け手法を使い分けた方が効果的であることが本研究の事例で示唆された。



クラスター C1 ($R = 0.70$)

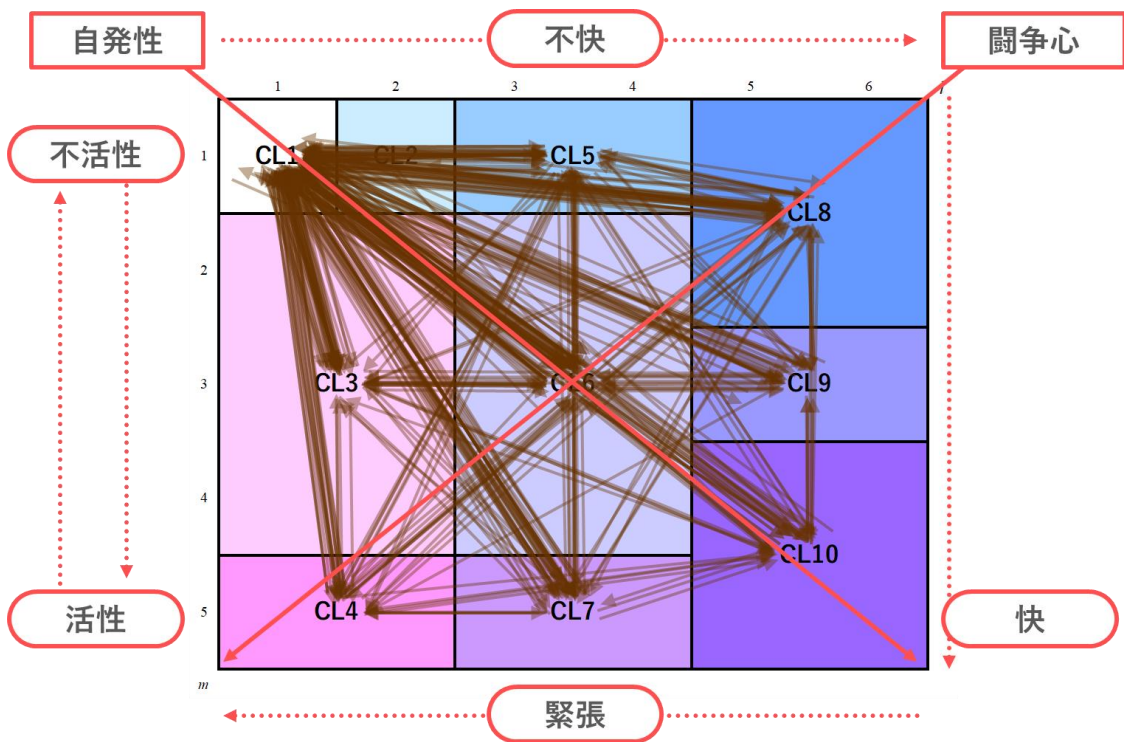


図 4.33 ベクトル I_c が形成するクラスター A1 ($R = 0.90$)

4.5. 結言

前章および本章を通じて提案された手法では、従来の統計解析に代わり、非線形解析の SOM とファジィクラスタ分析の結果を組み合わせることで、二次元平面上への心理状態の評価軸の設定を実現した。本章では、アスリートがマップ上で示す座標遷移から共通のパターンを抽出し、可視化を行った。特に、様々な心理的評価軸上での遷移に注目することで、アスリートの心理変化の傾向について、より深く理解することが可能となった。

言語データの分析および可視化を通じて、対象者の心理における意識的側面と、計り知れない無意識の側面とを統合しうる、心理的評価軸の抽出へとつながる可能性が見出されたといえる。ファジィクラスタ分析により、アスリートの心理状態と変化に基づくクラスタリングがなされたが、これは、個々人が自らの意識、あるいは無意識的な側面に注意を向ける度合いを反映した分類であると解釈できる。例えば、図 4.30 に示されたクラスタ B 1 ($R = 0.81$) の可視化の図からは、アスリートが自らの心理状態と変化の中でも、意識的な側面に焦点を当てていることが読み取れる。これに対し、図 4.31 に示されたクラスタ B 2 ($R = 0.81$) の可視化の図は、快および不快といった、無意識的な側面への関心を示している。このように、言語データに特化した解析と可視化の手法を適用することで、無意識の領域も含めて、個人および集団の心理状態とその変化を多面的にとらえることができた。

第5章. 結論

5.1. 緒言

本研究では、多次元ベクトルとして表され、対象者の心理状態を反映する言語データを、非線形解析により二次元平面上へマッピングすることで、個々人および集団の心理状態、ならびに心理変化の可視化を試みた。研究を通じて提案された手法は、「心理学的トポロジーの導入」、「集団における心理変化のパターン抽出・可視化」という2段階のプロセスからなる。本章では、第3章および第4章で述べた結論をもとに、これら2つのプロセスを包括する総合的な結論を示す。

5.2. 総合結論

第1段階（心理学的トポロジーの導入）では、TF-IDF法で重み付けした入力ベクトル \mathbf{H}_j および、TF-DF法で重み付けした入力ベクトル \mathbf{H}'_j の解析により、心理的評価軸の抽出を行った。得られた心理的評価軸は、対象者の心理変化としての感情を表す「快」および「不快」、「活性」、「不活性」、「緊張」の5種類であった。さらに、入力ベクトル \mathbf{H}'_j を用いた解析では、対象者の心理状態を表す「自発性」および「闘争心」という2種類の評価軸が抽出された。5種類の評価軸は、「3.4 言語データからの Modal word の選定」で設定された、3種類の感情の評価軸（ u : 快—不快, v : 活性—不活性, w : 緊張—弛緩）に対応するものである。これに対し、マップ上の Modal word の分布に注目することで新たに見出されたのが、心理状態を表す2種類の評価軸であった。

個々の Modal word が示す感情の評価軸は、SOMのマップ上において、5種類の評価軸として表現された。さらに、TF-DF値を成分にもつ入力ベクトル \mathbf{H}'_j の解析結果からは、マップ上の Modal word の分布と、感情間の関連性を明瞭に読み取ることができ、新たな評価軸の抽出につながった。このように、提案手法では、言語データに対して多数の心理的評価軸を設定することができる。さらに、個々人のデータがこれらの軸上を遷移する様子を、二次元空間上で視覚的に表現でき、直観的な理解につながるという点に、提案手法の有効性があるといえよう。

第2段階（心理変化のパターン抽出・可視化）では、集団に共通してみられる心理

変化の傾向として、3種類のパターンが明らかにされた。得られた3種類のパターンは、それぞれ、活性の高まりと、快および不快感情の高まりという心理変化を示すものであった。個々の心理変化を特徴づけるキーワードに注目すると、活性の高まりに際しては、発話中に「勝つ」や「目標」といったキーワードが多くみられるようになる。これは、競技成績や勝利といった、スポーツにおける外的事象への関心を示すとともに、対象者の意識的な精神活動を反映していると考えられる。これに対し、快や不快が高まる際には、「嬉しい」、「悪い」といったキーワードが多くみられるようになる。これは、自らの内面への関心を示すと同時に、快・不快という本能的な情動が優勢な状態であり、無意識的な精神活動を反映していると考えられる。このように、遷移ベクトルを用いたクラスタリングにより、対象者の示す座標遷移に共通のパターンを見出すことができた。また、それぞれのパターンについて、意識・無意識の両面から心理学的な解釈を加えることが可能であった。

質問紙法など従来の量的調査は、被験者が明確に意識している心理状態、すなわち意識的な側面を中心に測定するものである。これに対し、提案手法では、言語データの分析・可視化を通じて、心理状態を表すキーワードの分布とともに、多様な評価軸の抽出を可能とした。さらに、評価軸上での遷移を視覚的に表現することで、心理状態に関する多様な情報が得られ、意識・無意識の両面を含む幅広い解釈が可能になったと考えられる。

総合すると、本研究の提案手法により、言語データを用いた心理状態の定量化・可視化を実現することで、従来の心理測定・評価と比較して、心理の多様な性質の理解が促進されたと考えられる。言語データは、対象者の心理状態に関わる多くの情報を含むものの、その性質を定量的に扱うためには、多次元データとしての表現が必要となる。その上で、情報の損失を抑えながら、二次元平面上にデータ間の関連性を示すために、非線形解析手法としてのSOM、ファジィクラスタ分析を組み合わせることで有効であった。さらに、解析結果として得たマップのトポロジー的な性質に注目することで、マップ上での遷移を定量的にとらえ、多様な評価軸を設定することができた。特に、心理状態とその変遷を多軸上で可視化するという手法は、意識・無意識も含めた心理の多様な側面を解釈し、他者に伝達する手段として、適切であったと

いえよう。また、マップ上での座標遷移を遷移ベクトルとして定式化し、クラスタリングを行った結果、集団内で共通した遷移のパターンが見出された。これにより、言語データによる心理状態の定量化・可視化の手法を提案するとともに、個人・集団のそれぞれを対象とした分析について、一定の指針を示すことができた。

心理学的トポロジーの導入、集団における心理変化のパターン抽出・可視化という2段階のプロセスを実行し、結果に対する考察を経て、提案手法の有効性が裏づけられた。提案手法により作成された可視化の図は、従来の心理データの解析結果と比較して、より多くの情報を提供するものであったといえる。効果的な可視化が実現された要因としては、以下の2点が考えられる。

- 1) SOMのマップ上におけるキーワードの分布に、心理学的な意味を見出すことができ、適切な心理的評価軸の設定がなされた。
- 2) 各アスリートの遷移ベクトルを用いて、ファジィクラスタ分析を行った結果、座標遷移に基づくクラスタが形成され、心理変化のパターンが抽出された。

上記の‘1)’には、SOMの入力ベクトルとして用いた値と、代表ベクトルに対するファジィクラスタ分析の実施が関係している。心理学的トポロジーの導入においては、SOMの入力ベクトルの成分として、TF-IDF値およびTF-DF値を利用し、それぞれの解析結果について比較を行ったところ、後者の方がより多様な心理的評価軸を抽出することができた。TF-IDF法は、少数の発話に含まれるキーワードを重み付ける手法であり、発話間の差異を強調する役割を果たす。その結果、TF-IDF値を成分とする入力ベクトルは、多次元空間上で複雑な分布を示し、マッピングの結果から評価軸を見出すことが困難であったと考えられる。これに対し、本研究では独自に定義したTF-DF法は、多くの発話に共通して用いられる語に対して重み付けを行う。その結果、TF-DF値を成分とする入力ベクトルは、主に10語程度のキーワードによって特徴づけられ、発話ごとのまとまりを形成しやすくなった。すなわち、入力ベクトルのマッピング結果が適度に単純化され、キーワードの分布を読み取りやすくなったと推察さ

れる。その上で、ファジィクラスタ分析によりマップ上を複数の領域に分け、各領域の特徴を明確化することにより、多数の評価軸を抽出することができた。

また、上記の‘2)’には、ファジィクラスタ分析に用いた遷移ベクトルの値が関係している。遷移ベクトルは、SOMによる各ニューロンへの入力ベクトルの分類個数に対して、TF-IDF法およびTF-DF法による重み付けを行った値を成分としてもつ。ファジィクラスタ分析の結果、TF-DF値を用いた遷移ベクトルが、ほぼ1つのクラスタに収束したのに対し、TF-IDF値を用いた遷移ベクトルは、3種類のクラスタに分類された。これにより、マップ上での座標遷移のパターンを複数見出し、それぞれの示す心理変化の特徴について、解釈することが可能となった。TF-IDF法により適切なクラスタリングがなされた理由として、SOMに用いた入力ベクトルの多くがゼロベクトルであり、特定のニューロンに集中して分類される傾向にあったことが挙げられる。すなわち、TF-IDF値による重み付けを行わない場合、各アスリートの遷移ベクトル間の距離は小さくなり、座標遷移における差異を見出すことが困難となる。TF-DF値を成分とした遷移ベクトルであれば、それらの差異はより小さくなるであろう。このように、差異を見出しにくいベクトルデータについては、その成分に対してTF-IDF法を適用することで、複数のクラスタへの分類が生じやすくなると考えられる。

以上のように、本研究の提案手法を用いて、より有効な解析および可視化の結果を得るためには、入力ベクトルの算出方法にも配慮する必要がある。本研究で用いたTF-IDF法はデータの特異性を、TF-DF法はデータの共通性を強調するものであり、データの性質に応じて、適切な手法を選択する必要があるといえる。可視化のためのクラスタリングに関して、本研究で定義したTF-DF法は、TF-IDF法以外の選択肢を増やした点で、意義があり、さらに、実際の本研究の事例がその有効性を示す証左といえる。これも本研究の重要な成果の1つである。

言語データからは、特定のキーワードの頻度にとどまらず、様々な分析の指標を得ることができる。心理状態および変化の可視化を行う上では、言語データの解析手法の確立とともに、指標となる値の算出方法についても、さらなる検討を行っていくことが望ましい。

5.3. 今後の研究への提言

本研究を通じて、言語データを用いた心理状態の分析と可視化の手法が提案され、解釈を行う上での基準となる心理変化のパターンが抽出された。今後、提案手法の発展と応用を推進していく上では、以下の課題について検討する必要がある。

- 1) 本研究の解析に用いたデータは、60名のアスリートの発話に対応する、736のベクトルデータである。対象者数が60名と少なく、アスリートに限定されていることから、本研究で得られた心理的評価軸、ならびに心理変化のパターンは、対象者集団の特性に依存するものであるといえる。そこで、より普遍的な評価軸と、心理変化のパターンを抽出することが望まれる。そのためには、対象者の幅を広げ、より多くのデータ収集に努める必要がある。また、異なる特徴をもつ集団に対して、提案手法による分析と可視化を実施し、評価軸と心理変化の特徴を集団間で比較するという方法も考えられる。言語データと分析結果の蓄積を経て、多くの人々に適用可能な知見を得ることが期待される。
- 2) 提案手法による心理状態の可視化結果の妥当性について、検証を行う必要がある。本研究では、Web上から収集したインタビュー記事を使用したため、対象者本人から、言語以外のデータを得ることができなかった。今後は、対象者集団へのインタビューと合わせて、質問紙による心理状態の測定を実施するなど、比較対象となる量的データの収集が不可欠である。
- 3) 提案手法による心理状態の可視化結果の有効性について、利用者を対象とした検証を行う必要がある。臨床心理士などの心理職のほか、本研究のようにアスリートの心理状態に関する可視化を行う場合、アスリート本人や監督、コーチといった関係者にも、結果をフィードバックする可能性がある。このような、可視化結果の利用が想定される人々を対象に、可視化手法と結果に関する評価実験を行い、提案手法の有効性を検証することが課題である。

5.4. 結言

本章では，第3章・第4章で示された提案手法と，分析・考察から得られた結論に基づき，本研究を通じての総合的な結論を示した．また，提案手法のさらなる発展と応用を見据えて，検討が必要となる課題について述べ，今後の研究への提言とした．

謝辞

埼玉大学大学院理工学研究科博士後期課程で研究を行う機会を与えていただき、また、研究全般に渡って手厚いご指導をいただきました。主査の平原裕行教授に、心より感謝申し上げます。また、研究に関して貴重なご指導、ご助言をいただきました。埼玉大学大学院理工学研究科、副査の綿貫啓一教授、副査の島村徹也教授、論文審査委員の小原哲郎教授に、深く感謝致します。そして、研究を進めるにあたり、多くのご協力をいただきました。東洋大学総合情報学部の土田賢省教授、加藤千恵子教授に厚く御礼申し上げます。

参考文献

- あいおいニッセイ同和損保. AD Challenge Support, アスリートインタビュー.
https://www.challenge-support.com/athlete_interviews/
- Alsakran, J., Chen, Y., Zhao, Y., Yang, J., & Luo, D. (2011). Streamit: dynamic visualization and interactive exploration of text streams. *2011 IEEE Pacific Visualization Symposium*, Hong Kong, 131-138.
- Aoki, K., Hirahara, H., Kato, C., & Tsuchida, K. (2018a). A Consideration on Emotional Topology: Verbal Data Processing and Representation Applied to Athlete Statements. *Psychology*, **9**, 876-895.
- Aoki, K., Hirahara, H., Kato, C., & Tsuchida, K. (2018b). A Consideration on Emotional Topology: Representing Psychological Transitions on a Two-Dimensional Map. *Psychology*, **9**, 1917-1934.
- Aoki, K., Uehara, M., Kato, C., & Hirahara, H. (2017). Evaluation of Rugby Players' Psychological-Competitive Ability by Utilizing the Analytic Hierarchy Process. *Open Journal of Social Sciences*, **4**, 103-117.
- Back, M. D., Küfner, A. C. P., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, **21**, 1417-1419.
- Bajpai, R., Poria, S., Ho, D., & Cambria, E. (2017). Developing a concept-level knowledge base for sentiment analysis in Singlish. *CoRR*, abs/1707.04408.
- Cavicchiolo, E., Alivernini, F., & Manganelli, S. (2015). A Mixed Method Study on Teachers' Diaries: Teachers' Narratives and Value-added Patterns. *Procedia - Social and Behavioral Sciences*, **205**, 485-492.
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, **15**, 687-693.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*, New York: Plenum.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, **6**, 169-200.

- Ellis, D. A., & Merdian, H. L. (2015). Thinking Outside the Box: Developing Dynamic Data Visualizations for Psychology with Shiny. *Frontiers in Psychology*, **6**, 1782.
- Friggeri, A., Lambiotte, R., Kosinski, M. & Fleury, E. (2012). Psychological Aspects of Social Communities, *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, Amsterdam, 195-202.
- Hearst, M. A. (1999). Untangling text data mining. *the Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics*, University of Maryland, June, 20-26.
- Heer, J., & Kandel, S. (2012). Interactive analysis of big data. *XRDS: Crossroads, The ACM Magazine for Students - Big Data*, **19**, 50–54.
- 保坂 亨. (2000). “序章 人間行動の理解と面接法”, 心理学マニュアル 面接法. 保坂 亨, 中澤 潤, 大野木裕明 (編著). 北大路書房, 1-8.
- 井波真弓, 齊藤兆古, 堀井清之, 細井尚子, 山縣貴幸, 藤澤延行, 村井祐一, 山田美幸, 熊谷一郎. (2013). 『源氏物語』の可視化. 可視化情報学会誌, **33**, 97-107.
- Irving, J. A., Park-Saltzman, J., Fitzpatrick, M., Dobkin, P, L., Chen, A., & Hutchinson, T. (2014). Experiences of Health Care Professionals Enrolled in Mindfulness-Based Medical Practice: A Grounded Theory Model, *Mindfulness*, **5**, 60-71.
- 磯 博行. (2005). “第 8 章 3 節 情動の回路”, 新・心理学の基礎知識. 中島義明, 繁榎算男, 箱田裕司 (編). 有斐閣, 246-247
- 金 明哲. (2017). R によるデータサイエンス : データ解析の基礎から最新手法まで (第 2 版), 森北出版.
- 金 明哲. (2012). “第 1 章 コーパスとテキストマイニング”, コーパスとテキストマイニング. 石田基広, 金 明哲 (編著). 共立出版, 1-14.
- 那須川哲哉. (2006). テキストマイニングを使う技術／作る技術 : 基礎技術と適用事例から導く本質と活用法. 東京電機大学出版局.
- 日本スポーツ振興センター. アスリート育成パスウェイ, アスリートライフスタイル, インタビュー. <http://pathway.jpnsport.go.jp/lifestyle/interview.html>
- 城 佳子. (2009). 3次元感情状態尺度の作成. 人間科学研究, **30**, 57-66.

- Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., & Ikemura, T. (2001). Analysis of codon usage diversity for bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome, *Gene*, **276**, 89-99.
- 小平朋江, 伊藤武彦, 松上伸丈. (2007). テキストマイニングによるビデオ教材の分析: 精神障害者への偏見低減教育のアカウンタビリティ向上をめざして. マクロ・カウンセリング研究, **6**, 16-31.
- Kohonen, T. (1995). *Self-Organizing Maps, Springer Series in Information Sciences, Vol.30*. Springer, Berlin, Heidelberg, New York.
- 小峰 恒, 絹川博之, 中川裕志. (2002). 単語の文書頻度と文の長さを利用した抄録縮約方式. 情報処理学会研究報告自然言語処理 (NL), 2002(44(2002-NL-149)), 73-80.
- 工藤 拓. (2005). 形態素周辺確率を用いた分かち書きの一般化とその応用. 言語処理学会全国大会.
- Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer*, **30**, 1373-1393.
- Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of research and development*. **1**, 309-317.
- 松村真宏, 三浦麻子. (2009). 人文・社会科学のためのテキストマイニング. 誠信書房.
- Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, **81**, 17-42.
- McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. *Proceedings of AAAI-98 Workshop on Learning for Text Categorization*, 41-48.
- 宮下一博. (1998). “第1章 質問紙作成の基礎”, 心理学マニュアル 質問紙法. 鎌原雅彦, 宮下一博, 大野木裕明, 中澤 潤 (編著). 北大路書房, 10-21.
- 村田真樹, 馬 青. (2012). “第11章 自然言語処理技術と機械学習法を用いたテキストマイニング”, コーパスとテキストマイニング. 石田基広, 金 明哲 (編著). 共

- 立出版, 129-140.
- 中村 真. (2018). “第 3 章 感情理論の展開：基本感情説と次元説, その後の感情理論の展開”, 感情心理学：感情研究の基礎とその展開. 今田純雄, 中村 真, 古満伊里 (共著). 培風館, 59-83.
- 中澤 潤. (1997). “序章 人間行動の理解と観察法”, 心理学マニュアル 観察法. 中澤 潤, 大野木裕明, 南 博文 (編著). 北大路書房, 1-12.
- NI, X., HE, P., XU, W., GONG, Y., ZHU, Q., HUANG, W., & WANG, J. (2017). Research on Cigarettes Customer Needs Importance Algorithm Based on KJ / RAHP / KANO. *MATEC Web of Conferences*, **139**.
- 日本ユニシス. Club Unisys + PLUS, PROFESSIONAL GOLFERS プロフェッショナルの肖像. <http://www.unisys.co.jp/club/golf/golfers.html>
- 大平英樹. (2010). 感情心理学・入門, 有斐閣.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In Plutchik, R. & Kellerman, H. (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*. New York: Academic, 3-33.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**, 1161-1178.
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, **76**, 805-819.
- Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, **18**, 613-620.
- Seale, C., Ziebland, S., & Charteris-Black, J. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science & Medicine*, **62**, 2577-2590.
- Sharma, A., & Dey, S. (2012). Performance Investigation of Feature Selection Methods and

- Sentiment Lexicons for Sentiment Analysis. *International Journal of Computer Applications*, **3**, 975-8887.
- Shinkai, K. (2008). Fuzzy Cluster Analysis and its Evaluation Method. *International Journal of Biomedical Soft Computing and Human Sciences*, **13**, 3-9.
- Shiota, M. N., & Kalat, J. W. (2012). *Emotion* (2nd ed). Belmont, CA: Wadsworth.
- Spärck Jones, K. (1972). A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, **28**, 11-21.
- 菅原久嗣, Alena Neviarouskaya, 石塚 満. (2009). 日本語テキストからの感情抽出. 人工知能学会全国大会論文集.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, **37**, 267-307.
- 武田善行, 梅村恭司. (2001). キーワード抽出を実現する文書頻度分析. 情報処理学会研究報告自然言語処理 (NL). 2001(112(2001-NL-146)), 27-32.
- 東洋大学 . TOYO UNIVERSITY TOP ATHLETES 応援サイト 2016 .
<http://www.toyo.ac.jp/site/toyo-topathletes/>
- Wise, J. A. (1999). The Ecological Approach to Text Visualization. *Journal of the American Society for Information Science*, **50**, 1224-33.
- Wu, Y., Provan, T., Wei, F., Liu, S., & Ma, K.-L. (2011). Semantic-preserving word clouds by seam carving. *Computer Graphics Forum*, **30**, 741-750.
- Wundt, W. (1924). *An introduction to psychology*. London: Allen & Unwin.
- 八ッ塚一郎. (2007). 「ボランティア」と「NPO」の社会的構成プロセスに関する新聞記事分析研究：「助詞分析」の試み. 実験社会心理学研究, **46**, 103-119.

本研究に関する講演論文および講演発表

青木滉一郎, 塩野康徳, 平原裕行, 加藤千恵子, 土田賢省. (2017). テキストデータ解析によるアスリートの心理状態の定量化および可視化. 第 45 回可視化情報シンポジウム講演論文集.

Aoki, K., Shiono, Y., Hirahara, H., & Tsuchida, K. (2017). Extraction and Visualization of Mental Features Based on Athletes' Statements. *Abstract Book of the 14th Asian Symposium on Visualization*, May. 23, Beijing, China, 66-67.

加藤千恵子, 江原康生, 小山田耕二, 平原裕行, 土田賢省, 青木滉一郎. (2016). 心理の可視化 (ワークショップ). 可視化情報全国講演会 (日立 2016).

本研究に関する原著論文

Aoki, K., Hirahara, H., Kato, C., & Tsuchida, K. (2018a). A Consideration on Emotional Topology: Verbal Data Processing and Representation Applied to Athlete Statements. *Psychology*, **9**, 876-895.

Aoki, K., Hirahara, H., Kato, C., & Tsuchida, K. (2018b). A Consideration on Emotional Topology: Representing Psychological Transitions on a Two-Dimensional Map. *Psychology*, **9**, 1917-1934.

付録A. 対話的な可視化システム

A.1. 緒論

本研究では、第3章、第4章で述べた心理学的トポロジーの導入と、心理変化パターンの抽出・可視化に着手する前に、言語データによる心理状態の可視化の対話型のプロトタイプシステムの開発を試みた。多次元データがSOMのマップ上で示す動きを視覚的に体感するとともに、可視化手法の実際的観点からの要件や課題を洗い出し、後の研究に活用することが開発の目的であった。

本章では、対話的な可視化システムの実現に向けて取り組んだ、アプリケーション開発の成果について報告する。これは、第3章、第4章で述べた研究の以前に行われた試みであり、使用したデータおよび解析手法は、前章までのそれとは異なる。ただし、インタビューを通じて得た言語データの分析や、SOMによる解析結果の可視化という部分ではいくつかの共通点もある。システムの開発を通じて得た知見は、後の研究に活かされるとともに、提案手法の応用に関して具体的な指針を示すものであったといえよう。この点を踏まえ、開発されたアプリケーションの機能と特徴に加え、今後のシステム開発における課題について以下に述べる。

A.2. アプリケーションの概要

言語データから読み取れる心理的特徴を数量化し、SOMによる解析を行うことで、集団の心理的特徴や個人の心理変化を、二次元平面上で視覚的に表現できる。このような解析の結果をより効果的に可視化し、心理学的な解釈を行いやすくするために、Javaをベースとした可視化言語 Processing (ver. 2.2.1)を用いてアプリケーションを開発した。本アプリケーションでは、SOMによる解析結果を読み込み、入力ベクトルの分類結果、ならびに代表ベクトルの更新結果を含むマップを出力する。また、個々の対象者に関する入力ベクトルが、SOMのマップ上で示す座標遷移を、アニメーションで表現することができる。さらに、代表ベクトルの注目したい成分を選択的に表示したり、マップ上のニューロンをユーザが任意の配置に置き換えたりといった操作が可能な、対話的な可視化システムを実現している。アプリケーションの機能および実

行の詳細について、以下に示す。

A.3. アプリケーションの入力データ

本アプリケーションの起動時には、入力データとして CSV データを読み込む必要がある。入力データは、SOM の入力ベクトル、ならびに解析結果として、入力ベクトルの分類結果と代表ベクトルの更新結果を含む必要がある（図 A.1）。

本章で用いた入力データは、大学生アスリート 1 名（ラグビー選手）を対象に行ったインタビュー結果から作成したものである。インタビューは、調査者とアスリートの一対一で行われた。アスリートには、自身が出場した試合の映像（前半の 40 分間から 20 のシーンを抽出したもの）を観ながら、試合中に感じていたこと、考えていたことについて、自由にコメントしてもらった。録音した音声をもとに、アスリートのコメントは文字化され、20 のシーンごとに分割された。また、各アスリートのコメントのうち、心理状態を示すキーワード（“頑張る”、“緊張”、“プレッシャー”、“怖い”）の出現頻度を、シーン別に算出した。さらに、個々のキーワードの TF-IDF 値を算出し、これらの値を成分とするベクトルデータを、入力ベクトルとして用いた。

上記の手順により求めた入力ベクトルに対して、統計解析用ソフトウェア R の SOM パッケージによる解析を実施し、入力ベクトルの分類および代表ベクトルの更新に関する結果を得た。これらの解析結果を、図 A.1 のような CSV データの形式にまとめることで、アプリケーション側からの読み込みが可能となる。

マップの縦軸・横軸のニューロン数 (5×5=25)

5	5			-0.40471	-0.32489	-0.41205	-0.46591
4	代表ベクトルの成分数			-0.40471	-0.32489	-0.41205	-0.46591
0.147884	1.00001	1.000000	1.10573	0.40471	0.32489	0.41205	0.46591
更新された代表ベクトルの各成分				可視化を行う入力ベクトルの各成分			
0.843226	-1.4719	0.507409	0.006462	-0.40471	-0.32489	-0.41205	-0.46591
1.208617	-1.55032	-0.19851	-0.68081	-0.40471	-0.32489	-0.41205	-0.46591
1.601281	-1.65457	-0.92737	-1.39881	-0.40471	-0.32489	-0.41205	-0.46591
-0.29706	-0.59869	1.665653	1.379471	2.832995	-0.32489	-0.41205	0.308428
0.047908	-0.64455	0.923708	0.664991	-0.40471	2.924038	-0.41205	-0.46591
0.386456	-0.70111	0.228101	-0.00752	-0.40471	-0.32489	-0.41205	-0.46591
0.738775	-0.77379	-0.45745	-0.67451	-0.40471	-0.32489	0.373273	3.176804
1.119281	-0.86284	-1.17757	-1.37966	-0.40471	-0.32489	1.875551	-0.46591
-0.72601	0.156003	1.376269	1.355241	2.832995	-0.32489	-0.41205	-0.46591
-0.37429	0.089112	0.648834	0.646673	-0.40471	-0.32489	3.435219	2.020453
-0.0388	0.024629	-0.03286	-0.01737	1	1	4	1.736101
0.302562	-0.03921	-0.71141	-0.67658	2	2	2	0.77526
0.668776	-0.10577	-1.43415	-1.37764	マップ上における各入力ベクトルの分類座標			
-1.16186	0.913435	1.115327	1.356485	2	2	2	0.77526
-0.79693	0.826674	0.393335	0.646615	2	2	2	1.159213
-0.45573	0.754208	-0.29004	-0.02267	2	2	2	0.77526
-0.1185	0.697841	-0.97802	-0.68957	2	2	2	0.711325
0.23569	0.651816	-1.71352	-1.39666	7	2	2	0.77526
-1.62365	1.704282	0.867643	1.377957	8	2	2	0.77526
-1.24479	1.605561	0.1402	0.659728	9	3	2	0.475446
-0.89268	1.528446	-0.55952	-0.02624	10	3	2	0.475446
-0.55165	1.475969	-1.26935	-0.71274	11	2	2	0.77526
-0.20217	1.441852	-2.02338	-1.43377	12	2	2	0.77526
20	可視化を行う入力ベクトル数			13	2	2	0.77526
-0.40471	-0.32489	-0.41205	-1.471479	14	3	0	2.272545
可視化を行う入力ベクトルの各成分				15	2	4	1.549466
-0.40471	-0.32489	-0.41205	-0.46591	16	2	2	0.77526
-0.40471	-0.32489	-0.41205	0.308428	17	0	2	2.158377
				18	1	1	1.578601
				19	3	0	2.057196
				20	0	1	1.904934

図 A.1 入力データの形式

A.4. アプリケーションの実行画面

開発統合環境 Processing 上でアプリケーションを実行すると、コントロールパネル (図 A.2) および SOM の解析結果を示すマップ (図 A.3) が表示される。コントロールパネル上の“Read File”ボタンを押すことで、入力データとして使用する CSV データを選択し、読み込むことができる。

本章における SOM では、入力データを 5×5 の出力層へマッピングしているため、アプリケーション側も同様に、 5×5 のマップを結果として表示するようにしている。図 A.3 は、データを読み込む以前の初期状態のマップであるが、それぞれのニューロンには 1 から 25 の番号が ID として記載され、“n=0”の表示は、ニューロンに割り当てられた入力ベクトルの個数が 0 であることを示している。

“Read File”ボタンにより入力データを読み込むと、図 A.4 のように、いくつかのニューロンには数値が記載される。これらの数値は、各ニューロンに割り当てられた入力ベクトルの番号を示すものである。

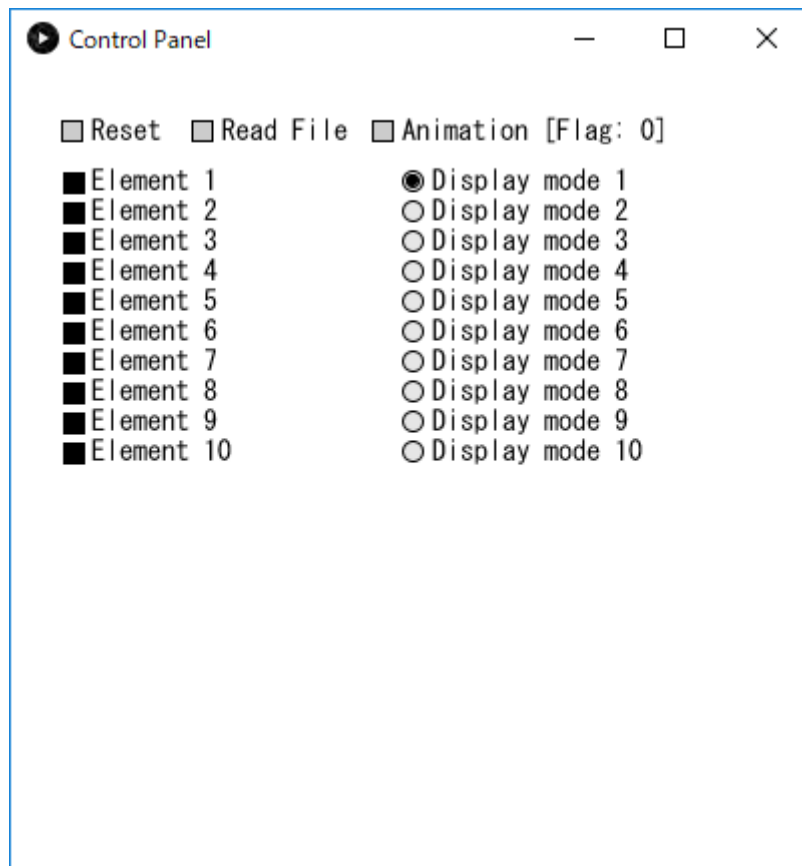
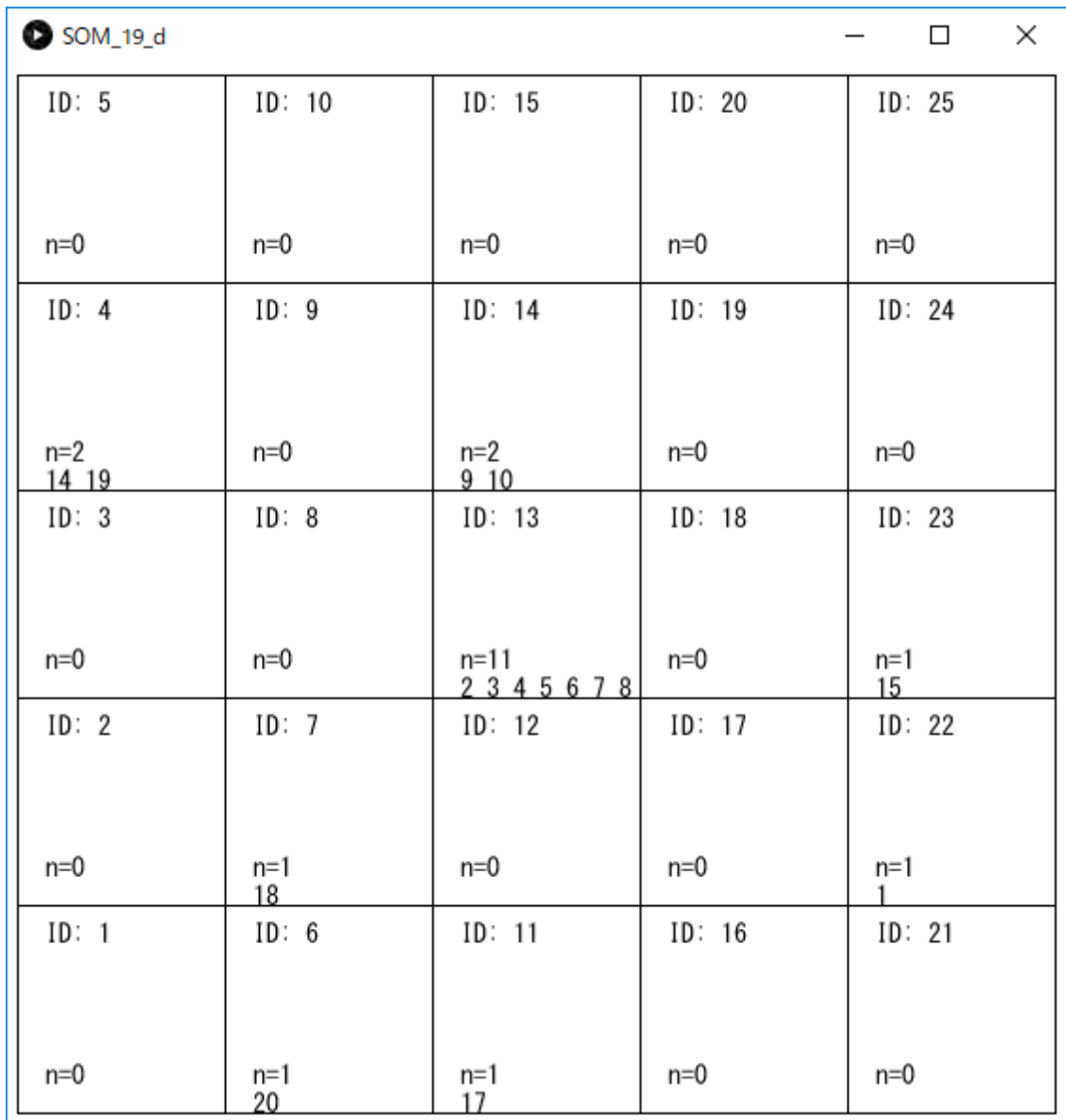


図 A.2 コントロールパネル

ID: 2 n=0	ID: 2 n=0	ID: 2 n=0	ID: 2 n=0	ID: 25 n=0
ID: 1 n=0	ID: 1 n=0	ID: 1 n=0	ID: 1 n=0	ID: 20 n=0
ID: 1 n=0	ID: 1 n=0	ID: 1 n=0	ID: 1 n=0	ID: 15 n=0
ID: 6 n=0	ID: 7 n=0	ID: 8 n=0	ID: 9 n=0	ID: 10 n=0
ID: 1 n=0	ID: 2 n=0	ID: 3 n=0	ID: 4 n=0	ID: 5 n=0

図 A.3 マップ



ID: 5	ID: 10	ID: 15	ID: 20	ID: 25
n=0	n=0	n=0	n=0	n=0
ID: 4 n=2 14 19	ID: 9 n=0	ID: 14 n=2 9 10	ID: 19 n=0	ID: 24 n=0
ID: 3 n=0	ID: 8 n=0	ID: 13 n=11 2 3 4 5 6 7 8	ID: 18 n=0	ID: 23 n=1 15
ID: 2 n=0	ID: 7 n=1 18	ID: 12 n=0	ID: 17 n=0	ID: 22 n=1 1
ID: 1 n=0	ID: 6 n=1 20	ID: 11 n=1 17	ID: 16 n=0	ID: 21 n=0

図 A.4 データ読み込み後のマップ

データの読み込みが完了すると、コントロールパネルを使用して、SOMの解析結果の表示方法を選択することができる。図 A.5 の“Element 1”～“Element 10”は、解析結果として表示する代表ベクトルの成分を示しており、本章では入力ベクトルの成分数 (=4) に合わせて、“Element 1”～“Element 4”を選択する。これらは、アスリートのコメントにみられる各キーワード (“頑張る”, “緊張”, “プレッシャー”, “怖い”) の TF-IDF 値に対応する。このとき、代表ベクトル中の特定の成分のみを選んで表示することも可能である。

また、解析結果の表示手法についても、複数のモードから選択可能とした。ここでは、既にアプリケーションに実装されている“Display mode 1”～“Display mode 1”について述べる。

まず、Display mode 1 (図 A.6) では、それぞれのニューロンがもつ代表ベクトルの成分を数値の形式で表示する。ニューロン内に記載された数値は、上から順に、Element 1, 2, 3, 4 の値に相当する。

次に、Display mode 2 (図 A.7) では、代表ベクトルの各成分の値が、円グラフを構成する扇の大きさによって示される。代表ベクトルの成分数が4であることから、円グラフは四分割され、それぞれの扇は右上から順に、Element 1, 2, 3, 4 の大きさを表している。

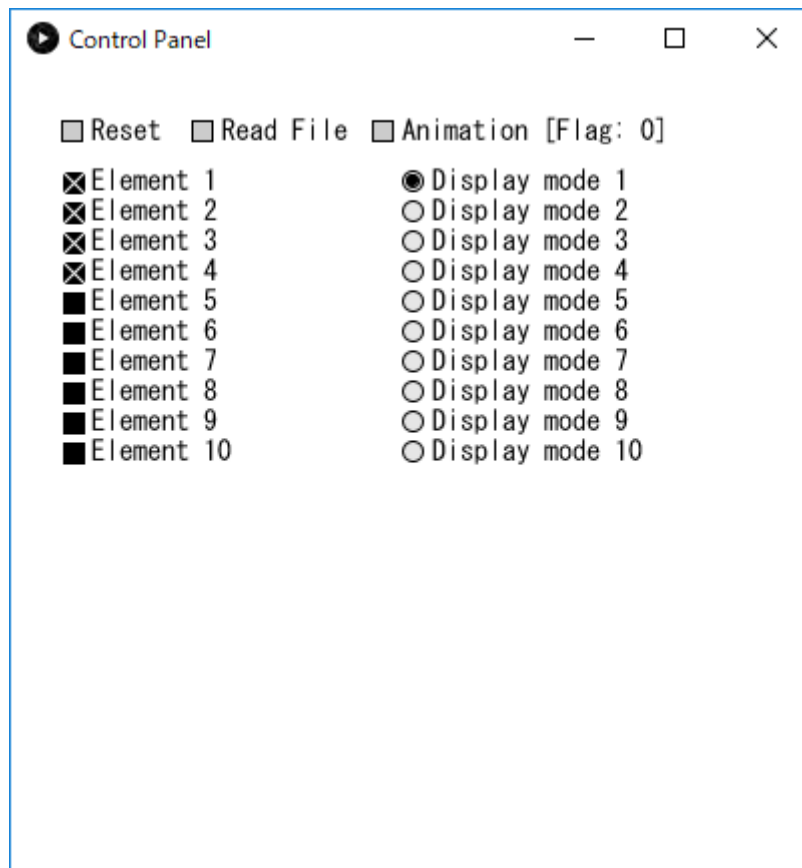


図 A.5 マップ上での表示成分の選択

SOM_19_d				
ID: 5 1. 60128061 -1. 65456524 -0. 92736787 -1. 398811003 n=0	ID: 10 1. 11928122 -0. 86284104 -1. 17756684 -1. 379663933 n=0	ID: 15 0. 66877577 -0. 105771 -1. 4341526 -1. 37763981 n=0	ID: 20 0. 23569 0. 65181617 -1. 71352483 -1. 396662896 n=0	ID: 25 -0. 20216973 1. 44185159 -2. 02337966 -1. 433768087 n=0
ID: 4 1. 20861723 -1. 55032004 -0. 19850812 -0. 680806862 n=2 14 19	ID: 9 0. 73877495 -0. 77379444 -0. 45744841 -0. 674513295 n=0	ID: 14 0. 30256216 -0. 03920927 -0. 7114079 -0. 676575762 n=2 9 10	ID: 19 -0. 11850377 0. 69784053 -0. 97801941 -0. 689567555 n=0	ID: 24 -0. 5516516 1. 47596925 -1. 26935416 -0. 712736374 n=0
ID: 3 0. 84322588 -1. 47189713 0. 50740909 0. 006462139 n=0	ID: 8 0. 38645643 -0. 70111002 0. 22810061 -0. 007515148 n=0	ID: 13 -0. 03880086 0. 02462907 -0. 03285838 -0. 017367255 n=11 2 3 4 5 6 7 8	ID: 18 -0. 45572949 0. 75420815 -0. 29003922 -0. 022674186 n=0	ID: 23 -0. 89267979 1. 52844608 -0. 55951716 -0. 026238429 n=1 15
ID: 2 0. 49482544 -1. 42027149 1. 22706467 0. 698711463 n=0	ID: 7 0. 04790847 -0. 64455057 0. 9237077 0. 664990885 n=1 18	ID: 12 -0. 37429252 0. 08911178 0. 64883439 0. 646672903 n=0	ID: 17 -0. 79693186 0. 82667394 0. 39333539 0. 646614603 n=0	ID: 22 -1. 24478554 1. 60556084 0. 14019983 0. 659728288 n=1 1
ID: 1 0. 14769381 -1. 38809677 1. 98836925 1. 425729834 n=0	ID: 6 -0. 2970564 -0. 59868662 1. 66565337 1. 379471145 n=1 20	ID: 11 -0. 72601471 0. 15600301 1. 37626915 1. 355240919 n=1 17	ID: 16 -1. 1618579 0. 91343528 1. 11532712 1. 356484515 n=0	ID: 21 -1. 62364675 1. 70428218 0. 86764337 1. 377956582 n=0

図 A.6 SOM の解析結果 (Display mode 1)

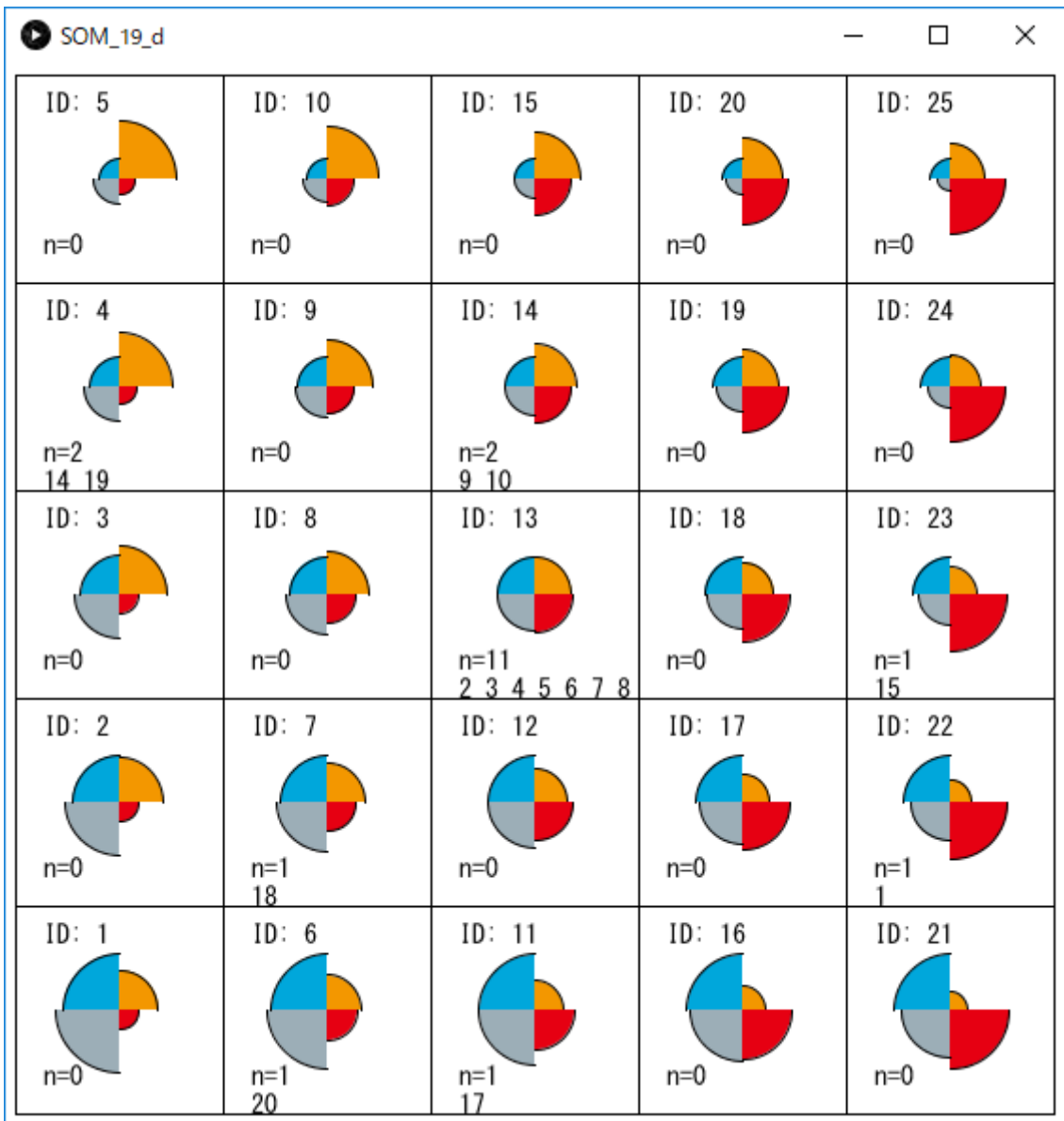


図 A.7 SOM の解析結果 (Display mode 2)

Display mode 2 において、選択する Element の数を 4 未満 (図 A.8, ここでは Element 1-3 を選択する) とすると、解析結果は図 A.9 のように、円グラフの一部が欠けた状態で表示される。

これに対し、Display mode 3 では、選択した Element の数に応じて円グラフが分割される。Element 1-3 を選択した場合、図 A.10 のようにグラフは三分割され、それぞれの扇が個々の Element の大きさを表す。

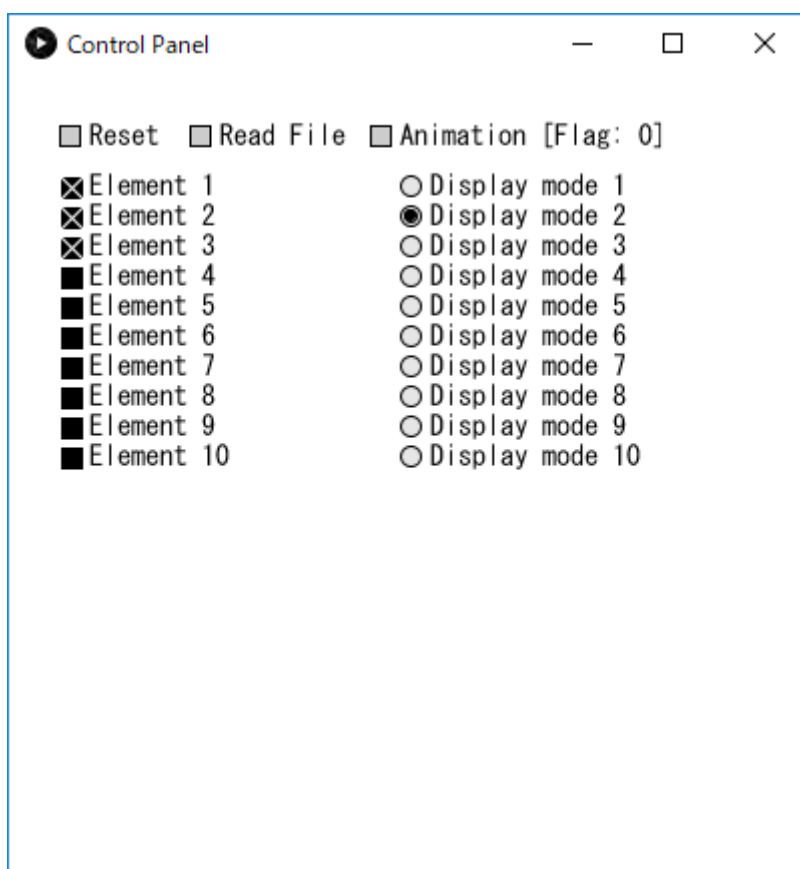


図 A.8 Element の選択的表示

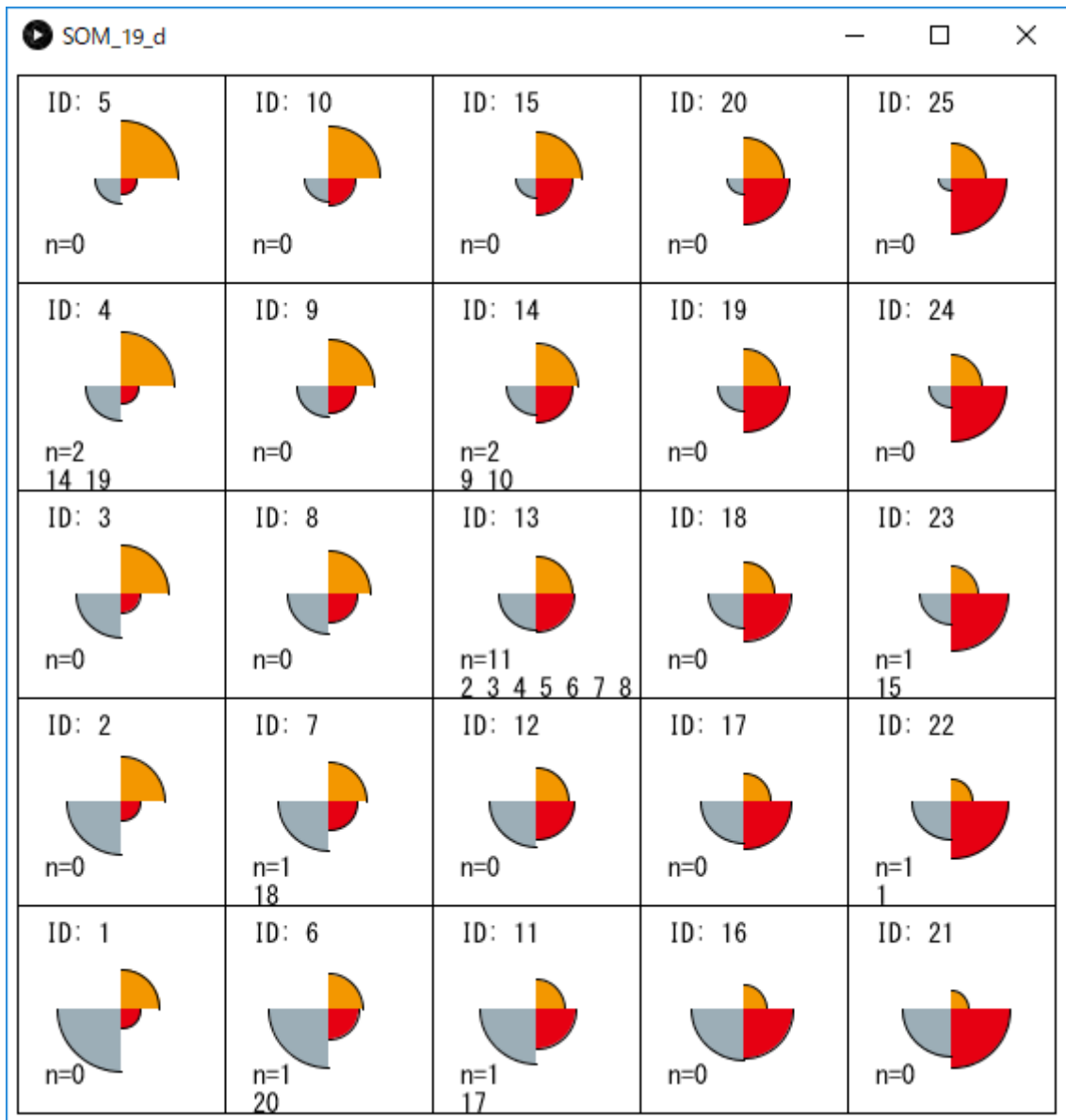


図 A.9 SOM の解析結果 (Display mode 2, Element 1-3 を選択)

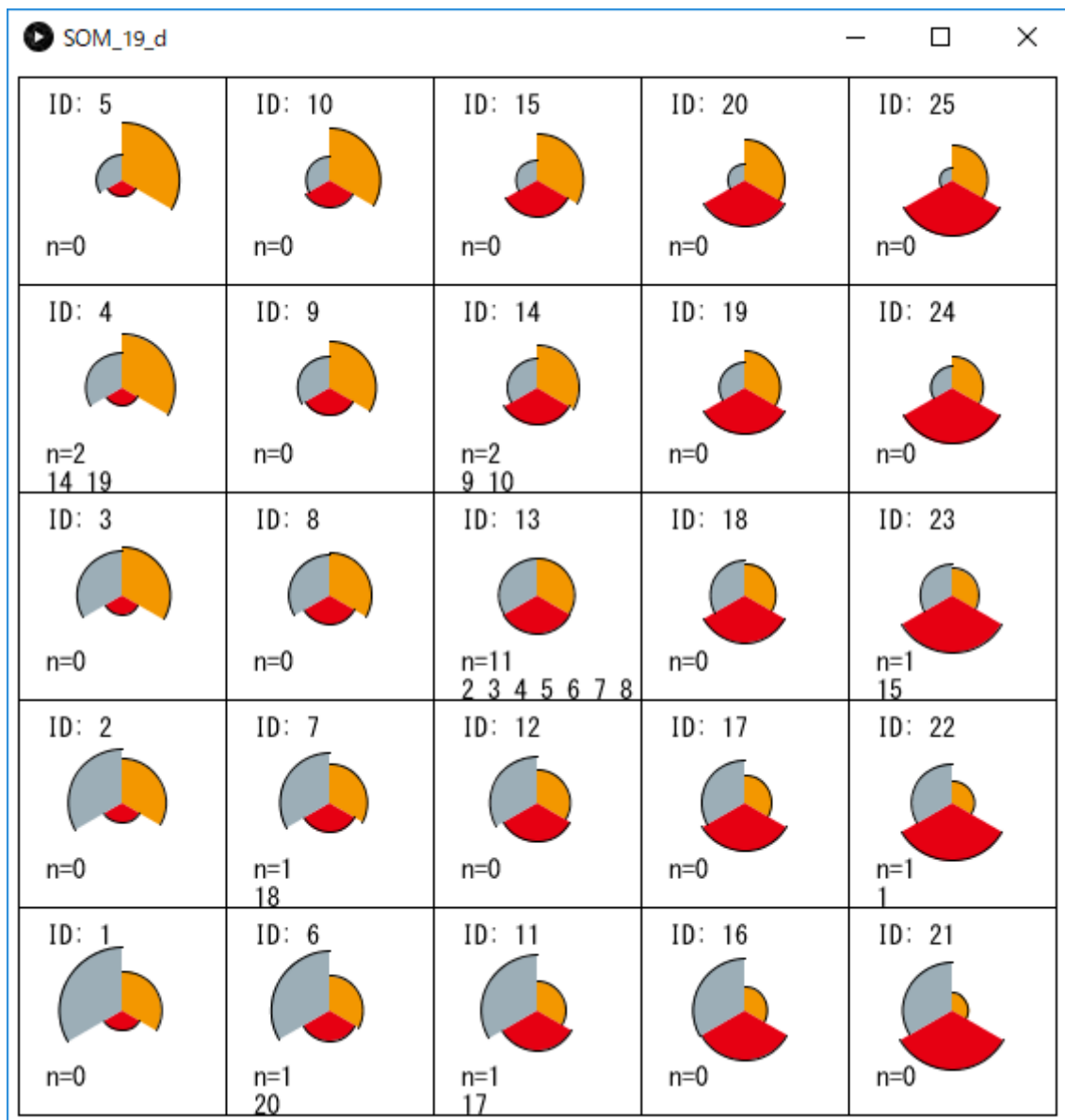


図 A.10 SOM の解析結果 (Display mode 3)

コントロールパネルにおいて，“Animation”のボタンを押すと，表示されているマップ上に，水色の球体がマップ上を動くアニメーションが再生される（図 A.11）。球体は，個々の入力ベクトルがマップ上で割り当てられたニューロンを，入力ベクトルの番号に従って移動する．すなわち，Animation の機能を利用することで，入力ベクトルのマップ上での座標遷移を，動的に表現することができる．座標遷移を表す静的な図と比較すると，遷移の特徴をより強調した可視化手法として，このようなアニメーション表現の活用が期待される．

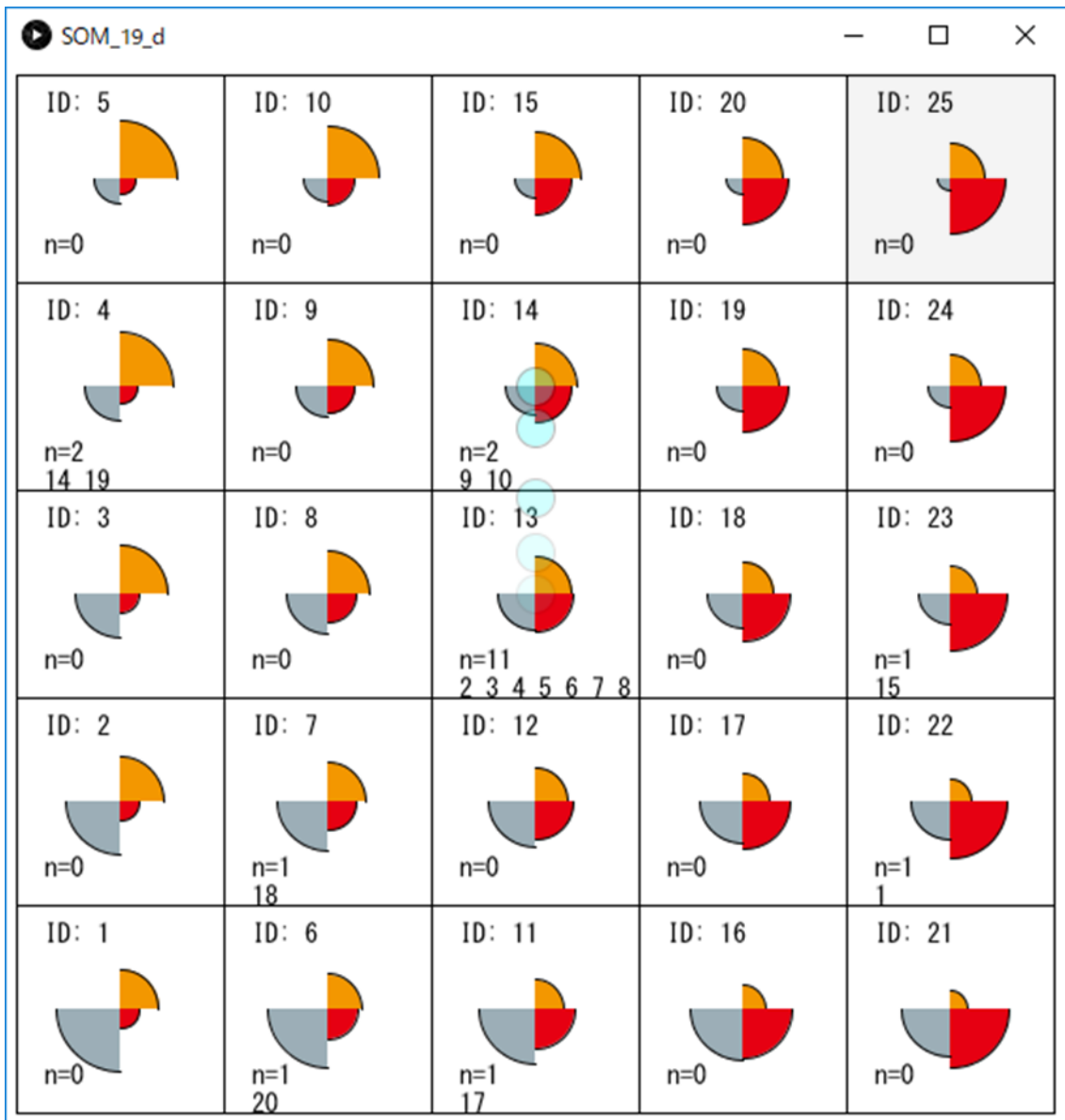


図 A.11 マップ上での座標遷移を示すアニメーションのイメージ

また、本アプリケーションでは、マップ上の任意のニューロンを選択し、配置を自由に移動させることができる（図 A.12）、この機能の実現は Processing のオブジェクト指向性の特徴を活かしたものである。このとき、前述の Animation の機能を使用すると、球体は配置の変化したニューロンの間を、入力ベクトルの順番に従って移動する。コントロールパネルの“Reset”のボタンを押すことで、ニューロンは初期の配置へに戻る。

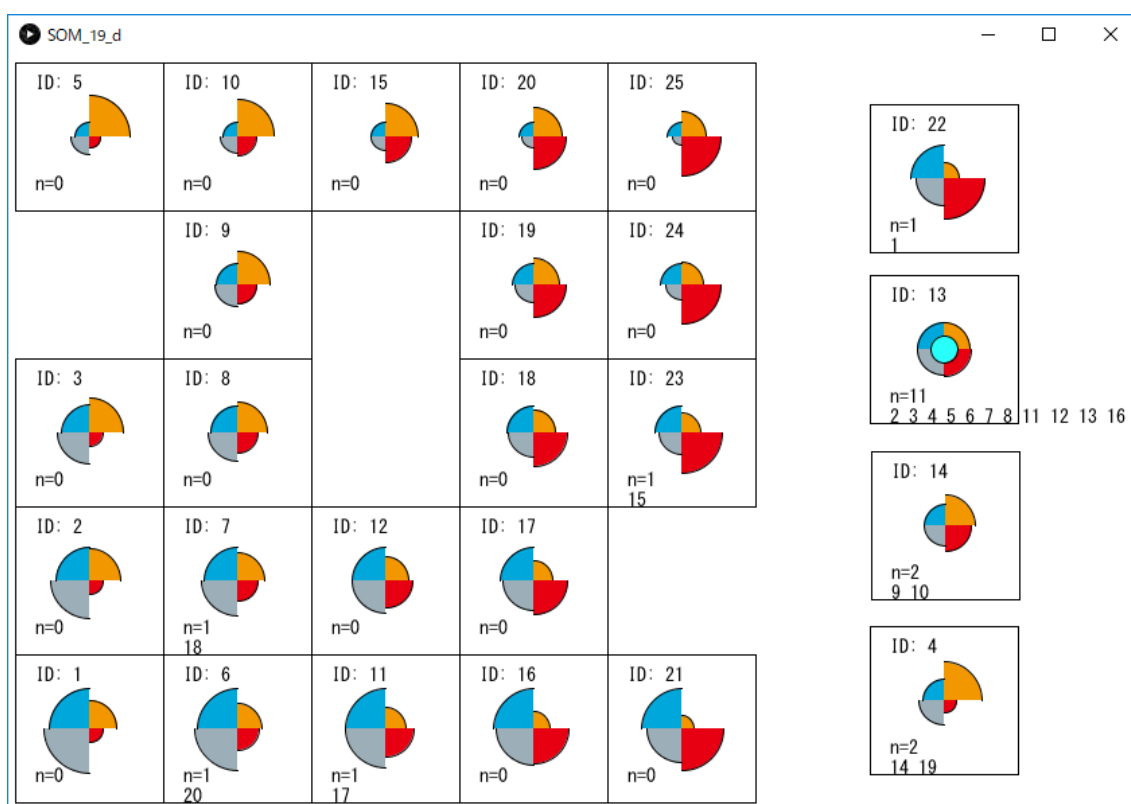


図 A.12 マップ上のニューロンの移動

A.5. 結論

本研究で提案された分析および可視化手法のシステム化を目指し、言語データの解析結果を視覚的に表示するアプリケーションを開発した。本アプリケーションの特徴として、動的な可視化表現を実現した点と、ユーザの視点に応じた対話的な操作を可能にした点が挙げられる。第6章および第7章で提示された可視化の結果は、いずれも静的な画像により、マップ上での入力ベクトルの座標遷移を表現したものであった。

これに対し、アプリケーションを利用して、マップ上での遷移をアニメーションの形式で表すことにより、対象者個々人の示す遷移の特徴が強調され、心理的な理解を促進することが期待できる。さらに、マップ上の配置を変えられるようにしたことで、特定のニューロンを抽出し、代表ベクトルの性質について考察したり、ニューロン間での移動の特徴を観察したりすることが可能となった。代表ベクトルのうち、特定の成分のみを表示する選択機能と組み合わせて、分析の目的に応じた可視化をもたらす機能であるといえよう。以上より、提案手法をアプリケーションに適用することで、より効果的かつ、分析者の目的に合致した可視化表現を実現できることが示唆された。

本章で報告した、アプリケーション開発の課題および改善案として、以下の3点が挙げられた。

- 1) 解析に用いた SOM は逐次学習型であり、また、代表ベクトルの初期値をランダムに設定している。そのため、入力ベクトルの内容が同じであっても、解析を行うたびに、マッピングの結果が変化する。解析結果の再現性を保証するために、解析を繰り返しても結果が変わらないような、一括学習型の SOM を用いることが適切である。
- 2) SOM により得られるマップは、直交座標系とは異なり、入力データの座標に意味を与える評価軸が存在しない。マップ上の各領域の性質を明らかにし、二次元平面上に評価軸を見出すために、代表ベクトルの初期値の設定方法に一定の基準を与える、マップ上を構成するニューロンのクラスタリングを行うといった対策が考えられる。
- 3) 開発したアプリケーションでは、集団の示す座標遷移を同時に可視化し、共通のパターンを見出すことが難しい。集団における心理的特徴や、心理変化の傾向を明らかにするためには、異なる可視化手法が必要となる。

これらの課題および改善案を踏まえて、心理学的トポロジーの導入、ならびに集団の心理変化パターンの抽出・可視化に関する研究を行い、言語データの分析と可視化の手法を提案するに至った。今後は、ファジィクラスタ分析の結果に基づいてニューロンを統合したり、個人や集団が示す座標遷移の軌跡をマップ上に表したりするなど、提案手法に沿った形で、アプリケーションの機能を拡張していくことが望まれる。