

博士論文 2015 年度(平成 27 年度)

インターネット環境における
探索的社会調査ツールの開発と実査

慶應義塾大学大学院 政策・メディア研究科

伊藤貴一

博士論文要旨 2015 年度(平成 27 年度) インターネット環境における 探索的社会調査ツールの開発と実査

本研究の目的は、ソーシャルネットワークを始めとするさまざまなインターネット上のデータを分析対象とし、それらデータから構成されるネット社会を分析するために多様なツールを開発し、かつそのツールを活用した 3 タイプの社会調査を実施することで、まったく新しい探索的な社会調査手法の可能性を提示することである。本研究は、分析者がデータと対峙しながら試行錯誤を通じて探索的に知識を発見することを目的としているため、新たな AI の分野といわれる知的インタラクティブシステムの研究と方向性を共有し、かつそれを社会調査のツール開発と実査という研究レベルに特定化したものである。

ツールは、データの取得から分析結果の可視化まで、総合的なプラットフォームとして開発されている。データ取得ツールは、インターネット上のデータを取得するスクレイピングツール「Rawler」、および画像イメージの調査アンケートサイト「GoocaBooca」の 2 つである。分析ツールは 3 タイプを開発した。第 1 は、インタラクティブな探索的テキストマイニングツール「ひっぱるくん」で、大量でかつアモルフなデータ群から多様な関係性を探索して、いわば暗黙知を新しい知識の地平に表出化することを可能にする。第 2 は、半教つきクラスタリングに基づいてデータを構造化する「こうぞうくん」で、全体（クラスタ間の差異性）と部分（クラスタ内の同一性）のバランスを探索的に調整して、最適な構造を確定する。第 3 は、構造化されたデータについてマイニングを通じてデータの解釈をする「ふかぼりくん」で、既存の構造を多様な条件指定（絞り込みや統合・削除など）により深層化を図り、隠された下位構造を炙り出すツールである。

ツール活用の実査段階では、探索的な社会調査ツールが分析者にたいして新しい創発性（閃きや気づき）をもたらす、という価値探索機能があることが判明した。それは、つぎの 6 つの対抗的で補完的なフレーム、すなわち「自由と秩序」「価値表明と目的達成」「自明性と創発性」「階層とネットワーク」「目的と手段」「一般と特殊」というフレームの中で見出される機能である。このフレームこそ、3 つの分析ツールについて、程度の差こそあれ共通してみられた価値探索機能を正当化する調査設計プラットフォームなのである。

キーワード：探索的社会調査、ツール開発（ひっぱるくん、こうぞうくん、ふかぼりくん）、価値探索機能（表出化、構造化、深層化）、対抗的相補的フレーム、インターネット環境

慶應義塾大学大学院 政策・メディア研究科
伊藤貴一

Abstract of Doctoral Thesis Academic Year 2015

Development and Inspection of Exploratory Social Survey Tools in Internet Environments

The objective of this study is to develop a variety of tools for analyzing various Internet data, such as the data found in social networks, in order to analyze the network society that this data constitutes. By conducting three types of social surveys using these tools, we will propose the possibility of a completely new method for exploratory social surveys. As the objective of this study is for the analyst to discover exploratory knowledge through trial and error while confronting data, it shares its direction with Intelligent Interactive Systems research, which is said to be a new artificial intelligence (AI) field, and has been specified as being at the research level in regard to research and inspection of a social survey tool.

These tools have been developed as a comprehensive platform, handling the acquisition of data to the visualization of analysis results. The two data acquisition tools are the Web scraping tool “Rawler,” used to acquire data from the Internet, and the image survey questionnaire site “GooCaBooca.” Three types of analysis tools have been developed. The first enables one to search for diverse relationships from groups of large-scale amorphous data, that is, for so-called tacit knowledge to be externalized. This is done by using the interactive, exploratory text mining tool known as “Hipparu kun”. Second, with “Kouzou-kun” structured with data based on semi-supervised classification, the balance between the total (disparity between clusters) and part (similarity within a cluster) can be adjusted in an exploratory way and the optimal structure can be determined. Third, using “Fukabori-kun,” which interprets data by mining structured data, deep stratification can be provided through specification of the existing structure in diverse conditions (filtering, integration, and deletion), thus revealing hidden lower level structures.

In the fieldwork stage of tool utilization, it was judged that the exploratory social survey tool had a value search function that brought about new creativity (inspiration and awareness) in the analysts. In other words, this is the function discovered within the following six opposing but complementary frames, i.e., “freedom and order,” “expressions of value and achievement of objectives,” “triviality and creativity,” “hierarchy and networks,” “objective and means,” and “general and particular”. This frame normalizes the value search function seen as being shared by the three analysis tools and serves as the design concept for the exploratory social survey tool.

Keyword: Social Survey, Tool Development ,Internet Environments

Takaichi Ito
Graduate School of Media and Governance
Keio University

目次

第一章 研究背景と研究目的.....	1
1-1 研究背景.....	1
1-1-1 社会構造とライフスタイルの変化.....	1
1-1-2 分析方法論の変化.....	2
1-1-2 分析方法論の新しい潮流.....	3
1-1-2-1 データマイニングと探索性.....	3
1-1-2-2 背景知識と形式知、暗黙知.....	4
1-1-2-3 可視化・メタ認知・FNS ダイアグラム.....	5
1-1-2-4 SECI モデル.....	7
1-1-2-5 知的インタラクティブシステム.....	7
1-2 先行研究と先行的社会調査ツール.....	8
1-3 研究目的.....	9
1-3-1 開発ツールの位置付け.....	9
Ralwer.....	9
ひっぱるくん.....	9
こうぞうくん.....	10
GoocaBooca・ふかぼりくん.....	10
1-3-2 期待される成果.....	10
第二章 データ取得ツールの開発と実装.....	12
2-1 概要.....	12
2-2 目的とコンセプト.....	14
2-2-1 入力補助の効くテキスト処理プログラミング言語.....	14
2-2-2 テキスト処理の合成基盤.....	14
2-2-3 データフロープログラミングとビジュアルプログラミング.....	15
2-2-4 XML でデータフローを記述する.....	15
2-2-5 祖先ノードを参照することによるパラメータの自動設定.....	17
2-2-6 テキスト変数.....	17
2-2-7 同時実行、並列処理.....	18
2-2-8 宣言型で関数型言語的なしくみ.....	18
2-3-1 開発の経緯.....	19
2-3-2 開発の目標.....	19
2-4 設計と実装.....	19
2-4-1 XAML を使った言語.....	19
2-4-2 RawlerBase と RawlerMultiBase.....	20
2-4-3 祖先参照の実装.....	20
2-4-4 テキスト変数の実装.....	20
2-4-5 実装したノード.....	22
Web.....	22
Data.....	22
Html.....	22
Text.....	22

IO	22
Condition.....	22
Enumeration.....	23
Utility.....	23
NPL.....	23
2-5 開発上の工夫.....	23
2-5-1 XAML ベースのシステム.....	23
2-5-2 Nuget 対応.....	23
2-5-3 XML の記述の短縮化	24
2-5-4 入力フォーム機構.....	24
2-5-5 拡張性.....	25
2-5-6 オープンソース	27
2-6 運用実績.....	27
2-7 開発活動を通じて得た知見.....	27
2-7-1 学習コスト.....	28
2-7-2 メタプログラミング・動的生成	28
2-8 まとめ.....	29
第三章 探索的表出化ツールの開発と実査.....	31
3-1 開発コンセプト	31
3-1-1 複雑ネットワークと柔らかい構造化	31
3-1-2 自己探求と動的な生成過程	32
3-1-3 実装	32
3-2 実査 1 ブログ解析：食の意味世界の気づき	34
3-3 実査 2 ツイッター解析：家族の解釈多様性	35
3-4 実査 3 エッセイ解析：S 組織のガバナンスの意味付け	37
3-4-1 基本キーワードの選択.....	38
3-4-2 構造化の生成ルールの設定	40
3-4-3 構造化の生成過程(1)ー全体構造化と階層構造ー	40
3-4-4 構造化の生成過程(2)ー下位構造化と下位階層構造ー	47
3-4-5 「環境」から「小さな成長戦略」へ	49
3-4-6 「イノベーション」から「小さな成長戦略」へ	51
3-4-7 発見と正当化の根拠.....	53
1) 自由と秩序.....	55
2) 価値表明と目標達成	55
3) 自明性と創発性.....	55
4) 目的と手段.....	56
5) 階層とネットワーク	56
6) 一般と特殊.....	56
3-4-8 正当化とデータ検証.....	57
3-5 評価と課題	63
3-5-1 評価 1：自己探求とメタ認知	63
3-5-2 評価 2：柔らかい構造化と解釈多様性.....	64

3-5-3 課題.....	64
第四章 探索的構造化ツールの開発と実査.....	67
4-1 開発コンセプト	67
4-1-1 べき乗分布と頻度の層（レイヤー）	67
4-1-2 インタラクティブな関係の表示とクラスタリング	68
4-1-3 概念化とメタ認知.....	68
4-1-4 制約付きクラスタリング.....	68
4-1-5 クラスタリングの失敗の可視化	69
4-1-6 クラスタ間の関係の可視化	69
4-1-7 属性情報の付加と可視化.....	69
4-2 実装.....	69
4-2-1 入力ファイル.....	69
4-2-2 画面の説明	69
4-2-3 クラスタリングのアルゴリズム	71
4-2-4 クラスタリングの評価指標	72
4-2-5 クラスタリングの失敗の検出.....	73
4-2-6 クラスタマップの作成.....	73
4-3 実査4：テレビドラマ（半沢直樹）のツイート分析	73
4-3-1 分析結果	73
4-3-2 考察	77
4-4 実査5 『ハチミツとクローバー』に見る、普通の若者のカルチャースタイル.....	77
4-4-1 はじめに	77
4-4-2 分析対象データ	77
4-4-3 データクリーニング	78
4-4-3-1 ユーザのクリーニング	78
4-4-3-2 コミュニティのクリーニング	78
4-4-4 構造化.....	79
4-4-4-1 デフォルトの出力結果.....	79
4-4-4-2 クラスタ数の調整.....	79
4-4-4-3 ユーザの共起情報に基づく調整	80
系統1：香り	80
系統2：本.....	81
系統3：芸能	82
系統4：アニメ.....	83
系統5：女性向けライフスタイル	84
系統6：自分らしさ.....	84
系統7：音楽	85
系統8：女性向けコンテンツ.....	86
ユーザの共起情報に基づいて調整したコミュニティ	87
ユーザ数共起に基づいて調整し作成した系統	87
4-4-4-4 属性情報に基づく調整.....	88
女性特化の調整	88
男性特化の調整	89
属性情報に基づいて調整したコミュニティ	89

属性情報に基づいて調整し作成した系統	90
4-4-5 構造の解釈.....	90
4-4-5-1 各系統の解釈.....	90
系統 1：香り	91
系統 2：本.....	91
系統 3：芸能	91
系統 4：アニメ.....	92
系統 5：女性向けライフスタイル	92
系統 6：自分らしさ.....	93
系統 7：音楽	93
系統 8：女性向けコンテンツ.....	93
4-4-5-2 解釈に基づいて命名した各系統	94
4-4-6 概念化.....	95
4-4-6-1 連関の調整.....	95
4-4-6-2 連関の解釈.....	97
4-4-7 結論：普通の若者にとっての「カルチャー」「スタイル」	101
普通の若者にとっての「カルチャー」	101
4-4-8 実例からみた「こうぞうくん」の価値探索機能	103
4-4-8-1 使用の全体.....	104
①セーブ／ロード機能	104
4-4-8-2 構造化.....	104
クラスタリング	105
②基準変更	105
③統合	107
④分割	107
相関ルール	108
⑤ソート.....	109
⑥彩色	110
⑦移動.....	111
⑧Lock	112
⑨削除	112
特化係数.....	113
⑩彩色	114
4-4-8-3 概念化.....	114
クラスタリング	114
⑪生成	115
相関ルール	116
⑫ソート.....	117
⑬結合	117
第五章 探索的深層化ツールの開発と実査.....	120
5-1 Asking ツール（GooCaBooca）の開発と実装.....	120
5-1-1 目的とコンセプト.....	120
5-1-2 実装	121
5-2 探索的深層化ツールの開発と実査.....	123
5-2-1 目的・コンセプト.....	123
5-2-2 解決する問題.....	124

5-2-3 開発の経緯と目標.....	124
5-2-3-1 設計	124
質問データ	124
回答データ	125
絞り込み機能.....	125
作成した機能コンセプト	125
5-2-3-2 実装	125
Save&Load	125
単純集計.....	125
クロス集計	125
彩色機能.....	125
グループ化集計	126
クラスタリング	126
データの調整.....	127
フリーアンサー分析	128
5-2-3-3 開発上の工夫.....	128
5-2-3-4 ソフトウェア開発として特徴的な点	129
5-2-3-5 開発活動を通じて得た知見	129
5-3 実査 6 女性ファッションスタイル（恋服）の解析	129
5-3-1 調査フレーム.....	130
5-3-2 価値探索プロセス.....	131
5-3-2-1 単純集計	131
5-3-2-2 クロス集計.....	138
5-3-2-3 クラスタリング	144
5-3-2-4 グループ化集計	152
カジュクール G スタイルの場合（図 151）	152
キュート G スタイルの場合（図 152）	153
ナチュキュート G スタイルの場合（図 153）	153
アンチナチュ G スタイルの場合（図 154）	154
5-3-3 結論	154
①カジュクール G スタイルの場合（21 名）	156
②キュート G スタイルの場合（34 名）（図 161）	158
③ナチュキュート G スタイルの場合（10 名）：（図 162）	158
④アンチナチュラル G スタイルの場合（23 名）：（図 163）	159
①カジュクール G スタイル（72 名）	160
②キュート G スタイル（85 名）	163
③ナチュキュート G スタイル（50 名）	165
④アンチナチュ G スタイル（85 名）	167
5-3-4 実例からみた「ふかぼりくん」の価値探索機能	170
5-3-4-1 使用の全体.....	170
①統合機能	170
②絞込機能	170
③削除機能	171
④彩色機能.....	171
⑤ソート機能.....	171
⑥スケール機能	172

⑦生成機能	172
⑧付加機能	172
⑨リンク機能.....	172
⑩セーブ（ロード）機能.....	172
5-3-4-2 比較多層分析.....	172
5-3-4-3 マルチデータ分析.....	177
第六章 結論と今後の課題	181
6-1 作成したツールの全体像	181
6-2 探索性、創発性	182
6-3 実査から発見された価値探索機能の創発性	184
6-3-1 「ひっばるくん」の表出化ツールの価値探索機能	184
6-3-2 「こうぞうくん」の構造化ツールの価値探索機能	187
6-3-3 「ふかぼりくん」の深層化ツールの価値探索機能	188
6-4 今後の課題	191
謝辞.....	192
参考文献.....	193

図表目次

図 1	KDD プロセス	4
図 2	FNS ダイアグラム	5
図 3	チャンス発見の 2 重らせんモデル	6
図 4	Visual Studio の XAML エディタによる入力補助	13
図 5	クライアント実行環境	14
図 6	Web スクレイピングのデータフロー	15
図 7	木構造の記述	16
図 8	図 6 を Rawler のコードで表現	17
図 9	テキスト変数の概念図	18
図 10	テキスト変数の使用例	18
図 11	並列実行を行うコード例	18
図 12	C# と XAML での同じ表現	20
図 13	祖先参照をつかい変数の設定と取得	21
図 14	ノードの変数を木構造で表現	21
図 15	テキスト変数を使った動的なファイル名の指定	21
図 16	フォーム機構の記述	25
図 17	作成されたフォーム	25
図 18	ツイッターからのデータ取得	27
図 19	ひっぱりくんの挙動プロセス	33
図 20	「おいしい」の単語と共起関係	35
図 21	「おいしい」のさらなる探求	35
図 22	学生 A が作成したグラフ	36
図 23	学生 B が作成したグラフ	36
図 24	単語頻度のランキングと削除語の指定	38
図 25	ソフトピアを起点に開始	39
図 26	図 25 から「ベンチャー」「支援」「iomas」を展開	40
図 27	図 26 から「ベンチャー」から派生した語をさらに展開	41
図 28	図 27 から「支援」から派生した語をさらに展開	42
図 29	図 28 から「iomas」から派生した語をさらに展開	43
図 30	図 29 を構造化	44
図 31	階層を超えて接続するものを着色	44
図 32	派生語の第二階層での赤ラインの自動生成	45
図 33	第三階層での関連	46
図 34	支援に絞った関係図	47
図 35	最終的に解釈として発見されたクラスター	48
図 36	「雇用」と「成長」から作成	48
図 37	図 24 で発見した語を構造化に追加	49
図 38	「環境」から「小さな成長戦略」への下位構造化	50
図 39	図 38 を階層構造に変換	51
図 40	イノベーションを起点とした下位構造化	51
図 41	最終的な階層構造化	52
図 42	「ソフトピア」「ベンチャー」「支援」を含む文章の表示	57
図 43	図 42 の条件に「iomas」を追加した時の文章	58
図 44	「成長」を含む文章	59

図 45	「成長」「戦略」「小さ」を含む文章	60
図 46	階層構造図をもとに作成	61
図 47	「iomas」を起点にした時	62
図 48	「イノベーション」「先端」「新産業」「iomas」を含む文章	63
図 49	Hipparu と McS の関係図	64
図 50	可視化の最終的な形	67
図 51	アイテム選択時、関係するアイテムの表示	68
図 52	「こうぞうくん」の機能説明	70
図 53	クラスタ（縦列）に名前をつけられる	71
図 54	クラスタ間の関係を表示	73
図 55	ロック機能	74
図 56	最終的なクラスタリング結果	74
図 57	クラスタ間の関係図	75
図 58	堺雅人に特化しているクラスタ	75
図 59	壇蜜に特化しているクラスタ	76
図 60	大和田常務に特化しているクラスタ	76
図 61	はじめにクラスタリングした結果	79
図 62	クラスタ数を8にしてクラスタリングした結果	80
図 63	解釈を元にアイテムを移動（1）	81
図 64	解釈を元にアイテムを移動（2）	82
図 65	「永作博美」でのアイテムの関係	83
図 66	「東のエデン」でのアイテムの関連	83
図 67	「LOWRYS FARM」でのアイテムの関係	84
図 68	「(∇)わー」でのアイテムの関係	85
図 69	「浅野にいい」「ソラニン」移動	86
図 70	「TRICK! TRICK! トリック!」の移動	87
図 71	属性「女性」での特化	88
図 72	属性「男性」での特化とアイテムの移動	89
図 73	調整後のコミュニティの全体構造	90
図 74	系統に名前を書き込んだ全体構造図	95
図 75	デフォルト（Lift 値 1.15 以上）でのクラスタ間の関係図	96
図 76	Lift 値 1.13 以上でのクラスタ間の関係図	96
図 77	クラスタ間関係図の解釈 1	97
図 78	クラスタ間関係図の解釈 2	98
図 79	クラスタ間関係図の解釈 3	99
図 80	クラスタ間関係図の解釈 4	100
図 81	クラスタ間関係図の解釈 5	101
図 82	各系統の位置づけ	102
図 83	各系統の位置づけ	103
図 84	セーブロード機能の概念	104
図 85	クラスタリング機能の概念	105
図 86	クラスタリング結果（評価指標 0.257）	106
図 87	クラスタリング（評価指標 0.256）	107
図 88	系統数6で作成したクラスタリング結果	108
図 89	系統数8で作成したクラスタリング結果	108
図 90	相関ルールの機能概念 1	108

図 91	相関ルールの機能概念 2	109
図 92	相関ルールの詳細値の表示.....	110
図 93	「よつぱと！」を起点とした際の彩色例.....	111
図 94	アイテムの手動による移動.....	111
図 95	ロック機能.....	112
図 96	削除機能.....	113
図 97	特化係数の機能の概念.....	113
図 98	特化係数による彩色とアイテムの移動	114
図 99	クラスタリング機能の概念.....	115
図 100	クラスタマップ.....	116
図 101	相関ルールの概念（ソート・結合）	116
図 102	クラスタマップのデータ詳細例	117
図 103	クラスタマップによるクラスタの結合例.....	118
図 104	GoocaBooca 全体像	120
図 105	質問設定の XML ファイル	123
図 106	「ふかぼりくん」 ツール画面.....	123
図 107	カラーパレット	126
図 108	彩色したクロス表	126
図 109	クラスタリング結果画面	127
図 110	「データの調整」画面	128
図 111	「恋服」調査の調査フレーム	130
図 112	単純集計の表示結果.....	132
図 113	ソート機能操作前	133
図 114	ソート機能操作後	133
図 115	データの調整（女性で絞込）	134
図 116	女子：422 名	134
図 117	男子：303 名	135
図 118	全体：725 名	136
図 119	女子：422 名	136
図 120	男子：303 名	136
図 121	画像の対応関係.....	137
図 122	女子、男子での画像の序列.....	137
図 123	年齢の単純集計	138
図 124	年齢のデータ調整（20 歳以下で変数をまとめる）	138
図 125	データの調整後の年齢のクロス表.....	139
図 126	6 段階評価から 3 段階に統合・縮約.....	139
図 127	データ調整後のクロス表一覧.....	140
図 128	年齢と恋人の有無のクロス表	141
図 129	女子の 4 スタイルのクロス表 1	141
図 130	女子の 4 スタイルのクロス表 2	142
図 131	女子の 4 スタイルのクロス表 3	142
図 132	女子の 4 スタイルのクロス表 4	142
図 133	女子の 4 スタイルのクロス表 5	143
図 134	女子の 4 スタイルの関連図.....	143
図 135	男子の 4 スタイルのクロス表.....	143
図 136	男子の 4 スタイルの関連図.....	144

図 137	女子 4 スタイルのクラスタリング設定画面	145
図 138	女子 4 スタイルによるクラスタリング結果	146
図 139	女子の 4 スタイルの関連図とクラスタとの関係のプロット	147
図 140	クラスタの人数分布	147
図 141	男子 4 スタイルによるクラスタリング結果	148
図 142	クラスタリング結果の人数分布	148
図 143	男子の 4 スタイルの関連図とクラスタとの関係のプロット	149
図 144	年齢での絞り込みと、無駄なクラスタの排除	149
図 145	女子スタイルクラスタと性別とのクロス表	150
図 146	男子スタイルクラスタとクロス表	150
図 147	年齢の統合 (20 歳以上、以下の 2 つにする)	151
図 148	女子スタイルクラスタと年齢のクロス表 (女子限定)	151
図 149	男子スタイルクラスタと年齢のクロス表 (男子限定)	152
図 150	絞り込み条件の指定	152
図 151	年齢、恋人、服でグループ化集計 (カジュクールでソート)	153
図 152	年齢、恋人、服でグループ化集計 (キュート G スタイルでソート)	153
図 153	年齢、恋人、服でグループ化集計 (ナチュキュート G スタイルでソート)	153
図 154	年齢、恋人、服でグループ化集計 (アンチナチュ G スタイルでソート)	154
図 155	女子スタイルと男子スタイルのクロス表 (女性限定)	154
図 156	男子スタイルと女子スタイルのクロス表 (男子限定)	155
図 157	男子スタイルと女子スタイルのクロス表 (男子限定、20 歳以上)	155
図 158	図 143・図 145 から、黄色の関係を図化	156
図 159	女性限定で、カジュアルクール G、カジュアル B で絞込	156
図 160	図 159 の条件に基づく画像の集計	157
図 161	キュート G スタイルで絞込での集計	158
図 162	ナチュキュート G スタイルで絞り込み集計	159
図 163	アンチナチュラル G スタイルで絞り込み集計	160
図 164	絞り込み条件 (女性限定、カジュクール G)	161
図 165	カジュクール G のフリーアンサー	162
図 166	カジュクール G の上位の画像 (年齢 25 歳以上)	163
図 167	キュート G スタイルのフリーアンサー	164
図 168	キュート G スタイルの上位画像 (キュートガール限定・21-24 歳に限定)	165
図 169	ナチュキュート G スタイルのフリーアンサー	166
図 170	ナチュキュート G スタイルの上位画像 (21-24 歳に限定)	167
図 171	アンチナチュ G スタイルのフリーアンサー	169
図 172	アンチナチュ G スタイルの上位画像	169
図 173	「ふかぼりくん」の価値探索機能	171
図 174	全体での女子 G スタイルの単純集計	173
図 175	女子 422 名による女子 G スタイルの単純集計	173
図 176	男子 303 名による女子 G スタイルの単純集計	174
図 177	女子 G スタイルと性別のクロス集計	174
図 178	グループ化集計 (カジュアルクールでソート)	174
図 179	グループ化集計 (キュート G でソート)	175
図 180	グループ化集計 (ナチュキュート G でソート)	175
図 181	グループ化集計 (アンチナチュ G でソート)	175
図 182	カジュクール G スタイルの支持の高い 4 ガールスタイル画像	176

図 183	カジュークルGスタイルのフリーアンサー（一部）	177
図 184	フリーアンサー集計.....	178
図 185	フリーアンサー集計（詳細表示）	178
図 186	絞り込み条件の設定.....	179
図 187	図 174 での絞り込み条件による画像	179
図 188	ツールの全体像.....	181
図 189	「ふかぼりくん」の価値探索機能概念図.....	189

表目次

表 1	参加コミュニティ数ごとに等分割して層化	78
表 2	各レイヤーのコミュニティ数と参加人数.....	78
表 3	移動コミュニティ一覧	87
表 4	調整後のコミュニティ名とユーザの人数と男女の割合	87
表 5	移動アイテム.....	89
表 6	システムのユーザ数と男女の割合の一覧表	90
表 7	仮システム名と本システム名.....	94
表 8	本システム名と短縮システム名	97
表 9	システム名とハレとケの解釈	102
表 10	「こうぞうくん」の機能一覧.....	104
表 11	質問データの列名とその意味と書式	124

第一章 研究背景と研究目的

第一章 研究背景と研究目的

1-1 研究背景

1-1-1 社会構造とライフスタイルの変化

1980年代、まだ、インターネットの兆候もない、消費社会が全盛の時代にあって、日本の若者たちは、それ以前の真面目な世代と異なり、楽しく生きることをモットーに、自分らしさを追い求めていた。私探しゲームをする女の子は新しい豊かな社会におけるアイデンティティを模索する典型として脚光を浴び、男の子は自分の世界へのこだわりから既成の知識体系から逸脱のレッテルを貼られることを厭わず、オタクと自称する道を歩んだ。その結果、鈴木謙介が「私たち消費論[1]」で語るように、「私たちが欲望することが私の欲望すること」となって、自分たちに固有の閉鎖的なライフスタイルを、それぞれの私たちが無数に創造しはじめ、それによって、他の私たちのライフスタイルには見向きもしないで自分の独自の世界に浸る、という社会文化現象が発生した。そのことを、宮台真司は「島宇宙化」と呼び、タコツボ化してしまった消費社会における若者文化の特徴を的確に表現した[2]。同じ日本語を話し、同じ地域に住み、同じ性別や年齢階層であっても、自分の島宇宙から離れた別の島宇宙については、わからない、したがって当然、無関心、という没関係性の島宇宙が無数に存在する時代として、消費社会が描かれるようになった。アニメオタク・アイドルオタク・腐女子・コスプレイヤーなど、自分のことなら十分すぎるほど理解可能であるが、一歩外れると、まったく理解不能という島宇宙が存立する状況が発生し、それが消費社会における若者のライフスタイルとして定着していった。消費社会は、もはや大衆すなわちマスという大きな一つの塊という一様なライフスタイルではなく、ライフスタイルの多様性を前提として豊かな社会を語るべき時代になり、同時に企業サイドからは、そのライフスタイルの多様性を射程に入れた戦略的マーケティングを企業存立の生命線として重視する時代になった。まさに企業も生活者も、私たちの消費を喚起させるニーズと満足をめぐって丁々発止のゲームを展開する時代になった。

さらに1990年代の後半になると、インターネット環境とモバイルツールが、すべての生活者にとって自明のものと認識される時代へと変化していった。それが情報社会の到来である。情報社会の前期は、マスメディアの登場によって発生した1950年代であるが、ここでの情報社会とは、インターネット以降の時代状況であり、従来の産業化と消費化がもたらした、貧しい社会から豊かな社会への変化を説明する軸とは異なり、もう1つ別のメディア軸を生成する契機であった。それがマスメディアからインターネットメディアへの変化であり、それは、ライフスタイルとして、まったく対照的な様相をもたらした。それが「サイレントマジョリティ」[3]から「おしゃべりなロングテール」[4]への変容である。マスの時代では、人々は自ら発信するメディア状況にないがために、一方的な情報の受信者としてただ沈黙して受容するだけの大きな1つの塊としてしか存在しえなかったが、ネットワークメディアの登場によって、生活者は、スマホなどのツールを駆使して多様で個別的な情報を発信し、同時に身近な生活者からの多様で個別的な情報を受信するという、従来のマスメディアではありえない相互的な情報行動を、ネットコミュニティを介して活発に遂行するようになった。これこそ、おしゃべりなロングテールという社会的存在そのものである。それは、それぞれの生活者が個別で多様なニーズをもって、弱くかつ柔らかく近似的な他の無数の生活者とつながることで、自分らしい存在を共感的かつ共存的に表明している小さな公共的コミュニティの生成である。ここでは、80年代の孤立する島宇宙の集合とは異なりで、島宇宙の境界をやや曖昧にし、その曖昧さを島宇宙間で共有するノードにすることで、弱くかつ柔らかにつながる関係性が島宇宙間で維持されている。これが、メディア軸において、ネットワークメディアへの変化がもたらしたライフスタイル上の変化であり、おしゃべりなロングテールの基本的なスタンス

である。

いま、おしゃべりなロングテールというライフスタイル上の特性をもった生活者がインターネット環境において、さまざまなモバイルツールを駆使して、社会的な情報行動を営んでいる。このようなおしゃべりなロングテールが示す社会的行動こそが、本論文が扱い、研究の対象とするものであり、そこを解明するための社会調査ツールを開発し、実装することが、本論文の目指すところである。

1-1-2 分析方法論の変化

現代社会における社会構造は、前述したように、産業化・消費化そしてネットワーク化という意味での情報化を契機にして、大きくその特性を変化させてきた。とすると、そのさまざまな社会構造を、社会調査という視点から理解しようとする場合には、その社会構造をもっとも的確に理解するための分析上の方法論が必要となる。従来の社会統計を基本とした社会調査法は、その多くを産業化と消費化に対応した社会調査として最適な手法であった。しかしその方法論がそれなりの普遍性をもつからといって、いかなる社会構造の理解においても最適であるという保証はない。本論文の立場は、従来の社会調査の理論と方法論と具体的な分析手法を尊重すると同時に、この新しいネットワーク環境におけるおしゃべりなロングテールの特性を理解するには、それなりに新しい社会調査手法の開発が不可欠であると主張するものである。

産業化・消費化・情報化を契機とする3つの社会構造とその構造のもとで典型的なライフスタイルについて、その理解や解釈に適合する社会調査の方法を想定すると、以下のような関係性が想定できると思われる。

かつて産業化の時代にあつて、その時代の担い手であつた大衆＝サイレントマジョリティという大きな1つの塊について、その基本構造を社会調査の対象としてデータ化する場合、それは、ブリュデュの表現を利用すれば、所得水準(経済資源)と学歴水準(文化資源)の2要因から取得すればよく、それによって産業社会の基本的な構造化が可能になった。ここでは、回帰分析のような従来の社会統計手法を活用することで、産業社会の基本である職業構造からなる社会的階層構造と社会移動のパターンが明確に解析されることになった。

その後、消費化が優先される時代状況になると、大衆は分衆化し、サイレントマジョリティは後退し、人々は、職業構造(所得と学歴)よりも消費構造(支出と趣味)を優先することで、自分のニーズと満足の基準から、自分らしいライフスタイルの表現を求めるようになり、その結果、ライフスタイルの多様性を認知し理解することが社会構造のテーマとなっていく。宮台真司は、この状態を島宇宙化[2]呼び、消費生活と消費文化こそが時代の主要な社会層であることを見事に叙述した。この社会層を社会統計的に解析しようとするとき活用されたのが、数量化理論などの多変量解析の手法である。クラスタという島宇宙を社会統計としての的確に解析する手法が開発され、消費化の時代はその明確な島宇宙＝クラスタの構造として可視化された。

このような産業化と消費化の段階では、従来の社会統計が有効である。そこでの多様な解析手法が、大衆＝マスや分衆＝島宇宙という対象の相違によって、線形モデルからクラスターモデルへと手法の変化がみられるものの、従来の社会統計の手法として非常に有効性を発揮していた。1つの大きなマス構造の分析から多数のやや小さなクラスタ構造群への分析という変化がここにはみられるが、社会統計手法の有効性は不動であった。

ここで共通しているのは、社会調査に限定すれば、マス調査であれクラスタ分析であれ、質問紙による調査だという点である。その意味では、質問紙自体に研究者の構造が刷り込まれており、その検証という形式でサンプリングによる調査が実施された、という共通点は、マスであっても、島宇宙であっても同様である。

しかしこのようなオーソドックスな社会調査は、社会構造が情報化の段階に入ると一気に変貌する。ネットワーク環境にビッグデータが散乱し、多様なモバイルツールが身体化され、そこからデータがネ

ットワーク環境に伝送されたり、取得できたりするような状態になると、従来の質問紙調査では想像もできないような社会調査環境が生成されることになった。とすれば、ネットワーク環境に散乱するデータを瞬時に取得し、そのデータをクリーニングし、それをもとに構造化を図るような調査設計が可能になり、それがデータサイエンスという専門領域を形成するまでに発達しはじめた。

現在、マーケティングの世界で流行のソーシャルリスニング[7][8]という発想は、まさに、この新しい社会調査の動向を察知し、ネーミングしたものである。確かに、従来の質問紙調査のように、質問紙を介して「問う (asking)」のではなく、インターネット上の多種多様なデータは解析対象の宝庫そのものであるから、そこに耳を立てて、十分に聴くだけで、新しい知見が発見されるはずである。ソーシャルメディアで聞き耳をたてる (Listening) 方法論がソーシャルリスニングである。このソーシャルリスニングを根源の思想は、群衆の叢智 (Wisdom of Crowds) という、多数の人の平均的評価は各個人の評価より正確であり、なおかつ専門家による評価より正確であるという事実に基づいている[9]。このことは、民主主義の原則がそうであるように古くから知られていることであるが、ソーシャルリスニングができる環境になった今、大量の意見を集めることができるようになったため、注目されている。東浩紀は、「一般意志 2. 0」において、ソーシャルリスニングによる意見の可視化は、大衆の無意識の可視化であり、新しい民主主義の合意形成には、この無意識の可視化を基礎にすべきだということを主張している[10]。このとき、合意形成は、可視化の結果に従うのではなく、可視化の結果を考慮すべきとしている。つまり、可視化はあくまで道具であり、分析者自身が、可視化の結果から、次のアクションを考えていかななくてはならないことを示している。このように、ソーシャルメディアのデータの可視化には、今までにない新しい社会的リアリティの発見と次のアクションを考えるためのツールになる期待が膨らんでいる。

小野田[11]・妹尾[12]・山崎[13]が開発し活用している「柔らかな構造化手法」は、ソーシャルリスニングの先行事例として代表的なものである。

本研究は、柔らかな構造化手法の基本構成を継承し、さらなる発展の形態を提案するものである。その発展の方向としては自動化や汎用化を求めながら、柔らかな構造化手法を構成する多様なツールをさらに開発して、実査での価値をより高めるような仕組みを作ることを目指している。次節では、分析方法論に関する新たな潮流を整理して、柔らかな構造化手法との関係を明確にする。

1-1-2 分析方法論の新しい潮流

1-1-2-1 データマイニングと探索性

1990 年代以降、IT 化が進み、記憶容量の大容量化により、様々なデータが蓄積されるようになった。加えて、コンピュータのメモリの増大、CPU 速度の向上、システムの並列化により、蓄積された巨大なデータを分析できるようになった。その中で生まれてきたのは、データの中から有益な知識を発見するデータマイニングという方法論である。データマイニングは、統計学と、機械学習、パターン認識という AI の分野と、データベースの 3 つの分野が融合した大規模データ分析の手法である。従来の統計学と比べて、データの量が増え、変数も増えているので、一意の答えを求めることは困難なので、近似解を求めることや、可視化を行うのが一般である。

データマイニングを通じて、知識発見を成功させるプロセスとして、KDD プロセスが提案されている[14] (図 1)。データからの知識発見のためには、データの「Selection (選択)」「Preprocessing (前処理)」「Transformation (変形)」「Data Mining (分析)」「Interpretation/Evaluation (解釈・評価)」というプロセスで進むことが必要であると述べられている。データマイニングは、単線的な処理をすることで、データから何かしらの知識が得られると考えがちだが、違う。KDD プロセスで一番重要な

は、「解釈・評価」が満足行くものでなければ、「選択」「前処理」「変形」「分析」に立ち戻る点である。適切なデータを入力しなければ、適切な答えはかえって来ない。適切な答えを導くためには、適切なデータを用意し、データからノイズを捨てるクリーニングしデータ分析をできるデータの形に変形しなくてはならない。そのため分析者が結果に納得できなかったとき「選択」「前処理」「変形」のプロセスに立ち戻る必要がある。このように、KDD プロセスは繰り返しを前提とした探索的プロセスである。本論文のツールでは、この KDD プロセスの立ち戻りを支援する仕組みとなっている。特に、第二章の Web スクレイピングツールの「Rawler」は、Web からの「選択」「前処理」「変形」のためのツールであり、分析プロセスは、ないものの、Web からのデータ取得に関する基盤ツールになりうるポテンシャルを持っている。

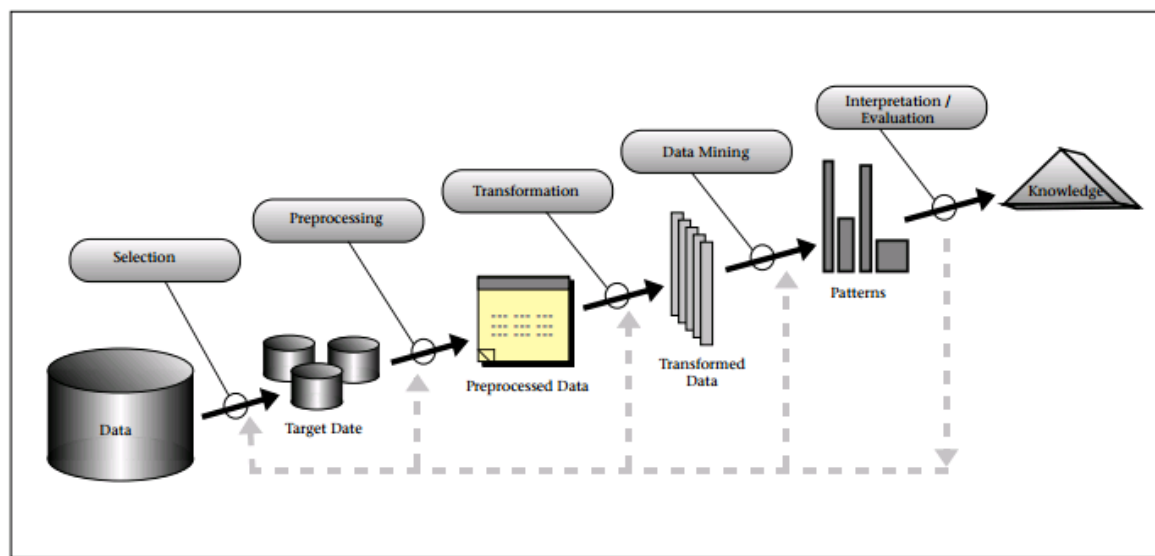


図 1 KDD プロセス

1-1-2-2 背景知識と形式知、暗黙知

KDD プロセスの最後の分析結果の評価、解釈は、人間が行うものである。その評価・解釈が適切であるためにも、データへの理解、データの背景知識の理解、データのアルゴリズムの理解が必要である。

分析対象のデータの理解、分析アルゴリズムの理解は、データ分析する上で当然である。これがわかっていなければ、データ分析ができない。では、データの背景知識の理解とはなにか。データそれ自体の知識ではなく、データの背景にある知識であり、データ分析の結果が確からしく解釈できるかを決める知識である。このような専門領域の知識を「ドメイン知識」という。そのため、データマイニングには、その対象に対する深い知識をもつ「専門家」が必要である。

専門家の背景知識への負担を減らすため、データ分析自体に背景知識を組み込めばいい、と考えられるが、しかし、実際には、なかなかうまくいかない。それを端的に示す例は、帰納論理プログラミング[15]である。帰納論理プログラミングとは、論理プログラミング言語 Prolog をベースとした、述語論理をベースとした機械に帰納推論をさせる仕組みである。特徴は、背景知識を使った知識発見を行うことができることである。そのため、人間が行っているような高度な帰納推論が可能になる。成功事例としては、化合物の中から、新しいタイプの化合パターンを発見したものがある。しかし、この背景知識が非常に難しい。背景知識は Prolog で記述できるので、様々なことを述語論理で柔軟に記述できるのだが、仕組み上、記述した知識の中から知識を探すので、自分が記述した知識の中から答えを探すということになってしまう。そのため、化学の原子結合といった事例ならまだしも、人の社会の中にある様々な知識すべてを記述することは困難である。知識表現の難しさは、ひとつに状況依存性にある。状況依存性とは知能自体がその場の状況に依存しており、環境と切り離れた知識はないということである。そ

のため、知識表現は状況とセットでなければならず、そのため記述しなければならない知識も増えていってしまう。

また、知識の記述が可能だとしても、何がデータ分析に必要な知識なのかがわからないため、些細な知識までも記述しなくてはならず、膨大な記述になってしまう。そして、当然、膨大な記述になればパフォーマンス低下につながり、出してくる答えは自身が記述したことで発見的ではないという、労多くして益なしとなってしまふ。この問題は、フレーム問題[16]と同型であり、AI 困難問題の一つと言われる。

また、記述できる知識は、知識の一部でしかないという本質的な問題もある。記述できる知を「形式知」といい、一方、記述できない知を「暗黙知」という。暗黙知を始めに提唱した、マイケル・ポランニー『暗黙知の次元』[17]では、「知識というものがあるとすると、その背後には必ず暗黙の次元の「知る」という動作がある」という概念であり、具体的には「顔の認識」や「自転車の運転」など、言葉では表現しづらいが、人間は習得している知のことをいう。

このように、データ分析のためには、背景知識が必要であるが、知識は多様であり、記述が難しい物すら含むので、知識を機械の中に埋め込んでいくことは、困難を伴う。そのため、人間側の知識を引き出し、人間が結果に関与することができる仕組みが必要である。本論文の、3章、4章のツールは、データを可視化することで人間の知識を引き出すことを目的としている。

1-1-2-3 可視化・メタ認知・FNS ダイアグラム

暗黙知を知ること、すなわち「暗黙的に知ることを知る」とは、「認知の認知」であり、メタ認知行為である。そのため、メタ認知行為は、暗黙知を知ろうとする挑戦的行為である。暗黙知は、従来、語りえぬもの、言葉で表現しづらいものだと言われている。しかし、諏訪正樹「身体知獲得のツールとしてのメタ認知的言語化」[18] (2005) では、積極的に暗黙知を言語化に試みることが身体能力の向上につながると主張されている。すなわち、言語化という行為によって、自身から外部化したものを自身で見ることによって、新たな高次変数を獲得し、それによって身体スキルが向上し、さらなる言葉を生む。この時、必ずしも言葉である必要はなく、シンボルであればいい。何かしらの形になっているもので外化することが重要である。外化したシンボルを見ることが、認知の認知、メタ認知になり、新しい発見を得て、次の思考につながる。

このメタ認知のサイクルを一般化したものが、FNS ダイアグラム[19]である(図 2)。FNS ダイアグラムとは、中島、諏訪、藤井らが提唱する、現象学哲学のノエマ・ノエシスという概念を用いて、新しいものが世に誕生するプロセスの一般的構造を説明したものである。

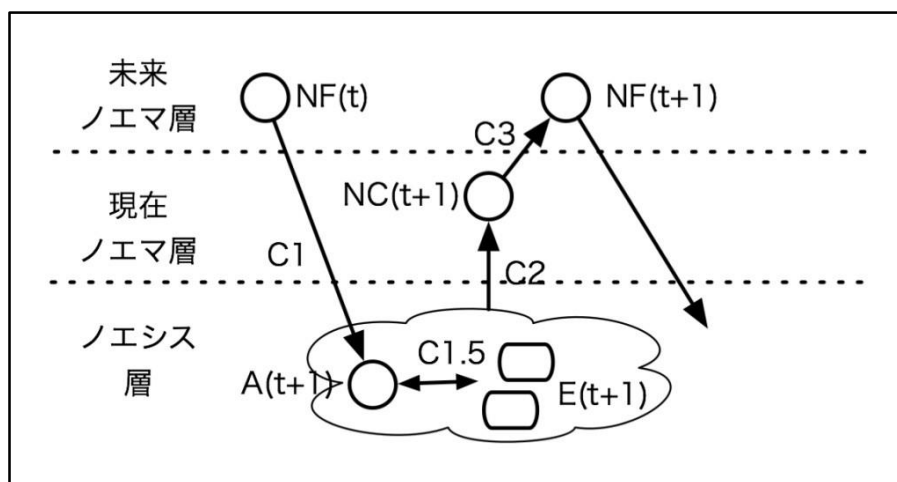


図 2 FNS ダイアグラム

ノエマとは、思考された対象を意味し、ノエシスは、思考作用を指す。そのため、未来ノエマ層は、対象への期待であり、現在ノエマ層は、対象そのものである。FNS ダイアグラムは、対象への期待から、思考し、その結果、対象が生まれ、その対象から、新たな対象への期待が発生するモデルである。イノベーションプロセスは、生成・評価の繰り返しプロセスであり、これは、探索的プロセスと同様である。この観測事実から、仮説を組み立てて検証していくという推論の形式は、アブダクションであるといえる。このダイアグラムで重要なのは、「思考すること」を分解したことであり、また、何かしらの具現化、現在ノエマ層がさらなる未来ノエマ、ノエシスと繋がっていく点である。言葉化、シンボル化がキー要素であり、それにより思考が働いていくことで、イノベーションといえる新しいものを創発していることである。

このプロセスをチャンス発見の文脈で捉えているのが、大澤による2重らせんモデルである [20](図3)。チャンス発見とは、「意思決定を左右する重要な事象、状況、またはそれらに関する情報を理解し活用すること」であり、2重らせんモデルは、そのためのプロセスフローである。2重らせんモデルは、機械が行う処理のラインと、人の行う解釈と理解のラインが2つあり、それが交互にインタラクションをしながら、チャンス発見をしていこうとするものである。また、チャンス発見のためには、人側に準備がなくてはならないという。すなわち、人がこのデータにはチャンスが潜んでいるという期待を元に、機械にデータを分析させるべきということであり、人側の意識は、FNS ダイアグラムの「未来ノエマ」に相当し、機械による処理は、「ノエシス」に相当する。2重らせんモデルは、KDD プロセスの前処理→分析→解釈のフローを、機械と人の視点でとらえたものでもある。

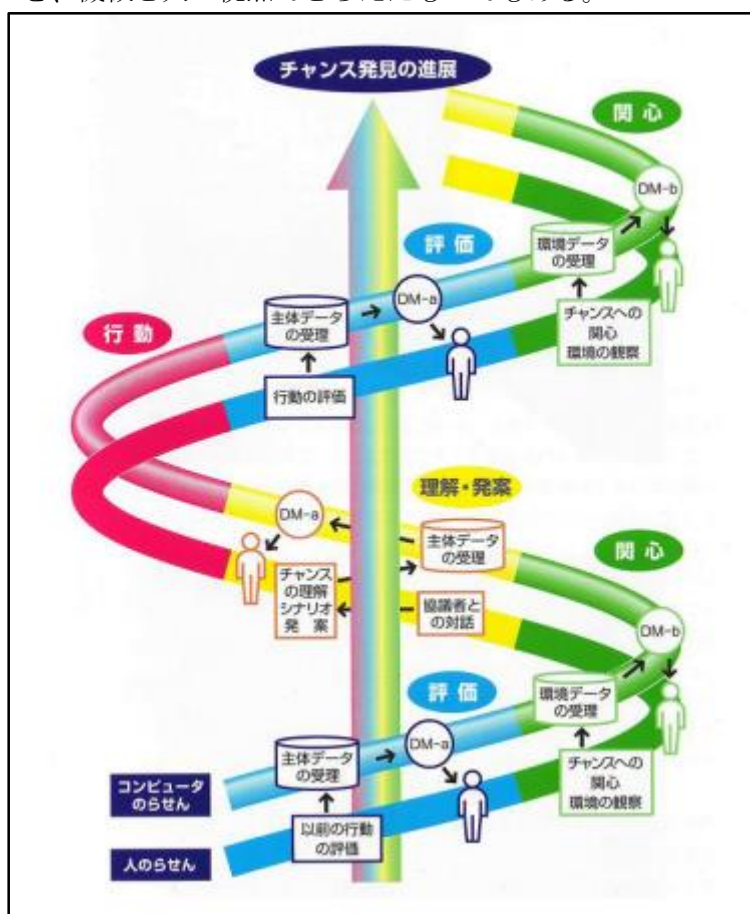


図3 チャンス発見の2重らせんモデル

データマイニングの目的は、データの中から有益な知識を発見することであり、その発見した知識を使って、意思決定することにある。そのために、発見した知識が人にとって、どれだけ納得できるのか？

が重視されるべきであり、そのためにも、納得できるまで、試行錯誤して探索的に行うことが必要であり、そのための仕組みも必要であるべきだ。

探索的に行っていたことから、さらに探索を生むためには、言葉として外化を行い、メタ認知をしていくことが重要である。

1-1-2-4 SECI モデル

このようなプロセス論は、人工知能の分野だけのものではなく、経営学にもある。暗黙知を含めた形で、組織内の知識マネジメントを取る方法として、SECI モデルが知られている。SECI モデル[21]とは、野中郁次郎によって提案された、知識を暗黙知形式知にわけ、暗黙知を形式知に変換していく知識創造におけるモデルを提示したものである。ここでの暗黙知は、「個々人の体験や特定状況に根ざす知識であり、信念・ものの見方・価値システムといった無形の要素を含む」ものとしており、マイケル・ポランニーの「記述不能な知識」としての「暗黙知」ではない。SECI モデルは「共同化」(Socialization)、「表出化」(Externalization)、「連結化」(Combination)、「内面化」(Internalization)を繰り返すモデルである。この頭文字をとって SECI モデルという。

- 共同化 共同体験などによって、暗黙知を獲得するプロセス
- 表出化 得られた暗黙知を共有できるように形式知に変換するプロセス
- 連結化 形式知同士を組み合わせる新たな形式知を創造するプロセス
- 内面化 利用可能となった形式知を元に個人が実践を行い、その知識を体得するプロセス

暗黙知に属するのは、内面化、共同化であり、形式知に属するのは、表出化、連結化である。SECI モデルは、組織内の知識共有プロセスを念頭においている。ここでも、サイクル構造である。諏訪の「メタ認知的言語化」は、表出化プロセスに相当し、その結果「言葉が言葉を生むに、高次の変数を獲得する」は、連結化に相当する。また、KJ 法[22]では、カード作りが表出化に相当し、連結化が、カードのグループ化に相当する。

データと機械によって支援できるのは、形式知側に属する、表出化→連結化のプロセスであり、その他のプロセスは、人間本人に任せるしかない。このためにも、人間と機械・データが一体となって探索するシステムが必要である。

1-1-2-5 知的インタラクティブシステム

このような、人間と機械が一体になるためには、人間側の一方的な入力ではいけないし、機械側からの一方的な出力でもいけない。両者がインタラクティブにやりとりをするシステムでなければならない。高速計算、網羅的探索、データの可視化といった、機械側が得意なことを、機械に任せ、人間にしかやれない、正誤判定、ヒューリスティック、能動学習は、人間側が行うという形が理想である。

そして、機械側も賢い結果を人間側に提示しないと、人間側の負担が大きくなってしまうので、知的なものにする。このようなシステムを「知的インタラクティブシステム[23]」と呼ばれており、そのインタラクションデザインは、新たな AI の研究分野だとしている。

「知的インタラクティブシステム」は、機械学習の中から出てきた。機械学習のためのアルゴリズムが高度化するに従って、その機械学習のためのパラメータ設定が難問になっていった。そのため、適切な答えを機械に出してもらうには、人間の介入が必要である。

機械学習において、学習の手法として、以前から知られているものは、教師あり学習、教師なし学習の2つがある。教師あり学習とは、教師データ、正事例、負事例があるデータを元に学習するものであり、例として、決定木、ベイズ分類器、SVM、帰納論理プログラミングなどが挙げられる。すでに教師として正解を与えているので、新しい入力に対して正解かどうかを判断するシステムになる。すべてのデータに対して、教師データを与えないといけないため、データが大きくなるとその教師データを与えるコストが高くなる。一方、教師なし学習とは、教師データがない学習であり、データの構造を明らか

にする性質がある。例としては、主成分分析、K-Means 法、自己組織化マップなどがある。教師なし学習の結果においては、人がその結果を解釈するものとなる。しかし、結果に対して人間がおかしいと思った点を修正する手段はなく、パラメータを替えて再計算させるなどの手段しかない。

この中で、半教師付き学習という不完全な教師データに基づき、機械学習を行わせるという研究が進んできた。

「知的インタラクティブシステム」は、この教師あり学習と、教師なし学習の中間をいくシステムであるといえる。この知的インタラクティブシステム的な考えた方を大きく取り入れているのが、3章の「ひっぱりくん」4章の「こうぞうくん」である。データに基づき機械により大枠の分析結果を提示してくれる。しかし、なにが重要であるかという情報は、人間が与えていくしくみを取っている。

1-2 先行研究と先行的社会調査ツール

社会、顧客の声を正しく聞くとことは、コトラーの「マーケティング・マネージメント」[24]にもあるように、マーケティングの基本要素である。ソーシャルリスニングは、社会、技術の変化によって、新しく出てきたマーケティング用語である。ソーシャルリスニングの概念と実際の調査事例が書かれている著書にティーン D. ラパポートの「リッスン・ファースト」[9]がある。

データ収集と可視化では、ソーシャルリスニングのための Web ツールとしては、ユーザーローカル社の「Social Insight¹」、ホットリンク社の「クチコミ@係長²」、トライバルメディアハウス社の「Boom Research³」などがあり、既に商用サービスが展開されている。Web 上でデータ分析が可能な形になっており、特別なソフトのインストール無しに使えるようにしている。しかし、探索的、発見的なアプローチを特に取っているわけではなく、データ収集やデータ集計に重きを置いたものになっている。

データマイニングでは、データ分析ツールとしては、R 言語がある。R 言語は、オープンソースの統計処理プログラミング環境である。R 言語⁴は、オープンソース故に無料であり、永続的にデータ分析ができる。また、活発に最新の統計手法がいち早く取り入れられたパッケージが公開されているため、人気が高い。インタプリタ型の言語で、逐次実行ができるため、試行錯誤を行うことができるようになっている。しかし、CUI（コマンドラインインタフェース）のため、マウスを使った直感的な操作である GUI（グラフィカルユーザインタフェース）ができず、コマンドの習熟を必要とする。

テキストマイニングでは、商用データ分析ツールである DIAMining⁵、Text Mining Studio⁶、TRUE TELLER⁷などは、GUI で作られており、直感的な操作ができるようになっている。商用のため、利用料が高いということもあるが、R と商用データ分析ツールは、信頼性の高い手法をメインにして、仮説検証に重きを置き、探索的発見を重視していない。

データ可視化とマイニングを目指したツールに、KeyGraph アルゴリズムを実装した Polaris⁸がある。KeyGraph アルゴリズムは、グラフを作成し、低頻度の重要語を発見するアルゴリズムであるが、このグラフ構造を可視化したのは Polaris である。しかし、KeyGraph アルゴリズムは、パラメータ設定後、自動的にグラフが作成され、そのグラフ構造をインタラクティブに動かせないものになっている。そのため、パラメータの設定と期待するグラフ構造とのミスマッチが起きやすく、探索的な思考を妨げるものになっている。

¹ <http://social.userlocal.jp/>

² <http://www.hottolink.co.jp/service/kakaricho>

³ <http://boomresearch.tribalmmedia.co.jp/>

⁴ <http://www.r-project.org/index.html>

⁵ <http://www.mdis.co.jp/products/diamining/>

⁶ <http://www.msi.co.jp/tmstudio/>

⁷ <http://www.trueteller.net/>

⁸ <http://www.chokkan.org/software/polaris/>

テキストマイニングのグラフィカルな統合環境を目指したプロジェクトに、TETDM プロジェクト [25]がある。人工知能学会の近未来チャレンジに 2010 年に採択されたプロジェクトであり、知的インタラクティブシステムの研究者が参加し、それを意識したものになっている。そして、探索的データ分析のために、データ分析技術をモジュール化し、柔軟に組み合わせることを可能にする仕組みを用意している。様々な発見的データ分析手法が使えるのは便利であり、Java での実装のため、マシン環境への依存度は低い。しかし、なんでも組み合わせられるというシステムが逆に、利用者を迷わせるものになっている。開発者視点、研究者視点になっていて、利用者視点ではないものになっている。

以上のように、先行研究として、ソーシャルリスニングのためのツール、データ分析ツールが挙げられる。

1-3 研究目的

以上を踏まえて、私の研究目的は、「私たち」の分析のために、ソーシャルネットワーク上にあるデータを、「私たち」が分析を行うためのツールとして開発することである。言い換えると、ソーシャルリスニングのためのツールである。

本研究では、データの取得から分析まで行い、分析のためのプラットフォームを作る。これは、「私たち」のための分析ツールであるため、自身の暗黙知と引き出す形になっている。そのため、仮説検証型ではなく、仮説探索型である。

データを可視化することは、外化の一種である。外化することで、分析者に、新たな発想をもたらす。そして、インタラクティブに有益なデータを提示する仕組みをもたせることで、データと向き合えるものになっている。インタラクティブ性を持たせることで、試行錯誤を容易にし、知識発見プロセスを加速させるようなものにする。

また、多くの人が簡単に使えるようにする、という点で、なるべく Web ブラウザ上でつかえるものにし、OS に依存しない利用環境を目指す。

1-3-1 開発ツールの位置付け

開発するそれぞれのツールの位置づけは以下のようにになっている。データ取得ツールは、Rawler および GooCaBooca であり、データ分析ツールが、ひっぱるくん、こうぞうくん、ふかぼりくんとなる。

Rawler

ソーシャルリスニングのためには、Web 上にあるデータの取得に関しての煩わしさが問題となる。試行錯誤をし、知識発見のためのサイクルを回すためには、Web 上のデータ取得 (Web スクレイピング) のためのコストを下げることは必須である。そのための Web スクレイピングツール「Rawler」を作成した。これは、プログラミング言語となっているため、柔軟性があり、再利用も可能なものとなっている。言語の側で、煩わしい処理を受け持っているので、Web スクレイピングに特化したものになっている。

ひっぱるくん

Web 上のデータは、テキストデータが多い。そのため、テキストデータの分析のためのツールを開発した。テキストデータを入力データとし、はじめは、単語のランキングが表示される。真っ白な画面に、単語のランキングから自身で選択した単語を画面に配置していく。そして、配置した単語をクリックすると、その単語と関係している単語を表示する。そこで、自身の興味に従って、気になる単語を「ひっぱる」ことで、単語間に線が引かれる。このように単語を引っ張っていくことで、グラフを自身の手で作っていく。ただ、引っ張っていくだけでは自身の考えを反映するだけで、データとの対話にならない。

そのため、自動的に既に表出させた語句とのリンク関係も表示するようにしている。このような形で、データから自身の考えを表出化させ、それとテキストデータとの関係性を分析者自身に考えさせるツールにしている。

こうぞうくん

バスケット形式のデータを分析するためのツールとして開発した。Sliverlight で作成し、Web ブラウザ上で使うことができる。対象とするデータは、バスケット形式を採用しているため、購買履歴のようなデータから、分かち書き済みのテキストなどまで対象としている。表形式でデータ全体の構造を見るというツールになっており、クラスタリングを行える。可視化としては、データの全体像を把握するために、横列を頻度の大小で、上から順に頻度が大きい物から小さいものへと並べる。そして、縦列は、相関関係が高いものが並ぶようになっている。このように表現することで、頻度の大小と、クラスタがどのような形になっているかを表現するツールになっている。クラスタリング結果を分析者がそのまま受け入れるのではなく、自分でそれぞれのアイテムを移動できるようにしており、その移動を反映してクラスタリングもできるようにしている。そのため、自動的にデータから生成された教師なしクラスタリングから、自分で調整した、半教師付きクラスタリングが行えるツールになっている。

GoocaBooca・ふかぼりくん

GoocaBooca は、画像のイメージを調査するためのアンケートサイトで、スマートフォンなどを使い、気軽に答えることができるツールである。40 枚程度の画像の Yes (Good)、No (Bad) を答えてもらう。画像を調査対象とするのは、「雰囲気」などの暗黙知の領域に関するものを調査対象とするためである。

ふかぼりくんは、アンケート分析ツールである。Sliverlight で作成し、Web ブラウザ上で使うことができる。名前の通り、アンケートのふかぼりを可能とするツールである。アンケート項目の属性での絞り込みを簡単に行えるようにし、その絞り込みにもとづき、各分析手法を扱えるようにしている。また、変数のクラスタリングも行えるようにしており、複数の変数の統合を行うことができ、その変数を新たに加えて、同様の分析が可能になっている。また、アンケートにつきものの自由回答文の分析も回答と絡めた形でできるようにしている。

このふかぼりくんは、GoocaBooca のアンケート結果も分析できるようにしている。

1-3-2 期待される成果

インターネットが普及し、人々のコミュニケーション手段として、Web、SNS が欠かせないものになってきた。そのようなコミュニケーションデータは大量に蓄積されている。しかし、そのようなデータの分析手段がまだ発達段階である。そこへの貢献が、本研究によって期待される成果である。

また、データ分析は、データだけですべてが分かるわけではない。人間が解釈することで、データ分析は価値を持つ。機械がサポートできることは、SECI モデルでいう「表出化→連結化」プロセスのみである。知識創造プロセスとして、「表出化→連結化」プロセスのみでは不十分であり、それ以外のプロセスを回すための探索的にデータ分析を行えるデザインが必要である。このような問題に対してのソリューションとして、知的インタラクティブシステムという形で、インタラクティブにデータ分析を行えるツールを作成した。

これらのツールを使うことにより、社会の声を掴み、探索的な社会分析をすることを可能にする。



第二章 データ取得ツールの 開発と実装

第二章 データ取得ツールの開発と実装

Web を使ったコミュニケーションが当たり前となっている。そのため、社会を分析するためのデータは、既に Web 上に多くあるのだが、分析するためには、まずそのデータを取得しなくてはならない。Web からのデータ取得を、Web クロールあるいは Web スクレイピングという。2つはニュアンスが異なり、Web クロールはサイト全体の取得というイメージであり、Web スクレイピングは、必要なところを取り出すというイメージで用いられている。ここでは、主に、Web スクレイピングの方に主眼を置き、ページにある欲しいところを取得するというを考える。

今日のプログラミング言語には、HTML の解釈を行うライブラリが既にあり（Ruby の Nokogiri⁹、Python の Scrapy¹⁰など）、プログラミング言語に習得した人なら簡単に Web からのデータ取得を行えるだろう。しかし、データの取得のためには、ハードルが高い。Web サイトであるので、HTML/CSS の習得が必要であり、プログラミング技術の習得も必要である。プログラミングの中でも、制御構造、繰り返し構造の基本から、通信とそれに伴う例外、テキスト処理関係、取得したデータの保存方法、長時間データ取得には、スレッドプログラミングの知識が必要である。そのため、ただ特定の Web サイトのデータがほしい、という欲求に対して、水準が高すぎるスキルが要求される。探索的な社会調査のためには、データ取得に関してもトライアンドエラーの試行錯誤が必要である。そのためにも、簡単にデータ取得できるツールが必須であり、そのためのツール Rawler を作成した。これは、ツールとして位置づけているが、実行内容の記述を行うため、プログラミング言語である。

2-1 概要

Rawler は XAML (XML 派生系) で記述するテキスト処理を得意とする Web スクレイピング用のプログラミング言語である。Rawler のプログラミングスタイルは、テキスト処理の連鎖を XML の木構造で記述していくスタイルである。

Rawler のプログラムコードである XAML の編集は、Visual Studio¹¹を使う。Visual Studio を使うのは、Rawler 自身が、Visual Studio での入力補助が機能するように設計されているためである。(図 4)

入力補助により、使える機能の一覧が表示され、その中から選ぶだけなので、XAML の記述量が多いが、ユーザ自身の記述量は少ない。

⁹ <http://www.nokogiri.org/>

¹⁰ <http://scrapy.org/>

¹¹ <https://www.microsoft.com/ja-jp/dev/default.aspx>

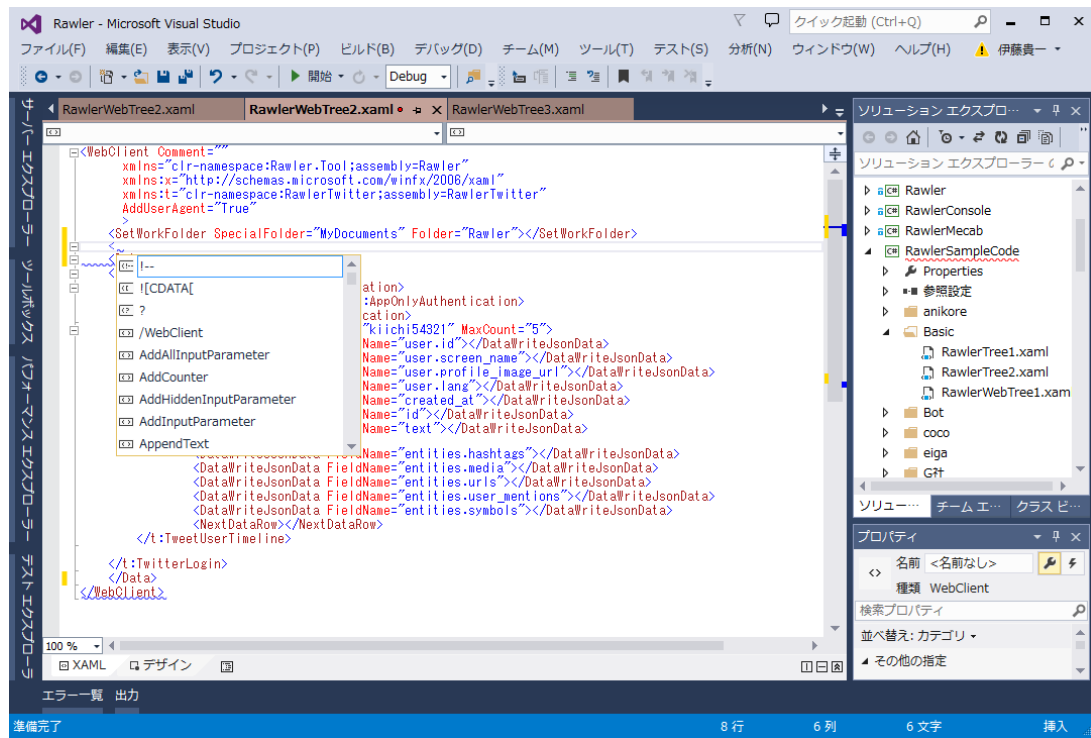


図 4 Visual Studio の XAML エディタによる入力補助

Web スクレイピングに関しては、頻出するパターンには対応しているため、対象の Web ページからのデータ取得方法のみの記述で済む。Web サイトのページングにも対応し、次に読むべき URL を指定すると、一連のその HTML に対しての処理が終わった時、その次の URL を元に同じ処理をするようになっている。また、Rawler は、複数の結果が得られた時、それを繰り返すというルールで、プログラムの for 文などの繰り返し構文を排除している。そのため、対象の Web ページの取得したいところのテンプレートを記述するだけで、そのテンプレートにマッチしたもののすべて取得することができる。また、一覧ページと詳細ページで構成される Web サイトよくあるが、そのような形にも Rawler は、一連のデータフローを記述するだけでよく、本質的な挙動のみで記述できるので、見通しのいいものになる。Rawler の実行は、専用クライアントを使い実行する (図 5)。XAML のところに Rawler のコードを記述し、「Run」ボタンを押して実行する。このソフトは、Web サイトにアクセスすること¹²で起動するようになっている。

¹² <http://kiichi.azurewebsites.net/Rawler/publish.htm>

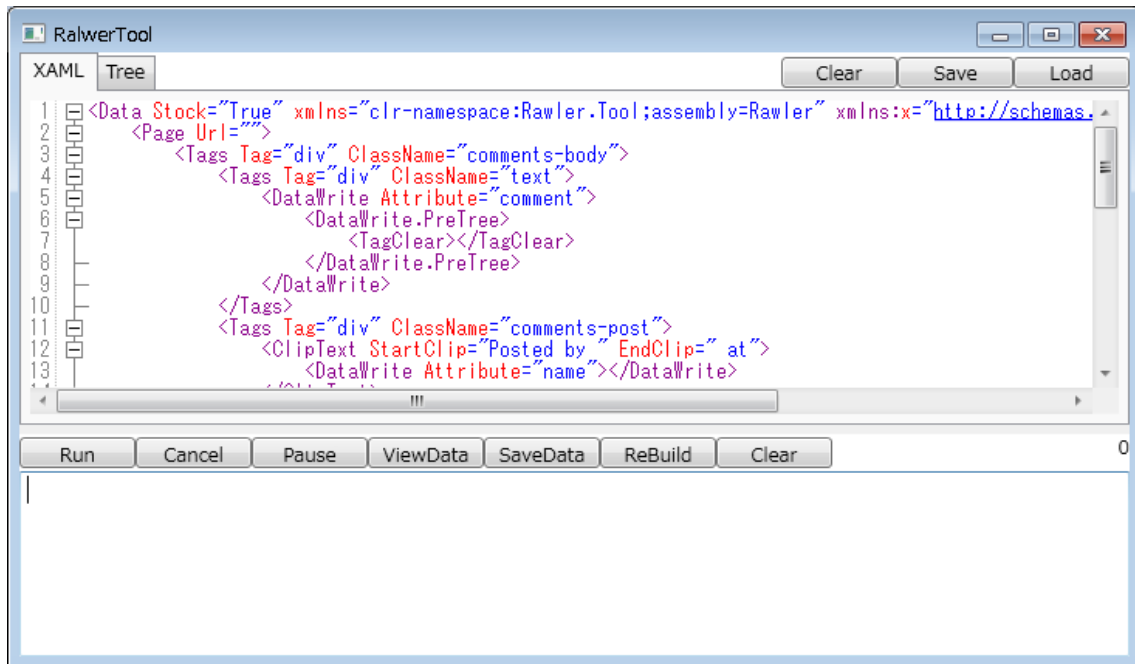


図 5 クライアント実行環境

2-2 目的とコンセプト

2-2-1 入力補助の効くテキスト処理プログラミング言語

Web サイトスクレイピングにおいて、行うことは限られている。Web ページの取得を行い、その HTML の解析を行い、その結果に基づき、ファイルに書き出すというのが基本的な流れである。そのため、Web サイトスクレイピングとは、「Web ページの取得」「HTML の解析・取得」「ファイルの読み書き」の 3 種から成立しているといっていよい。しかし、例えば、「HTML の解析・取得」にしても、対象とする Web ページの HTML が違えば、取得のための指定は、対象ごとに違うので、「HTML の解析・取得」の内容は、それぞれ違う。それぞれ違うことへの適応のために、プログラミング言語的な柔軟性が必要である。

この時、Web スクレイピングという形で、私の理想としては、入力補助の効くことが望ましい。今まで、C#と Visual Studio による入力補助の効くプログラミング環境でプログラミングをしてきたため、入力補助ができないことは、あり得ないと思っている。そのため、これを取り入れたいと思った。少しの入力で、候補が現れて、その選択だけでプログラミングができるなら、それに越したことはない。

2-2-2 テキスト処理の合成基盤

Web サイトスクレイピングはテキスト処理であるので、テキスト処理の連鎖で記述できるようなプログラミング言語であることが望ましい。例えば、「Web ページの取得」は、URL（文字列）を変数として、HTML（文字列）を返す関数として考えることができ、「HTML の解析・取得」は、HTML（文字列）を変数として、指定の解析結果（文字列）を返す関数である。また、「ファイルの読み書き」は、ファイル名（文字列）を変数に、そのファイルの内容（文字列）を返せば、ファイルの読み込みであるし、文字列を変数にして、それをファイルに書き込めば、それがファイルへの書き込みである。文字列の入力が文字列を返すという、テキスト処理の連鎖で、Web サイトスクレイピングは記述できる。これは図

示すると、図 6 のようなデータフローになる。

Rawler の基本的な発想は、テキスト処理の連鎖を記述していくための基盤がほしいというものである。既存のテキスト処理の連鎖を記述するものとしては、Unix のパイプが広く知られている[26]。プログラミング手法としては、メソッドチェーンがある。Rawler と、Unix のパイプ、メソッドチェーンとの違いは、木構造と単線との違いである。Unix のパイプ、メソッドチェーンは、一つのテキストソースに対して、ひとつの関数しか記述できない。一方、Rawler は、Web スクレイピングの常として、一つの HTML ソースから、複数の箇所を削り取っていく（スクレイプ）ものなので、一つのテキストソースに対して、複数の関数を実行できるようにしている。Unix のパイプやメソッドチェーンは、プログラミング上で、変数として格納することで、一つのテキストソースに対しての複数の処理を実現している。

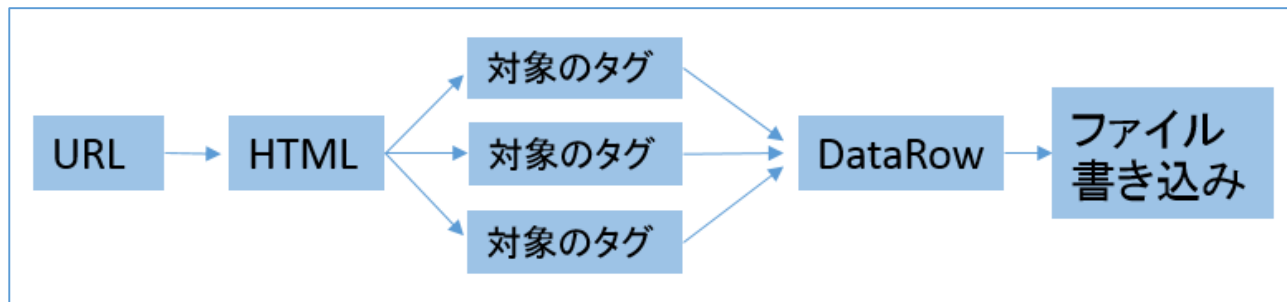


図 6 Web スクレイピングのデータフロー

2-2-3 データフロープログラミングとビジュアルプログラミング

このようなテキスト処理の連鎖、データ処理の流れ（データフロー）を有向グラフで記述していくプログラミング手法をデータフロープログラミング[27]という。データフロープログラミングの利点は、機械の手続きの記述ではなく、データを主体にして記述できるため、本質的なデータの処理の記述に集中できることである。これは、関数型プログラミング言語の利点を含んでいる。関数型プログラミング言語の利点とは、「なぜ関数プログラミングは重要か」[28]で述べられているように、モジュール化と、関数の合成にある。小さな関数を組み合わせて、再利用性を高く、柔軟性を持つということであり、Web スクレイピングを行う Rawler では、Web スクレイピングのための小さな関数を予め用意しておけば、その組み合わせでプログラミングできるということである。

データフロープログラミングは、処理のフローの有向グラフを記述するため、グラフィカルに表現した、ビジュアルプログラミングと相性がよく、しばしば利用され、LabVIEW¹³、Simulink¹⁴などがある。データフローの有向グラフを作るためには、「ノードの配置」「ノードの設定」「ノード間の接続」を設定していく必要があるが、グラフィカルな表現では、ノードとノード間の接続の表現がわかりやすく表現できる。そのため、ビジュアルプログラミングは直感的な操作でプログラミングしていくことは利点であるが、GUI 操作によってすべてをプログラミングしなくてはならない。そのため、テキストによるプログラミング言語のようなコピーアンドペーストはできず、GUI 操作を繰り返さないといけないという欠点がある。

2-2-4 XML でデータフローを記述する

データフロープログラミングの有向グラフを作るためには、「ノードの配置」「ノードの設定」「ノード間の接続」を行わなくてはならない。この記述を行うために、Rawler は、XML を使う。XML とは、汎用的なマークアップ言語であり、テキスト記述により、木構造のデータを表現できるものである。この XML により、「ノードの配置」「ノードの設定」「ノード間の接続」を自然に表現できる。特に、「ノード

¹³ <http://www.ni.com/labview/ja/>

¹⁴ <http://jp.mathworks.com/products/simulink/>

間の接続」について、タグの親子関係によって、暗黙的に決まることが大きい。図 7 のように、通常のプログラミング記述では、ノード間の接続のためのコードを必要とする。しかし、XML の記述では、ノード間の接続について記述は必要としない。このことは、データフローの記述を簡素化することに寄与し、プログラマ自身への負担を減らし、プログラム全体の見通しの良さをもたらす。

C#での木構造の表現	XAMLでの木構造の表現
<pre>Node root = new Node() { Name = "root" }; Node node1 = new Node() { Name = "node1" }; Node node2 = new Node() { Name = "node2" }; Node node3 = new Node() { Name = "node3" }; root.Add(node1); node1.Add(node2); node1.Add(node3);</pre>	<pre><Node Name="root"> <Node Name="node1"> <Node Name="node2"/> <Node Name="node3"/> </Node> </Node></pre>

図 7 木構造の記述

しかし、XML でのデータフローの記述には問題がある。前述のように、XML は一つの親が複数の子を持つこと木構造のデータである。そのため、一つの祖先から、多数の子孫の記述はできるが、複数の親を一つの子にするような記述はできない。

これは、一つのデータソースから、広がっていくデータフローを記述できても、複数のデータソースを集約するデータフローの記述はできない。しかし、木構造は、一つの根から、向きを逆向きに捉えようと、複数のノードを集約するノードを記述できている。そのため、Rawler では、XML だけの記述でデータフローを表現するために、複数のデータフローを集約するノードを XML の祖先に近いところに配置し、そのノードに対応する子は、自身の祖先をたどり、その集約ノードにたいして、アクションを命令するルールを設けた。これは、子-祖先関係に予め対応関係を持たせることで、自動的に、ノード間の接続が行われるようにしている。このようにして、木構造の XML で記述するだけで、拡散と収束のデータフローを表現することが可能となる。

<pre><!--DataRowを受けるData。最終的に内容を保存する。--> <Data FileName="foo.txt"> <!--Urlで指定したWebページを取得--> <Page Url="http://www.hogehoge.com"> <!--liタグのことをすべて取得し繰り返す--> <Tags Tag="li"> <!--終了時、祖先のDataに列情報を送る--> <DataRow> <!--aタグの全体を取得する--> <Links VisbleType="Tag"> <!--aタグで囲まれたところを取得する--> <Links VisbleType="Label"> <!--内容をDataRowに「Label」という名前を付けて登録--> <DataWrite Attribute="Label"></DataWrite> </Links> <!--aタグのUrl部分の取得--> <Links VisbleType="Url"></pre>
--

```

<DataWrite Attribute="Url"></DataWrite>
</Links>
</Links>
</DataRow>
</Tags>
</Page>
</Data>

```

図 8 図 6 を Rawler のコードで表現

2-2-5 祖先ノードを参照することによるパラメータの自動設定

Rawler では、収束のためのノード間の接続以外にも、祖先ノードをたどるという処理をすることで、その文脈に沿ったパラメータの自動的な設定も一部のノードに関して行われる。例えば、Web ページの Link の取得をする処理である。Web ページのリンクは、HTML での記述の簡便さから、相対パスで記述されていることが多い。これを Web ブラウザは、読み込んだページの URL と組み合わせて、絶対パスを作り出している。この仕組みと同様のことを Rawler も行っている。リンクの取得ノード「Link」は、祖先にある、Web ページ取得ノード「Page」を参照して、その現在の URL を取得し、相対パスを絶対パスに変換して返すようにしている。このようにすることで、文脈としてわかりきっていることの記述をしなくて済むようにしている。同様に、ノード「Page」は、祖先にノード「Page」があるとき、その「Page」の URL を、HTTP リファラに自動的に入力し、Web サーバから HTML の取得を行う。これは、極力ブラウザと同じ挙動をするようにするためである。

また、祖先ノードを参照するとことで、設定の共有も行っている。祖先ノードを参照することで、共通の設定を使い挙動することができるようになっている。設定が複数の箇所に散らばってしまう自体を避ける働きがある。例えば、Web ページを取得する設定を司るノード「WebClient」は、必ずノード「Page」より祖先に配置しなくてはならないものであるが、ノード「Page」による、Web ページ取得時に使う、UserAgent の設定や、連続してアクセスしないようにするための Sleep 設定を設定できるようになっている。この祖先ノードを参照するという挙動は、オブジェクトプログラミングにおいては先に new(インスタンス化)をしたものを扱っているという感覚に近いものであり、直感的な挙動であるといえる。

2-1-6 テキスト変数

Rawler では、ノードは、一つの親の文字列を受け取り、それを変形して、子に渡すということを基本とする。これは、一つのテキストを変数とし、テキストを返す関数であるといえ、単線的なデータフローしか表現できないことを意味する。多くのテキスト処理はこれで十分であるが、一つの入力だけでは対応できない問題も多い。テキスト変数は、一つのノードに対して、複数の入力を可能とする記法である。(図 9)

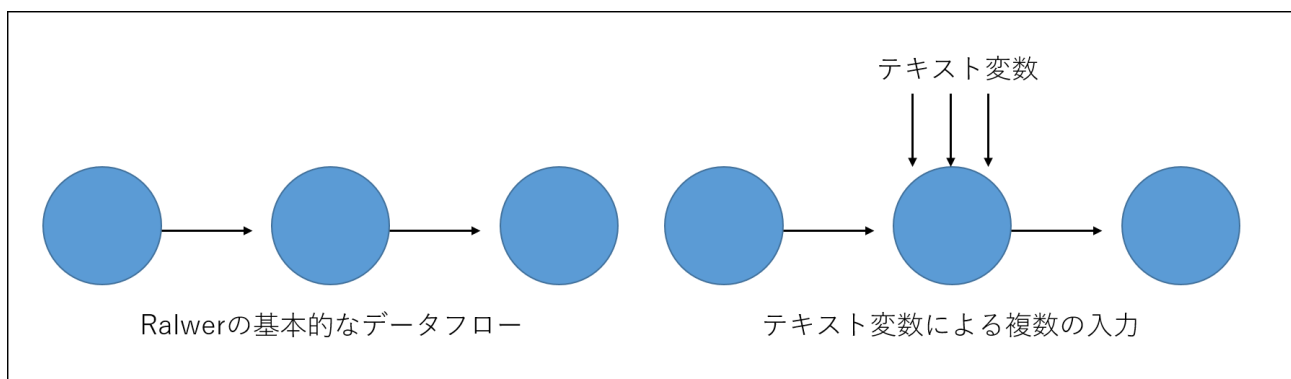


図 9 テキスト変数の概念図

テキスト変数は、記法としては、[変数名]という形で[]で囲ったものを変数とし、その名前で、祖先に格納されている変数にアクセスし、それを変数として入力する（図 10）。あくまで[]で囲ったところが、変数として展開されるので、テンプレートとしての使い方ができる。テキスト変数は、ほかのところで生成した文字列を使うことができるため、これは、複数のデータフローを一つのノードにつなげていることを意味する。

```
<KeyValueStore>
  <SetKeyValue Key="foo" Value="bar"></SetKeyValue>
  <FileSave FileName="[foo].txt">
    <Page></Page>
  </FileSave>
</KeyValueStore>
```

図 10 テキスト変数の使用例

2-2-7 同時実行、並列処理

マルチコア CPU が普及する中、マルチスレッドによる並列実行プログラミングは、当たり前ものになっている。データフロープログラミングモデルは、もともと並列実行のために考えられてきたプログラミングモデルであり、同様のプログラミングモデルである Rawler も並列実行できる。構文もシンプルで、ノード「Concurrent」で囲んだ、子ノードが同時実行される。図 11 では、「Concurrent」で囲まれた、複数の「Page」ノード以下が、同時実行される。Rawler では、任意のノードは、親のノードのテキストを変数とした関数であるので、このようなことが可能である。同時実行において問題となるのは、共有メモリに対する書き込みである。これは、Rawler においては、データを蓄積する「Data」ノードのような集約ノードが問題になる。初期設計において、シングルスレッドを前提としていたため、集約ノードは、並列実行に対応していない。そのため、新たにアクターモデル[29]で設計しなおした「RawlerParallel」を追加し対応した。このような並列処理機構を持つことにより、Rawler は、並列して複数の Web サイトからのデータ取得や、巨大なデータに対する並列処理を簡素な記述で行うことが可能である。

```
<!--Concurrentの子は同時実行される-->
<Concurrent>
  <Page Url="http://www.hoge.com"></Page>
  <Page Url="http://www.foo.com"></Page>
  <Page Url="http://www.bar.com"></Page>
</Concurrent>
```

図 11 並列実行を行うコード例

2-2-8 宣言型で関数型言語的なしくみ

以上述べたように、Rawler は XML でデータフローを記述する DLS（ドメイン特化言語）であり、宣言型で、関数型言語の性質を持っている。記述の組み合わせで Web サイトのデータ取得から、データのクリーニングを行うことができる。出力が次の入力になるというパイプラインを作るため、変数の一時作成の必要がなく、複数出力されるものは複数繰り返すので、For 文のような構文は必要としない。そのため取得したデータの変形の連鎖という形でデータを主体として表現できる。これはデータフロープログラミングの利点を取り入れた結果であるといえる。XML による記述、祖先参照とテキスト変数の仕組みにより、柔軟にデータフローの接続を少ない記述で表記できるようにした。このような仕組みを持つことで、具体的な手続きを隠蔽し、設定のみを記述できるので、全体の見通しが良く作成することができる。

2-3-1 開発の経緯

これまで mixi や Pixiv などのウェブサイトからのスクレイピングツールを作成してきた中で、個別のスクレイピングツール開発ではなく、汎用性の高いツール開発の需要が高まった。それは、個別のツール開発のままでは、スクレイピングは単純なパラメータの設定であるのに対し、プログラムをプログラムとして動かすための記述が多いためである。そのため、プログラミングとして見通しが悪く、修正がやりにくい。以上を踏まえ、効率的にプログラミングが出来るライブラリの必要性が高まった。

様々な機能を追加し洗練させていく中で、C#の中で使うライブラリではなく、独立した DSL（ドメイン特化型言語）になった。ツール単体で、Web からのデータ取得、加工、ファイルへの書き出しが行えるものである。

以上のように、当初はライブラリという形で開発を始めたが、開発が進むにつれ、フレームワーク、DSL というというものになっていったという経緯がある。

2-3-2 開発の目標

開発の目標は、Web からのデータ取得を簡単に行える手段を持つことである。現代において、Web データの取得はデータ分析の基本的な能力の一つである。なるべくシンプルな記述で、入力補助が機能し再利用性の高いものが求められる。そして、Web からのデータ取得だけでなく、ファイルへの保存、データの加工までも行え、さらに、各種 Web サービスの API にも対応し、欲しいデータをすぐに取りに行けるというものを目指す。データ取得には繰り返し処理も含まれる。そのため柔軟な繰り返し構文を使うことができるものを作ることを目指す。

2-4 設計と実装

Rawler の設計と実装について述べる。

2-4-1 XAML を使った言語

C#は、開発環境の Visual Studio の組み合わせがあつてのプログラミング言語であり、他のプログラミング言語との優位性は Visual Studio に寄るところが大きい。C#は強い型付けの言語であり、それゆえに、入力補助が強力に機能する。そのため、開発ツールには、同様の入力補助を機能させたいと考えた。しかし、入力補助を機能させるための大量のコードはデバッグの手間を考えると書きたくない。そのため、自力でコンパイラを作ることはせずに、Visual Studio の XAML エディタに乗る形でプログラミング言語を設計した。XAML とは、Microsoft 社が開発した、XML の派生系である。XML の記述により、主に UI のデザイン（WPF）を行うことができる、宣言言語である[30]。XAML は、実体としては、C#のオブジェクトの設定の記述であり、オブジェクトの木構造を効果的に記述する手法である。図 12 のように、木構造を記述しようとする、通常、ノードの宣言と、ノード間の関係を記述する必要があるが、XAML による記述では、XML が木構造なので、ノードの宣言だけでよく、直感的な記述になる。XAML のテキストデータと C#のオブジェクトに相互に変換でき、XAML のタグは、C#オブジェクトの new を宣言しているだけである。そのため、XAML は C#と同じく強い型付けの言語である。強い型付けのため、入力補助も効果的に機能し、XAML エディタは文法エラーも検出するので、実行前に文法的エラーを直すことができる。XAML では、このような強力なエディタ機能を使える上、実装自体は、構文解析を必要とせず、C#のクラスを作ればよいのである。そのため、Rawler は、C#のクラスを作ることが、関数を作ることになっている。

C#での木構造の表現	XAML での木構造の表現
------------	---------------

<pre> Node root = new Node() { Name = "root" }; Node node1 = new Node() { Name = "node1" }; Node node2 = new Node() { Name = "node2" }; Node node3 = new Node() { Name = "node3" }; root.Add(node1); node1.Add(node2); node1.Add(node3); </pre>	<pre> <Node Name="root"> <Node Name="node1"> <Node Name="node2"/> <Node Name="node3"/> </Node> </Node> </pre>
--	---

図 12 C#と XAML での同じ表現

2-4-2 RawlerBase と RawlerMultiBase

Rawler のすべてのノードは、基底クラスの RawlerBase からの継承をしている。RawlerBase は、親ノードへの接続と、子ノードへの接続情報を持ち、これによって木構造を表現している。RawlerBase にある、Run メソッドを実行することで、親から子に Run を深さ優先的に実行していくという仕組みである。これは、オブジェクト指向の多態性を利用した設計である。

RawlerBase は、一番初期に設計したもので、単数での処理を前提としており、複数ではない。そのため、複数を処理するために、RawlerMultiBase がある。これが複数処理を行うもの基底クラスとなる。RawlerMultiBase には、クエリを使うための機構をもち、複数生成された文字列群から必要とする文字列を絞り込むことができるようになっている。

諸機能の実装は、RawlerBase を継承し、親のテキストから、子に渡すテキストを作ることによって達成される。つまり、Run メソッドを override すればいい。そのため、新しい機能を実装するのは容易である。この RawlerBase を継承すれば、Rawler の木構造を使えることができるので、機能の組み合わせが容易になる。その意味では、RawlerBase は機能の組み合わせをするための基板であるといえる。

2-4-3 祖先参照の実装

Rawler では、ノード「Data」とノード「DataWrite」といった祖先-孫関係の対応関係を多用している。祖先孫関係の対応関係は、C#に備わっている型システムを利用している。Rawler の木構造から、祖先を辿っていく中で、指定した「型」に初めにマッチしたノードのメソッドを実行することで実装している。ノード「DataWrite」は、祖先にある、クラス「Data」を探索し、そのメソッドを実行することで、ノード「DataWrite」の親にあるノードの文字列をノード「Data」に書き込む。このような形で実装している。そのため、クラス「Data」を継承したクラスを作ると、それも同様にノード「DataWrite」の接続先になる。クラス「Data」このような形で、型システムを利用している。基本的には、はじめに見つけた指定の型という制約を持たせることで、接続に関する指定を省略できるようにしている。しかし、指定のノードに接続してほしい時があるので、そのときは、ノードにつけられた Name を使い結びつける対象を探すようにしている。

2-4-4 テキスト変数の実装

Rawler での変数の設定、取得も、祖先参照を利用して実装している。祖先にノード「KeyValueStore」を配置し、ノード「SetKeyValue」で親ノードのテキストを「KeyValueStore」に、Key をつけて保存し、ノード「GetKeyValue」でそのテキストを取得できるようにしている。(図 13)

<KeyValueStore>

```

<!--変数の設定-->
<SetKeyValue Key="foo" Value="bar"></SetKeyValue>
<!--変数の取得-->
<GetKeyValue Key="foo">
    <Report></Report>
</GetKeyValue>
</KeyValueStore>

```

図 13 祖先参照をつかい変数の設定と取得

初期段階では、保存するファイル名を動的にしたいときなどの、変数を使用したいときは、ノード変数を **Rawler** で書けるようにしていた。しかし、これでは、記述が冗長すぎるという問題があった。FileNameTree の最後の値、「bar.txt」をノード「FileSave」の FileName に入力している。この処理のために、5 行を使っている。また、プロパティごとに対応した Tree を用意しないといけないため、プロパティ自体が増えるという問題を抱えていた。(図 14)

```

<KeyValueStore>
    <SetKeyValue Key="foo" Value="bar"></SetKeyValue>
    <FileSave>
        <!--ノード変数を指定-->
        <FileSave.FileNameTree>
            <GetKeyValue Key="foo">
                <AppendText Footer=".txt"></AppendText>
            </GetKeyValue>
        </FileSave.FileNameTree>
    </FileSave>
</KeyValueStore>

```

図 14 ノードの変数を木構造で表現

この問題を解決するため、プロパティに直接、GetKeyValue と同等の変数取得を記述できるようにした。記法としては [Key] という形で、変数名を [] で囲む (図 15)。この記法で、祖先のノード「KeyValueStore」からテキストの値を受け取り、その部分に当てはめる。テンプレートのように機能し、2 つ以上の変数でも書き換えられるようにしている。このようにして、URL の生成などに使えるものになっている。このテキスト変数は、WPF の DataBindings 構文に影響を受けており、その簡易表記であるといえる。

```

<KeyValueStore>
    <SetKeyValue Key="foo" Value="bar"></SetKeyValue>
    <FileSave FileName="[foo].txt">
        <Page></Page>
    </FileSave>
</KeyValueStore>

```

図 15 テキスト変数を使った動的なファイル名の指定

このような形で、一行に収まる形で、変数を使うことができるようになり、かつ、テキストの生成をできるようなものになっている。

2-4-5 実装したノード

Rawler では、目的を達成するため約 200 種類のノードを実装している。Rawler の最も基本となる Rawler.dll には、大分類として、「Web」「Data」「Html」「Text」「IO」「Condition」「Enumeration」「Utility」「NPL」の 8 種のノードを実装している。

Web

Web ページの取得、その周りに関することを実装している。「WebClient」ノードは、Web アクセスの基本的な設定を司り、ログイン情報などのクッキーを保持し続ける。「Page」ノードは、親のノードのテキストを **Url** として、その Web ページのテキストを自身のテキストとする。「NextPage」ノードは、親ノードのテキストを **Url** として、祖先の **Page** にその **Url** を送る。これにより、Web ページのページングに対して効果的にデータ取得ができるようになっている。「GetCurrentPage」ノードは、祖先にある **Page** にアクセスし、そのテキストを取得する。これは、元の HTML が必要となったとき使用する。

Data

データを書き込むためのノード群である。「DataWrite」ノードは、親のノードのテキストを、**Attribute** をつけて、祖先の「DataRow」に送る。「DataRow」は、複数の「DataWrite」ノードの入力を受けて一行分のデータを蓄積し、終了時に、祖先の「Data」に送る。

「Data」ノードは、データを蓄積するためのノードであり、子孫の「DataRow」からの一行分のデータを受け入れる。そして、終了時に、ファイルへの書き込みをする。「Data」ノードは、蓄積型であり、メモリにため続ける。そのため、対象データが大きいときは、「Data」を継承したクラスの「FileSave」ノードを使う。こちらは、「DataRow」からの命令が入ったときに、ファイルに書き込みを行う。このような機構をもって、Rawler では、Data の書き込みを行う。

Html

Html の解釈を行う。「Tags」ノードでは、指定した **Tag** にマッチする部分を複数返す。タグの **Id** 名、**Class** 名で絞り込むことができる。これを組み合わせることで **Html** からほしい部分を取得できる。Rawler.Tool.Html を XAML の参照に加えると、**Html** のタグの名前で **Tags** を指定できるようにしている。これにより、インテリセンスによる入力を補助が効くようになり、取得したい **Html** に近い形で取得できるようにしている。「Link」ノードは、**Html** にあるリンクを取得するものであり、**Html** の中の特に、<a>タグに特化したものである。祖先の「Page」ノードから現在の **Url** を取得し、取得したリンクの相対パスを絶対パスに自動的に変換し返す。

Text

テキスト処理を行うノード群である。「Document」ノードは、テキストを Rawler で扱えるようにするノードである。「ApendText」ノードは、親の文字列の前後に文字列を追加する。「Replace」ノードは、置換、「Trim」ノードは、前後の空白文字の削除を行う。「Split」ノードは、テキストを複数に分割する。「ReadLines」ノードは、改行区切りでテキストを読み込むことをする。

IO

主にファイル操作を行うノード群である。ファイルの存在の有無のチェックや、フォルダの作成、ファイルの書き込み、読み込みを行う。「FileReadLines」は、テキストファイルを一行ずつ読み込む。「TsvReadLine」は、タブ区切りのテキストファイルを読み込むときことをサポートする。子孫に、「GetTsvValue」ノードで、列名を指定すると、指定した列名の値を取得できる。

Condition

条件分岐を行う。指定の条件にマッチしたら、子を実行する、しないを行う。「Switch」ノードでは、子に配置した「Case」ノードから当てはまるものを実行するという制御を行う。そのほか、文字列に変化があると実行される「ChangeText」ノード、同じ文字列の場合には実行しない「Once」ノードなどがある。

Enumeration

繰り返しを制御する。「Loop」ノードは、子を無限に繰り返す。「Range」は、プロパティの **Start,End** で指定した間の数値を生成する。「Iterator」ノードは、「Iterator.SourceTree」で追加されたテキストを繰り返す。「GroupBy」ノードは、プロパティの **Key** で指定したもので、まとめられ、子では、その集合を扱えるようにしている。「GroupBy」は、祖先にある、RawlerMultiBase が終了したとき、繰り返しが終了したときに実行される。

Utility

Utility は、その他の機能になる。標準出力を行う、「Report」ノードや、Sleep 処理を行う「Sleep」ノードがある。Rawler で変数を扱うことができるようにしている「KeyValueStore」ノードがある。

NPL

Rawler は、テキスト処理に特化しているので、簡単な自然言語処理ができるようにしている。

2-5 開発上の工夫

2-5-1 XAML ベースのシステム

開発上の工夫としては、この Rawler は、もともと、ビジュアルプログラミングのツールを作ることと考えて作っていた。それが今の XAML で表記のみで完結することが最大の工夫である。祖先参照の仕組みを取り入れたことで、木構造の XML で、発散と収束のデータフローを表現できるようにした。XAML というテキストベースになったことで、コピーアンドペーストが可能になった。これがビジュアルプログラミング的な UI での設定をするものになると、実装は大変だし、それに見合った効果はえられない。すべてがテキストになったことで、全体の見通しが良くなった。Web スクレイピングは、パラメータの設定が肝であるが、そのパラメータを見やすくなっている。また、Rawler においては、コピーアンドペーストは強力に機能する。Rawler は、どこを切り取っても、親を変数とする関数であるので、タグの開始終了を適切にコピーアンドペーストすると、それ自体が関数となる。そのため、関数の再利用が可能となる。

Visual Studio には、強力な XAML エディタがあり、実装したクラスを解析して、入力補助に自動的に補完要素にしてくれる。Visual Studio は、本来は有料のソフトウェアであるが、現在、無料で利用できる仕組みがある。そのため、Windows マシンに限定されるが、導入コストは小さくなる。XML は、記述の正確性を要求され、終了タグを必要とするため、機械が読むには最適であるが、人間が記述するには難点があるが、記述としては長くなり、Visual Studio の XAML エディタを使うことで、ユーザの負担を減らすことができる。

Web ページからのデータ取得の常として、対象の Web ページの構造が変わると取得できないという問題が孕む。そのため、Web ページからのデータ取得をハードコーディングするのは、よくなく、スク립ト的に、外部化するのがいい。その点でも、XML 形式の Rawler は、最適である。

2-5-2 Nuget 対応

また、Nuget 対応も行った。Nuget とは、Visual Studio のパッケージバージョン管理ツールであり、Net からパッケージを Visual Studio に導入することができる。そのため、Nuget で「Rawler」と検索するとパッケージが見つかり、そのまま導入できるようになった。それ以前は、Github に上げたソースをダウンロードして、コンパイルするか、不定期にあげていていた Wiki からダウンロードして、参照に加える必要があった。しかし、Nuget に導入することによって、検索するだけでできるようになり、簡単にできるようにした。

2-5-3 XML の記述の短縮化

Rawler は、XML で記述するため、どうしても記述が長くなりがちである。そのための短く記述できる方法を探ってきた。祖先・孫関係の対応関係は、ルールを分かっているならば、接続のための記述を必要としないため、記述の短縮化とコードの読みやすさにつながっている。これは変更にも強く、一か所の変更だけで済む。また、祖先参照の機能は、祖先・孫の対応関係さえあればよく、その間に、適切なデータフローを記述すればいいため、適切なテキストの変形の記述に注力すればいい。これはスピーディに Web からのデータ取得に寄与する。この自身のノードから、親、祖先と辿っていくことは、文脈を調べていることに相当するし、実行時になって、その文脈から挙動を決めていることになる。そのため、任意のノード以下をカットアンドペーストしても、祖先参照の対象は、自動的に切り替わり、祖先のノードを切り替えが容易であることを意味する。これは、あるページに対して適切に取得できるコードを組み立ててから、繰り返し方式、保存方法などを切り替えることが簡単になっている。

また、複数の返り値があるものに対してのクエリ機能を追加は、集合に対しての処理を記述できるようになったので、記述の短縮に貢献した。元々、全てを繰り返すことをルールにしたので、複数の返り値が多い、Web からデータ取得において、有効だった。しかし、この XAML という形式で記述する上で、構造化プログラミングの三大要素である、順次、繰り返し、分岐のうち、順次、繰り返しは容易な記述であるが、分岐が特に苦手であった。分岐する内容が増えれば増えるほど木構造が深くなっていき、記述が長くなってしまふ。しかし、クエリ機能の追加によって、分岐を表現することになる上、コード上の見通しが良くなることになった。

テキスト変数の追加も記述の短縮に効果を発揮している。本来は、変数の取得・変形は Rawler による記述を必要としていたのだが、テキスト変数の追加により、コード全体の見通しをよくしたうえで、変数を扱うことができるようになった。しかし、これは、XAML には存在しない機能なので、入力補助が機能せず、実行時になってバグが判明するもののため、記述にプログラマ側に負担を強いるものである。

2-5-4 入力フォーム機構

学部生たちに Rawler を試験的に使わせてみたところ、XAML コードのどこを書き換えればいいのかかわからず、間違ったところを書き換えて、実行ができないということが発生した。このような事態をなくすために、Rawler で入力フォーム機構を記述できるようにした。入力フォームにより入力された情報は、テキスト変数を使い使うことができる。図 16 のようなコードを実行すると図 17 のフォームが立ち上がり使用者は必要なところを自身で入力し、Submit ボタンを押すと実行する。RawlerForm.Properties の内容に基づきフォームが生成されている。RawlerForm ノードは、KeyValueStore を継承しているため、自身に入力した変数を蓄積し、そのアクセスには、GetKeyValue ノードや、テキスト変数によって行うことができる。このようにしたこと、安全に特定の部分を書き換えることを可能にし、スクレイピングとテキスト処理のアプリケーションを作ることができることとなる。多数の使用者にとっては、Rawler をつかいプログラミングをする必要がなく、フォームを入力することだけでよくなり、開発者にとっては、入力フォームの作成から、スクレイピングを一つのソースファイルに記述できるので利便性が高い。

```
<Root Comment=""
  xmlns="clr-namespace:Rawler.Tool;assembly=Rawler"
  xmlns:x="http://schemas.microsoft.com/winfx/2006/xaml"
  xmlns:form="clr-namespace:RawlerView.Form;assembly=RawlerView"
  Title="フォーム機構の使用例"
>
```

```

<SetWorkFolder SpecialFolder="MyDocuments" Folder="Rawler"></SetWorkFolder>

<form:RawlerForm SettingFileName="setting">
  <form:RawlerForm.Properties>
    <form:TextProperty Key="InputText" Value="text" Name="Text"
Help="help1"></form:TextProperty>
    <form:FileProperty Key="OpenFile" Name="使うファイル"
FileDialogType="OpenFile"></form:FileProperty>
    <form:FileProperty Key="SaveFile" Name="保存先"
FileDialogType="SaveFile"></form:FileProperty>
  </form:RawlerForm.Properties>
  <!--RawlerFormは、KeyValueStoreを継承しているため、GetKeyValueで変数の取得が行える-->
  <GetKeyValue Key="InputText">
    <!--内容を標準出力する-->
    <Report></Report>
  </GetKeyValue>
  <!--テキスト変数を使い、フォーム変数へのアクセス-->
  <Data FileName="[SaveFile]">
    <FileReadLines FileName="[OpenFile]">

    </FileReadLines>
  </Data>
</form:RawlerForm>
</Root>

```

図 16 フォーム機構の記述

図 17 作成されたフォーム

2-5-5 拡張性

XAML は、外部の DLL を取り込める。

図 16 の「`xmlns:form="clr-namespace:RawlerView.Form;assembly=RawlerView"`」の記述は、dll を取り込む記述である。このようなことができるため、拡張的な機能に関しては、別の DLL を作ることにした。Rawler 本体には、基本機能のみとし、なるべく参照する DLL を減らすようにした。

RawlerBase を継承することで、Rawler に乗っかることができるので、簡単に実装できる。親のテキストを使い、自身のテキストを生成すればいいので、自然言語処理の形態素解析もその枠組で実行することができる。

また、TwitterAPI などの WebAPI を使う環境も作ることができる。RawlerTwitter プロジェクトが

それであり、これを使うと、面倒な Twitter のデータの取得がシンプルな形で行うことができ、その上で繰り返し処理などへの対応が可能になる。図 18 のコードは、フォーム機構と連携をし、フォームから保存先のファイル名と、ツイッターの ScreenName（複数）を入力すると、その入力に従い、対象の ScreenName のつぶやきを取得し、ファイルに保存される。TwitterAPI では、多くの情報を返してくれるが、すべてを分析の時に必要とするわけではない。このような形で、宣言的に必要とするところを指定できることは強みである。

```
<Root Comment=""
  xmlns="clr-namespace:Rawler.Tool;assembly=Rawler"
  xmlns:x="http://schemas.microsoft.com/winfx/2006/xaml"
  xmlns:tw="clr-namespace:RawlerTwitter;assembly=RawlerTwitter"
  xmlns:p="clr-namespace:RawlerParallel;assembly=RawlerParallel"
  xmlns:form="clr-namespace:RawlerView.Form;assembly=RawlerView"
  Title="ScreenNamesのリストからつぶやきを取得する"
>
<SetWorkFolder SpecialFolder="MyDocuments" Folder="Rawler"></SetWorkFolder>
<!-- フォームを出現させ、保存先と対象のScreenNamesを入力させる -->
<form:RawlerForm SettingFileName="setting.txt">
  <form:RawlerForm.Properties>
    <form:FileProperty Key="FileName" Name="File名" DefaultExt=".txt" Value="tweet.txt"
FileDialogType="SaveFile"></form:FileProperty>
    <form:TextProperty Key="ScreenNames" Name="ScreenNames" Lines="5" Help="ScreenNamesの
改行区切りデータ" DoSave="False"></form:TextProperty>
  </form:RawlerForm.Properties>
  <!-- 指定したファイル名で保存 -->
  <FileSave FileName="[FileName]">
    <!-- 取得する対象をQueueで管理する。 -->
    <p:ParallelQueue CompletedFileName="[File:FileName]_CompletedIdList.txt"
IsSingle="True" >
      <!-- 取得する対象をQueueに送る -->
      <ReadLines Lines="[ScreenNames]">
        <p:Enqueue></p:Enqueue>
      </ReadLines>
      <!-- ツイッターにログインする。アプリ認証を使う -->
      <tw:TwitterLogin>
        <tw:TwitterLogin.AppOnlyAuthentication>
          <tw:AppOnlyAuthentication></tw:AppOnlyAuthentication>
        </tw:TwitterLogin.AppOnlyAuthentication>
        <!-- 上流のQueueから一つのScreenNameを取り出す。 -->
        <p:Dequeue>
          <Report Header="Start:"></Report>
          <!-- 親のScreenNameからつぶやきの取得を行う -->
          <tw:TweetUserTimeline ParentUserIdType="ScreenName" SleepSecond="1.2">
            <tw:TweetUserTimeline.CompletedTree>
              <p:DequeueCompleted></p:DequeueCompleted>
            </tw:TweetUserTimeline.CompletedTree>
          </tw:TweetUserTimeline>
          <!-- 一行分の定義。Json形式のテキスト -->
```

```

        <DataRow>
            <!--必要なところを抜き出し-->
            <DataWriteJsonData FieldName="user.id" />
            <DataWriteJsonData FieldName="user.screen_name" />
            <DataWriteJsonData FieldName="user.profile_image_url" />
            <DataWriteJsonData FieldName="user.lang" />
            <DataWriteJsonData FieldName="created_at"/>
            <DataWriteJsonData FieldName="id"/>
            <DataWriteJsonData FieldName="text"/>
        </DataRow>
    </tw:TweetUserTimeline>
</p:Dequeue>
</tw:TwitterLogin>
</p:ParallelQueue>
</FileSave>
</form:RawlerForm>
</Root>

```

図 18 ツイッターからのデータ取得

2-5-6 オープンソース

Rawler は、オープンソースの分散バージョン管理システムである GitHub で公開している。今のところ、多くの人と共同作業的に開発を行っているわけではなく私一人で開発を行っているのだが、オープンソースにしているのには理由がある。一つは、中長期的なサポートを考えると、オープンソースにしたほうが、自分がすべてのメンテナンスをしなくてはならないことから開放されることである。また、ソースコードが最大の挙動に関するドキュメントであるので、詳しい挙動を示すためにもオープンソースにしている。Rawler は、一つ一つのクラスが様々な細かいことをやっている。それをブラックボックス化しないためである。さらに、Rawler は、拡張性で述べたように RawlerBase を継承することで、自作のクラスを Rawler 上に乗っけることができ、XAML では、外部 DLL を読み込めるため、その自作のクラスを活用できる。そのような開発のためにも、オープンソースとして、作成サンプルの提示と自身による拡張をできるようにしている。

2-6 運用実績

ブログの記事の収集、ブログのコメント欄の収集、mixi、Pixiv など SNS サービスからのデータ取得、Twitter のつぶやきデータの取得など、様々な形で利用した。また、YahooAPI を使って、テキストのキーワード抽出データの作成も行った。第四章の Mixi からのデータ取得もこれによって行っている。

2-7 開発活動を通じて得た知見

この Rawler の開発動機は、Web スクレイピングのためのツールを作りたいというのがまずあったが、プログラマとしては、オブジェクト指向言語の特徴である多態性を使ったプログラミングをやってみたというのがあった。初期のアイディアとしては、その多態性を使い、テキスト処理の連鎖を起こすライブラリを作りたいというものであった。その狙いはブレなかったのだが、開発段階では、記述をどのようにすべきなのか？ということでは、試行錯誤し続けた。祖先参照は、データフロープログラミング

的には、データの流れを記述としては、イレギュラーであるため、不適切な記法であるといえる。例えば、「Data」は、子孫の「DataWrite」を受けるのだが、本来のデータフローの記述では、様々な処理の末に「Data」にデータを送るため、終着点にすべきである。しかし、「DataWrite」で「Data」に送るべきものは多数あるため、複数のものをまとめる形で、「Data」を書かなくてはならない。そのため、本来は、グラフで処理の連鎖を記述するという形になる。

しかし、XAML は XML であり、木構造しか扱えない。グラフ的なものも、仕様上、適切な名前をつけることで、XML での記述は対応可能だが、そのような記述はすごく面倒である。そのため、祖先参照という方たちで、対応した。これにより、1 から多、多から 1 へのフローを混在させる記述であるといえる。このような記述をすることにより、記述の複雑さは削減される。祖先参照は、木構造の祖先をたどるため、その文脈を利用しているといえ、自動的に接続するため、記述の手間を省くものである。これは利点であるが、欠点でもある。それは、どこに接続するかは、実行時になって今までの構築した木の構造によって決まり、明示的指定をしていないため、どこに接続するかがわかりにくいことは欠点である。そのため、データフローが複雑化したとき、可読性が落ちると思われる。明示的指定なしに接続することは便利であるが、トレードオフ関係である。Web スクレイピングにおいては、そこまで複雑なデータフローを扱わないことを前提に設計している。また、この祖先参照の記述は、変数として蓄積するところを初めに宣言するという点で、普通の手続き型プログラミングにおける、初めの変数の初期化と同じである。そのため、手続き的な直感に沿うものであるといえる。

2-7-1 学習コスト

Rawler は、よくも悪くも Web スクレイピング専用の DSL (ドメイン特化型言語) である。入力補助の恩恵があるとはいえ、特殊な言語をイチから学ぶというのは、学習コストが高い。その意味で、当初目標の非プログラマーが使うためのツールという目標は達成できているとはいえない。

しかし、私自身、対象の Web ページに対して、スクレイピングをするために Rawler を書くと 30 分程度でできてしまうので、教えるためのコストと、自分で作ってしまうコストが釣り合っていない。1 つ 2 つの案件なら自分一人で作ってしまうほうが教えるより圧倒的に速い。

やはり、プログラミング的な経験がないと使えない代物かも知れない。学部生たちに初めて覚える言語としては、Rawler は相応しくなく、Java、Ruby といった汎用プログラミング言語を覚えたほうが将来性という点でもよいと思われる。これは、DSL が常に抱える問題だろう。

しかし、Web からのデータ取得ということを考えた時、一つ一つのサイトで自分が勉強して書かなければならないというより、スクリプトの共有環境を整えるというのが正しい道だとも思う。Rawler は、Windows マシンであれば、Web にアクセスするだけで実行が可能のため、環境を整えることには優位がある。その意味で、Rawler スクリプトの Web での共有の仕組みを作ることが必要であると考ええる。

2-7-2 メタプログラミング・動的生成

設計段階、作成中は特に意識していなかったが、使っているうちにわかってきたことだが、Rawler にはメタプログラミング能力が備わっている。メタプログラミングとは、プログラミングのプログラミングであり、プログラミングでプログラミングを生成することをいう [31]。メタプログラミングは、なにが利点かというと、実行時、動的に動作を作ることができることである。そのため、Rawler の場合、HTML の解析からの動的な木構造の構築により、自身の挙動を決める、データの取得の仕方を切り替える、などが可能である。実装で述べたとおり、実態は、C# のオブジェクトの木構造である。そのため、プログラミング上で、その木構造を作ることがメタプログラミングである。Web スクレイピングに必要な雑務を Rawler 側が行ってくれるので、本質的な処理のみの構築だけすむということになる。

このような Rawler のメタプログラミングの性質を活用したのが、RawlerExpress¹⁵というサービス

¹⁵ <https://rawler.net/>

である。これは、ブラウザ上でスクレイピングして欲しい対象の Web サイトの URL のを指定すると、ボタンひとつでスクレイピングしてくれるサイトである。このサービスは、内部的には、対象の Web ページを解釈し、よく繰り返しているものを判別し、それを取得するための **Rawler** を動的に組み立てている。この組み立てた **Rawler** を実行することで、対象のサイトに対してのデータ取得が実行される。ユーザが入力するのは、対象のサイトの URL だけで、完璧ではないものの、ある程度の精度で、データを取得できるので、**Rawler** の使い方を学習する必要がなくなる。もちろん、ある程度の精度でしかないので、適切な形でデータを取得するには、動的生成された **Rawler** にたいして適宜修正をする必要がある。しかし、一から **Rawler** を組み立てるよりもある程度の形は既に出来ているので、修正は小さくていい。この修正の時、**Rawler** は、宣言的言語であり、本質的記述のみにしているのもので、動的生成とはいえ、修正は容易である。

このような形での、プログラミングの動的生成と組み合わせることが、学習コストと利便性とのバランスを取る方法であると思われる。

2-8 まとめ

テキストのデータフローを XML で記述することができる、Web スクレイピングに特化したプログラミング言語 **Rawler** を作成した。Web スクレイピングにおいて必要な、Web ページの取得、Web ページの解釈、データの保存に関して、多数の命令を用意しているため、それを組み合わせることで、素早く対象の Web サイトに対して柔軟にデータの取得ができる。XML のため、終了タグを必要とするため、記述量は、多いが、入力補助に対応しており、プログラマは、命令を選んでいくことで、記述できるため、負担は小さい。また、変数の省略、繰り返しの省略、データフローの自動的な接続をおこなうため、簡潔に記述できる。Web スクレイピングのための Web ページの取得→リンクの取得→Web ページの取得といった複雑化したときに威力を発揮する。欠点は、以下の点があげられる。XML のため記述自体が長くなることであり、if 文のような分岐は、ネストが深くなり可読性が落ちること。テキスト処理に特化しているため、数値計算が全くできないこと。オリジナルの言語のため、ドキュメントがすくなく、習得コストが高いことである。ライブラリの拡張により、ツイッターの API を実行することや、並列集計処理が行えるようになっている。このような性質のため、オープンソースにしている。しかし、**Rawler** 自身による記述能力が低いため、定型処理を大きく外れる場合は C# で記述したノードを用意しなくてはならない。そのため、ノードの数がどんどん増えていくという問題点がある。

Rawler の可能性としては、プログラミングをするためのツールというよりも、メタプログラミングによる動的生成やあらかじめ作っておいたプログラムをフォームからの入力による利用が良いと思われる。このような需要にこたえるため、**Rawler** のコードを共有化するような仕組みが必要だと思われる。

第三章 探索的表出化ツールの の開発と実査

第三章 探索的表出化ツールの開発と実査

本章では、探索的社会調査ツールを構成する第1番目のツールである、探索的表出化ツール「ひっばるくん」について、その目的・コンセプトから、実装そして、実際の社会調査の解析結果について、検討する。

3-1 開発コンセプト

現在、多くのテキストマイニングツールが開発されているが、それらのツールを利用するユーザは、ツールが解析した客観的で絶対的な結果を前提に、その結果を懸命に解釈するだけの受動的な他者ではない。しかしユーザの多くは、マイニング過程をツールに全面的に依存するのではなく、自らマイニング過程に参加することを望んでいる。それでなければ、望ましい成果は出ないと強く実感している。とすれば、ユーザは解説する主体として自らの問題意識に従って、ツールが合理的に判断した素案と対話しながら自分なりに納得する成果を導き出す、という「自己探求的で対話的な関与」を可能にするツール開発が必要とされる。本論文は、このような新しいコンセプトに基づいて開発した、探索的表出化ツール（その名称を『ひっばるくん』と呼ぶ）について、その実装と評価を考察するものである。

新しいテキストマイニングツールである『ひっばるくん』の開発コンセプトは何か。ここでは2点を明確にする。

3-1-1 複雑ネットワークと柔らかい構造化

あるテキストを品詞ごとの形態素に分けて、その頻度を数えると、その分布は一部の高頻度語と多数の低頻度語で構成されるロングテール[32][34]の分布になる。また頻度に基づいて機械的にグラフを作ると、多数のノードとリンクで構成され、少数の巨大なハブがある複雑ネットワーク[33]になる。これは特殊なデータでないかぎり発生する一般法則である。しかしこの複雑ネットワークがデータとして客観的事実だとしても、それは、人間がそのデータを解釈するときの評価基準である主観的実感に適用とは限らない。

実感として、ロングテールに分布される低頻度語群が重要な評価基準になる場合も少なくない。したがって従来のマイニングツールを利用する場合、事実としての複雑ネットワークを所与として、その事実だけを対象にして、そこに実感を込めて解釈することは、奇跡的な偶然の一致を除くと意外と難しい。ここに従来のマイニングツールの限界が潜んでいる。そこで本研究は「事実と実感」という異なったリアリティの評価基準が共に機能するツールという開発コンセプトを採用した。

つぎに言葉の関係ネットワークが複雑ネットワーク化してしまう事実について、そこでは、複数の言葉の連環が多層的に積み重なっているので、集約化すると巨大なハブと多数のリンクで絡み合うネットワークになってしまう[35]、という問題点がある。そのため、複雑ネットワークを、そもそも多層的に存在する複数の言葉の連環として解釈しやすいように、再び腑分けする作業が重要課題になる。つまり従来の客観的事実としての複雑ネットワークの全体（固い構造）をそのまま解説の対象に設定するのではなく、ユーザの問題意識を基にして、全体を構成する多層性を手がかりに、全体をさらに腑分けする視点が不可欠になる。これが自己探求の視点である。この自己探求の視点で全体の固い構造を腑分けして設定した新しい対象領域において、自己探求の実感とデータの事実とが相互に参照しあいながら、多様な関係性（共起グラフ）がダイナミックに生成されていく過程が「柔らかい構造化[11]」である。この柔らかい構造化こそが、事実と実感のリアリティを融合させる新しいマイニングツールに不可欠な開発コンセプトである。

3-1-2 自己探求と動的な生成過程

本ツールが生み出す共起グラフは、従来のマイニングツールにみられるような機械（合理）的に自動作成されたグラフではない。では、なぜ機械的に作成されたグラフでは不十分なのか。その理由は2つある。まずは、機械的に生成されたグラフでは、ツールが提示する解は絶対かつ不動なもので、それがユーザの実感に合わないとしても、ユーザはその解に妥協せざるをえない。したがってどうしてもそこで思考が止まってしまう。だからこそその妥協を超えるために、人間が解析過程に直接関与する仕組みが必要となる。

次に、自然言語処理的な問題として、形態素分析に基づく共起グラフを作った場合、多数の意味を成さない不要語が含まれる。多くの不要語は一目でわかるが、言語の関係性において構造（文脈）的に不要語だと判定するには、ユーザの解釈が不可欠である。つまり不要語の判定は機械処理との相性が悪いので、現状では、多くのユーザは何度もグラフを作り直す試行錯誤を繰り返さなければならない。とすれば、最初から試行錯誤を組み込んだ自己探求の方向で構造化を試みる方が賢明となる。

自己探求する行為は、本ツールが提示する言語群（高頻度ランキング）からある特定の言語を、自己の問題意識をもとに選択することから開始される。そこから動き出す自己探求は、一つ一つの言語とその関係性を選択する過程であり、またそこに徐々に複雑化・多様化・重層化する共起グラフを生成させる過程でもある。つまり自己探求の過程は同時にダイナミックに共起グラフを生成する過程そのものである。したがって柔軟かい構造化ツールは自己探求とダイナミックな生成過程を表裏関係として共起グラフを可視化するツールである。

3-1-3 実装

探索的表出化ツール（ひっぱりくん）は、手軽に使ってもらうために、特殊な形式のデータファイルを作る必要がない、ただのテキストデータを読み込むだけで、処理をしてくれるように実装した(図 7)。ツールの画面領域はメニューバー、サイドバー、キャンバス、データビュー、単語ランキングの5領域からなる(図 7-A)。メニューバーは入力データの指定や **SAVE/LOAD** 機能、サイドバーはノードの削除や表示の調整などの機能、キャンバスはノードとリンクを描く場所、データビューは文章データを見る場所、単語ランキングは注目単語を指定する場所である。

以下、挙動プロセスを示す。なお、挙動プロセスの番号は図 19-B 以降に対応する。

1) メニューバーの「入力データ」「テキストファイル」を選択し、単語の頻度ランキングを作る。さらに自身の問題意識に基づいて、ランキングから適切な単語を選択し、キャンバスにノードをおく。ユーザ自身で言葉のネットワークを探索するために、ノードはマウスで自由に動かせるようにしてある。そしてノードを選択したとき、その周辺に7個の関係の強い語が時計回りで表示され、データビューにその単語を含む文章が表示される。関係の強さは **Jaccard**¹⁶指標を使い、基本設定では一行分の文章での単語の関係を計算している。単語の後ろにある数字は単語の頻度である。また7個だけの表示では探求する単語にたどり着けないことがあるので、「P」のところをクリックすると次の候補7つを出すことができる。(図 19-B)。

¹⁶
$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

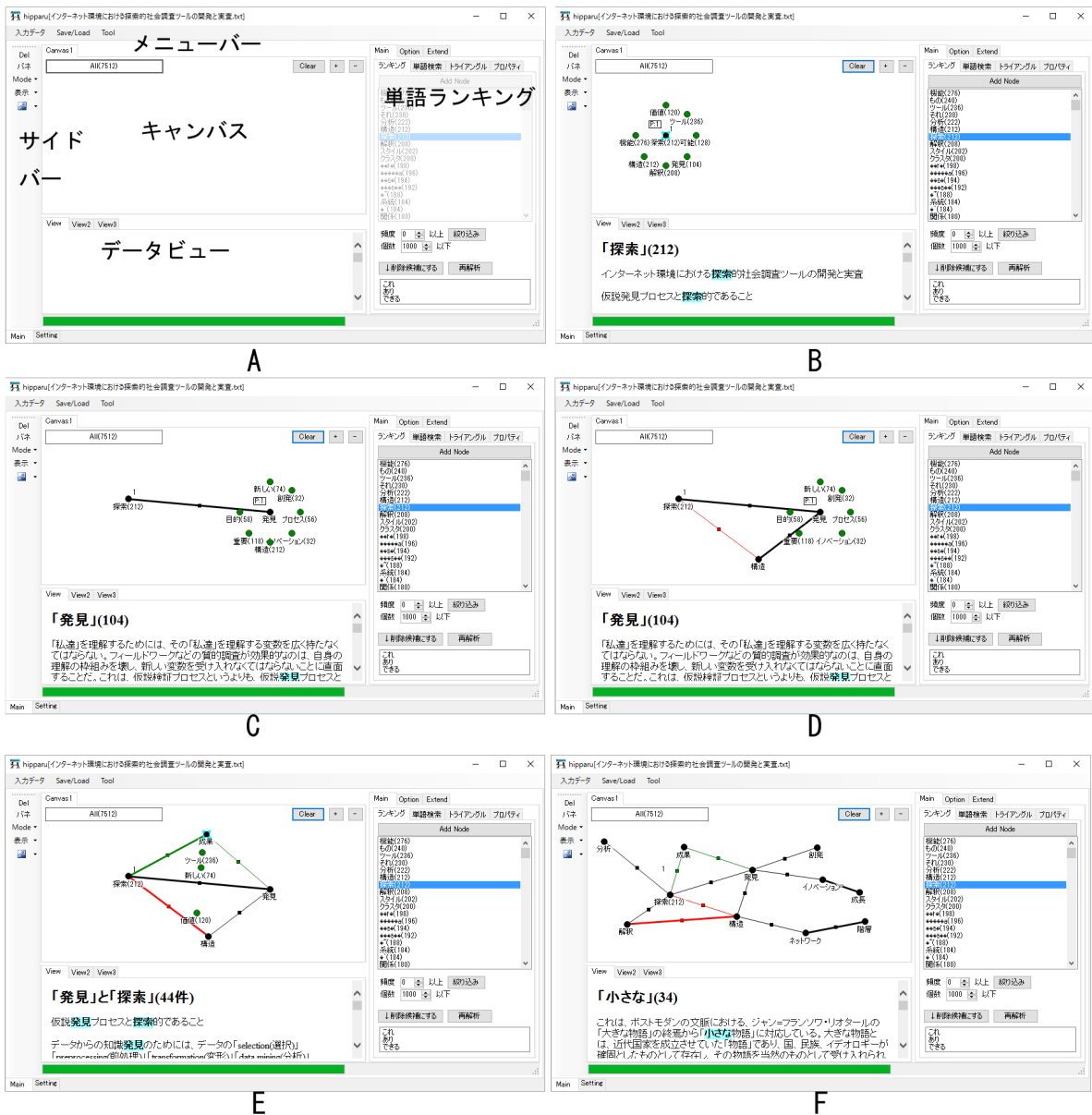


図 19 ひっばるくんの挙動プロセス

- 候補語の中で適切だと思った単語をマウスでつかんで「引っ張る」とノードとして採用され、黒リンク（図上では太線）が作られる（図 19-C）。この新たに引っ張られたノードを選択すると、その周囲に関係の強い語が表示され、同様に引っ張ることができる。
- ネットワークを探索する過程で、すでにノードとした単語と関係の強い単語を選択したとき、新たに赤リンク（図上では細線）が自動的に引かれる。黒リンクが自己探索の線であるのに対して、赤リンクはツールが合理的判断に基づいて引いた線である（図 19-D）。
- リンク上にある黒いボタンを選択すると、リンクで結ばれる 2 つのノード単語と関係が強い単語の頻度ランキングの上位 4 つが表示される。この 4 つの候補ノードも選択すると緑リンク（図上では点線）が引かれる。この単語は、頻度は低いが、ユーザが探索したリンクと両ノードに共通する単語なので、さらに深い探索を誘発する機能をもつ（図 19-E）。
- 上記の探索行為の繰り返しによって、複雑で多様な構造図が作成される。（図 19-F）。

3-2 実査 1 ブログ解析：食の意味世界の気づき

ここでは、探索的表出化ツール（ひっばるくん）の独自性である「柔らかい構造化に基づく気づき」と「自己探求過程がもたらす解釈多様性」を端的に示す3つの事例を紹介する。

実査1では、ある女子学生が自分の食べ歩きブログ98記事を素材に、「ひっばるくん」を利用して、彼女自身の食意識を探索した柔らかい構造化の成果を取り上げる。彼女は、「おいしい」という高頻度単語を中心に探求を行った結果、まず自分の食意識には3つの階層構造があることに気づき、さらに階層ごとに食意識に関する新しい関係性を発見して、今まで漠然と考えていた自分の食意識がいかに構造化されているかを知った。以下、具体的に説明する。

第一階層では、図20のように、「おいしい」という単語と共起関係の強い単語が、①味に関するもの、②香りに関するもの、③食感に関するものに集約され、彼女の「おいしい」という食の意味世界が味覚、嗅覚、触覚という3つの感覚から構成されている事実を発見し、それらが彼女に固有なおいしさの原点であることを知って、十分に納得した。

第二階層では彼女の「おいしい」を構成する味覚、嗅覚、触覚にどのような特徴があるかを知るために、「味」「香り」「食感」それぞれの単語を中心に探求を行った。そのプロセスの中で、「味」と共起関係の強い8個の単語、シンプルさ、豊かさ、辛さ、旨さ、甘さ、酸っぱさ、苦さ、生臭さを発見し、「香り」と共起関係の強い4つの単語、生臭さ、香ばしさ、優しさ、贅沢さを発見した。そして「食感」と共起関係の強い単語はある素材Aとある素材Bの食感の対比が良いもの（例：皮のパリパリと身のプクプク）、反対に食感の対比が悪いもの（例：皮のパリパリと身のペチャペチャ）、素材の食感がアクセントになるもの（例：納豆のネバネバ）の3つの概念が存在することに気づいた。

さらに第三階層では8つの「味」概念、4つの「香り」概念、3つの「食感」概念が、それぞれどのような具体的な食材と共起関係が強いかを探索した。その結果、例えば「味」の8概念の場合では、図

21のように、「味-酸味-レモン」「味-苦味-サザエ」「味-辛味-キムチ」というような関係性が発見され、どのような素材がどのような味・香り・食感であるときに、彼女の「おいしい」に影響しているのかが明確になり、彼女自身の食意識の解釈について非常に大きな気づきとなった。

このように、「おいしい」という高頻度単語を中心に、それに共起する単語を探索していくと、抽象と具象の軸に沿って三層構造の食意識が発見された。彼女自身なにげなく書いているブログであっても、それについてツールを利用して食の意味世界を探索していくと、このような柔らかい構造化が発見されることになった。普段、彼女が漠然と意識している食の意味世界は、ツールが提供する合理的な単語群と対話しながら、一つ一つ自分が納得する意味世界を探索し確認していくと、自分でも思っていなかった関係性や構造が発見され、「自分はそんなことを考えていたのか」という新鮮な気づきをもたらされた。

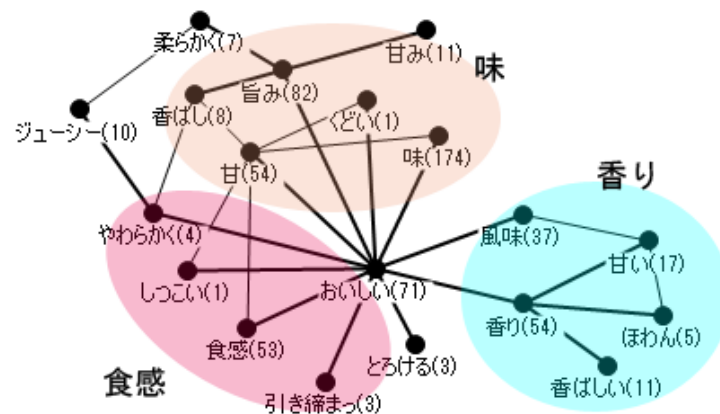


図 20 「おいしい」の単語と共起関係

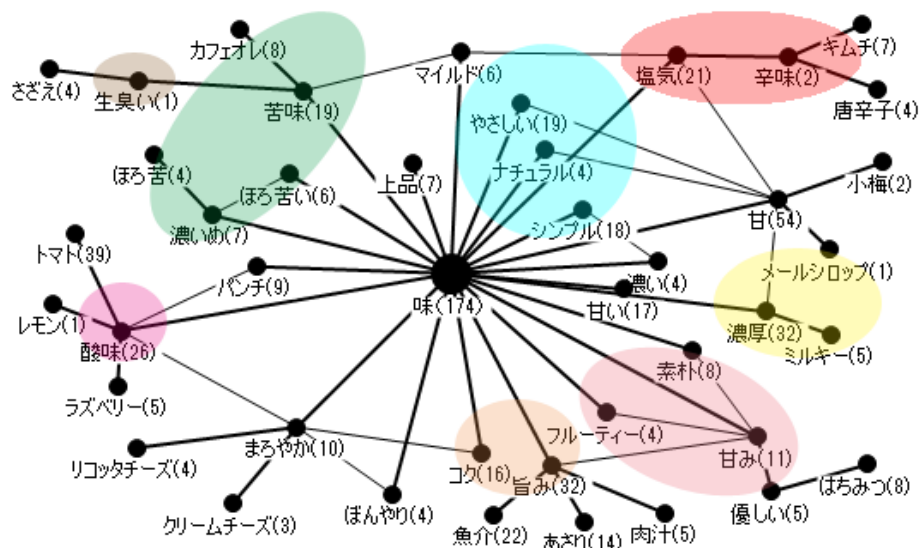


図 21 「おいしい」のさらなる探求

3-3 実査 2 ツイッター解析：家族の解釈多様性

実査 2 では、熊坂賢次¹⁷の 1,000 を超えるツイートを素材に、熊坂研究室に所属する学生に「ひっぱるくん」を利用して、自由に彼の日常世界の解釈を試みさせた。ここでは比較の視点から、2 人の学生の柔らかい構造化の成果を取り上げる。2 人とも、家族関係を中心にして探求を開始している。しかし同じ社会領域を扱っているにもかかわらず、2 つの構造化図は全く異なった成果になっている。

学生 A は、家族の内部構造を詳細に探求して、図 22 のように、多くの家族役割の関係性を発見している。彼にとって最大の発見は「弱いー娘ー可愛いー反逆ーお嬢さんー芝居」の意味連合であり、しかもこの意味連合は赤リンクが多く、それらは彼には予想を超えた意外な発見であった。彼は「先生は、今の家族をこんな風に考えていたのか」を知って納得していた。したがって、ここから、この連合を根拠に、「威厳ある父ー夫」を中心にした核家族が崩壊の方向で変容しているという解釈を施し、学生 A さんの家族関係にみる家族変容論を展開した。

¹⁷ <https://twitter.com/kumakenG/>

他方、学生 B は、図 23 に示すように、家族と外部との関係性に強い関心をもっている。したがって家族役割については常識的な関係性の探求で止め、それが外部のジェンダー論や流行の草食男子論とどのように関係しているかを探求している。その結果、外部での大きな社会的変容が従来の核家族の家族役割の再編成をもたらすはずだ、と解釈し、学生 B などの外部環境からの家族変容論を展開した。

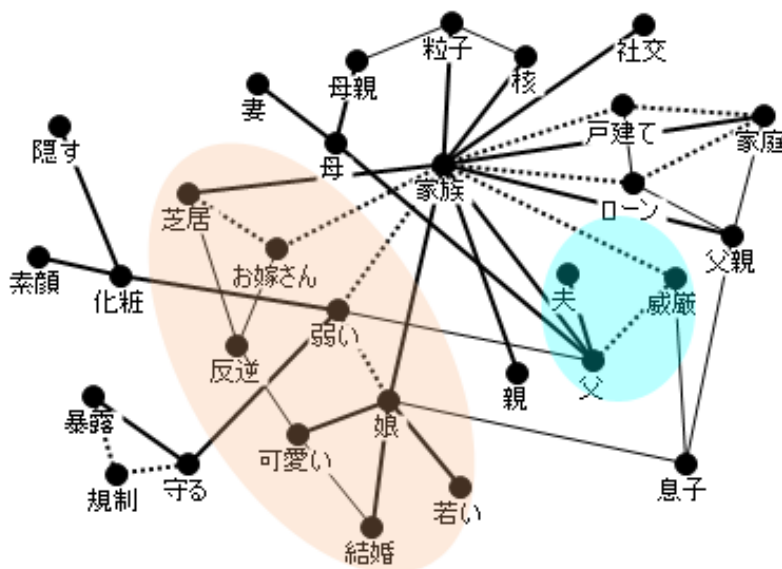


図 22 学生 A が作成したグラフ

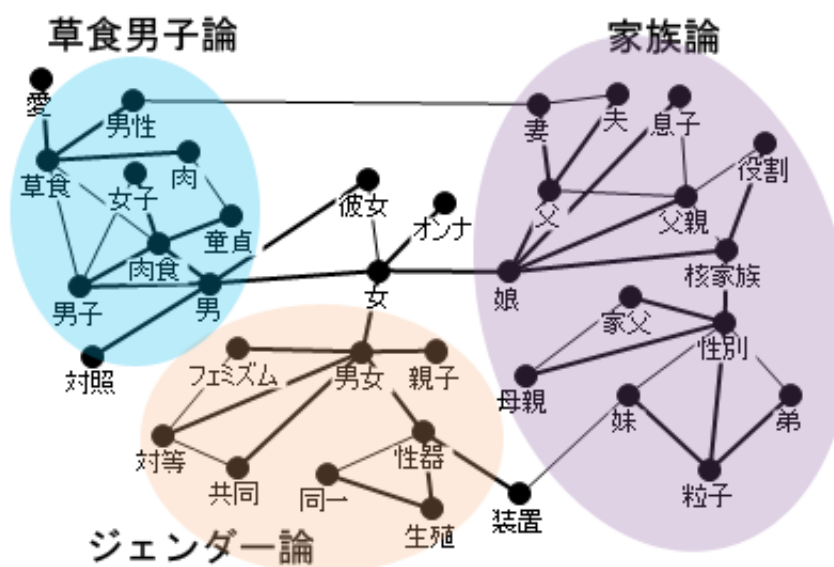


図 23 学生 B が作成したグラフ

このように、同一の素材とツールを活用しても、探求する主体の問題意識によって、探求し解釈される柔らかい構造化には大きな差異がみられる。ここには、本ツールがもつ解釈多様性が1つの重要な特性であることが示唆されている。もちろん、この場合の解釈多様性については、この2つの図を統合する視点を想定すれば、家族と家族役割を焦点にして、その内部構造と外部社会との関係性を一体的に解釈して、さらに大きな柔らかい構造化を実現することもできよう。

3-4 実査 3 エッセイ解析：S 組織のガバナンスの意味付け

「ひっばるくん」は、大量のテキストデータをもとに、ある方向性を定めた上で、そのツールの使用者が、納得するゴールを達成するまで、何度となく試行錯誤を繰り返しながら、データ探索を行うための価値表出化の探索ツールである。

このツールは、2つのゴールをもっている。

1つは、構造生成型であり、これは、事前に予測された構造を、テキストデータから生成させることを目的としたものである。膨大なテキストデータを前に、使用者は、データのフレーム（コンテンツと量の規模）から、どのような構造を引き出したいのか、ある程度の暗黙知とも呼べるイメージをもっている。そのイメージに近似する構造をいかに抽出したり、生成させたりするか、の試行錯誤を繰り返すが、ここでのゴールである。

もう1つは、チャンス発見型である。これは、ある程度の構造が予測されていても、テキストの関連性を探索する過程で、予想もしない関連性やその集合である構造が発見されることである。いい意味で、事前のイメージが否定され、新しい関連性や構造が生成されることを認知する状況がある。これを、大澤幸生はチャンス発見と呼称しているが、それと同一の成果が認知されることである。

この2つのゴールは、排他的関係にあるものではなく、多くの場合、相互補完的な関係にある。テキストマイニングから探索される価値は、1つは構造化の可視化であり、もう1つがチャンス発見の可視化である。

ここでは、テキストデータの規模としては小さく、同一人物が書いたエッセイを解析の対象として、そこから、事前に予測されたイメージが、テキストデータのマイニング過程でどのように探索され、価値の確定がなされるかを検討する。

対象のデータは、岐阜県の公益財団法人理事長熊坂賢次が、2013年から2014年の間で書いた2種類のエッセイである。1つは岐阜新聞の素描というコラムに掲載された9回のシリーズ¹⁸もので、内容はソフトピアジャパンの成果について、市民に分かりやすく情報提供するものである。もう1つのテキストは、BCN¹⁹というIT系のビジネス用の情報誌であり、過去2年間、ほぼ隔月で書いた記事で、内容は、岐阜という地方での情報産業がどのような活躍をしているかを、ソフトピアジャパン関連の事業を基に書き記したものである。現在まで21回の連載をしている記事である。

両方の記事をまとめると、文字数は約3万字であり、テキストの規模としては小さい。しかも同一の人物が書いた記事なので、方向性が一致しており、事前の構造化イメージについては、さほどのブレがあるとは予想されない。無難に、公益財団法人の理事長としての役割から発せられたメッセージであり、かなり固定した視点からのテキスト内容である。

そこで、以下、このように十分に構造が予想されるテキストデータについて、いかなる方法論で、具体的な価値探索がなされるかを、個別の事例として詳細に検討してみる。

¹⁸ 岐阜新聞朝刊 2013年 7月4日・7月11日、18日、25日・8月1日、8日、15日、22日、29日に掲載

¹⁹ <http://biz.bcnranking.jp/list/serial/eye/index.html>

3-4-1 基本キーワードの選択

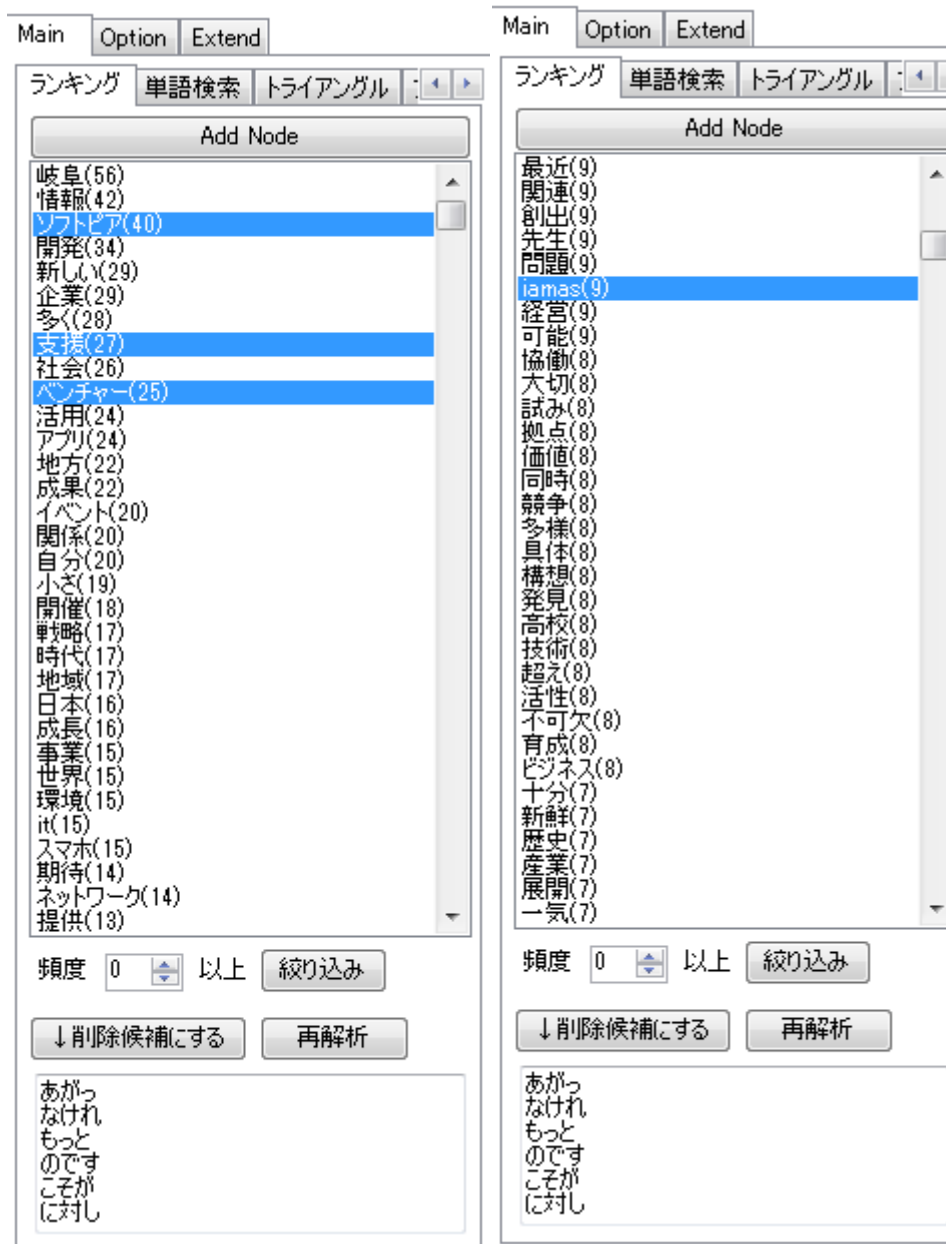


図 24 単語頻度のランキングと削除語の指定

解析の最初にすることは、まず準備作業としては、図 24 の下部にある「削除候補にする」の箇所
で、解析に不必要な用語を設定して、それを削除することである。そのことで、解析がしやすくなる。
ここから解析が実質的にスタートする。

まず、解析の起点となる用語の選択である。ここでは、ソフトピアが起点となる。ソフトピアという
組織の特性を構造化する試みであるから、ここを起点にするのは当然である。その上で、この用語に関
連する用語として、何を選択するか、が重要である。それは、この構造化の方向性を決めるキーコンセ
プトとしてどの用語を設定するか、という問題であり、もっとも重要な作業である。

ここでは、支援・ベンチャー・IAMAS の 3 用語を選択する。この選択は、解釈者の恣意的な選択であり、解釈者の事前の構造化イメージに適合する用語を選択することであり、ツールの「ひっぱりくん」は、そのためのサポートをするだけである。それが、図 25 における関連語（緑色の円環）の提示である。これは、用語の単純集計での頻度構成とは違い、ソフトピアとの関連性が強い序列から随時円環で提示される仕組みを参考にして、選択されるものである。ツールは、純粋に用語の関連度の強い順に提示されるだけで、解釈者は、その序列をみながら、意味あると想定される用語を選択する。

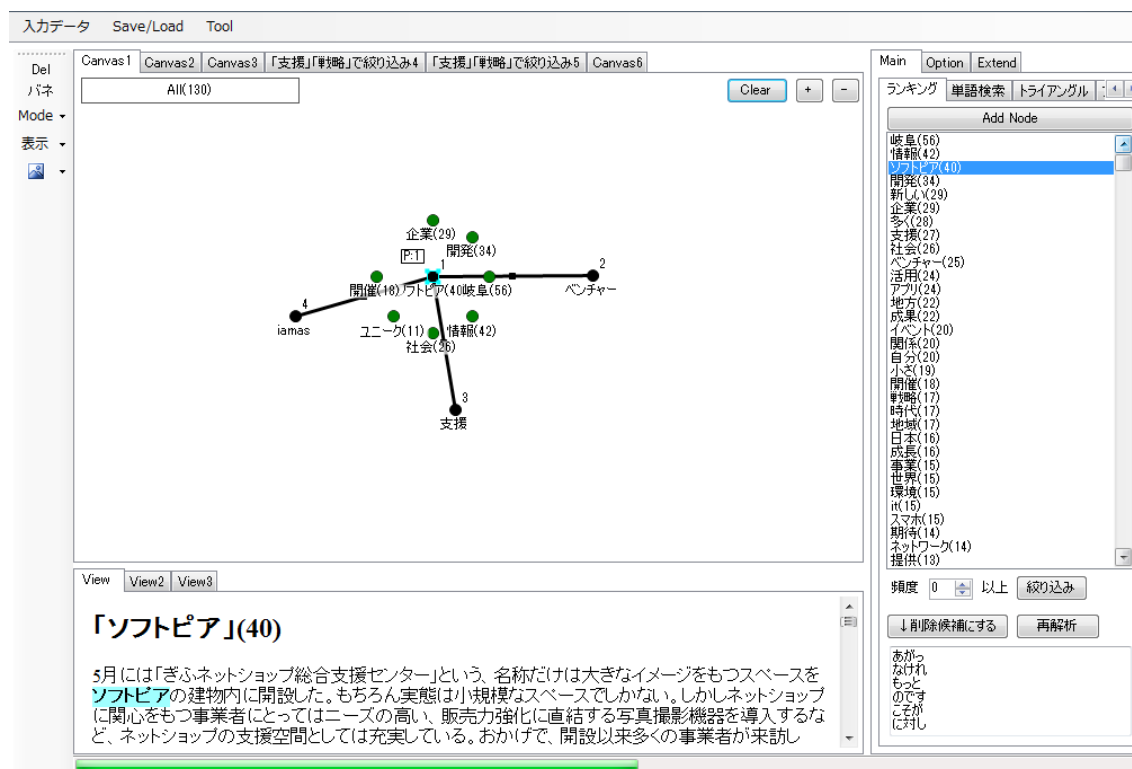


図 25 ソフトピアを起点に開始

ここでは、頻度順位もかなり高い箇所から、「支援」と「ベンチャー」を選択する。「支援」との関連では、「情報」とか「開発」の用語の頻度が高いが、ソフトピアという組織のミッションを考えると、「情報」は一般的すぎるし、「開発」は抽象度としては適切であるが、組織ミッションとしては、ベンチャー企業の「開発」を「支援」することが最重要なテーマなので、ここでは「開発」ではなく「支援」を設定する。そして、「開発」に関連する「ベンチャー」という組織スタイルを重要語として採用する。

もう 1 つのキーワードとして、ここでは、単純な頻度ではかなり低い、関連度としては緑円環の 2 順目に提示される、その意味では、特化係数が高い用語である「IAMAS」を選択することにする。ここでのツールの貢献は、単純頻度（低い序列）と特化頻度（相対的に高い序列）とのデータを提供することで、どの用語を選択することが許容されるか、その許容範囲を提示している点にある。単純頻度では、序列が低いので、それを根拠に「IAMAS」をキーワードに設定することには、かなりの恣意性が想定され、構造化が解釈者のやや強い恣意性に依存することになり、解析の正当性という観点からすると、やや問題ではある。それに対して、特化頻度が提示されることで、2 順目の円環で提示されるまでに、序列が特化頻度では上昇していることから、この用語をキーワードとして設定することが、これによって十分に正当化されるといえよう。このように、解釈の恣意性と正当化という間に、1 つの評価基準を示しているのが、このツールの「単純頻度と特化頻度」という指標である。

3-4-2 構造化の生成ルールの設定

キーワードを起点に、さらに用語の関連性を探索する過程で必要なことは、どのようなルールで関連性を追求するか、である。ここでは、幅優先探索的に、つぎの生成ルールを設定することで、構造化を求めることにする。

ルール 1：キーワードは、それぞれ 3 つの用語を選択する。

ルール 2：選択される用語は、特化頻度が高い用語を優先する。

ルール 3：その特化頻度の高い用語は、可能な限り用語の抽象度を多様化する。

3-1：キーワードと比較して、より抽象度を高くする。これは、他の 2 つのキーワードと基本用語の「ソフトピア」との関連が赤ノードでつながるように仕組む。

3-2：キーワードと比較して、より抽象度を低くする。これは、キーワードをさらに深く説明することを仕組む。

ルール 4：ルール 1 を、さらにもう 1 階層、用語の追求を深める。

これによって、基本用語 (1)、キーワード (3)、ルール 1 ($3 \times 3 = 9$)、さらにルールを展開 ($9 \times 3 = 27$) を求める。以上から、40 用語を選択する。

ルール 5：40 用語を構造化するために、トライアングルの関連性が生成されていない用語は削除する。それによって、構造化が完成すると判定し、その構造を解釈の対象とする。

3-4-3 構造化の生成過程(1)ー全体構造図と階層構造ー

構造化の生成ルールを活用すると、まずは、つぎの図が生成される。

1) 「ベンチャー」には、「岐阜」「開発」「情報」を選択する。「情報」は抽象度が高く、「開発」と「岐阜」は抽象度が低い。

2) 「支援」には、「地方」「成長」「戦略」を選択する。「成長」は抽象度が高く、「開発」と「地方」は抽象度が低い。

3) 「IAMAS」には、「先端」「イノベーション」「新産業」を選択する。「先端」と「イノベーション」は抽象度が高く、「新産業」は低い。

その結果は、図 26 になる。

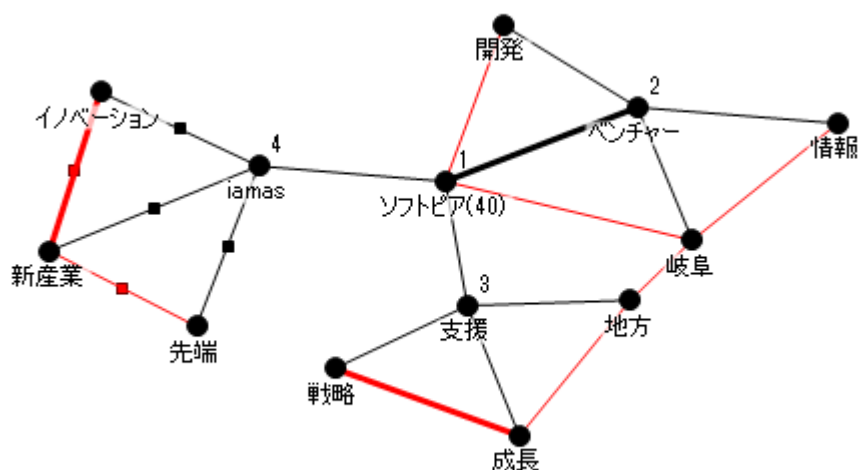


図 26 図 25 から「ベンチャー」「支援」「iamas」を展開

ここでは、キーワードの内部での関連性が、赤ラインで示されるように、5つ生成されている。さらに、基本用語（ソフトピア）との関連性が「ベンチャー」の2つから生成され、またキーワード間については、「岐阜」と「地方」がつながる。このように、意図的に選択した9つの黒ラインに対して、8つの赤ラインが自動的に引かれ、構造化への第1歩が生成されている。

さらに、次のステップとして、9の派生用語について、各3つの用語を選択する。ここでの選択のルールは、上記のものと同一である。

まず、ベンチャーに関連する3用語についての関連性を生成させる。

1)「開発」は、「雇用」「人材」「アプリ」を選択する。ここでは、ベンチャーが「アプリ」開発を行うことで、地域の拠点化を目指しているのがソフトピアであり、それは、新しい「人材」の「雇用」施策と連動している、という視点を重視することで選択した。

2)「情報」は、「スマホ」「環境」「イベント」を選択する。ベンチャーにとって、情報は、具体的には「スマホ」のアプリ開発であり、その拠点化のための「環境」づくりと施策としての「イベント」を開催するのがソフトピアの役割だ、ということである。

3)「岐阜」は、「地元」「ユニーク」「小さ(な)」を選択する。「地元」は開発の「雇用」につながり、岐阜の「ユニーク」な存在として「IAMAS」と連動し、さらに、「小さ(な)」はソフトピアが展開する「成長戦略」に連動して、岐阜ならではの小さな成長戦略拠点として機能している、という関連性が生成されている。以上を図15に示す。

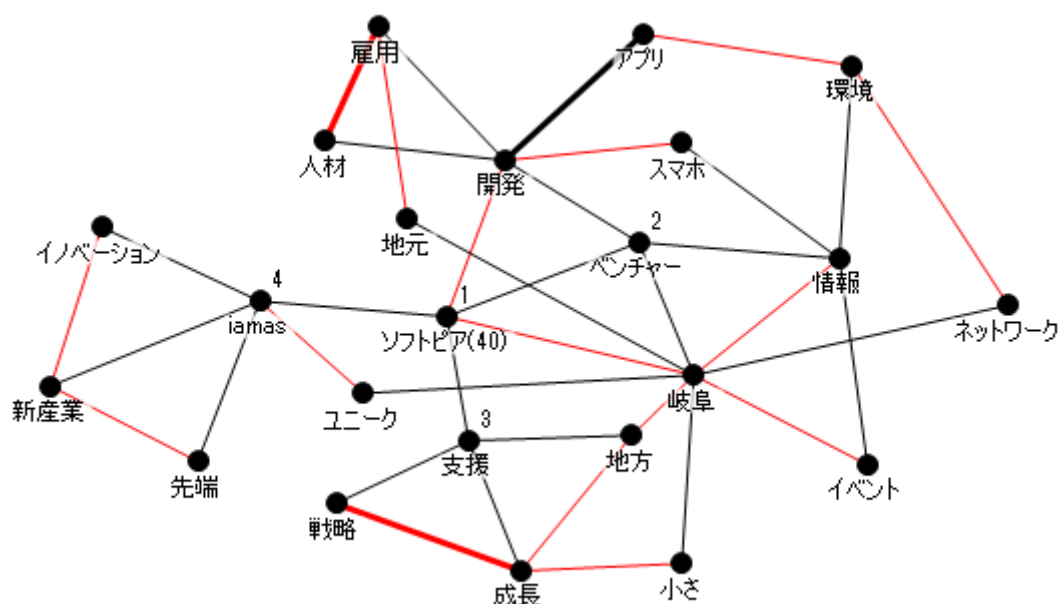


図 27 図 26 から「ベンチャー」から派生した語をさらに展開

つぎに、キーワードの「支援」系の3つの派生用語「地方」「成長」「戦略」の関連性を検討すると、以下ようになる。

1)「地方」は、「コミュニティ」「公的」「地道」を選択する。ここでは、岐阜のような地方では、ソフトピアの支援は、「公的」で、「地道」で、かつベンチャーをも含んだ広義の「コミュニティ」に対して実施されるものだ、ということである。

2)「成長」は、「アベノミクス」「緩和」「創出」を選択する。成長が求めるものは、それが国レベルでの戦略ならば「アベノミクス」であり、手法ならば規制「緩和」である。他方、それが岐阜版である

ならば、新しい雇用を「創出」するものでなければならぬし、それがソフトピアのミッションである、ということが、ここでの関連性で明確になっている。

3)「戦略」は、「デザイン」「商工」「ネットショップ」を選択する。ここで示されていることは、ソフトピアが求める小さな岐阜版成長戦略は、IAMAS と連携した「デザイン」であり、そのデザインは「商工」関係の企業を支援することで達成されるものであり、その成果物は、「ネットショップ」を媒介として全国展開をして、地元企業の小さな成長を実現しようとするものである。これこそソフトピアが実現を求めるミッションである。

以上をまとめると図 28 となる。

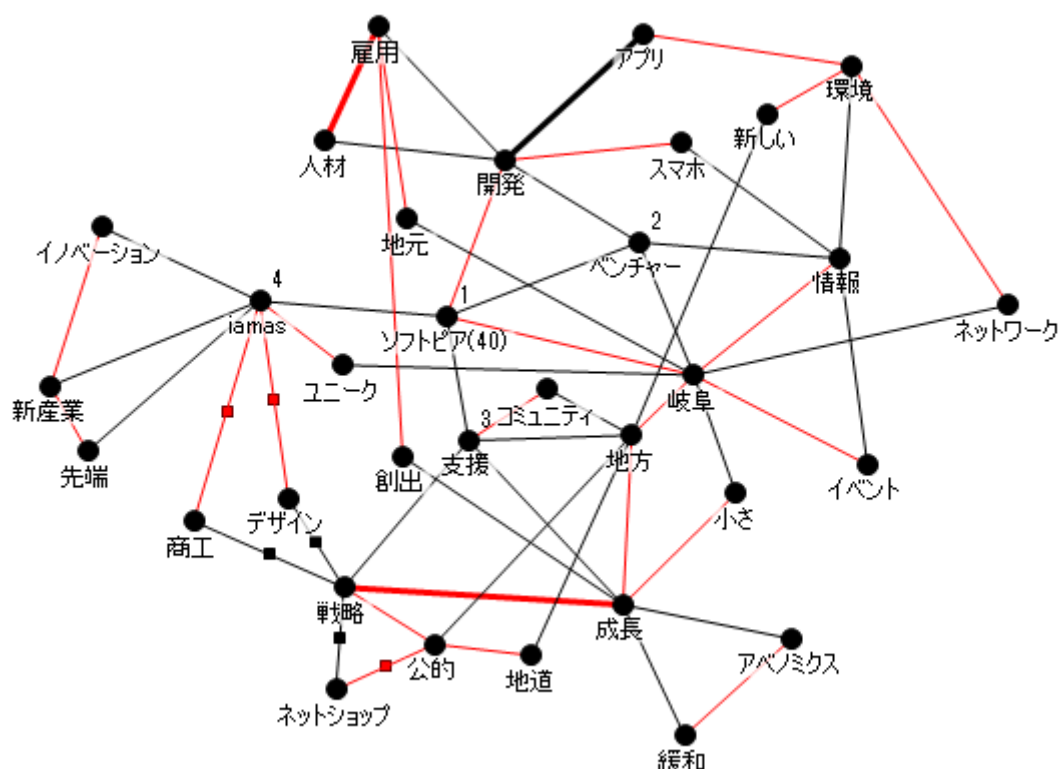


図 28 図 27 から「支援」から派生した語をさらに展開

最後に、IAMAS のキーワードから派生する 3 用語「イノベーション」「先端」「新産業」に関連する用語を生成させると、つぎようになる。

1)「イノベーション」は、「世界」「社会」「鎌倉」を選択する。ここでは、イノベーションが展開する空間表現が 3 階層で選択されている。「世界」では、IAMAS の学術の世界観を示す「融合」と「先端」に連動し、「社会」はその世界観を実態化する主体であるソフトピアと連動し、「鎌倉」は、先端の時代表現である 3D プリンタと連動した「ファボラボ」に連動する。

2)「先端」は、「ファボラボ」「融合」「芸術」を選択する。ここでは、IAMAS の大学院の特性である、先端と「融合」のコンセプト、具体的な表現手法としてオーソドクスな「芸術」そして先端での具体的な表現である「ファボラボ」が選択され、それぞれが、他の用語と連動している。

3)「新産業」は、「想像」「科学」「産官」を選択している。この 3 用語は、すべて先端が生成させた「芸術」と連動して、IAMAS の具体的なイメージを鮮明にしている。とくにここでは、産官学の志向性が重視され、だからこそソフトピアとの関連性が不可欠であることが発見される。

以上を示すと、図 29 になる。

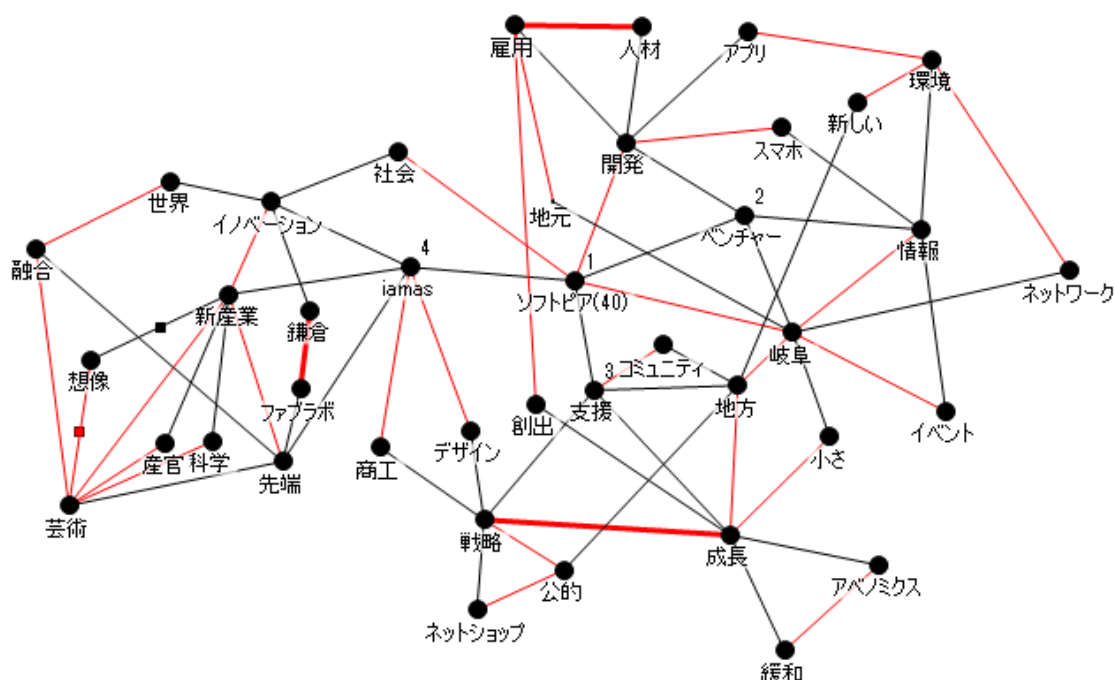


図 29 図 28 から「iamas」から派生した語をさらに展開

これが、全体構造図である。2 つのメディアに記載した 30 程度のエッセイ、約 3 万字のテキストデータをもとに、「ひっばるくん」を活用して、起点ワードのソフトピアから、全体の構造を生成させた成果が、この全体構造図である。

そこで、この構造図にかんする意味的な解釈をする前に、構造化の状況を説明する。

1) 構造化は、黒ラインによって、階層的に生成されている。これが図 30 である。これは、解釈者の構造化イメージとツールから提出される特化頻度（緑円環）との対話と探索の過程を経て生成されたものである。

ここでは、起点ワードの「ソフトピア」を基に、3 つのキーワード（ベンチャー、IAMAS、支援）からなる第 1 階層を特定し、そこから派生する用語を各 3 つ探索し、それを第 2 階層として特定した。その数は 9 である。さらに、第 2 階層の 9 派生用語から、それに関連する用語を各 3 つ探索し、それを第 3 階層として特定化した。その数は 27 である。以上から、合計して 40 の用語と 3 階層から構成される黒ライン系列での構造化を生成させた。

つぎに、この黒ラインと同時に、自動生成される赤ラインが用語の関連性を特定化するので、黒ラインでの階層構造は、そこでの階層が赤ラインによって横断され、かつ縦断されることで、全体の構造は結果としてネットワーク化する。ツールのもつ自動生成機能によって、解釈者が自己の恣意性（予期した構造イメージ）とツールの特化頻度との対話と探索の成果である階層構造は、さらに新しい発見のタネを提供してくる赤ラインの用語の横断性と縦断性によって、ネットワーク化される。

起点ワード	キーワード	派生用語 (1)	派生用語 (2)		
ソフトピア	ベンチャー	情報	環境	イベント	スマホ
		開発	人材	雇用	アプリ
		岐阜	地元	小さな	ネットワーク
	支援	地方	新しい	公的	コミュニティ
		成長	緩和	創出	アベノミクス
		戦略	デザイン	商工	ネットショップ
	IAMAS	イノベーション	世界	社会	鎌倉
		新産業	想像	科学	産官
		先端	融合	芸術	ファボラボ

図 30 図 29 を構造化

赤ラインの効果を詳細に検討すると、まず、起点ワードとキーワードに対して、階層を超えて直接関連する用語が派生用語にはいくつかある。それを示すと図 31 になる。

起点ワード	キーワード	派生用語 (1)	派生用語 (2)		
ソフトピア	ベンチャー	情報	環境	イベント	スマホ
		開発	人材	雇用	アプリ
		岐阜	地元	小さな	ネットワーク
	支援	地方	新しい	公的	コミュニティ
		成長	緩和	創出	アベノミクス
		戦略	デザイン	商工	ネットショップ
	IAMAS	イノベーション	世界	社会	鎌倉
		新産業	想像	科学	産官
		先端	融合	芸術	ファボラボ

図 31 階層を超えて接続するものを着色

ここに示されているように、起点ワードの「ソフトピア」には、キーワード以外に 3 つの派生用語（「開

発」「岐阜」「社会」)が縦断的に関連している。

キーワードの階層については、「ベンチャー」に縦断する用語はなく、「支援」に1つの派生用語(「コミュニティ」)、また「IAMAS」には2つの派生用語(「デザイン」「商工」)が縦断し、かつこの2つは支援系の用語なので、その意味では系を横断してキーワードに関連している。その意味で、この2つは、キーワードを連動させる機能を持つ点で、非常に重要な機能をもっている。

つぎに、派生用語の第2階層での赤ラインの自動生成をみると、図32になる。

起点ワード	キーワード	派生用語 (1)	派生用語 (2)		
ソフトピア	ベンチャー	情報	環境	イベント	スマホ
		開発	人材	雇用	アプリ
		岐阜	地元	小さな	ネットワーク
		地方	新しい	公的	コミュニティ
		成長	緩和	創出	アベノミクス
	支援	戦略	デザイン	商工	ネットショップ
		イノベーション	世界	社会	鎌倉
		新産業	想像	科学	産官
		先端	融合	芸術	ファボラボ
	IAMAS				

図 32 派生語の第二階層での赤ラインの自動生成

これから、ベンチャー系と支援系が横断的に連携する。特に、「岐阜」が媒介になって、ベンチャー系の「情報」と連動し、さらに支援系の「地方」と連動する。支援系の3つは「成長」を媒介に連動し、それによって、ベンチャー系と支援系が強く連動する。また縦断的連携については、「岐阜」が「イベント」に連動し、「成長」が「小さな」に連動し、「戦略」が「公的」に連動する。

また、ベンチャー系では、「開発」が「スマホ」に連動している。この関連性は、前述した上位階層との連動と連結させると、「ソフトピア」を媒介として、岐阜ー情報系と開発系がさらに連動することになる。その結果、キーワードの「ベンチャー」と「支援」が派生用語の下位階層を媒介として、強く連動していることが分かる。

他方、IAMAS系については、3つが「新産業」を媒介にして横断的に連動し、さらに「新産業」が「芸術」と縦断的に連動している。

最後に、第3階層での連動をみると、つぎのようになる。(図33)

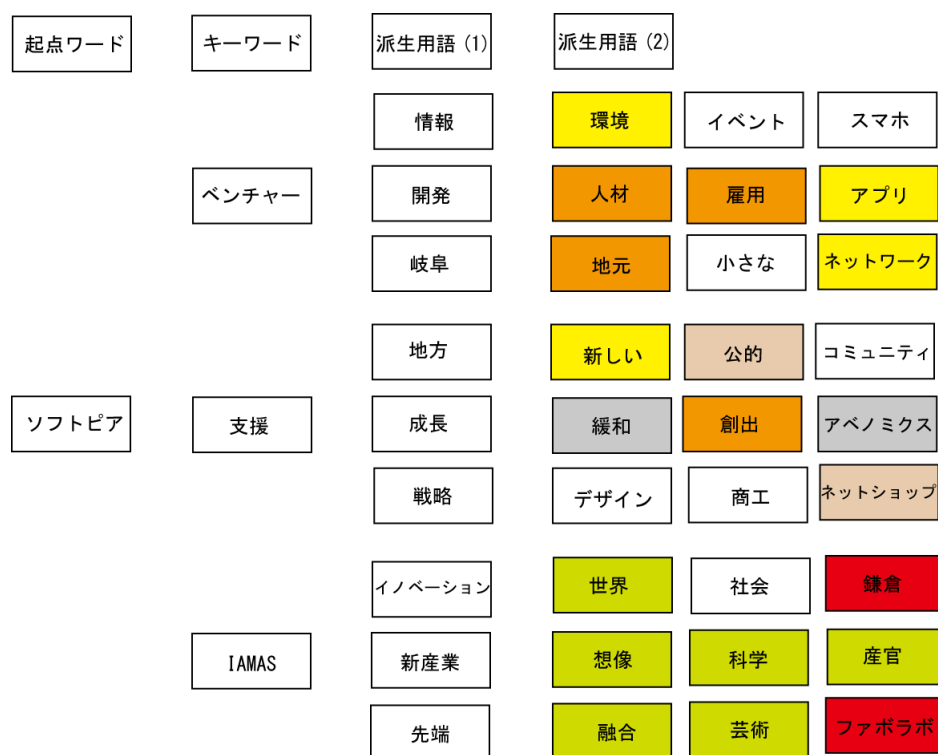


図 33 第三階層での関連

ここでは、系内で完結する系内連動パターンと系間で連動する系間連動のパターンがある。まず、系内連動については、支援系に 2 つある。1 つが「公約」と「ネットショップ」であり、もう 1 つが「アベノミクス」と「緩和」である。IAMAS 系でも、2 つの連動がある。

1 つは、「鎌倉」と「ファブラボ」で、もう 1 つが「芸術」を媒介として、「科学」「産官」「想像」「融合」さらに、「融合」は媒介になって「世界」に連動している。この芸術を核とした連動パターンは大きなクラスタを構成している。もう 1 つが「鎌倉」と「ファブラボ」である。

これにたいして、系間連動のパターンがベンチャー系と支援系との間で生成されている。その 1 つが「環境」を媒介とした「ネットワーク」「アプリ」「新しい」の連動である。もう 1 つは「雇用」を媒介として「人材」「地元」「創出」が連動しているパターンである。

以上をもとに、構造図の解釈とすると、つぎのような観点からの解釈が重要であることがわかる。

1) この構造図で最重要な目的用語は、「岐阜」「地方」「小 (な)」「成長」「戦略」「支援」のセットである。要は、「ソフトピア」が果たすべき組織としてのミッションは、岐阜県という地方における小さな成長戦略を支援すること、である。(図 34)

同じ「公的」であっても、国レベルで実行される「アベノミクス」の規制「緩和」という法規制を核とした成長戦略とは違い、それと補完するかたちで、岐阜県固有の小さな成長戦略をいかに実行していくか、そのためにソフトピアのミッションは何かを模索するのがこの構造図なのである。

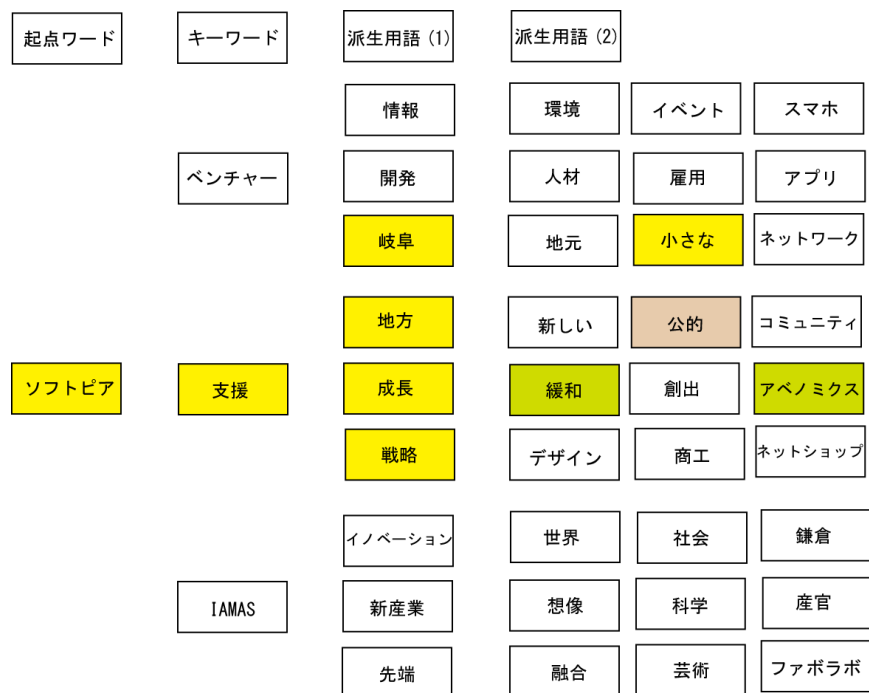


図 34 支援に絞った関係図

2) この目的を達成するための手段として、ここでは3つのルートが示唆されていることがわかる。それが、派生用語(2)の階層で示されている大きな3つのクラスタである。

第1は、「雇用」「人材」「創出」「地元」のクラスタである。ここでは、小さな成長戦略を成功させるには、地元での人材雇用の創出が不可欠であり、そこをいかに実現するかが、ソフトピアのミッションである、という施策方針の提案である。

第2は、「新しい」「アプリ」「ネットワーク」「環境」からなるクラスタである。ここでは、小さな成長戦略の実行のためには、岐阜県が誇るネットワーク環境を活かした新しいアプリ開発の拠点化が、ソフトピアの支援によって達成されなければならない、という施策方針の提案である。

第3は、「芸術」「世界」「融合」「科学」「産官」「ファボラボ」からなるクラスタである。ここは、IAMASの世界であり、芸術と情報技術の融合と先端を掲げ、産官学一体になって地域に貢献する学術機関であり、ファボラボのような実験スペースをソフトピアと共同で立ち上げるなど、ソフトピアのミッションの実現に、学術の視点から大きな貢献が期待されている。

以上3つのクラスタが、ソフトピアのミッションを実現するための手段的・施策的な用語セットである。(図 35)

3-4-4 構造化の生成過程(2)ー下位構造図と下位階層構造ー

つぎに、各3手段的クラスタが、ミッションの目的クラスタにどのように関連しているかを、再度「ひっぱりくん」を活用して検討する。

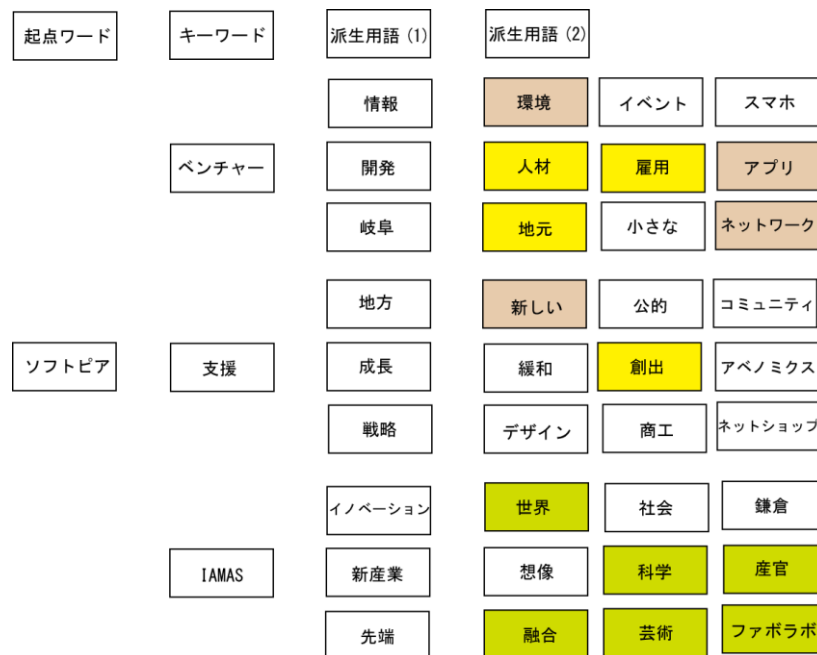


図 35 最終的に解釈として発見されたクラスタ

1) 「雇用」から「小さな成長戦略」へ

ここでは、つぎの下位構造図 (1) が生成される。(図 36)

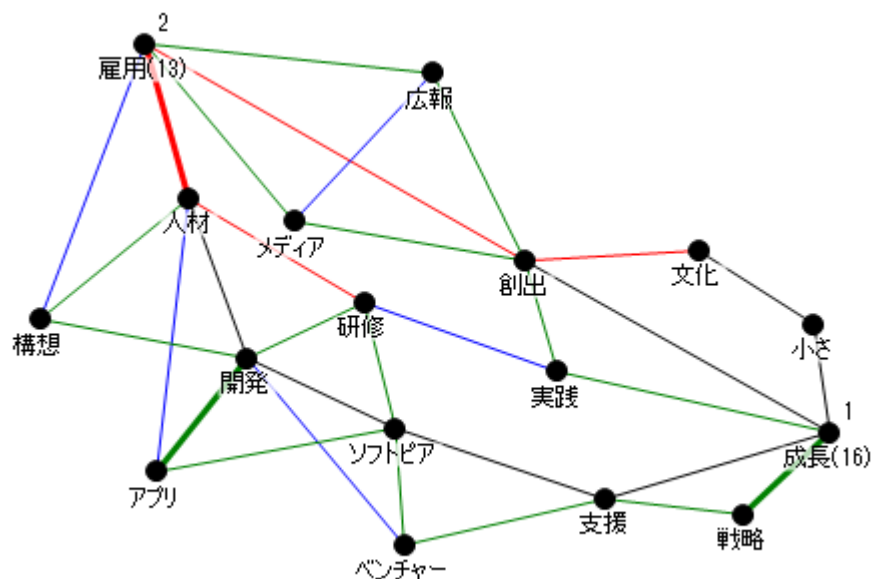


図 36 「雇用」と「成長」から作成

ここには、2つのルートがある。

第1は、ストレートに、「成長」は「雇用」の「創出」である、という言明である。そのために、「広報」「メディア」が、単なるビジネスの拠点化ということではなく、「文化」の「実践」という方向で、この雇用問題を扱うことが方針として重要である、という補足情報を出している。小さな成長戦略は、地方

の文化創出であり、単なる経済的な雇用問題ではない、という視点が明確にされている。

第2は、上記の直接的な言明を、より具体的に示したルートである。つまり「人材」「雇用」「構想」は、「ソフトピア」からすれば、「ベンチャー」の「アプリ」開発の「支援」として「研修」を打って、新しい人材の確保に貢献することである、という主張である。地域文化の醸成のために人材の確保が必要であるが、直接的には、その人材はアプリ開発で自立的にビジネスが岐阜で立ち上げることができるまでにしなければ、成長戦略とはいえない、ということである。この2つのルートが補完しあうことで、雇用から成長への筋道は通る、という言明のセットである。

ここでは、緑ラインと青ラインが、黒ラインと赤ラインの他に生成されている。緑ラインは、黒・赤ラインを媒介する用語を生成させる機能で、なぜ、黒・赤ラインが生成されているのか、その根拠を示す用語であり、青ラインは、その効果として別の用語との間に自動的に生成される関連性を示すものである。

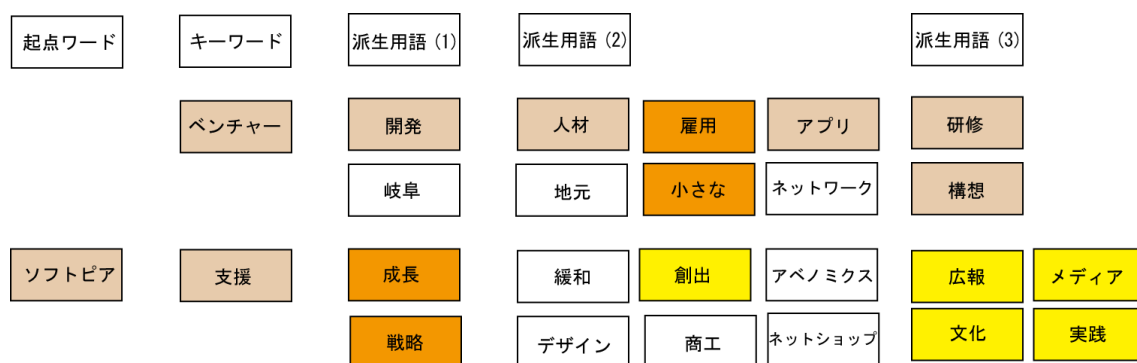


図 37 図 24 で発見した語を構造図に追加

以上を、階層構造で示すと、図 37 になる。新しい派生用語 (3) を生成させることで、雇用から小さな成長戦略までの用語連鎖のルートをより明確にするには、どのような派生用語の探索が必要であるかが分かる。特にここでは、雇用と成長戦略にとって、文化的な実践の意義をアピールすることが重要である、という発見は、解釈者にとって大きな気づきであったといえよう。

3-4-5 「環境」から「小さな成長戦略」へ

新しい情報環境はどのようにして小さな成長戦略を成功に導くための状況要因になるのか、それがここで確認する第2の下位構造図である。全体構造図では、やや混沌としていた状況は、この下位構造図によって、2つのルートから生成されていることが分かる。

第1は、「新しい」「ネットワーク」「環境」の系列から生成する、新しい成長戦略の方向性である。ここでは、派生用語が数多く、緑と青ラインから生成される。「ビジネス」「専門」「貧しい」「豊か」「多様」「地域」「文化」の7つの派生用語が生成される。これによって、新しいネットワーク環境では、小さな成長戦略のために、つぎのようなアイデアが必要とされている。つまり地方の小さな成長戦略には、新しい地域文化の覚醒が必要であるが、現状では、貧しいだけの地域文化には、新しいビジネスの流れの融合が必要で、それが豊かで多様なネットワーク環境のもとで生まれる専門的なビジネスである。

このビジネスが、第2のルートである競争環境にあるスマホ開発のビジネスである。ここでは、派生用語は「競争」だけで、後は、すべて全体的な構造図にみられるものである。要は、スマホのアプリ開発コミュニティの中で頑張るベンチャーとそれを支援するソフトピアとの協働によって、具体的な小さな成長戦略のモデルとなる、という主張である。

これを示す下位構造図（ネットワーク図）が、図 38 である。

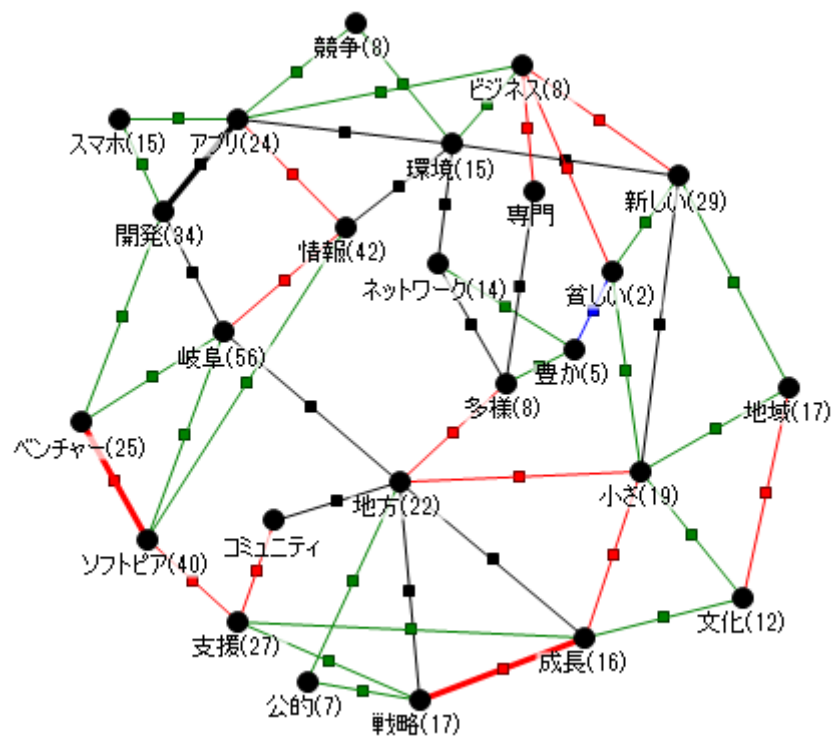


図 38 「環境」から「小さな成長戦略」への下位構造図

この下位構造図について、階層構造に変換すると、つぎのようになる（図 39）。

ここで分かるように、全体構造図で語られていたのはルート 1 であって、それを補完するには、ルート 2 による説明が必要になっていた。とすると、全体構造図だけで、解釈が完成されるものではなく、つねに、下位の構造化を探索することが大事な作業である。そのためにも、階層構造として用語の構造を再構築することは、必要な作業である。ネットワークする構造図は、全体のイメージを可視化するには適した表示方法であるが、同時に、階層構造と併存させることが、重要であることが予想されよう。

この場合は、ルート 1 における 7 つの用語が緑と青ラインによって探索されたが、この過程を付加することで、解釈のための深堀が可能になり、より豊かな用語の連関が可視化されてくる。ここでは、ネットワーク環境の豊かさと地域文化の貧しさの対照性が可視化されることで、その貧しさを乗り越えることに、次の成長戦略の意図が隠されていることが発見される。このような発見こそ、階層構造をさらに下位階層化すること操作によって可能になるのであり、このツールが期待する成果を獲得する視点なのである。

の「先端」を走りながら「イノベーション」を通して「新産業」の立ち上げに貢献しよう、という関連性であり、まさに、これはソフトピアの理事長の視点からみた望ましい IAMAS の姿そのものである。「ファボラボ」を共同で立ち上げ、今話題の 3D「プリンタ」を活用して、「鎌倉」のファボラボに対抗するものづくりの「世界」を作りたいという構想そのものが、ここで提示されている。

このヴィジョンを実現して、小さな成長戦略に繋げるルートが 2 つあり、1 つの IAMAS ルートでは、地元の「小規模」の「商工」企業とのワークショップを実施して、そこで「デザイン」「思考」の発想法によるものづくりの手法を学び、新しい成長戦略に繋げよう、とする「施策」を展開している。ここで生成されたデザイン思考こそ、イノベーションを実現する手法として期待されているものであり、それがここに出現していることは、重要なことである。

もう 1 つのソフトピアルートは、「イノベーション」との繋ぎの用語が「社会」であることに象徴されるように、IAMAS ルート（デザイン思考）を背後で支える社会的な環境整備についての用語群である。ここでは、ソフトピアが主催する「技術」関連の「研修」によって、「ベンチャー」「支援」を行い、成長戦略に繋げたいという主張である。

以上を、階層構造図に変換すると、つぎのようになる（図 41）。

ここで分かることは、前述した 2 つの手段的手法では、起点になる用語の説明はなかったが、ここで起点となる「イノベーション」については、どのような意味でイノベーションなのかの説明に多くの部分が割かれている。これは、IAMAS という学術機関の特性を説明しないかぎり、岐阜とソフトピアが志向する成長戦略の根源が理解されないだろう、という思惑があったからである。その意味で、この 3 番目の手段的手法にこそ、他県にはない岐阜県ユニークな手法である、というアピールがここに見られる。

以上、全体構造図の生成から、さらに 3 つの下位構造図の生成にいたるまで、いかなる方法で、用語の選択と探索が 4 タイプのライン（黒・赤・緑・青）との対話と自動生成を通して遂行され、その成果として上述のような解釈が可能になった。

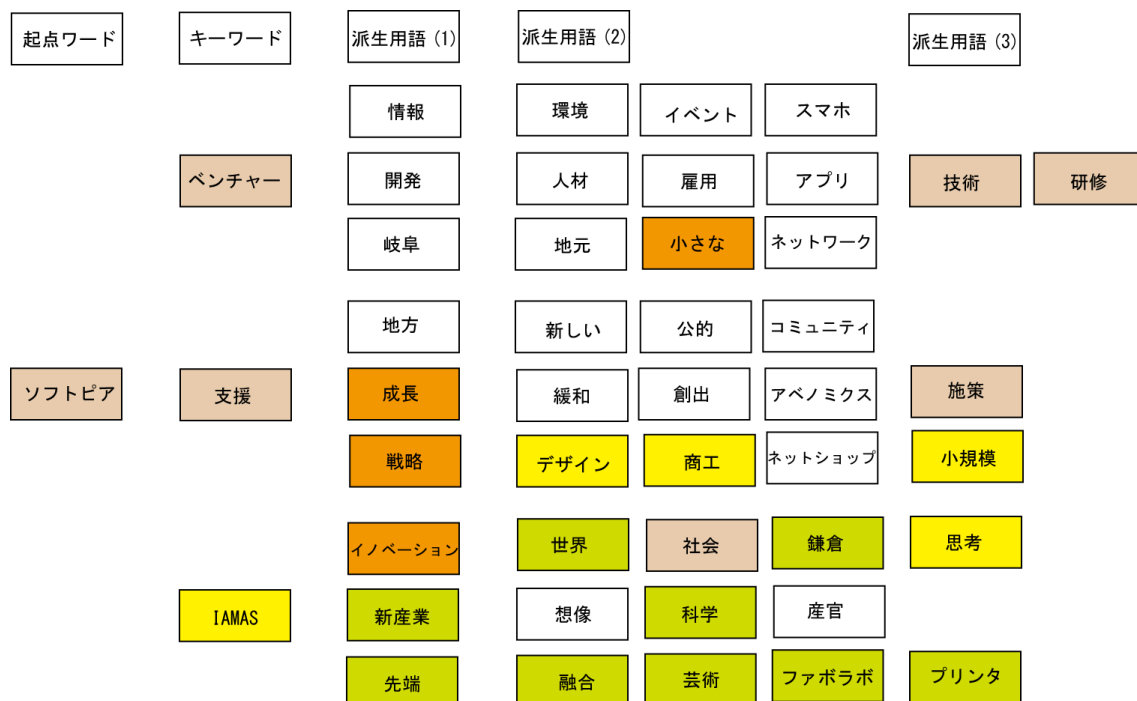


図 41 最終的な階層構造図

3-4-7 発見と正当化の根拠

このような解釈は、いかにして正当化されるのであろうか。その根拠が問われよう。そこで、この点について、解釈者とのインタビューを通じて、上記の解釈において、自明性の確認と創発性の発見について、元のエッセイのデータを参考にしながら、構造図の正当性について確認する。

まず、解釈者とのインタビューをまとめたものを記載する。解釈者が「ひっぱりくん」を活用した過程で感じたことは、以下の点である。

1) 最初、このツールを使用したとき、制約がなさすぎるので、自分（解釈者）の好きなような用語をスラスラとひっぱり、つまり自分の直感に任せて好きなように引っ張るだけで、その自由度の高さゆえに、なんでも構造化ができると同時に、構造図を生成させた達成感がなかった。つまりここには、深い思考過程が関与するという印象がなく、単に用語を自由勝手に引っ張るだけで、その自動化・機械化に自分の思考のコミットがあまりにも薄いという印象で、これでは納得できる構造図の作成は無理かな、という印象をもった。

2) 確かに、黒ラインを引くと、それに対応して、赤ラインが自動生成されることで、用語の関連性がダイナミックに生成される感があり、そこには自分の思考を刺激する何かが存在したが、にもかかわらず、自分の思考が強く反映しているとは思えず、引っ張る行為だけが自動化される気分になり、自分も黒ラインとして自動生成するマシンになっている気分であった。したがって、構造図が生成されても、自分のコミットがあるとは思えなかった。

3) いかに自分の思考が引っ張る過程においてコミットしているか、を考えることにした。その結果、思いついたのがルールの設定であった。ルール化は、一方では自由度の低減であり、外的な秩序を導入することであるから、自分のコミットする領域が制限されることを意味する。ここでは完全自由は拒否されるわけだ。しかし他方、秩序の方向性が設定されることで、確かに、その分自分の自由は剥奪されるが、だからこそ、その制限された領域の中で、思考を深くコミットさせることが強く求められるようになった。これは、自分の思考を活性化するものであった。要は、秩序と自由の適度のバランスが保たれるとき、自分の思考は引っ張る行為に深く価値コミットするのだ、ということを実感した。

4) そこで、以下のようなルールを設定した。

第1のルールは、起点ワードを1つ確定する、である。起点ワードは、構造図を作成するにあたっての、解釈者の価値表明である。今回のケースでは、「ソフトピア」を設定したが、これは解釈者がこの組織の理事長であり、その役割の関連でエッセイの依頼がきたのであるから、ソフトピアをある意味広報する意図でエッセイを書いているわけである。とすれば、その意図（価値）を、明示的であれ暗示的であれ、表明しているはずである。このように、起点ワードを設定することで、構造図を生成させるメタゴールがこれによって確定されることになる。

第2のルールは、起点ワードを3つの方向で特定化するキーワードを設定する、というものである。ここは、3つが重要である。起点ワードを確定すると、それを体現する用語はある意味エッセイの中で使用されたすべての用語である。とすると、ここで最初から、アドホックにいくつでも並行して、用語を引っ張ることは可能である。しかし、それでは、方向性が散漫になり、構造化ができるが、茫漠とした結果になることが分かった。そこで、第2のルールとして、起点ワードを3つの方向性で示すキーワードを設定することをルール化した。今回のケースならば、「ベンチャー」「IAMAS」「支援」である。候補はたくさんあっても、3つに限定するというルールを設定することで、深い思考（熟慮と決断）が関与することになり、この段階で、何を生成させたいのか、をいう、自分なりの構造化のイメージを具体的に保有することが可能になった。

第3のルールは、キーワードからの派生語の選択は3つとする、である。これ以降、派生語を選択する場合には、3つを選択することを自分に課した。こうすることで、論理的に思考する準備ができるようになった。しかも、選択する黒ラインの数を限定することで、そこから自動生成する赤ラインが適度に生成することを可能にした気がする。

5) このようなルールを設定することで、思考が深くなり、熟慮と決断の重要性を意識するようになっ

た。3つを確定するまでに、緑の円環をみながら、何度も試行錯誤を繰り返すようになり、安易に黒ラインを確定することにならず、しかも試行錯誤の過程で、赤ラインがどこに生成されるかも確認しながら確定できるので、黒ラインと赤ラインのバランスを保つことにも注目するようになった。ここには、マシンと自分が試行錯誤の過程で対話を繰り返している気分になったことは実感である。

6) 全体構造図ができると、それはネットワーク図であって、40の用語がネットワーク化されているので、それを見ただけで、すんなりと解釈ができるようなものではなかった。そこで、思いついたのは、用語の階層化である。まずは、起点ワードからキーワード、そして2階層の派生用語が生成されたので、それぞれの階層がどうなっているのか、を確定することが、解釈する事前の処理としては必要だな、と思った。この階層図を作成することで、どの用語が、この階層をもとに、縦断し、横断して、別の用語とネットワークしているか、が確認できるようになった。つまりキーワードの3つが示す系列とそれぞれの4つの階層から構成される階層図と、それが縦断し横断する赤ラインの自動生成機能によってネットワーク化がなされるわけであるから、黒ラインとしての自分の主張と赤ラインが生成させた全体構造図（ネットワーク図）を比較することで、解釈する糸口が鮮明にみえてくるようになった。

7) その結果、このケースならば、「小さな成長戦略」という目的用語群が発見され、それを説明する手段的用語が3つ（「雇用」「環境」「イノベーション」）とそれを起点とするルートがいくつか発見されることになった。この発見は、「やった！」という驚きであった。30余のばらばらのエッセイから、ここまで明確な関連性と構造化が生成されるとは思わなかったし、そこで、しっかりと目的用語と手段用語が発見された、というのも驚きであった。よく考えれば、その発見は妥当であり当然ともいえようが、30余のエッセイの背後に、このような構造が隠されていたか、という事実の発見は、解釈者としても十分に新鮮な驚きであり、「ひっばるくん」のツールとしての可能性を強く確信した。

8) さらに、目的手段関係を深堀する必要があると感じたので、下位の構造図を生成させることにした。基本は全体構造図なので、そこでの生成された用語を尊重しながら、その意味では、再度、個別の手段目的関係ごとに下位構造図を生成させる作業を実施した。当然その過程で、さらなる深堀が期待されるので、全体構造図で確定された用語を媒介する用語を選択する操作を実施した。それが緑ラインである。これによって、黒と赤ラインそれぞれの両端の用語がいかなる意味でつながっているかが確認できるし、さらにそこからの派生として青ラインが緑ラインの制約を超えて、別の用語との関連性を自動生成する機能を発揮した。その結果、下位構造図は、単に、全体構造図の部分的な切り取りではなく、その部分を基に拡張された新たな構造化が生成される。これによって、下位レベルでのネットワーク化が生成されるので、この構造を解析するために、上述した手続きと同様に、用語の階層構造化を実施することになる。その結果、下位構造レベルでの意味の深堀が可能になる。こうして、解釈は、多層構造のもとで、深まっていった。

9) 用語の抽象度に気を使った。緑の円環から選択する場合、ある用語を説明するのか、他の用語に関連づけるのか、さらに一般的な意味をもたせるのか、さまざまな条件のなかで、最適なものを選択しなければならない。その場合、3つを選択するルールにしているのだから、抽象度をバランスよく散らすという方法は、それなりに有効だったかな、と思う。ただ、それも状況次第で、構造図をイメージしていて、それに近づけるにはどうすればいいか、を考えると、抽象度を多様にすればいいというわけではないこともあり、ルールと状況判断は、いつも悩みではあった。

以上の気づきのコメントから推測されることは、構造図を生成させる（主体的には探索する）過程で、その過程を正当化する根拠はいかなるものか、である。その正当化の根拠は、つぎの6点である。

- 1) 自由と秩序
- 2) 価値表明と目標達成
- 3) 自明性と創発性
- 4) 目的と手段
- 5) 階層とネットワーク
- 6) 一般と特殊

1) 自由と秩序

「ひっぱりくん」は、基本的には、ツールを活用する主体（解釈者）の自由な選択と探索を重視しており、ほぼ完全なる自由を支持している。この完全なる自由は、ツールが投げかける緑の円環が示唆する特化頻度の序列を無視して、徹底して黒ラインだけを重視して構造化を図ることを可能にしている。ここでは、ささやかに赤ラインが自動生成されて、主体的に選択された用語の関連性を支援するだけである。ただ、この赤ラインの支援がないとネットワーク化ができない、という制約は、もしも解釈者が用語の階層のみで解釈は十分であるという立場ならば、完全自由の実現は維持されるが、そうではなく、ネットワーク化構造化には不可欠だとするならば、赤ラインの貢献は不可欠である。

他方、このツールを活用する場合、ルールを確定して、たとえば緑の円環で提示された上位3つの用語のみを選択する、とすると、解釈者の主体的な関与は消失し、単なる引っ張るマシンになってしまう。これは、「ひっぱりくん」の基本的精神とは異なるが、しかし活用次第で、解釈者をマシンに変えるものでもある。

要は、「ひっぱりくん」は、主体の自由（選択と探索）とツールの秩序（制約と提示）のバランスをいかにとるか、をケースごとに確定することが不可欠であり、そのバランスを求めることで、生成された構造化図が、主体とツール（＝もう一人の主体）との間主観的な了解を通して獲得された成果となる。したがって、解釈者は、構造化図を作成する時に、いかなる秩序を主体に課すか、そのためのルールをしつかりと確定しなければならない。そのルールが提示する「制約と提示」の秩序を前提として、自由な「選択と探索」を繰り返す時、その過程で生成される構造化図は、間主観的な了解という観点から十分に学術的に正当化される成果の位置を獲得しよう。

2) 価値表明と目標達成

起点ワードは、いかにして設定されるのか。これは、解釈者の価値の表明である。テキスト全体から、1つの用語を選択する基準は、外部から決定されることはない。ランダムに決めるものでないし、客観的に決定するルールは存在しない。ここは、解釈者の主観であり、事前にもつ構造化図のイメージをシンボリックに表現するであろう用語を選択することである。しかも構造化図のイメージは、解釈者が最終的に生成されたいイメージであるから、主体が望ましいと確信する価値そのものである。その価値表明が起点になって、自由と秩序のバランスのもとで、探索が始動する。

しかし価値表明は、ある意味、起点ワードを確定するまでの行為で完結してしまう。それ以降の選択と探索の行為は、価値を意識しながら、その価値実現を目標として、その達成を求めて試行錯誤を繰り返す多難な行為の連続である。その成果が生成された構造化図である。この成果は、価値表明（イメージ）との間で評価される。その評価過程は、イメージが不動の基準であり、その基準に沿っているかどうかで、成果が評価される、という一方向的な評価ではない。イメージ（予期された構造化図）と成果（生成された構造化図）は、成果を追求する過程で、何度も、お互いを修正することが繰り返される。イメージも、成果を追求する過程で、より具体化され、より多様化され、より多層化される。イメージと過程での成果は、その乖離を互いに埋めることで、間主観的（予期する主体と試行する主体）な了解に達するまで、試行錯誤を展開する。その結果、予期する構造化図イメージと生成された構造化図との間の乖離が縮小し、納得できる評価の妥当性が獲得されると、成果はここで確定し、構造化図が完成する。要は、この時点でやっと目標は達成された。

3) 自明性と創発性

価値表明と目標達成の間で、納得いく合意・了解が得られた時、その構造化図は生成され成果となって産出されたといえよう。しかしその成果を評価するもう1つの基準がある。それが、自明性の確認と創発性の発見のバランス問題である。もしも、予期した構造化図がそのまま成果の構造化図になったとしたら、それは、事前の構造化イメージがツールによってそのまま反映されたことになる。つまりこれは100%自明性の確認である。これは、確認だけを求めたツール活用ならば、それなりに評価されようが、本来のツールの活用法ではない。このツールを活用する根拠は、事前の予期のイメージが、そこそこに支持さ

れながら、他方、新しい何かの発見が求められている。つまり適度の裏切りが必要なのである。その裏切りが創発性の発見である。とすると、自明性の確認と創発性の発見のバランスが重要である。

単なるネットワークが生成されたのでは意味がなく、そのネットワークの中に、自明な部分と発見の部分が適度のバランスをもって存在するとき、この構造図は価値があるし、目標を達成したと評価されるのである。

4) 目的と手段

構造図は、そのネットワークの中に、なんらかの意味を創発させている。その意味解釈をする場合、ネットワーク全体はいかなる要素（意味集合）から構成されているのか、さらにその要素は目的と手段の関係として理解できるか、が重要である。特に、目的－手段関係が発見できるならば、ネットワークの意味解釈は容易である。

意味解釈の問題は、ネットワークの構造がいくつかの要素に分解でき、さらにその要素間に関係性が発見され、さらにはその関係が目的－手段とか因果関係として理解可能ならば、より良い解釈の方法論が創発されたといえよう。

今回のケースならば、「小さな成長戦略」の目的要素が発見され、さらにそれを説明する手段として、3つの要素（「雇用」「環境」「イノベーション」）が手段として発見された。このような方法論が構造図のネットワークのなかに発見できるならば、解釈は正当に可能になったといえよう。

構造図は、解釈されなければならない。起点ワードをはじめとして探索された用語の集合が単に存在する、というだけでは、解釈にはならない。価値表明との関連で、なぜ構造図は目標を達成するか、というと、それは、構造図のネットワークの中に、目的手段関係が顕在化するからである。それが意味の体系として存在するとき、はじめて構造図は価値を獲得する。

5) 階層とネットワーク

構造図が完成（解釈者が納得する）すると、その用語を階層構造に変換してみるといい。そうすることで、用語の階層が明示され、どの階層レベルで、どの用語を使用し、その同一階層では、どのような用語を探索しているか、が一目で理解可能である。ネットワークのままでは、理解不能な用語のつながりが、ここで整理されることで、全体の意味の体系を一気に理解することが可能になる。しかも、ここでは、手段と目的の用語がどの階層で使用され、しかもそのような経路をたどりながら、手段から目的への推移がなされたかが容易に判別可能である。しかも下位構造にまで言及すると、基本的用語以外の派生用語がどこで探索されているかも判明し、納得し理解することが容易になる仕組みになっている。こうすることで、構造にポジショニングされた用語が、全体の構造化にとって、どれくらい貢献するのか、またしないのか、という個別用語の評価も容易にできるようになる。対照的に、あまり深いコミットなしに、何気なく選択した用語が、構造化の段階では重要なコンセプトになることもあるはずである。

構造図におけるネットワークは、この階層構造に準拠しながら再検討すると、階層の中の経路として浮かび上がってくるので、当初の解釈とは異なる再解釈も可能となり、新しい視点からの解釈が再構成されることにもなる。こうして、ネットワークと階層が表裏の関係になることで、解釈は深化し、より納得される解釈への変貌するのである。

6) 一般と特殊

これは、用語の抽象度のレベルが高いか低いか、である。より一般的であるほど、他の多くの用語と連動し、より特殊であるほど、他の多くの用語との連動の可能性は低く、ある特定の用語のみとの連動が強まる。この抽象度の高低を活用して、探索行為は、深化したり、多様化したりする。なので、探索行為の遂行において、どの階層での探索なのかを十分に意識して、用語の選択をしなければならない。安易に一般的な用語を選択すると、今までのネットワークを勝手に超えて、意味のない連動に発展し、解釈が不能になる事態を招きかねない。また、ある場合に、特殊すぎる用語を選択すると、その特殊性ゆえに、他の言葉への赤ラインがひ

かれず、単に孤立するような事態が発生し、解釈の対象としては意味のない事態を引き起こすことになる。

このように、用語の選択と探索においては、つねに用語の抽象度に気遣いながら、実行しなければならない。そうしないかぎり、解釈は困難を極めるばかりになる。だからこそ、用語の選択と探索には、解釈者の熟慮と決断が求められる。解釈者は、その時の用語のネットワークと階層の状況をしっかりと熟慮し、さらにさまざまな試行をしながら、徐々に一手を慎重かつ確実に打つ決断しなければならない。その熟慮と決断の成果が、納得できるネットワーク構造図と階層構造なのである。

3-4-8 正当化とデータ検証

上記の正当化の根拠について、最後にデータと照合させることで、検証してみよう。

「ひっぱりくん」には、そのツールの下部に、ビューワがあり、その領域を拡大して、そこでのデータのありようを解析しながら、正当性の根拠を論じることにする。

図 42 を検討する。

これは、起点ワードを選択し、そこから 2 つのキーワードを選択したものである。

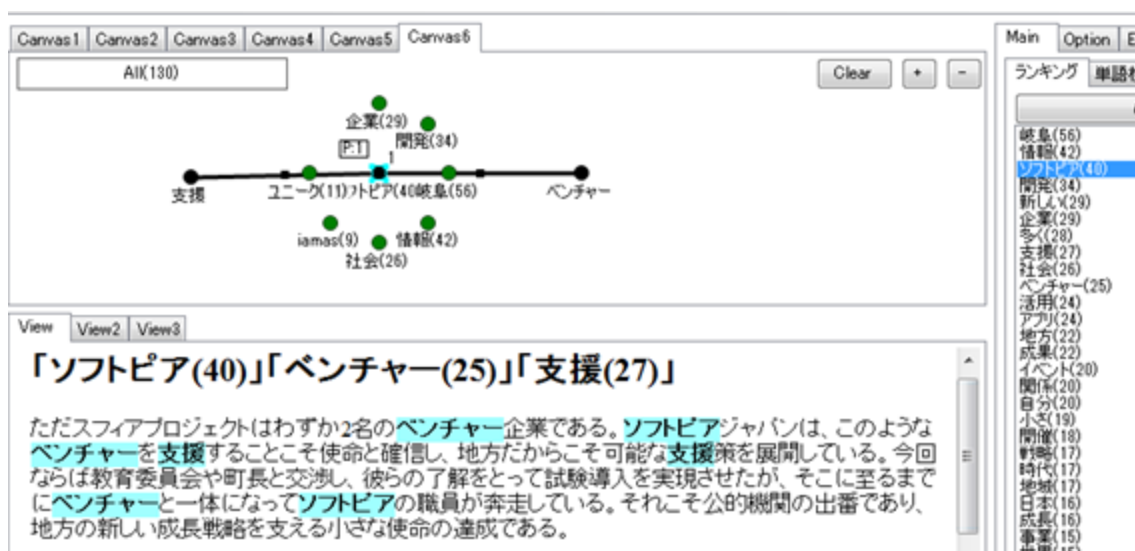


図 42 「ソフトピア」「ベンチャー」「支援」を含む文章の表示

右部のランキングにあるように、ソフトピア（40）は 40 回の出現頻度をもち、上位 3 位の用語である。これが、ここでの価値表明である。この特殊な用語の頻度の高さが、このテキストの固有性であり、シンボリックなポジションを示している。だからこそ、この用語を起点として選択したのである。この特殊用語の価値が、構造図によって明確にされたならば、それが目標達成ということである。

このために、解釈者の自由とツールの秩序のバランスとして、今回は、3 つの用語を選択するルールを設定した。では、最初の 3 つのキーワードはいかにして選択されたのか。まず、ランキングの単純頻度でも上位に位置し、かつ起点ワードの緑の円環での特化頻度でも上位に位置する、「ベンチャー」と「支援」をほぼ自動的に選択した。これは、ソフトピアのミッションの 1 つにベンチャー企業支援が唱えられているので、ほぼ自明なこととして選択された。

しかし 3 つ目の用語の選択については、ランキングでは第 1 面に登場していない用語 IAMAS を結果として探索し選択した。その根拠は、図の緑の円環にあるように、上位 5 番目に提示されており、それより上位の用語は抽象度が高く、ベンチャーのレベルに対抗できる抽象度の用語を探索していたので、それを選択することにした。

すると、ソフトピアは、キーワードの 3 つから、「ソフトピアは、IAMAS と協力してベンチャーを支

援する」というミッションをもつことになる。これは、ソフトピアを表現する文として最適であり、キーワードの選択はこれで完成する。とすると、この文こそが価値表明であり、それを他の用語から説明できるならば、目標は達成される、ということになる。

図 43 は、3 つのキーワードと起点ワードの 4 つの用語が共起する文を、パラグラフの規模で抽出したものである。ここには、僅か 2 つのパラグラフが抽出されているだけである。しかも 1 つ目は、ソフトピアは、IAMAS とベンチャーと協力して、地元の高校の情報化を支援している、という内容で、上記のミッションとは微妙に異なっている。要は、ここではベンチャーの位置づけが支援されるのではなく、支援する主体になっている。とすると、最重要な事例文の 1 つがミッションと違っているのだから、これは構造化とミッションの適合性への反証事例ではないか、という疑問が提示されよう。しかしこれは、性急でかつ単純すぎる解釈である。高校生を対象とするかぎり、つまりこの事例文に限定するかぎりでは、ベンチャーは支援する立場なのであるが、それは特殊なケースであり、だからこそエッセイの内容として価値をもっていたのである。この特殊性は、あくまでも、このエッセイの部分における価値であり、エッセイ全体の価値からすると、これはあまりにも特殊な事例ということである。だからこそ、「ひっふるくん」での構造化では、ベンチャーが支援するという形式は提示されないのである。要は、構造化の階層レベルの問題であり、このレベルをもっと下位層にまで深堀するならば、このようなベンチャーの特殊例も例示されるはずである。しかし今回のエッセイ全体の構造化のレベルでは、4 つの用語が共起している事実だけが重要な価値を創出するということである。

この証拠に、2 番目のパラグラフでは、まさに一般論が語られており、そこでは、上記のミッションをやや詳細に論じる文になっている。この文こそ、この解析と解釈で求めている構造化イメージである。

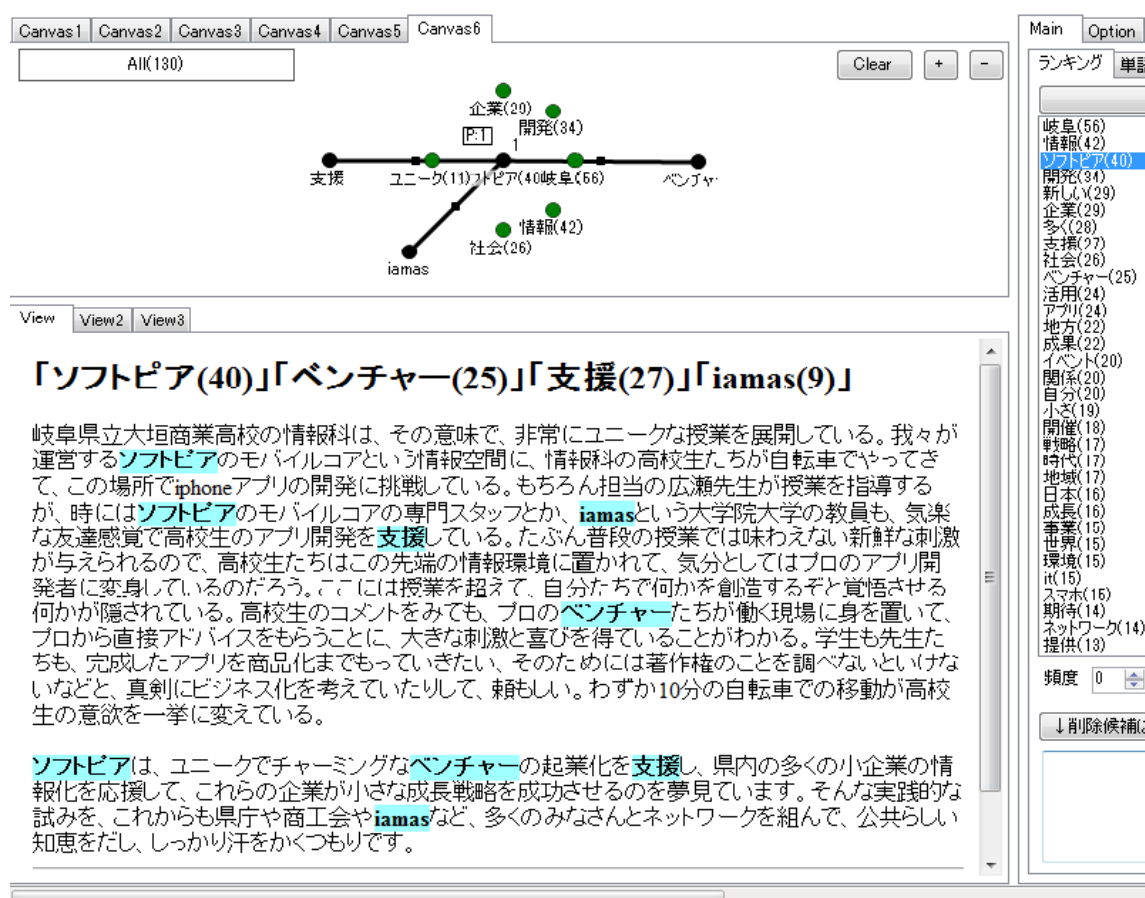


図 43 図 42 の条件に「iamas」を追加した時の文章

つぎに、図 44 と図 45 を検討する。これは、構造図の創発としてもっとも重要な目的用語の発見である。それが「小さな成長戦略」である。この用語群は、図 44 に示すように、「成長」とのつながりがトップにある「小さな」戦略」である。また、成長のみでビューアをみると、最初に登場するので、古田岐阜県知事の県内版成長戦略の話であって、決して「小さな」成長戦略の話ではない。

しかし図 45 になると、小さな成長戦略の話しそのものであり、8 つのパラグラフのなかで登場している。また、この 8 つについても、具体的な事例を列挙しているのが、図 34 にもあるように、スフィアプロジェクト（ベンチャー企業）の給食アレルギー対策アプリの話とか、それ以外でも FC 岐阜の応援アプリを作成したベンチャーの話など、特殊事例の話が 5 つと、一般論としてのパラグラフ（エッセイの結論部分）が 3 つである。エッセイを書いた解釈者も、ここまで書いた記憶はまったくなく、このツールで、このような探索することで、解釈者にとってここが一番重要な目的なのか、を強く実感した。30 余のエッセイで、8 カ所も、小さな成長戦略がアピールされている事実は、気づきとしては、非常に価値のあるものであった。

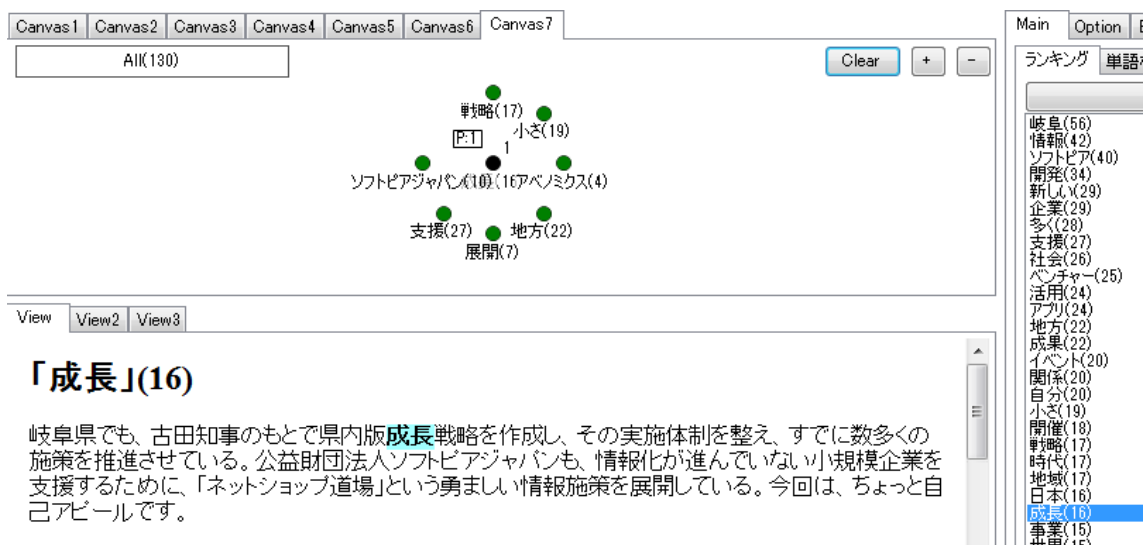


図 44 「成長」を含む文章

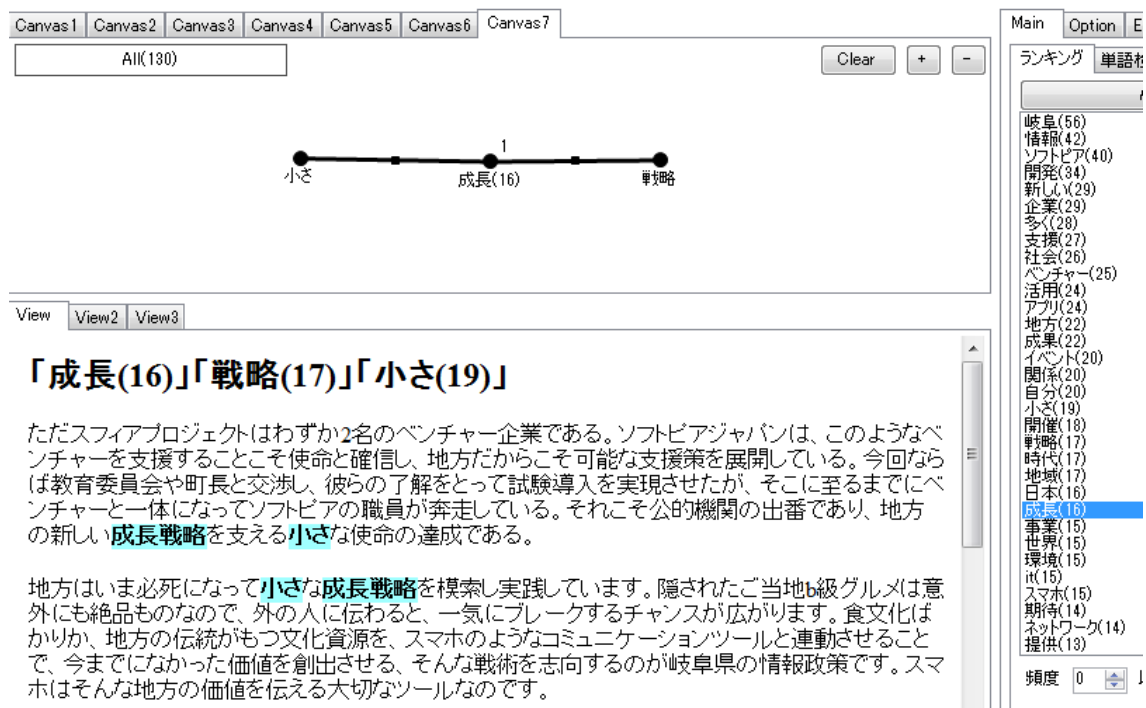


図 45 「成長」「戦略」「小さ」を含む文章

こうして、目的項目が納得されるかたちで探索された。もちろん、これを目的項目として設定したのは、全体の構造図が作成された後でのことである。それまでは、何も分からず、用語の階層を深堀していただだけである。ただ直観的に、このあたりは重要だと思いながら探索を進めていただだけである。あくまでも解釈が本格化するの、は、構造図が確定された後のことである。

しかし、「小さな成長戦略」の用語セットは、正確には、図 45 のように生成されたのではなく、図 46 のような階層構造のもとで生成されたものである。この図の違いは、致命的に重要である。「成長戦略」は「支援」系で選択されているが、「小さ (な)」は、「ベンチャー」系で選択されたものである。したがって、黒ラインの階層構造では、つながらない用語が、階層を横断する赤ラインのネットワーク機能によってつながるとい創発性がここで見事に発揮されている。

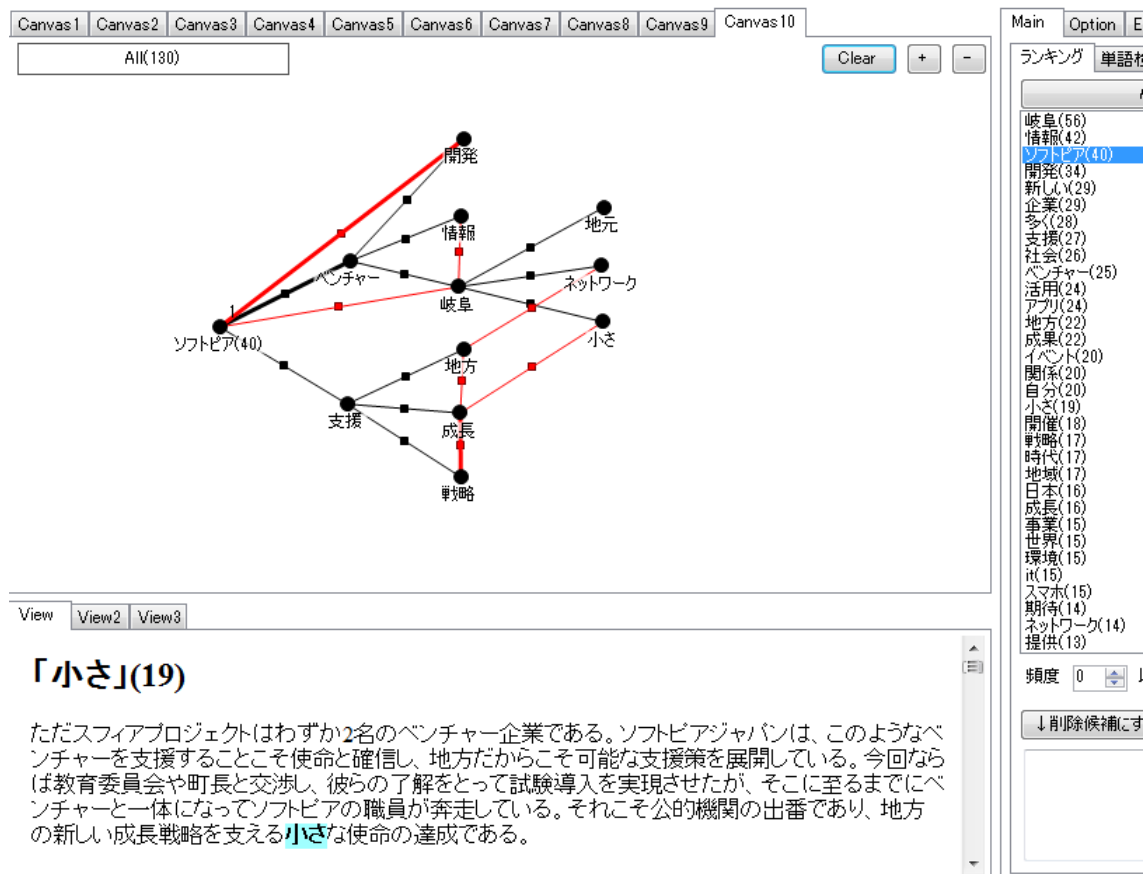


図 46 階層構造図をもとに作成

構造図を生成させる過程では、「支援」の3つの選択肢として、「小さな・成長・戦略」が選択可能ではあったにもかかわらず、それを選択せず、「地方・成長・戦略」を選択している。つまりその時点では、用語の関連性の広がりを考えて、「地方」を選択しており、目的項目になるはずの「小さな成長戦略」がここでは選択されていない。にもかかわらず、別系列の「ベンチャー」から「岐阜」が選択され、さらにそこから「地元・ネットワーク・小さな」が選択された。つまりこの階層構造をみるかぎり、「小さな・成長・戦略」は気づかれていない。ということは、解釈者の主体的判断では、「小さな・成長・戦略」は眼中にない。しかしこの階層構造を超えて、ツールのネットワーク機能によって、創発性が発揮されて、系列を横断して、「小さな・成長・戦略」が創発される。この瞬間、この構造図の成功が約束されたといっても過言ではない。階層構造とネットワーク機能の相互性が、いかに創発性に関して重要な役割をはたすか、これで実証されている。

つぎに、この目的項目を説明する手段項目の検証をしよう。手段項目は3つあるが、ここでは「イノベーション」についてのみ言及する。

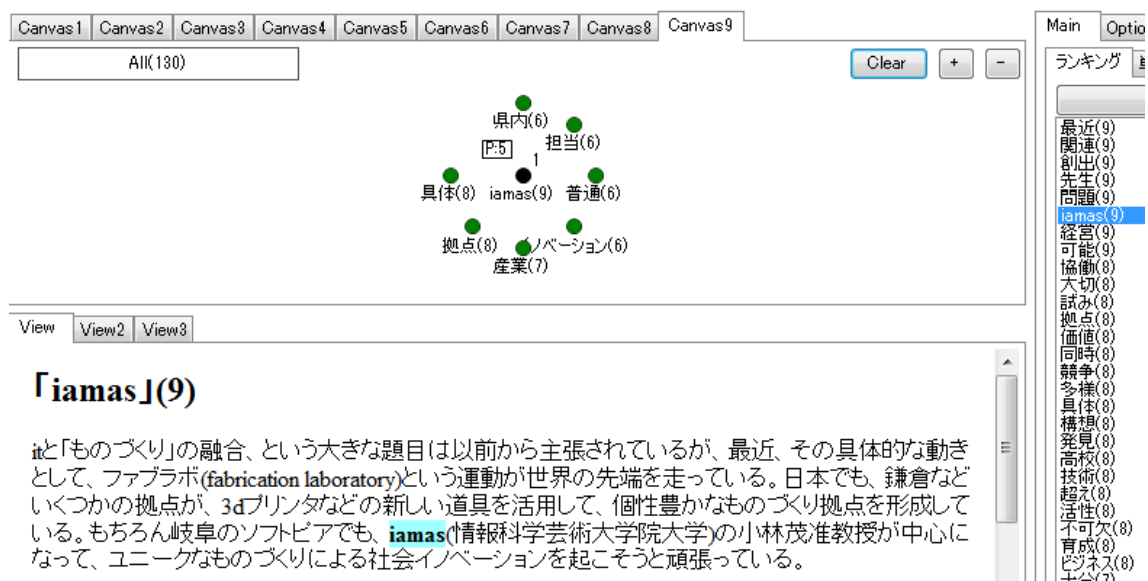


図 47 「iasmas」を起点にした時

「イノベーション」の発見は、強く主体的つまり恣意的である。図 47 に示すように、「イノベーション」は、「IAMAS」から探索される用語であるが、ここではかなり緑の円環を回している。5 巡目で、やっと「イノベーション」に出会っている。それまでは、特化用語は IAMAS の単なる特性を語る用語（大学院、科学、芸術など）の羅列で、それ自体の重要性は、その後、派生用語のレベルで採用するが、ここでは、それらを採用する決断ができず、何度も巡送りを繰り返した。「イノベーション」を見つけた瞬間、これだと直感した。まさに、これは解釈者の恣意的決断である。ツールからの提供される秩序に基づいた用語をすべて排除し、これに出会って、それだと思って決断したわけであるから、主観的な判断そのものである。その意味では、自由で勝手な思い込みである。これが、しかし、このツールを活用する場合の妙である。このツールを使いこまないかぎり、発揮できないパワーだともいえよう。よく言えば、名人芸の世界である。しかしこの巡りあわせをもたらしたのが、元はこのツールである、ということは意義深いことである。

ここでは、用語の一般と特殊の問題をいかに扱うのか、また、その扱いは状況によっていかに違うかを検証した。と同時に、ここでの深い熟慮と決断がいかに不可欠であるかを具体的に検証した。

イノベーションが発見されたことで、ここから IAMAS の特性を理解することはかなり時自動的に実行された。図 48 に示すように、「イノベーション」に特化する用語として、上位の「新産業」と「先端」が選択され、同時に、そこでは、先端が IAMAS と赤ラインで連動する形が生成された。そこで、この 4 つの用語が共起するパラグラフを抽出すると、1 つだけ提示された。そこでは、まさに、先端を走る IAMASこそ、新産業の拠点としてイノベーションを起こしてほしい、という文になっている。これこそ、構図で期待した手段項目である。

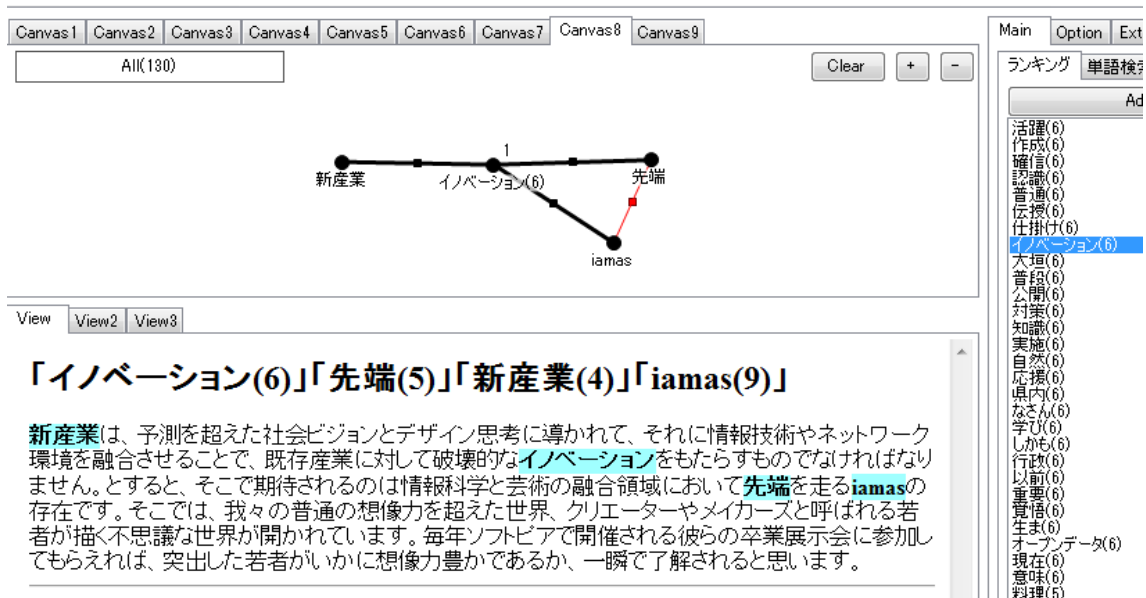


図 48 「イノベーション」「先端」「新産業」「iamas」を含む文章

このように、イノベーションを発見することで、ブレイクスルーができたと思う。その後の選択の操作は、ツールが提示する秩序に従って、かなり自動的に進められた。こうして手段項目が発見され、そこから、さらに、IAMAS と成長戦略をリンクするルートを模索することで、目的－手段関係は成就していった。

以上のように、構造図における自明性と創発性については、ビューにおけるエッセイデータのグラフ部分との照合によって、目的項目や手段項目の発見、またより特定化されたレベルでは、イノベーションの探索までの探索の結果から生成された創発性の発見など、多くの点で、预期されたイメージとは異なった、その意味で新鮮な驚きに通じる発見が見出された。と同時に、预期したイメージに導かれ、またそれを誘導するツールからの提示にほぼ自動的に従うことで、自明性の確保も十分に実現できた。自明性と創発性のバランスはどこなのか、という厳密な判定は不可能であるが、主観的な評価としては、十分に納得できる成果として構造図を産出することができたし、それを、ビューのデータで検証することも、十分ではないが、許容できる範囲で実現できたのではないかと、思う。

3-5 評価と課題

本ツールは3つの実査事例からどのような評価を受けたのか、その評価から見出された今後の課題は何かを明確にする。

3-5-1 評価1：自己探求とメタ認知

本ツールの実験を通して、ユーザから多くの評価が提示された。その中でも、本ツールの新規性を示すうえで、次のような重要な指摘があった。それは「これを使って、グラフを作成しようとすると、何度も行ったり来たり、試行錯誤を繰り返しながら、だんだん、自分の考えていることがはっきりしてくるようになるし、ああ、こういうことだったのか、という新鮮な気づきがある時はっきりみえてくる」という評価である。従来のマイニングツールと違って、本ルールがユーザを「自己探求する解読者」に仕立て上げる機能をもつことを示唆している。図 49 に示すように、ユーザ（探求主体：ひっぱりくん:Hipparu）は、ツール（合理他者：McS）が示す合理的な分析結果の提案に対して、その多くの選択

肢からもっとも意味的に適合する選択をし、その行為を通して、ユーザ自身が文脈的な解釈的探求を繰り返す主体である、と自覚するようになる。これこそが認知科学で主張されるメタ認知[18]である。メタ認知とは、自己の認知を認知する行為であり、気づきの連鎖過程を通して自己認識のフレームを変更し、さらには身体感覚の変化をもたらし過程であり、本ツールが求める自己探求過程そのものである。

このように、本ツールの第1の評価は、本ツールが探求主体（ユーザ）に対して次から次へと分析案を提示する合理他者であることで、探求主体にメタ認知を誘発するのに十分な機能をもつことを実証した点である。この評価は従来のツールにはない新規性である。

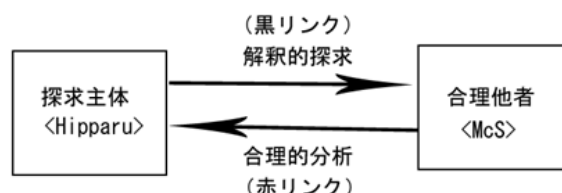


図 49 Hipparu と McS の関係図

3-5-2 評価2：柔軟い構造化と解釈多様性

本ツールの評価として2番目に重要な指摘は、実験2に参加したメンバーたちが述べたように、「同じデータでも、みんなが作る構造図は、それぞれこんなに違うんだ。できの良し悪しはあるけど、良い構造図が複数あるのは実感として納得だね。」というコメントに表れている。上述した実験2での2つの構造図に示されているように、本ツールは、解釈の多様性を自明なものとして作成されている。探求主体の問題意識が違えば、また合理他者の提示案を選択する視点が違えば、結果として作成される構造図は自ずと異なる。その差異こそが「柔軟い」構造化と呼ぶ所以である。従来のツールと違って、ユーザの問題意識と探求視点によって、探求主体が本ツールにコミットする度合いが非常に高まる仕組みになっており、そこから作成される構造図は多様な成果をもたらすことになる、という点がここでははっきりと実証されている。

以上から結論として、本ツールが求めた開発コンセプトは、上述の2つの実験において相当程度納得される評価を獲得している、といえよう。

3-5-3 課題

本ツールには、共起グラフへのコメント機能がないなど、細かな点について解決すべき課題が多いが、以下では、より本質的な課題を2つ挙げる。

第1に、実験2に示すように、本ツールでは解釈多様性を積極的に支持している。しかし解釈は多様だ、というだけで終わっては単なる相対主義（主観主義）であり、その多様性を統合する視点をいかに仕組みとして組み込むか、という重要な問題が残る。実験2で具体的に指摘したように、2つの構造図をツールとして自動的に統合するような仕組みの開発が必要である。そこで必要となるのが、同じデータを活用するならば多様な共起グラフを統合する機能、つまり上位の構造化を行う機能の開発である。これは多様性を補完する重要な統合機能である。

第2に、現状のツールでは共起グラフの正当化を示す客観性の担保がどこまであるのか、という重要な課題がある。現状では、探求主体の思いが強ければ、合理他者の影響を限りなく排除することは不可能ではない。ここでも多様性が強く支持されるがために、客観性が無視される事態が想定される。そこで客観性を支持する合理的な評価基準の付加機能、たとえばある基準以下の低頻度語群のリンクの色を薄めるといった機能などの開発が必要である。これによって、本ツールの客観的で合理的な手法の説得

力が増すと確信する。

第四章 探索的構造化ツールの 開発と実査

第四章 探索的構造化ツールの開発と実査

本章では、探索的社会調査ツールを構成する第2番目のツールである、探索的構造化ツール「こうぞうくん」について、その目的・コンセプトから、実装、そして実際の社会調査での解析結果について考察する。

4-1 開発コンセプト

「こうぞうくん」で扱うのは、バスケット分析の可視化である。商品の購買履歴や、自然言語を形態素解析後のデータを用いた分析である。共起関係に基づきアイテム間の関係を可視化する。これにより、商品購買なら、購買の関係図、自然言語なら、言葉の関係図を作り、データにある構造を読み解くことができるというものである。

このようなデータを分析するために、縦列をクラスタ、横列を頻度のレイヤーの二軸を使い、表形式でインタラクティブに表すツールを作成した。これは、「探索的構造化分析」 [13][36]において使われた手法のツール化である。

	< スポーツ情報 >	< 社会 >	< 一般受け >	< リア充エンタメ >	< 車 >	< ジャニーズ >
1	日本シリーズ 47news 博多 西武 朝日新聞デジタル	hide	朝ドラ シター 八重の桜 くまモン	miwa	ニノ 相模 spec arashi 翔くん	山田涼介 やまちゃん ジャニ びんとこな 知恵南季 知恵 もしかしてだ ー...
2	マー君 カープ 工藤 内田 稲葉 時事通信 デイリースポーツ	iphone 東大 安倍首相 ハムスター遠征 稲池 山本太郎	リーガル・ハイ 大奥 クドカン radiko 連続のモブ siko サマニ アムトーク	あつちゃん 忍び新究 きゅー ミステル 松尾 サマニ アムトーク	二宮 大野智 変態ゲーム 日産 松尾 anan	kitty ジャニーさん あいぼん apple 審判 なつちゃん バーナ
3	グリー ねとらぼ シネマトゥデイ 競馬 群馬県 まーくん 恵野家	みのもんだ 山口県 やなせたかし 岩手県 金髪庁 中日新聞 こはひびい	日曜劇場 ルパン 鶴姫 斎藤和義 菅野美穂 大河ドラマ	コプロ ワークマン 大島優子 西野カナ いきものがかり 高澤まさみ	ビルナンデス 北川景子 ノースアップ キムタク ショムニ 100均 キチちゃん	やっくん 太鼓の達人 ジャニオタ てへべろ 高田 田中聖 山下智久

図 50 可視化の最終的な形

最終的には、図 50 のような可視化を行うことが、探索的構造化ツールの開発の目的である。以下、作成したコンセプトを示す。

4-1-1 べき乗分布と頻度の層（レイヤー）

商品購買履歴のようなバスケット形式のデータを分析するとき必ず発生するのは、少数の高頻度のものと、多数の低頻度のもので構成されるべき乗分布になることである。自然言語処理の世界では、ジップの法則[34]と呼ばれるものであり、マーケティングの世界では、ロングテール[32]と呼ばれるものである。このような分布は必ず発生するものとして、分析に予め組み込む必要がある。べき乗分布の性質として、両対数グラフを作ると線形に近似するというものがある。（べき乗分布は反比例に近似し、 $x \cdot y = \alpha$ の両対数にすると、 $\log(y) = -\log(x) + \log(\alpha)$ となり直線となる。）これを利用して、最大の頻度の対数と、分析に使う最小の頻度の対数の差をとり、それを等分割することで、頻度の層（レイヤー）を作るということをする。これは、上のレイヤーからピラミッド状にアイテムの個数が増えていくものとなる。これを行にして、上の行は頻度の高いものであり、下にいくと頻度が小さいものと、直感的にわかるものとした。

4-1-2 インタラクティブな関係の表示とクラスタリング

関係性の表示のために、アイテムのクリック時、関係が強いアイテムに色を付けるということをしている。これは、グラフにおける、エッジを、インタラクティブに見せていることに相当する。そのため、複雑な模様になってしまいがちなネットワーク図と同じ情報をすっきりと見せるようにしている。縦の列で、なるべく関係の強いもので固めるという形で、クラスタリングを行う。教師なしで、関係のみを用いで行うため、自律的秩序形成であり、自己組織化させているともいえる。

	< スポーツ情報 >	< 社会 >	< 一般受け >	< リア充エンタメ >	< 萌 >	< ジャニーズ >
1	日本シリーズ 47news 博多 西武 朝日新聞デジタル	hide	朝ドラ ミッキー 八重の桜 くまモン	miwa	ニノ 相葉 spec arashi 翔くん	山田涼介 やまちゃん ジャコ びんとこな 知念侑李 知念 もしかしてだっど
2	マー君 カープ 工業 内田 稲葉 時事通信 デイリースポーツ	iphone 実入 安芸音羽 ハムスター達様 韓流 山本太郎	リーガルハイ 大受 あつちん 怒り研究 きりう ミスチル aiko サマニ アメトーク	あつちん 怒り研究 きりう ミスチル aiko サマニ アメトーク	ニノ 相葉 spec arashi 翔くん	kitty ジャニーさん あいもん apple 横アリ なっちゃん バーナ
3	グリー なとりば シネマトゥデイ 競馬 群馬県 まーくん 吉野家	みのもんた 山口県 やなせたかし 岩手県 金巻子 中日新聞 こけしむい	日曜劇場 ルパン 銭形 斎藤和義 宮野真悠 大河ドラマ	コクロ ウオーマン 大島優子 西野カナ いきものがかり 高澤まさみ	ヒルナンデス 北川景子 ノンストップ キムタク ショムニ 100円 キティちゃん	やっくん 太鼓の達人 ジャニオタ てへぺろ 高田 田中聖 山下智久

図 51 アイテム選択時、関係するアイテムの表示

図 51 において、赤は選択したアイテムであり、橙色は、関係しているアイテム。薄い橙は、弱い関係である。ネットワークを可視化しているといえる。

4-1-3 概念化とメタ認知

このようなクラスタリングによる、データの可視化だけでは不十分である。社会学的分析するためには、概念化が必要である。概念化とは、分析の全体像を考えることであり、それにしたがって、クラスタを表す言葉を探し、名付け、その塊を分析者が把握することである。そのような概念化をすることで、事象の解釈が可能になる。この行為は、認知科学的に言えば、メタ認知による言葉での外化である[18]。その促進ため、クラスタに名前をつけられるようにしている。

4-1-4 制約付きクラスタリング

名前をつける時に困るのは、データに忠実で機械的な処理に基づくクラスタリングの結果では、人間が考える概念とは、しばしばズレることである。このズレを解決するためには、人間の背景知識をクラスタリングの結果に反映する制約付きクラスタリング[37]という手法を用いる。制約付きクラスタリングとは、MustLink、CanNotLink を予め指定し、その情報を付加した上でクラスタリングする手法である。ここでは、グループ指定による、制約情報を UI の上で加えるようにする。ユーザが直感的に行えるように「固定する」というメタファで説明している。UI により固定化したアイテムは、クラスタリング時には、動かなくなる。当然、動かないとしても、そのアイテムの関係情報は使って、他のアイテムには影響している。

4-1-5 クラスタリングの失敗の可視化

このクラスタリングは、必ずどこかのクラスタに所属させるハードクラスタリングのため、うまくクラスタリングできていないアイテムが発生することがある。このようなものは、複数のクラスタと関係をもつものであり、ネットワーク構造的にはハブである。ネットワーク分析ではハブの重要性はしばしば指摘される[33][35]。特に、低頻度のハブは、**KeyGraph** で言う、赤ノードに相当し、**KeyGraph** の考案者である大澤の主張では、そのようなものにはチャンスが眠っているとされる[20]。このようなクラスタリングに失敗しているハブ的なものは、縦列に並んでいるものは、同じクラスタであるという、可視化のルールから外れるため、赤丸をつけ可視化する。

4-1-6 クラスタ間の関係の可視化

縦と横の二次元の可視化では、クラスタ間がどのような関係になっているかがわからない。概念化により、概念同士の関係がどのようなになっているかを知るためにも、クラスタをノードとして、グラフとして可視化した。この際、クラスタに対して適切な名前を与えていないと、機械的な名前になり、イメージ出来ないものになってしまう。このことでも、名前をつけることを促進させている。

4-1-7 属性情報の付加と可視化

データには、いつどこでだれがといった、5W1Hの情報が本来的にある。このような属性情報をテーブルに重ねあわせる仕組みを用意する。

4-2 実装

実装は、C#で行い、**Silverlight** というブラウザのプラグイン上で実行できるようにした。そのため、Windows と Mac のブラウザ上で実行できる。**Silverlight** にしたのは、Windows と Mac 両方で実行できるということと、最新版への更新が簡易なこと、ブラウザ実行とはいえ、ローカルファイルを扱え、通常のアプリケーション同様のことができるからである。

4-2-1 入力ファイル

こうぞうくんは、バスケット分析の可視化のツールであり、入力データもバスケットの形のデータである必要がある。そのため、**UserId** 文字列と **Item** 名をタブ区切りで一行としたに **UTF-8** でエンコードテキストデータを入力データとした。属性データも同様に、**UserId** 文字列と属性名と値のタブ区切りを一行としたテキストデータである。

4-2-2 画面の説明

ツールの基本画面は以下のようにになっている（図 52）。

	< 地名食べ物 >	< スポーツ情報 >	< 社会 >	< 一般受け >	< リア充エンタメ >	< 風 >	< ジャニーズ >	< ラタク >
1		日本シリーズ 47news 博多 西武 朝日新聞デジタル	hide	朝ドラ ミッシー 八重の桜 くまモン	miwa	ニノ 相葉 spec arashi 翔くん	山田涼介 やまちゃん ジャニ びんとこな 知念侑李 知念 もしかしてだっど ...	バズドラ コメケ ジョジョ 銀魂 ガンダム
2	下北沢 立川 八王子 吉祥寺 日本酒	マー君 カープ 工藤 内田 稲葉 時事通信 デイリースポーツ ...	iphone 東大 安倍首相 ハムスター速報 韓流 山本太郎	リーガル・ハイ 大奥 クドカン radiko 潮騒のメモリー 福山雅治 ...	あつちゃん 怒り新党 きやろ ミスチル aiko サマシ アメトーーク	二宮 大野智 家族ゲーム 日産 松浦 中居 anan ...	kitty ジャニーさん あいぽん apple 横アリ なつちゃん バーナ ...	初音ミク ボカロ アニメイト エヴァ 腐女子 仮面ライダー プリキュア ...
3	新宿駅 ドトール 渋谷駅 心斎橋 二郎 らーめん 伊勢丹 ...	グリー ねとらぼ シネマトゥデイ 競馬 群馬県 まーくん 吉野家 ...	みのもんた 山口県 やなせたかし 岩手県 金沢市 中日新聞 こぼろとい ...	日曜劇場 ルパン 鶴瓶 斉藤和義 菅野美穂 大河ドラマ ...	コブクロ ウオーグマン 大島優子 西野カナ いきものがかり 長澤まさみ ...	ヒルナンデス 北川景子 ノンストップ キムタク ショムニ 100均 キティちゃん ...	やっくん 太鼓の達人 ジャニオタ てへべる 高田 田中聖 山下智久 ...	ウルトラマン ニコニコ動画 アキバ ヤマト じゃがりこ マリオ ピカチュウ ...

図 53 クラスタ（縦列）に名前をつけられる

4-2-3 クラスタリングのアルゴリズム

クラスタリングのアルゴリズムは、可視化に合わせて作成した。「こうぞうくん」におけるクラスタリングとは、縦の列において、関係が強いものが固まることである。これは、アイテム間の関係のグラフに基づくクラスタリングであり、グラフクラスタリングを行っていることになる。「こうぞうくん」では、K-Means 法と自己組織化マップ[38]のアルゴリズムと同様にランダムな初期値からクラスタを形成していく手法をとっている。ある任意のアイテムを選択時に、表示される関係が強いクラスタに移動することを繰り返す。

予め指定する変数は、「クラスタ数」と、「選択関係数」、上位いくつまでをアイテム間の関係として採用するかを決める数と、「試行回数」の3つである。クラスタ数は、UIにより指定し、「選択関係数」、「試行回数」は、クラスタリング用のパネルで指定する。

準備段階として、アイテム間の関係を共起関係に基づき、補正信頼度²⁰をもとめ、重み付き隣接行列(a)を作成する。このとき、一つのアイテムにつき、補正信頼度上位「選択関係数」個分を採用し、それより小さいものを0とする。このように指定個数以下の切り捨てを行うのは、多くのアイテムとの関係があると、クラスタとして分離が難しくなるためであり、人間側の直感的なクラスタから離れてしまうためである。これは計算処理量を減らすことにもつながる。

1. 初めに、すべてのアイテムをランダムにクラスタを割り当てる。固定化されたアイテムに関しては、UIにより指定されたクラスタを使う。

²⁰ 補正信頼度 = 信頼度 - 支持度

$$\text{Lift 値} = \frac{\text{信頼度}}{\text{支持度}}$$

補正信頼度、Lift 値は、信頼度（条件付き確率）を支持度（出現確率）との関係でどのように評価するか？という指標である。ともに、信頼度と支持度の違いが大きいと評価が高まる。違いは、補正信頼度は、Lift 値が割り算である代わりに引き算になっていることである。Lift 値は、割り算を使うので、値域が0から無限大までとるが、補正信頼度は、-1から1の間で実データでは、-0.1~0.3 ぐらいの値に収まるため、扱いやすい。

2. すべての非固定化アイテムに対して、重み付き隣接行列を使い、どのクラスタが一番強い関係であるかを判別する。式は数式①を使い、これが最大値を持つクラスタにアイテムを移動する。(ただし、移動は同一頻度レイヤー間で行い、移動情報は一時表に保存)

$$\text{アイテム}\alpha\text{のクラスタ}\beta\text{の所属度} = \frac{\text{アイテム}\alpha\text{とクラスタ}\beta\text{に所属するアイテムとの重みの総和}}{\text{0より大きい重みのクラスタ}\beta\text{に所属するアイテム数}} \quad \text{①}$$

3. 一時表に保存したものを本表にアップデート。固定化アイテムは同じ位置のままである。
4. 移動したアイテムが 0 なら終了、あるなら、2 に移動。
5. 1~4 を、指定回数繰り返し、評価値が最も高いものを表示させる。

このような重み付きグラフのクラスタリングとした。このようなアルゴリズムにしたのは、ツールを使う人間にとって直感的である方法だからである。このツールは元々、クラスタリング機能は持っておらず、可視化ツールであり、アイテムをクリックした時に、それに関係するアイテムが表示されるだけであった。その関係するアイテムが多いクラスタに移動させることは、自然な行為である。これをすべてのアイテムに対して行い、収束するまで繰り返すということを行っている。

移動情報を一時表に保存するのは、移動したことによる変化の影響を与えないためである。仮に、移動情報の保存なしでは、移動する順番によって結果が変わってしまう。これは、自己組織化マップにおける、バッチ型 SOM の手法から来ている。

4-2-4 クラスタリングの評価指標

初期配置を乱数で作るため、その初期配置に依存して、クラスタリング結果は変わる。そのため、多数の初期配置から実行して、その中で、最もいいクラスタリング結果を探す必要がある。クラスタリングの評価指標にはジニ係数を使う。ジニ係数は、格差を示す経済指標として有名だが、機械学習でも使われている。(例えば決定木[39])。性質としては、値の格差が大きいと 1 に近づき、格差が小さいと 0 に近づく。このジニ係数を使い、2 つの軸を持って評価する。

1. すべてのアイテムが縦列で、まとまっていることがいいクラスタリングである。
 - (ア) すべてのアイテムで、指定個数分の補正信頼度の高い順にとり、それをクラスタごとに総和を求める。これを変数としてジニ係数を求める。
 - (イ) 求まったそれぞれのジニ係数の平均値を出す。1 に近いほどいい。
2. 可視化として、それぞれのクラスタに入っているアイテムの数が均等に近いのがいいクラスタリングである。
 - (ア) 頻度レイヤーごとに、クラスタごとのアイテム数を数え、これを変数としてジニ係数を求める。
 - (イ) 求まった各頻度レイヤーのジニ係数の相乗をだす。0 に近いほどいい。
3. 1 と 2 を 2 つを掛けあわせたのを最終的な指標とした。ただし、1 と 2 は向きが違うので、向きを揃えた。

1 は、ツールでアイテムそれぞれをクリックした時、関係しているアイテムが表示されるところが、縦の列でなるべくまとまっているというのを表現している。少ないクラスタでまとまっているとき、評価値は、1 に近づく。しかし、1 だけでは、不十分だった。この指標を最大化するには、クラスタの空欄を増やせば増やすほど高くなるため、クラスタ数の指定が意味を成さないことが判明した。そのため、なるべく空欄をださないように、それぞれのクラスタに均等になったもののがいいクラスタリングと評価されるように 2 の指標を追加した。

4-2-5 クラスタリングの失敗の検出

1 の評価指標は、一つ一つのアイテムにおいて、1 に近づけば近づくほど、クラスタとしてまとまっていることを意味する。逆に、0 に近いものは、複数のクラスタと関係を持っているアイテムであるということ、すなわち、ネットワーク構造上、ハブになっているものと思われる。縦の列でクラスタを作っていることが可視化のルールなので、そのルールから外れているので、このようなものを可視化する。具体的には、1 の平均値を求める前のデータで、ジニ係数が低い物順に指定個に対して、赤い丸をつける。

4-2-6 クラスタマップの作成

表形式では、クラスタ間の関係がわからない。そのため、図 54 のようにクラスタをノードとした、グラフを作成する。クラスタ間の関係には Jaccard 係数を使い、エッジの足切りには、Lift 値を使った。Lift 値の足切りにより、エッジの数を増やしても、グラフは完全グラフにはならないようにした。また、ノードの名前は、名付けていないと、素っ気ない機械的な名前になるので、積極的にクラスタに名前をつけることを促進している。また、関係性の実データが見えるようにもしている。

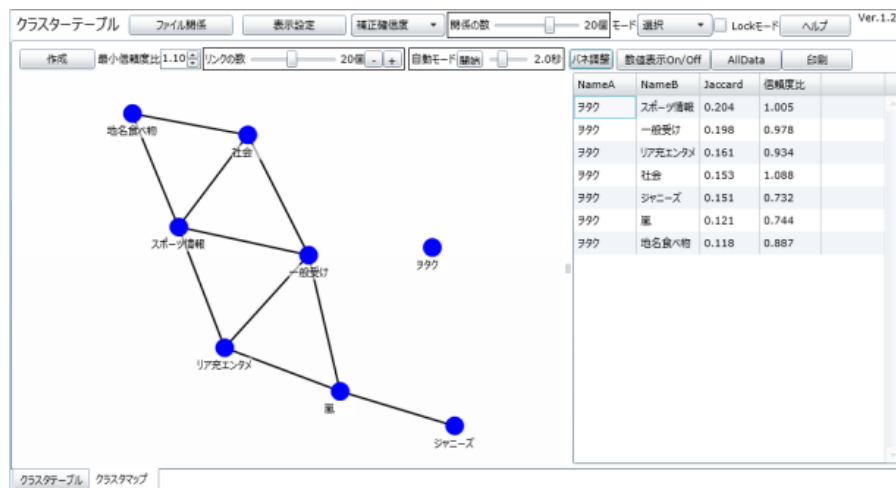


図 54 クラスタ間の関係を表示

4-3 実査 4：テレビドラマ（半沢直樹）のツイート分析

4-3-1 分析結果

2013 年、テレビドラマの「半沢直樹」は、最終話の視聴率が 42.2%（ビデオリサーチ調べ）という空前のヒットを飛ばした。この「半沢直樹」についての調査を行った。

分析データは、ツイッターで、「半沢直樹」の公式アカウント(@Hanzawa_Naoki)をフォローしているユーザ（45,315 人）のツイートを 2013 年 11 月に取得した。クリーニングとして、オープンであり、言語が日本語であり、ツイート数が 2,000 以上のユーザを使った。約 11,000 人に絞られた。そのツイートの中から、形態素解析を行い、頻出語 250 語を抽出し、その頻出語を用いて、バスケットを作成した。

また、そのユーザが特につぶやいた「半沢直樹」の俳優名と役名を属性とし、このデータを用いて、本ツールを使い分析を行った。

まず、はじめにクラスタリングを行った時、ジャニーズと嵐が混ざった感じの大きいクラスタを形成していた。この2つは当然結びつきが強いが、数として大きすぎるので、2つを分けるために、次のように固定化を行った(図 55)。嵐のメンバーと、中居くんなどの他のジャニーズのメンバーを分けるようにした。(図 56)

< ジャニーズ >	< 嵐 >
☐ もしかしてだけど	☒ 相葉
☐ 有岡大貴	☒ ニノ
☐ 裕翔	☒ 翔くん
☒ 知念侑李	☒ arashi
☐ 山田涼介	
☐ ジャニヲタ	
☐ ジャニショ	
☐ ジャニ	
☐ マリウス	
☐ ぴんとこな	
☐ 中島裕翔	
☐ 手越	
☐ 北斗	☒ 二宮
☒ 中居くん	☐ エイト
☐ パーナ	☐ あいぼん
☒ 中居	☐ 日産
☐ 錦戸	☒ 大野智
☐ 亀梨	☐ 家族ゲーム
☐ 横アリ	☒ 松潤
☐ ジャニーさん	☐ 大倉
☐ kitty	☐ anan
☐ apple	☒ 櫻井翔
☐ なっちゃん	☐ 安田

図 55 ロック機能

1	ミュージック	おっさん向けテレビ	スポーツ	若者向けテレビ	ニュース	ジャニーズ	嵐	おたく
1	☒ miwa	☐ 朝ドラ ☐ くもん ☐ 八重の桜 ☐ ミッシー	☐ 日本シリーズ ☐ 西武 ☐ 博多	☐ spec	☐ 47news ☐ 朝日新聞デジタル ☐ hide	☐ もしかしてだけど ☐ 有岡大貴 ☐ 裕翔 ☒ 知念侑李 ☐ 山田涼介 ☐ ジャニヲタ ☐ ジャニショ ☐ ジャニ ☐ マリウス ☐ ぴんとこな ☐ 中島裕翔 ☐ 手越	☒ 相葉 ☒ ニノ ☒ 翔くん ☒ arashi	☐ ガンダム ☐ コミケ ☐ バズドラ ☐ 燃魂 ☐ ジョジョ
2	☐ 八王子 ☐ ミスチル ☐ 吉祥寺 ☐ サマンソ ☐ 立川 ☐ alko ☐ 下北沢 ☒ サザン	☐ ガッキー ☐ 堀山雅治 ☐ オフ会 ☐ 大奥 ☐ じえし ☐ 楽屋のメモリー ☐ リーガル・ハイ ☐ クドカン ☐ 綾野剛 ☐ 能年玲奈	☐ ヤクルト ☐ 工藤 ☐ ティリースポーツ ☐ 内田 ☐ 稲葉 ☐ 日ハムスポーツ ☐ 長谷部 ☐ 美濃 ☐ マー君 ☐ イチロー ☐ カーブ	☐ オールナイトニッポン ☐ クモリ ☐ アムトーク ☐ スマスマ ☐ radiko ☐ 有吉 ☐ あつちゃん ☐ びり研究 ☐ きりー ☐ 安室ロイド	☐ 山本太郎 ☐ 安倍晋三 ☐ 産経新聞 ☐ 日本酒 ☐ ハムスター道場 ☐ 時事通信 ☐ 経済 ☐ 報道 ☐ 日本経済新聞	☐ 北斗 ☒ 中居くん ☐ パーナ ☒ 中居 ☐ 指原 ☐ 亀梨 ☐ 横アリ ☐ ジャニーさん ☐ kitty ☐ apple ☐ なっちゃん	☒ 二宮 ☐ エイト ☐ あいぼん ☐ 日産 ☒ 大野智 ☐ 家族ゲーム ☒ 松潤 ☐ 大倉 ☐ anan ☒ 櫻井翔 ☐ 安田	☐ 初音ミク ☐ 仮面ライダー ☐ モンハン ☐ アニメイト ☐ 高女子 ☐ プレミア ☐ ホロロ ☐ エヴァ
3	☐ いきものがかり ☐ 小田急 ☒ ウォークマン ☐ コパコ ☐ 表参道 ☐ ドトール ☐ 伊勢丹 ☐ ミュージックステーション ☐ 心斎橋 ☐ ハーゲンダッツ ☐ きりーばあばあ ☒ じゃがりこ	☐ リラクマ ☐ 上戸彩 ☐ 徹子の部屋 ☐ 日曜劇場 ☐ 小栗旬 ☐ 大河ドラマ ☐ 向井理 ☐ 三浦春馬 ☐ 斎藤和義 ☐ 石原さとみ ☐ やなせたかし ☐ 王様のブランチ	☐ 東スポ ☐ 競馬 ☐ 近鉄 ☐ 早稲田 ☐ 世界理上 ☒ 金コン ☐ プレレス ☐ サンクイズスポーツ ☐ まーくん	☐ 海城 ☐ シムニ ☐ 戸田恵梨香 ☐ woman ☐ めちゃイケ ☐ スキちゃん ☐ ファースト ☐ 高橋孝之 ☐ リンカーン ☐ タモさん ☐ 前田歌子 ☐ まゆゆ	☐ 8-めん ☐ 山口県 ☐ パチンコ ☐ 渋谷駅 ☐ 金剛寺 ☐ 新橋駅 ☐ 静岡県 ☐ 中日新聞 ☐ 大阪駅 ☐ みのもんだ ☐ センター試験 ☒ ロフト	☐ 激おこ ☐ フラベチーノ ☐ 山下智久 ☐ うえーい ☐ サマーヌード ☐ 嵐ヲ ☐ つらたん ☐ シェアハウス ☐ ぶり ☐ 太鼓の達人 ☐ ネプリーグ ☐ 高田	☒ 100均 ☐ セルナンデス ☐ 北川景子	☐ 二郎 ☐ ヤマト ☐ 剛力彩芽 ☐ ハリーポッター ☐ ドラクエ ☐ ニコ動 ☐ マリヲ ☐ ウルトランマン ☐ ニコニコ動画 ☐ 耳をすませば ☐ ハリー ☐ Up

図 56 最終的なクラスタリング結果

最終的には、クラスタに名前をつけて、図 57 のような結果になった。「半沢直樹」はテレビドラマであり、テレビドラマ的な要素が大きいことがわかる。その中でも、「おっさん向けテレビ」の大河ドラマ、朝ドラが好きな層と「若者向けテレビ」の若手のお笑いタレントが集まるクラスタと、「嵐」「ジャニーズ」のクラスタが発生したことがわかる。

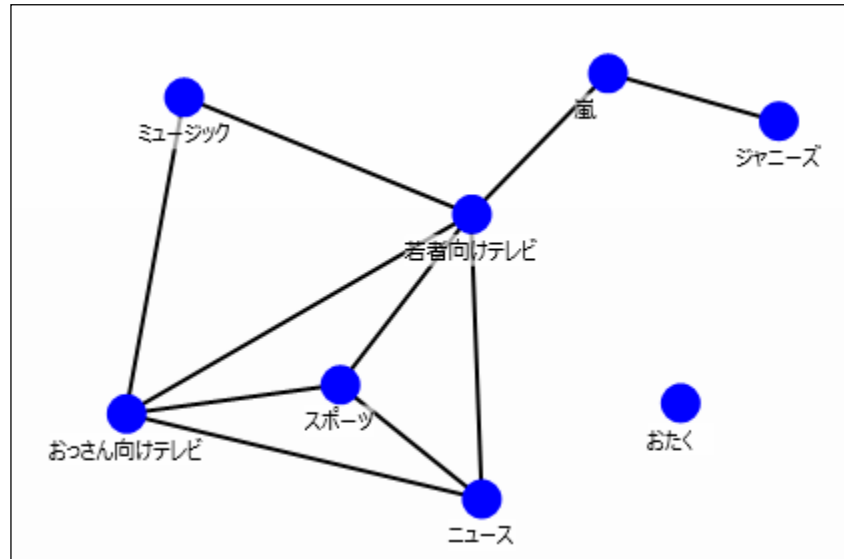


図 57 クラスタ間の関係図

クラスタマップを作るとこのようになる。全体がつながるようにと Lift 値 1.1 で作成した。しかし、「おたく」のクラスタは、テレビ関係のクラスタとは強い関係性はなく、独立関係となった。ジャニーズも嵐を媒介項として全体像とつながっていることがわかる点も面白い。

次に、属性で見る。これらのクラスタは、どの俳優役名と関係が強いかをみる。赤いところが、その俳優で特化しているところで、青色が特化していないところである。堺雅人(図 58)、壇蜜(図 59)、大和田常務(図 60)の結果を示す。主役の半沢直樹は、タイトル名であるため、分析にはそぐわない。

	< ミュージック >	< おっさん向けテレビ >	< スポーツ >	< 若者向けテレビ >	< ニュース >	< ジャニーズ >	< 嵐 >	< おたく >
1	●miwa	朝ドラ くもモン 八重の桜 ミッチー	日本シリーズ 西武 博多	spec	47news 朝日新聞デジタル hide	もしかしてだ 有岡大貴 裕翔 知念侑李 山田涼介 ジャニタタ ジャニシヨ ジャニ マリウス	相葉 二ノ arashi	ガンダム コメケ バズドラ 銀魂 ジョジョ
2	八王子 ミスチル 吉祥寺 サマニ 立川 aiko 下北沢 ●サザン	ガッキー 福山雅治 オフ会 大奥 じょしあ 潮騒のメモリー リーガル・ハイ クドカン 綾野剛	ヤクルト 工藤 デイリースポーツ 内田 極楽 日刊スポーツ 長谷部 姜媛 マー君	オールナイトニッポン タモリ アメトーク スマスマ radiko 有吉 あつちゃん 怒り新党 きゅりー	山本太郎 安倍晋相 産経新聞 日本酒 ハムスター速報 時事通信 韓流 東大 テレビ東京	北斗 中居くん バーナ 中居 錦戸 亀梨 横アリ ジャニーさん kitty	二宮 エイト あいばん 日産 大野智 家族ゲーム 松潤 大倉 anan	初音ミク 仮面ライダー モンハン アニメイト 腐女子 プリキュア ボカロ エヴァ
3	いきものがかり 小田急 ●ウオーグマン コブクロ 表参道 ドトール 伊勢丹 ミュージックステーション 心斎橋	ムック 香川照之 ルパン 鶴瓶 富野美穂 まんたんウェブ 堀北真希 大泉洋 ●やられたらやり返す	東スポ 競馬 近鉄 早稲田 世界陸上 ●会コン プロレス サンケイスポーツ まーくん	海猿 ショムニ 戸田恵梨香 woman めっちゃイケ スギちゃん アメトーク 長澤まさみ リンカーン	らーめん 山口県 パチンコ 渋谷駅 金銭庁 新宿駅 ●静岡県 中日新聞 大阪駅	激おこ フラバチーノ 山下智久 うそーい サマーヌード 昼ドラ つらたん シェアハウス ゴリ	100均 ビルナンデス 北川景子	二郎 ヤマト 剛力彩芽 ハリウッド ドラクエ ニコ動 マリオ ワルトロマン ニコニコ動画

図 58 堺雅人に特化しているクラスタ

	< ミュージック >	< おっさん向けテレビ >	< スポーツ >	< 若者向けテレビ >	< ニュース >	< ジャニーズ >	< 嵐 >	< おたく >
1	●miwa	朝ドラ くまモン 八重の桜 ミッチー	日本シリーズ 西武 博多	spec	47news 朝日新聞デジタル hide	もしかしてだけど 有岡大貴 裕翔 知念侑李 山田涼介 ジャニタタ ジャニシヨ ジャニ マリウス	相葉 ニノ 翔くん arashi	ガンダム コミケ バズドラ 銀魂 ジョジョ
2	八王子 ミステル 吉祥寺 サマソニ 立川 aiko 下北沢 ●サザン	ガッキー 福山雅治 オフ会 大奥 じえい 潮騒のメモリー リーガル・ハイ クドカン 綾野剛	ヤクルト 工藤 テイルススポーツ 内田 稲葉 日刊スポーツ 長谷部 愛媛 マー君	オールナイトニッポン タモリ アメトーク スマスマ radiko 有吉 あつちゃん 怒り新党 きやうり	山本太郎 安倍晋相 産経新聞 日本酒 ハムスター速報 時事通信 韓流 東大 テレビ東京	北斗 中居くん バーナ 中居 鏡戸 亀梨 横アリ ジャニーさん kitty	二宮 エイト あいぼん 日産 大野智 家族ゲーム 松浦 大倉 anan	初音ミク 仮面ライダー モンハン アニメイト 腐女子 プリキュア ボカロ エヴァ
3	いきものがかり 小田急 ●ウオクマン コプロ 表参道 ドトール 伊勢丹 ミュージックステーション 心斎橋	ムック 香川照之 ルパン 鶴瓶 菅野美穂 まんたんウェブ 堀北真希 大泉洋 ●やられたらやり返す	東スポ 競馬 近鉄 早稲田 世界陸上 ●会コン プロレス サンケイスポーツ まーくん	海猿 ショムニ 戸田恵梨香 woman めっちゃイ スギちゃん アメトーク 長澤まさみ リンカーン	らーめん 山口県 パチンコ 渋谷駅 金融庁 新宿駅 静岡県 中日新聞 大阪駅	激おこ フラベチーノ 山下智久 うそーい サマーヌード 昼ドラ つらたん シェアハウス ゴリ	●100均 ヒルナンデス 北川景子	二郎 ヤマト ●劇力彩芽 ハリーポッター ドラクエ ニコ動 マリオ ウルトラマン ニコニコ動画

図 59 壇蜜に特化しているクラスタ

	< ミュージック >	< おっさん向けテレビ >	< スポーツ >	< 若者向けテレビ >	< ニュース >	< ジャニーズ >	< 嵐 >	< おたく >
1	●miwa	朝ドラ くまモン 八重の桜 ミッチー	日本シリーズ 西武 博多	spec	47news 朝日新聞デジタル hide	もしかしてだけど 有岡大貴 裕翔 知念侑李 山田涼介 ジャニタタ ジャニシヨ ジャニ マリウス	相葉 ニノ 翔くん arashi	ガンダム コミケ バズドラ 銀魂 ジョジョ
2	八王子 ミステル 吉祥寺 サマソニ 立川 aiko 下北沢 ●サザン	ガッキー 福山雅治 オフ会 大奥 じえい 潮騒のメモリー リーガル・ハイ クドカン 綾野剛	ヤクルト 工藤 テイルススポーツ 内田 稲葉 日刊スポーツ 長谷部 愛媛 マー君	オールナイトニッポン タモリ アメトーク スマスマ radiko 有吉 あつちゃん 怒り新党 きやうり	山本太郎 安倍晋相 産経新聞 日本酒 ハムスター速報 時事通信 韓流 東大 テレビ東京	北斗 中居くん バーナ 中居 鏡戸 亀梨 横アリ ジャニーさん kitty	二宮 エイト あいぼん 日産 大野智 家族ゲーム 松浦 大倉 anan	初音ミク 仮面ライダー モンハン アニメイト 腐女子 プリキュア ボカロ エヴァ
3	いきものがかり 小田急 ●ウオクマン コプロ 表参道 ドトール 伊勢丹 ミュージックステーション 心斎橋	ムック 香川照之 ルパン 鶴瓶 菅野美穂 まんたんウェブ 堀北真希 大泉洋 ●やられたらやり返す	東スポ 競馬 近鉄 早稲田 世界陸上 ●会コン プロレス サンケイスポーツ まーくん	海猿 ショムニ 戸田恵梨香 woman めっちゃイ スギちゃん アメトーク 長澤まさみ リンカーン	らーめん 山口県 パチンコ 渋谷駅 金融庁 新宿駅 静岡県 中日新聞 大阪駅	激おこ フラベチーノ 山下智久 うそーい サマーヌード 昼ドラ つらたん シェアハウス ゴリ	●100均 ヒルナンデス 北川景子	二郎 ヤマト ●劇力彩芽 ハリーポッター ドラクエ ニコ動 マリオ ウルトラマン ニコニコ動画

図 60 大和田常務に特化しているクラスタ

主役を演じた、堺雅人が特化しているのは、ドラマクラスタだとわかる。一方、女優の壇蜜が、特化しているのは、ニュースクラスタであり、普段ニュースについてつぶやいている層に壇蜜は受けたというのが想像できる。当然、ジャニーズクラスタには不人気である。

クラスタマップにおいて、独立だった、「おたく」クラスタのみ特化していたのは、大和田常務である。大和田常務は、その顔芸がネットでヒットして、ネット上のまとめサイトで、いろいろな形でまとめられており、その影響だと思われる。

このように考えると、ドラマ「半沢直樹」は、国民的ヒットになっていったことは、おぼろげながら見えてくる。つまり、普段からドラマを見ている人たちを惹きつけ、ジャニーズ出演でジャニーズ好きな人たちを惹きつけ、壇蜜で、普段、ニュースをつぶやいている人たちを惹きつけ、大和田常務で、テレビドラマを見ない、ネットだけを見ている層を惹きつけることに成功したことが、大ヒットに繋がった、ということが推察される。

このような分析ができるのも、このツールだからこそである。

4-3-2 考察

クラスタリングは、ランダムな初期配置から作成していくアルゴリズムであるため、同じデータであれば、同じ結果を必ず保証するものではない。また、クラスタリングとしては、大雑把なクラスタリングのため、精密なクラスタリングのために、制約付きクラスタリングの枠組みを利用しているものとなっている。クラスタリングには正解はないとはいえ、分析者の能力に依存するところが大きい。

知的インタラクティブシステムのために、システムとして最小のユーザフィードバックで済むことが望ましいとされる[23]。このツールの場合、機械的な仕組みとして、最小のユーザフィードバックをサポートするような仕組みは存在しない。しかし、意味的な側面と可視化としてのサポートはある。それは、クラスタに名前をつけるのだから、意味合いとして大きい高頻度のアイテムを固定化すべきという意味的な要請と、アイテムをクリック時の関係の表示で、すでに相互に関係があって塊を形成しているものに対して固定化をしてもナンセンスであるということである。そのため、固定化すべきものは、UI 的におぼろげながら示していると言える。しかし、これも、分析者の能力への依存が大きく、初めて使うユーザにとっては不親切であり、何かしらの改善の余地はあるだろう。

とはいえ、ツールの使用者に聞くと、初めに出力される結果にある程度満足してしまうようである。そのため、固定化による制約付きクラスタリングは、アドバンストな機能であるといえる。

4-4 実査 5 『ハチミツとクローバー』に見る、普通の若者のカルチャースタイル

4-4-1 はじめに

羽海野チカ原作の『ハチミツとクローバー』（以下、ハチクロ）というマンガがある。2000 年から 2006 年にかけて連載され、第 27 回講談社マンガ賞少女部門を受賞、2006 年および 2007 年にはマンガ紹介ムック『このマンガがすごい！』にてオンナ編第 1 位を獲得するなど、2000 年代のマンガ界を語るにあたってなくてはならない存在だ。さらに、マンガという枠にとどまること無く、2005 年からはアニメ番組が、そして 2006 年には映画化、2008 年にはテレビドラマが放送された。美術大学を舞台とした青春群像劇というストーリーは、テレビドラマや普段積極的にマンガを読むことの無い、いわば「オタクではない」普通の若者たちに広く浸透していった。

かつて大学を舞台にした青春群像劇といえば、1980 年代にシリーズを開始しそのさきがけとなった「ふぞろいの林檎たち」や、1990 年代前半に月 9 ドラマの地位を確立した「あすなろ白書」などが挙げられる。若者たちはテレビの前で、ドラマから、その年代の人間関係の有り様やその時代の文化を学んだことだろう。このように、マンガやドラマのような様々なコンテンツは、その時代の文化と深い結びつきを持っていると考えられる。

時代との深い結びつきという意味においては、2000 年代は、映画化と多角的に展開された、まさにハチクロの時代と言えるだろう。そこで、本節では、ハチクロを対象として、2000 年代以降の普通の若者のカルチャースタイルを紐解きたい。分析可視化ツール「こうぞうくん」を用いて、mixi にてハチクロを支持するユーザのコミュニティを探索的に分析することで、現代社会の普通の若者はいかにして生きているのかを検討する。

4-4-2 分析対象データ

分析対象とするデータは、mixi コミュニティ「ハチミツとクローバー」に参加するユーザ、およびそのユーザらが参加する全てのコミュニティである。データ収集は、2013 年 9 月 30 日である。全ユ

ーザの中から、まずランダムサンプリングで 52,044 人を抽出し、彼らが参加する全コミュニティデータを取得した。

4-4-3 データクリーニング

4-4-3-1 ユーザのクリーニング

収集したユーザらの参加するコミュニティ数を見ると、そこにはばらつきがある。極端に参加数の少ないユーザはつながりの情報が乏しく、極端に多いユーザはノイズとなる可能性があるため除外する必要がある。そのため本分析では、極端に参加コミュニティ数の少ない第 1 層と、参加コミュニティ数の多い第 10 層を除き、第 2～8 層にあたる 40,108 人を分析対象とした。

参加コミュニティ数に基づいた層化の基準を表 1 に示す。

表 1 参加コミュニティ数ごとに等分割して層化

層番号	参加コミュニティ数	ユーザ数
1	1～23	4,870
2	24～39	5,010
3	40～53	4,924
4	54～71	5,134
5	72～91	4,921
6	92～115	5,105
7	116～154	5,003
8	155～217	5,025
9	218～369	4,986
10	370～1000	5,023

4-4-3-2 コミュニティのクリーニング

分析対象とするコミュニティは、ハチクロを支持するユーザの特性だけでなく、mixi ユーザとしての全体特性も含まれる。具体的には、「早い、簡単、旨い料理♪」「心に響いた名言集」「世界中を旅行したい!」をはじめとする全 mixi ユーザが最も参加しているコミュニティ群は、mixi に参加しているユーザの多くが支持する性質のものであるため、ハチクロを支持するユーザのスタイルを適切に映し出しているとは言いがたい。そのため、mixi ユーザ全体での参加コミュニティ数上位 500 件は、分析対象外としてクリーニングする。

コミュニティのクリーニング後、コミュニティを参加ユーザ数降順で並べ、上位 30 コミュニティをレイヤー1、次いで 60 コミュニティをレイヤー2、次いで 90 コミュニティをレイヤー3とした。レイヤーと各レイヤー内コミュニティ数、参加ユーザ数を表 2 に示す。

このようにして抽出された、計 180 コミュニティを分析対象として定め、ユーザの共起情報に基づき分析を行っていく。

表 2 各レイヤーのコミュニティ数と参加人数

レイヤー	コミュニティ数	参加ユーザ数
1	30	1,472 人以上
2	60	1,127 人以上
3	90	898 人以上

4-4-4 構造化

4-4-3 節にてクリーニングされたデータ、32,000 人のユーザと 180 コミュニティを分析対象として、ハチクロを支持する若者のカルチャースタイルを探索していく。

4-4-4-1 デフォルトの出力結果

クラスタリングのデフォルト設定は、クラスタ数 6、試行回数 10、選択関係数 20 となっている。出力結果を図 61 に示す。各系列内にコミュニティ数が多く、系統の特徴がわかりづらい状態となっている。



図 61 はじめにクラスタリングした結果

4-4-4-2 クラスタ数の調整

デフォルトの試行では系統内にコミュニティ数が多すぎるという問題があったため、クラスタ数を 6 から 8 に増やし再試行する。出力結果を図 62 に示す。デフォルトと比較するとばらつきはできたものの、依然として系統内にコミュニティ数が多い状態となっており、適切な系統化が行われていない。

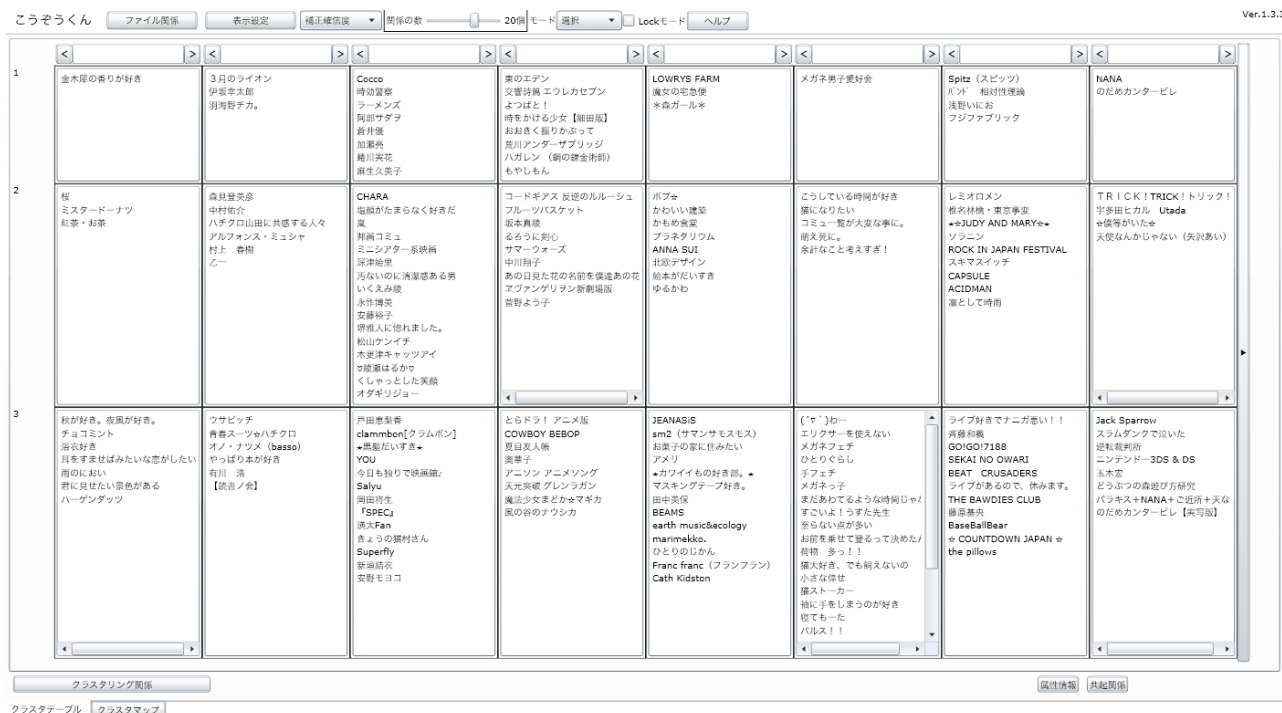


図 62 クラスタ数を8にしてクラスタリングした結果

4-4-4-3 ユーザの共起情報に基づく調整

続いて、ユーザの共起情報に基づき、コミュニティ1つ1つを精査し、系統間の調整を行う。通常、1つのコミュニティに着目する際、関連の高いコミュニティが同一系統内に多ければ多いほど、系統内での関係は強いと考えられる。そのため、同一系統内での関係がそれほど強くないコミュニティ群を、データおよび解釈に基づき調整していくことが必要となる。なお、レイヤーは参加ユーザ数に基づいているため調整は行わず、系統のみ移動によって調整を行う。

以降の系統名は仮のものであり、解釈を進めていく過程で修正を行っていく。

系統1：香り

一番左に位置する系統は、レイヤー1に「金木犀の香りが好き」、レイヤー2に「桜」「紅茶・お茶」、レイヤー3に「秋が好き。夜風が好き」「雨のにおい」など、香りに関するコミュニティが中心となって構成されている。

ここで、調整が必要なコミュニティを精査していくと、レイヤー3の「耳をすませばみたいな恋がしたい」は、図 63 のように他の系統と広く関係があり、かつ、解釈的にも香りにはあまり関連がない。そこで、同コミュニティは、系統5、レイヤー1に「魔女の宅急便」がある系統へと移動する。2コミュニティ間で強い関連があり、かつ、同じジブリアニメという意味における移動である。

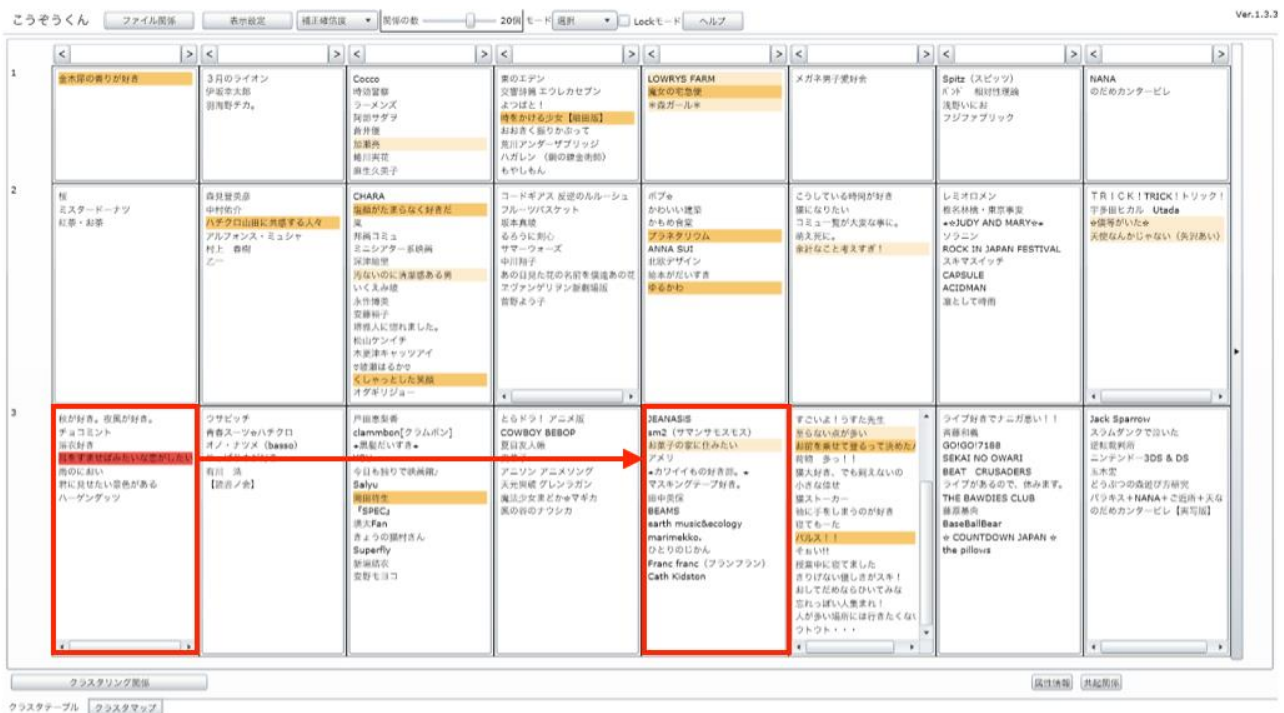


図 63 解釈を元にアイテムを移動 (1)

系統 2 : 本

左から 2 列目に位置する系統は、図 64 に示す通り、レイヤー1 に「3 月のライオン」「伊坂幸太郎」、レイヤー2 に「森見登美彦」「村上春樹」、レイヤー3 に「オノ・ナツメ(basso)」「やっぱり本が好き」など、本に関するコミュニティで構成されている。

ここで調整したいコミュニティは、レイヤー2「アルフォンス・ミュシャ」、レイヤー3「ウサビッチ」である。アール・ヌーヴォーを代表するデザイナーの「アルフォンス・ミュシャ」は、「Cocco」や「蜷川実花」と関連が強く、アートの側面が強いことから、系統 3 に移動する。「ウサビッチ」は同じアニメコミュニティが多い系統 4 へ移動する。

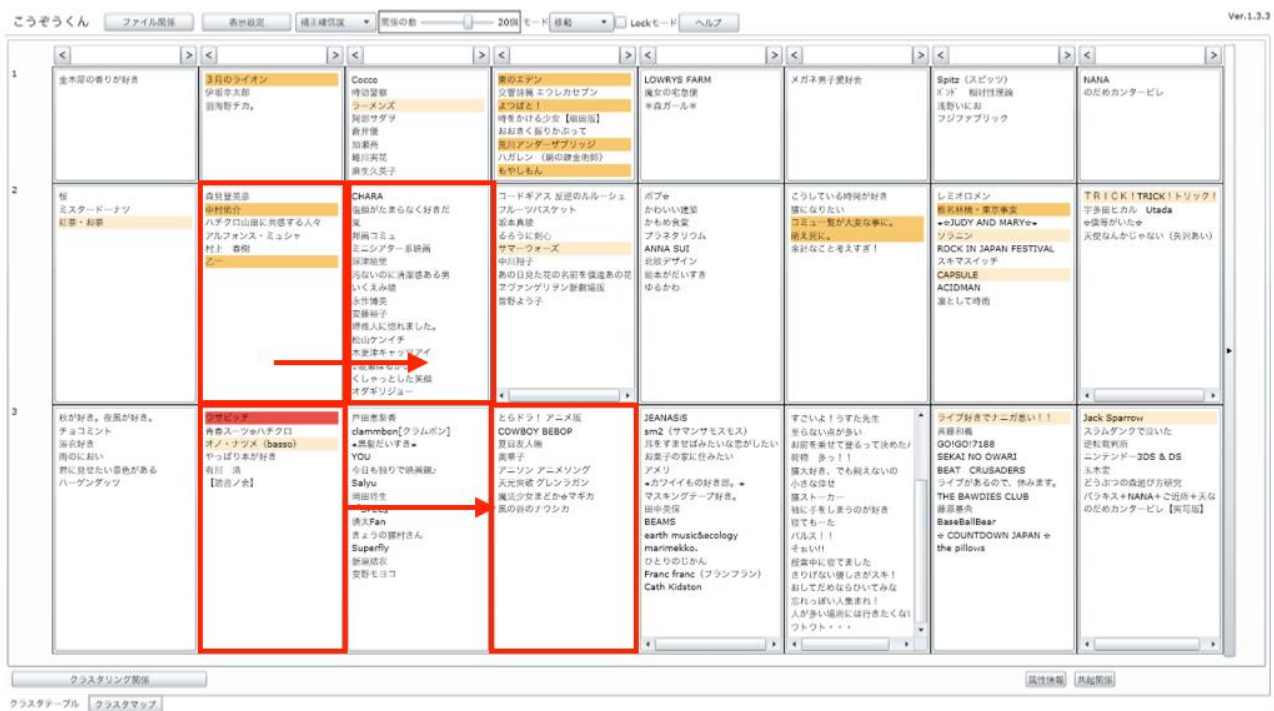


図 64 解釈を元にアイテムを移動 (2)

系統 3：芸能

左から 3 列目に位置する系統は、図 65 に示す通り、レイヤー1 に「時効警察」「蒼井優」、レイヤー 2 に「深津絵里」「オダギリジョー」、レイヤー3 に「戸田恵梨香」「clammbon[クラムボン]」と、芸能人やテレビ番組等で構成されている。

コミュニティ数が多いものの、系統内のコミュニティの連関が強くまとまりがあり、全体は芸能関係として一貫性があるため、ここでは調整をしない。

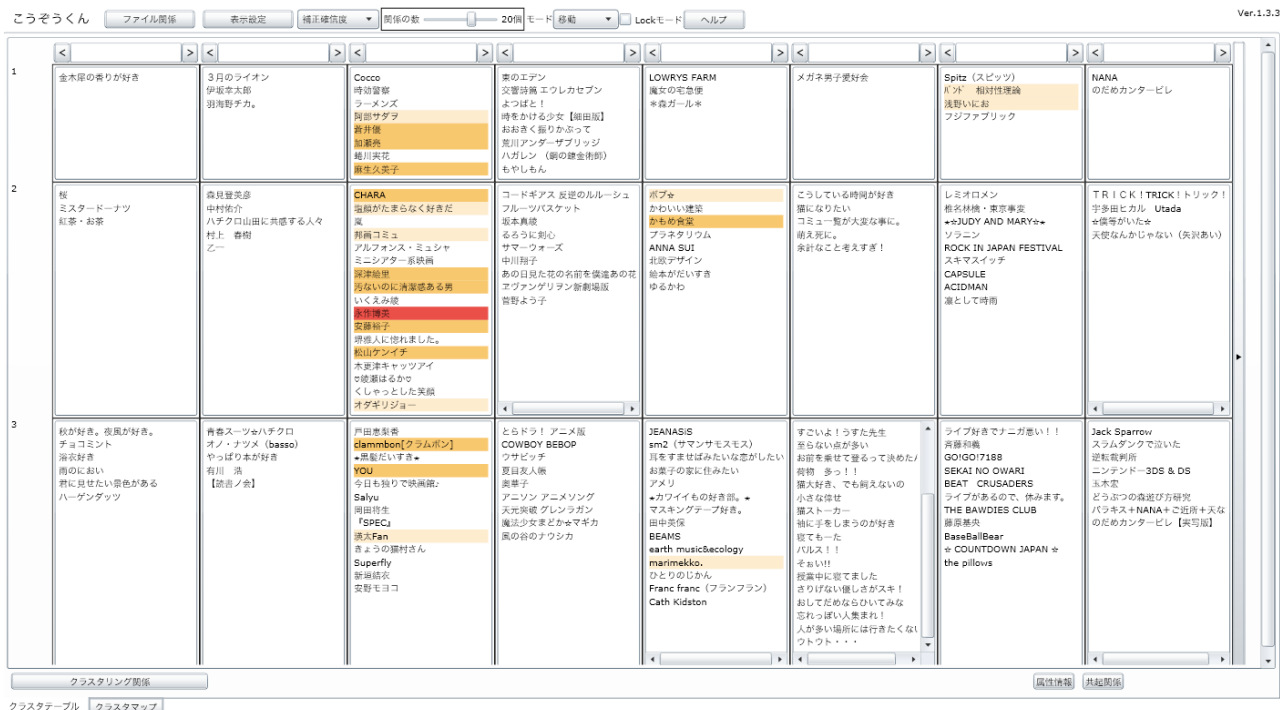


図 65 「永作博美」でのアイテムの関係

系統 4：アニメ

左から 4 列目に位置する系統は、図 66 に示す通り、レイヤー1 に「東のエデン」「よつばと!」、レイヤー2 に「コードギアス 反逆のルルーシュ」「サマーウォーズ」、レイヤー3 に「夏目友人帳」「アニソン アニメソング」と、アニメに関連するコミュニティで構成されている。

この系統も、系統 3：芸能と同様、系統内でのコミュニティ連関が強く、全体に一貫性があるため、ここでは調整を行わない。

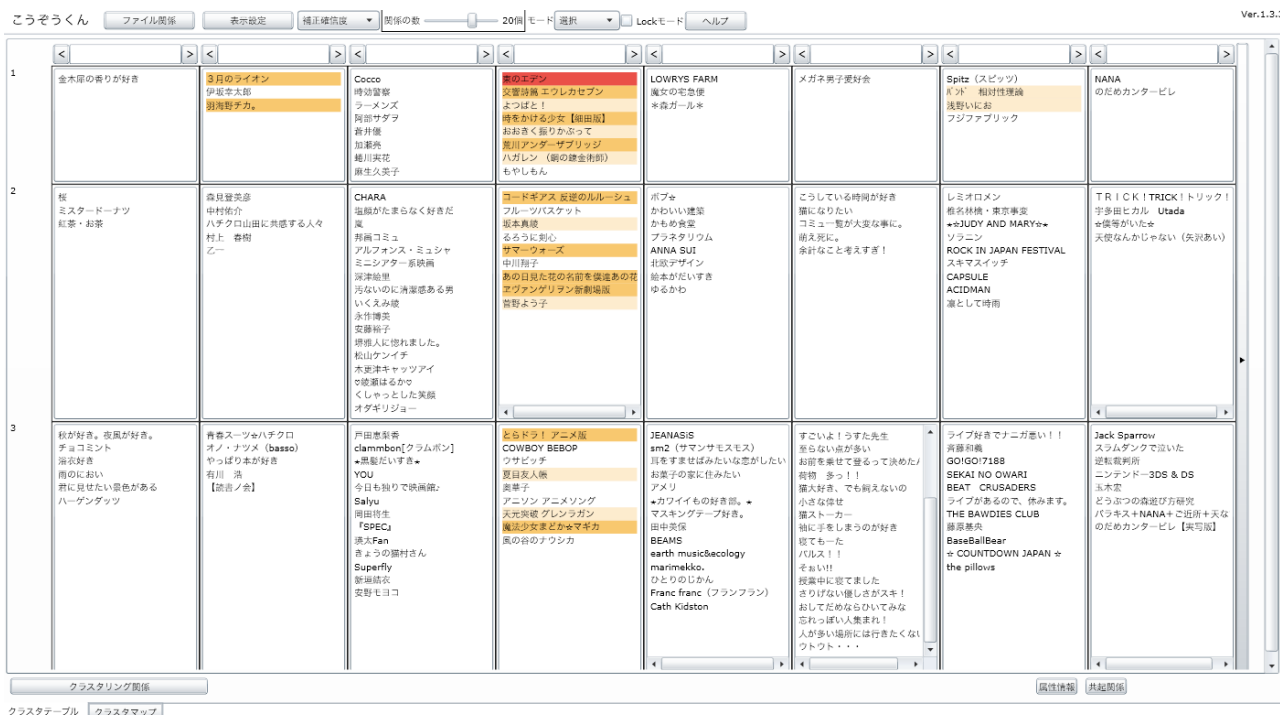


図 66 「東のエデン」でのアイテムの関連

系統5：女性向けライフスタイル

左から5列目に位置する系統は、図67に示す通り、レイヤー1に「LOWRYS FARM」「*森ガール*」、レイヤー2に「ANNA SUI」「北欧デザイン」、レイヤー3に「sm2(サマンサモスモス)」「earch music&ecology」と、女性向けのファッションをはじめとするライフスタイルに関するコミュニティで構成されている。

コミュニティの内容は、ファッションブランドから映画、デザイン等多岐にわたるが、ライフスタイルという枠で統一感を持って説明できるため、ここでもコミュニティの調整は行わない。特に、本来ならば、レイヤー2「かもめ食堂」レイヤー3「アメリ」「田中美穂」は系統3：芸能への移動も考えられるが、映画に関しては作品が描き出す世界観がライフスタイルの手本となるようなものであり、「田中美穂」に関しても芸能人としてでなく、あこがれのライフスタイルの体現者としての位置づけであると解釈できるため、調整は行わない。

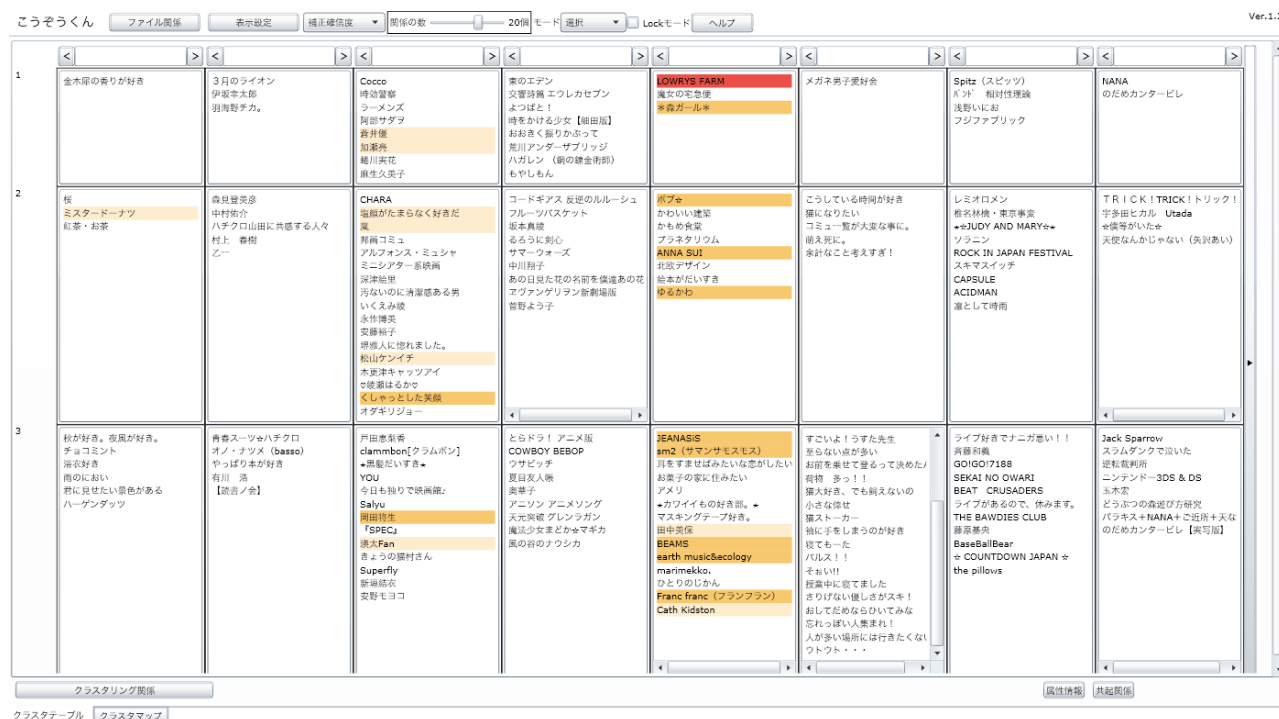


図 67 「LOWRYS FARM」でのアイテムの関係

系統6：自分らしさ

左から6列目に位置する系統は、図68に示す通り、レイヤー1に「メガネ男子愛好会」、レイヤー2に「猫になりたい」「萌え死に。」、レイヤー3に「手フェチ」「さりげない優しさがスキ!」と、ややネタっぽくもある自分らしさに関するコミュニティで構成されている。

コミュニティの内容は、自分の内面についてが多く、ファッションやコンテンツ以外に自分らしさを表現できる mixi コミュニティの特徴をよく表した系統であると言える。この系統では、「(▽`)わー」「うとうと・・・」は、元ネタが不明であることや、自己の内面性の表現としては内容に乏しいと考えられるため、削除した。

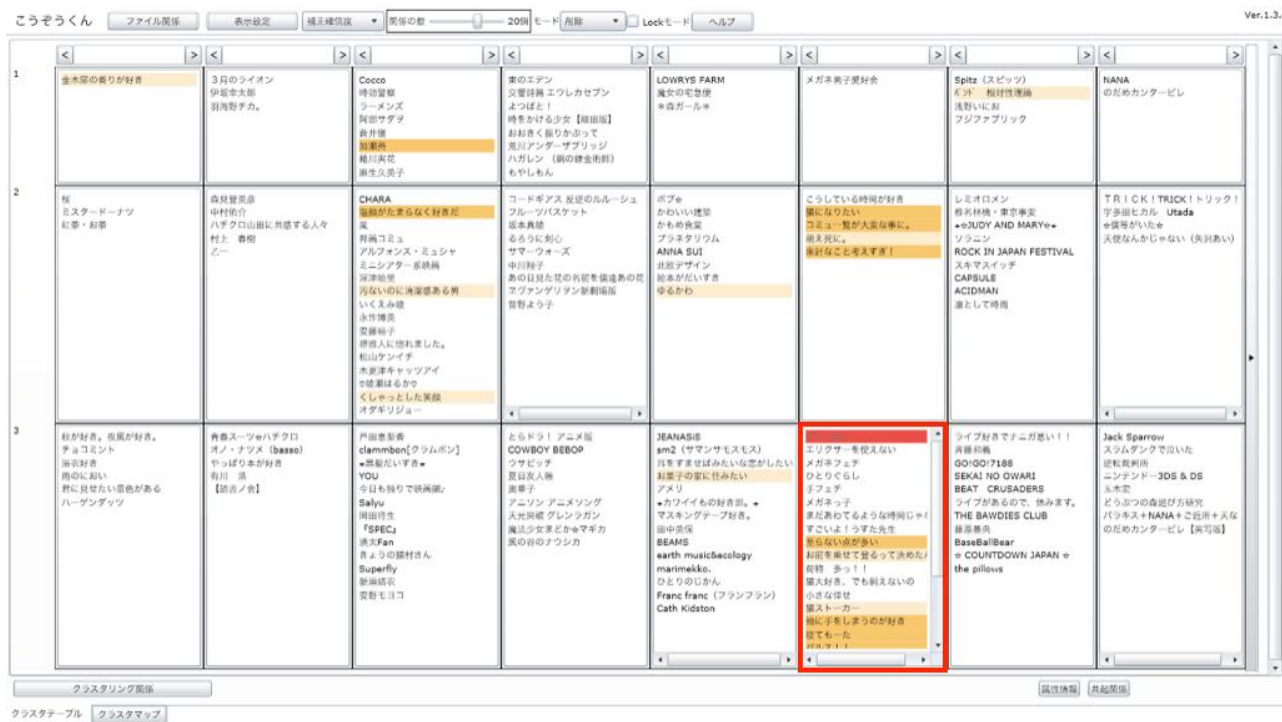


図 68 「(▽ `)わー」でのアイテムの関係

系統 7：音楽

左から 7 列目に位置する系統は、図 69 に示す通り、レイヤー1 に「Spitz(スピッツ)」 「バンド 相対性理論」、レイヤー2 に「レミオロメン」「椎名林檎・東京事変」、レイヤー3 に「SEKAI NO OWARI」「ライブがあるので、休みます。」といった音楽に関するコミュニティで構成されている。

調整が必要と考えられるコミュニティは、レイヤー1「浅野いにお」、レイヤー2「ソラニン」である。「浅野いにお」はマンガ家であり、「ソラニン」は彼の代表作である。作品の内容は音楽と関連性が高いが、コミュニティ間のつながりを見ると、どちらも本および芸能の系統に連関を持っていることがわかる。そのため、両コミュニティを、データおよび内容に基づき系統 2：本へ移動する。

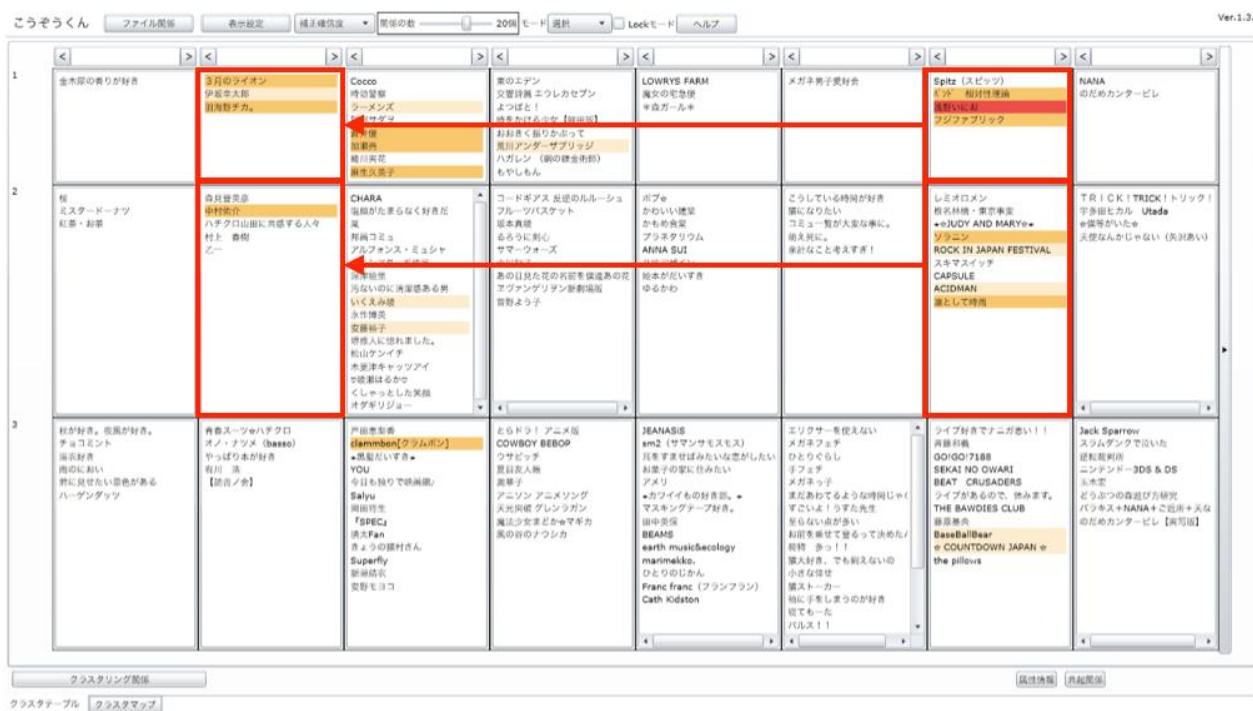


図 69 「浅野にいお」「ソラニン」移動

系統 8：女性向けコンテンツ

左から 8 列目に位置する系統は、図 70 に示す通り、レイヤー1 に「NANA」「のだめカンタービレ」、レイヤー2 に「宇多田ヒカル Utada」「*僕等がいた*」、レイヤー3 に「動物の森遊び方研究」「パラキス+NANA+ご近所+天ない」と、女性向けのマンガや映画等のコンテンツに関するコミュニティで構成されている。

調整が必要と考えられるコミュニティは、レイヤー2「TRICK! TRICK! トリック!」である。ユーザの共起を見ると、系統 3：芸能および系統 4：アニメの両者につながりがあるが、コミュニティがテレビドラマや映画に関するものであることを鑑み、系統 3：芸能へと移動する。その他のコミュニティに関しては、「Jack Sparrow」「玉木宏」「ニンテンドー3DS & DS」「逆転裁判所」など、一見女性向けコンテンツとは関連が薄く見えるコミュニティがあるが、「Jack Sparrow」「玉木宏」は女性向けコンテンツの映画化の際に関係があり、「ニンテンドー3DS & DS」「逆転裁判所」は女性向けコンテンツのプラットフォームおよびそこでの展開として位置づけられるため、移動はしない。

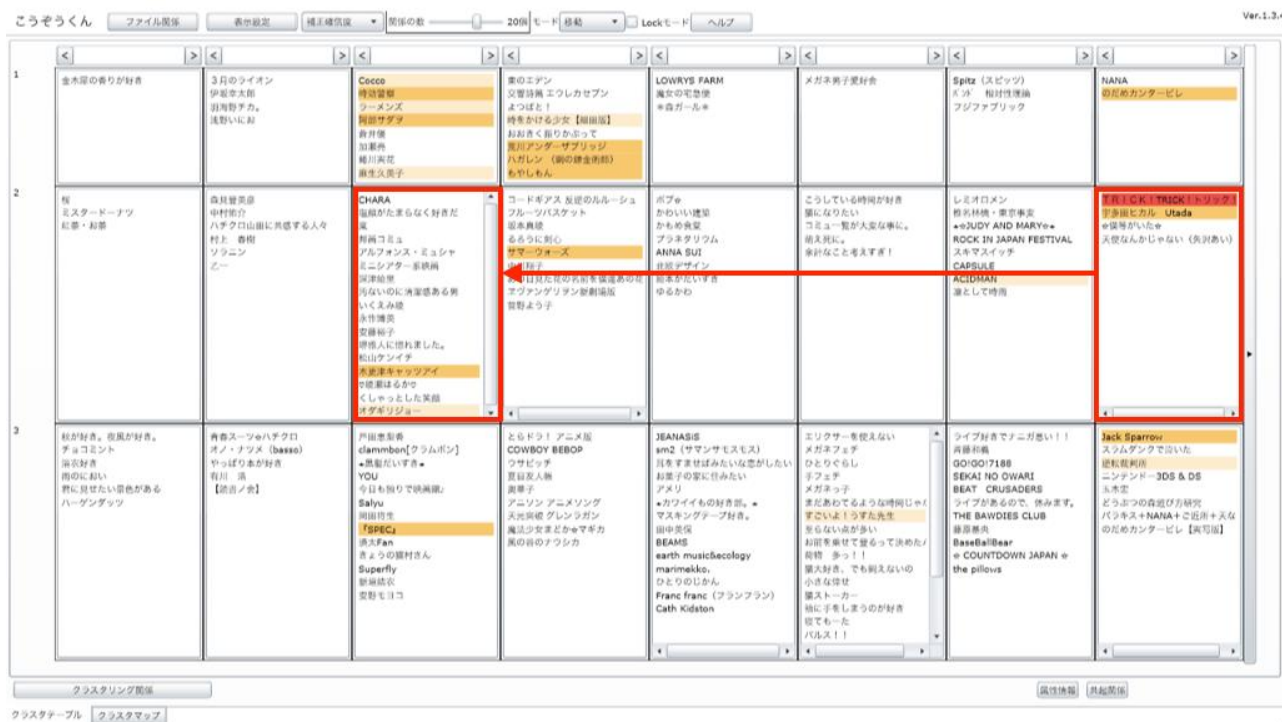


図 70 「TRICK! TRICK! トリック！」の移動

ユーザの共起情報に基づいて調整したコミュニティ

以上、8 系統に関して、個別のコミュニティを、ユーザの共起と系統内の統一性を基準とし解釈しながら調整を行った。調整したコミュニティ一覧を表 3 に示す。

表 3 移動コミュニティ一覧

レイヤー	コミュニティ名	移動前系統	移動後系統
3	耳をすませばみたいな恋がしたい	1: 香り	5: 女性向けライフスタイル
2	アルフォンス・ミュシャ	2: 本	3: 芸能
3	ウサビッチ	2: 本	4: アニメ
1	浅野いにお	7: 音楽	2: 本
2	ソラニン	7: 音楽	2: 本
2	TRICK! TRICK! トリック！	8: 女性向けコンテンツ	3: 芸能

ユーザ数共起に基づいて調整し作成した系統

これまでの調整に基づき作成した系統一覧を表 4 に示す。コミュニティには、各ユーザは任意の数参加できるため、支持ユーザ数は重複ありと見なされる。

表 4 調整後のコミュニティ名とユーザの人数と男女の割合

系統 ID	仮系統名	支持ユーザ数	男性ユーザ割合	女性ユーザ割合	不明ユーザ割合
1	香り	8,362 人	17.91%	63.23%	18.86%

2	本	16,560 人	25.46%	55.92%	18.62%
3	芸能	21,799 人	19.23%	62.84%	17.93%
4	アニメ	16,485 人	31.07%	49.88%	19.05%
5	女性向けライフスタイル	15,218 人	11.11%	71.08%	17.81%
6	自分らしさ	15,810 人	21.32%	60.58%	18.10%
7	音楽	13,955 人	24.59%	58.37%	17.03%
8	女性向けコンテンツ	12,639 人	20.93%	61.07%	18.00%

4-4-4-4 属性情報に基づく調整

続いて、属性情報に基づき全体構造を調整していく。本分析で対象とするユーザ 40,108 人のうち、分析対象としたレイヤー1 から 3 まで、全 180 コミュニティのどれか 1 つ以上に参加しているユーザは 37,627 人となっている。そのうち、男性は 9,130 人(24.26%)、女性は 21,774 人(57.87%)、不明が 6,723 人(17.87%)と、半数以上を女性が占めている。不明ユーザに関しては、性別の公開範囲によって情報取得できなかったものである。

性別特化の基準値は特化係数によって求めているため、半数以上が女性で構成されている本分析においては「女性に特化しない」コミュニティ群の抽出が難しいことが前提にある。属性の特化に関しては、選択属性に特化するクラスタをピンク色で、非特化クラスタを青色で示している。

女性特化の調整

属性を女性に定めた図を、図 71 に示す。ユーザの過半数が女性ということもあり、全体的に女性に支持される傾向の強いコミュニティ群で全体が構成されているため、女性特化と系統を合わせて見た際の不一致は特にない。したがって、ここでは調整を行わない。

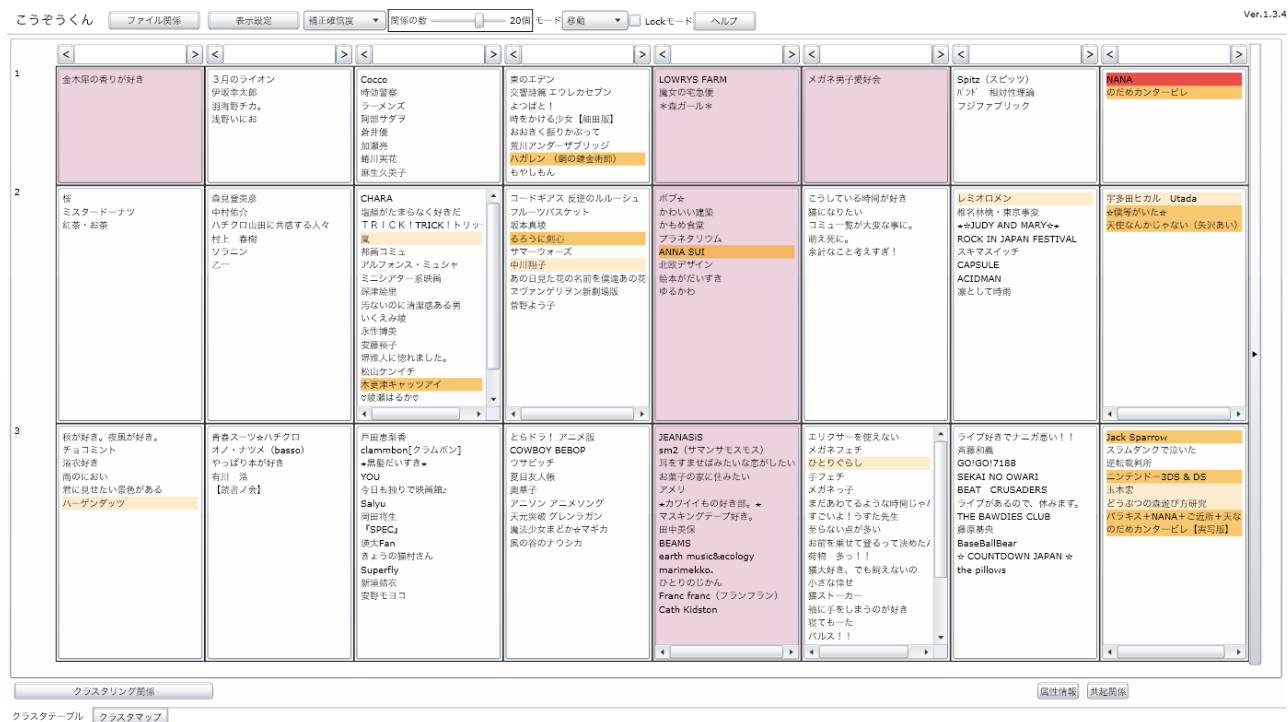


図 71 属性「女性」での特化

3	エリクサーを使えない	系統 6：自分らしさ	系統 4：アニメ
3	すごいよ！うすた先生	系統 6：自分らしさ	系統 4：アニメ

属性情報に基づいて調整し作成した系統

これまでの調整に基づき作成した系統一覧を表 6 に示す。

表 6 系統のユーザ数と男女の割合の一覧表

系統 ID	仮系統名	支持 ユーザ数	男性 ユーザ割合	女性 ユーザ割合	不明 ユーザ割合
1	香り	8,362 人	17.91%	63.23%	18.86%
2	本	16,560 人	25.46%	55.92%	18.62%
3	芸能	21,799 人	19.23%	62.84%	17.93%
4	アニメ	17,149 人	31.07%	49.94%	18.99%
5	女性向けライフスタイル	15,218 人	11.11%	71.08%	17.81%
6	自分らしさ	15,197 人	20.18%	61.73%	18.09%
7	音楽	13,955 人	24.59%	58.37%	17.03%
8	女性向けコンテンツ	12,639 人	20.93%	61.07%	18.00%

4-4-5 構造の解釈

これまで探索的に分析を行ってきたコミュニティ構造に関して、以降ではコミュニティ名に基づき解釈を行っていく。プロセスとしては、各系統の解釈、系統名の作成となる。

4-4-5-1 各系統の解釈

コミュニティの全体構造を図 73 に示す。左の系統 1 から順に解釈を行っていく。

系統 ID	仮系統名	支持 ユーザ数	男性 ユーザ割合	女性 ユーザ割合	不明 ユーザ割合
1	香り	8,362 人	17.91%	63.23%	18.86%
2	本	16,560 人	25.46%	55.92%	18.62%
3	芸能	21,799 人	19.23%	62.84%	17.93%
4	アニメ	17,149 人	31.07%	49.94%	18.99%
5	女性向けライフスタイル	15,218 人	11.11%	71.08%	17.81%
6	自分らしさ	15,197 人	20.18%	61.73%	18.09%
7	音楽	13,955 人	24.59%	58.37%	17.03%
8	女性向けコンテンツ	12,639 人	20.93%	61.07%	18.00%

図 73 調整後のコミュニティの全体構造

系統 1：香り

レイヤー1の「金木犀の香りが好き」を筆頭に構成される系統1は、支持するユーザ8,362人のうち、男性が17.91%、女性が63.23%、不明が18.86%と、過半数以上が女性に支持されている。レイヤー2には「桜」があり、金木犀同様に季節の香りを大切にする姿が見て取れ、さらに日常生活では「ミスタードーナツ」で「紅茶・お茶」を楽しむというスタイルが窺える。レイヤー3に続いていくと、「秋が好き。夜風が好き」「雨のにおい」「君に見せたい景色がある」と季節や自然を大切にするという特徴と、「チョコミント」「ハーゲンダッツ」のように日常生活における彩りを大切にする姿が見受けられる。

このように、系統1は、季節の移り変わりやほっと一息つく瞬間を大切に日常を生きる系統だと言える。本系統を「大切にしたい季節と日常」と名付ける。

系統 2：本

レイヤー1の「浅野いにお」「3月のライオン」「伊坂幸太郎」「羽海野チカ。」を筆頭に構成される系統2は、支持するユーザ16,560人のうち、男性が22.46%、女性が55.92%、不明が18.62%と、系統1に比べると男性支持が増えている。レイヤー1の「羽海野チカ。」はハチクロ作者であり、「3月のライオン」は同作者の別作品であることから、ハチクロ以外にも、同作者の世界観を支持する姿が見て取れる。また、他には、同じ青春をテーマにしながらも若者特有の精神世界を丹念に描く「浅野いにお」や、ミステリという切り口からの「伊坂幸太郎」など、同年代がテーマとなっていて感情移入のしやすい作品群を好む傾向がある。

レイヤー2には、「ハチクロ山田に共感する人々」とハチクロに関する共感をより具体的に表明するコミュニティや、レイヤー1 浅野いにおの「ソラニン」、同じように文化系の男女に支持される「森見登美彦」やバイブルの「村上 春樹」と、やはりレイヤー1 同様に青春時代に誰もが一度は通る道となる作家群が連なる。続いてレイヤー3にも「青春スーツ☆ハチクロ」とハチクロへの飽くなき憧れと、「オノ・ナツメ(basso)」「有川浩」と若者支持の高いサブカル作家がある。ユーザらは、「やっぱり本が好き」「【読書ノ会】」と、自身が上述した作品群を読書をしていることを胸を張っているという点も重要である。

このように見ると、系統2は、「羽海野チカ。」とその世界観を楽しむ姿と、同じように他の作品も楽しむ姿が混在して形成されている。本系統を「青春を物語に重ねる」と名付ける。

系統 3：芸能

レイヤー1の「Cocco」「蜷川実花」といったアーティスト、「ラーメンズ」「阿部サダヲ」の舞台を中心に活躍するサブカル系芸能人、「蒼井優」「加瀬亮」「麻生久美子」の独特の世界観を持つ若手俳優を筆頭として、系統3は構成される。男女比は、支持するユーザ21,799人のうち、男性が19.23%、女性が62.84%、不明が17.93%となっている。レイヤー1の芸能人は、単にテレビドラマや映画のスクリーン上で活躍するにとどまらず、独自のサブカル活動を行い、発信していくという共通項を持っている。また、活躍の場に関しても、広く一般受けするようなメジャーな作品というよりも、ある一定のサブカル的素養があるからこそ理解の進むような作品群となっている。

レイヤー2には「TRICK! TRICK! トリック!」「邦画コミュ」「ミニシアター系映画」と邦画でもひと味違うものを求める姿や、「CHARA」「深津絵里」「永作博美」「安藤裕子」と、空気感のある女優・アーティストを支持するスタイルが見て取れる。また、「堺雅人に惚れました。」「松山ケンイチ」「オダギリジョー」といった俳優を好み、彼らのことを「塩顔がたまらなく好きだ」「汚いのには清潔感ある男」「くしゃっとした笑顔」と、普通の好きとは違うアピールも欠かさない。

レイヤー3には「戸田恵梨香」「YOU」「Superfly」「新垣結衣」と、憧れの対象としての女優群があり、加えて「*黒髪だいすき*」と、彼女達の魅力を分解し自身に還元しようと試みている。次いで「今日も独りで映画館♪」「『SPEC』」とサブカルとの距離は常に近く保っている。また、「岡田将生」「瑛太 Fan」と、レイヤー2の俳優の若手互換に位置する俳優も情報収集を怠っていない。

このように、系統3は、女性ユーザにとって自身の憧れの対象として重ね合わせるアーティスト・女優群と、その女優群と物語の中で関係を構築する俳優群から成る。本系統を「憧れの女性と物語の中の恋人」と名付ける。

系統4：アニメ

レイヤー1に「東のエデン」「よつばと！」等アニメ作品を筆頭に形成される系統4は、支持するユーザ17,149人のうち、男性が31.07%、女性が49.94%、不明が18.99%と、全体の中で最も男性支持が高いことが特徴である。「東のエデン」はキャラクターデザインに羽海野チカが関わっていることもあり、作者への支持が強いことがよくわかる。また、他のコミュニティを見ていくと、「よつばと！」「荒川アンダーザブリッジ」「もやしもん」など、それぞれテーマは違えどゆるい日常をベースにしたまんが・アニメがあり、次いで「時をかける少女【細田版】」「大きく振りかぶって」と世界観に青春を設定しているもの、「交響詩篇エウレカセブン」「ハガレン(鋼の錬金術師)」とロボットやSFおよびファンタジーのいわゆる元祖オタクアニメの王道的なテーマを扱うものなど、アニメに関しても様々な方向性が見て取れる。

レイヤー2では、ゆるい日常と青春の「フルーツバスケット」や「あの日見た花の名前を僕達あの花」、「サマーウォーズ」があり、さらに、エウレカセブンやハガレンとの強いつながりを持つ「コードギアス 反逆のルルーシュ」「エヴァンゲリオン新劇場版」が連なる。また、アニメソングの作曲・プロデューサー「菅野よう子」や、声優の「坂本真綾」、オタク趣味で有名かつ自身も歌手等で幅広く活躍する「中川翔子」と、単に特定の作品内にとどまらず、周辺事象も含めてアニメを楽しむスタイルが見て取れる。

レイヤー3では、日常系に関してはラブコメディの「とらドラ! アニメ版」やシチュエーションコメディの「ウサビッチ」、日常系から枝分かれして伝奇色の強い「夏目友人帳」がある。ゆるい日常系「よつばと！」からの流れでは、レイヤー3では「すごいよ! うすた先生」に強い連関がある。エウレカセブンやコードギアスからのつながりとして「COWBOY BEBOP」「天元突破 グレンラガン」があり、さらに世界系が強まると「魔法少女まどか☆マギカ」へと続く。レイヤー3でも作品周辺事象として「アニソン アニメソング」やオープニングテーマを担当した「奥華子」も忘れない。

このように見てくると、系統4は特定のジャンルにとどまらない様々なアニメ作品と、その周辺事象から成る。ライトオタクという言葉が登場して久しいように、これまでに挙げられた作品はアニメやマンガを語る上では欠かせない、教養としての位置づけとも言える。したがって、本系統を「教養的アニメ作品群とその周辺事象」と名付ける。

系統5：女性向けライフスタイル

レイヤー1に「LOWRYS FARM」「魔女の宅急便」「*森ガール*」を据え構成される系統5は、支持するユーザ数15,218人のうち、男性11.11%、女性71.08%、不明17.81%と、女性に特化した傾向を持つ。「LOWRYS FARM」は「*森ガール*」御用達のファッションブランドであり、ハチクロの世界観をそのまま自身のファッションスタイルにも映し出したいという願望がここにあると言えるだろう。

レイヤー2には、レイヤー1の森ガールファッションを補強するコンセプトである「ゆるかわ」や、ファッションブランド「ANNA SUI」、ファッションに合わせた「ボブ☆」と、ユーザらは一貫した外見スタイルをSNS上にも構築している。さらに、「北欧デザイン」が好きで、そのお手本として映画「かもめ食堂」があり、デートに行くならば「かわいい建築」巡りか「プラネタリウム」と、ファッション以外のスタイルに関しても確固たるスタイルへの意識が存在していることがよくわかる。

レイヤー3にもファッションブランドの流れは続いており、「JEANASiS」「sm2(サマンサモスモス)」「BEAMS」「earch music&ecology」「Cath Kidston」と、自身のスタイルを実現するブランドをよく理解している。また、既製品によるスタイル確立にとどまらず、「ひとりのじかん」には「マスキングテープ好き。」でカスタマイズも行っている。手本となる映画は「耳をすませばみたいな恋がしたい」「アメリ」とゆるふわな女の子が登場するストーリー。さらに生活に関して言えば

「marimekko.」や「Franc franc(フランフラン)」のインテリアも大切にする。それもこれも「*カワイイもの好き部。*」で「お菓子の家に住みたい」という夢のライフスタイルの実現のためのパーツと言えるだろう。

このように、系統5は森ガールファッションを支持する女性の、ファッションブランドや手本となる映画等から形成されている。本系統を「ゆるかわスタイルの手本と実践」と名付ける。

系統6：自分らしさ

レイヤー1に「メガネ男子愛好会」持つ系統6は、支持するユーザ数15,775人のうち、男性20.18%、女性61.73%、不明18.09%から構成される。これまで見てきた系統と大きく違う点として、コンテンツや事象名ではなく、自身の内面を表現するコミュニティが多いことが挙げられる。

レイヤー2には、レイヤー1との連関で「萌え死に。」「猫になりたい」があり、女性ユーザにとっての憧れの男性像と、自身の関係の理想像が見て取れる。続いて「コミュ一覧が大変なことに。」「余計なこと考えすぎ!」と、理想と現実のギャップに空回る自身を理解してほしいという願望が想像できる。

レイヤー3には、「メガネ男子」に「萌え死に。」からのつながりとして「メガネフェチ」「手フェチ」「メガネっ子」がある。メガネ男子の「さりげない優しさがスキ!」だから、そんな「小さな偉せ」を守りたいという恋愛観がここにある。しかし、現実はそううまくはいかず、「至らない点が多い」「荷物、多っ!!」「寝ても一た」「忘れっぽい人集まれ!」と、自分のダメな側面もわかっている。しかし落ち込むことなく、それなりに楽観的に、全てを自分らしさと捉えている点が特徴的である。

このように、系統6は、憧れの恋愛像とちょっとダメな自分という二つの面から形成されている。本系統を「メガネ男子とヘタレな私」と名付ける。

系統7：音楽

レイヤー1に「Spitz(スピッツ)」「バンド 相対性理論」「フジファブリック」と音楽系のコミュニティを据え構成される系統7は、支持するユーザ数13,955人のうち、男性24.59%、女性58.37%、不明17.03%から構成される。比較的男性支持の高い系統である。

レイヤー2には、スピッツと同じ年代に活躍した「★☆JUDY AND MARY☆★」、フジファブリックのように「ROCK IN JAPAN FESTIVAL」等のフェスで注目を浴びる「スキマスイッチ」「レミオロメン」「凜として時雨」、相対性理論からはサブカルつながりで「椎名林檎・東京事変」がある。

レイヤー3では、フェス系のつながりで「ライブが好きでナニガ悪い!!」「ライブがあるので、休みます。」「☆COUNTDOWN JAPAN☆」と、音楽への情熱が見てとれる。また、自身の感情の代弁者としての「藤原基央」や「SEKAI NO OWARI」がここにある。また、レイヤー1および2からの流れとは別に、独自の路線を貫く「斉藤和義」や、世代を超えた支持を受ける「the pillows」も支持されている。

このように、系統7はフェス系文化と、その時代その時代を作ってきた誰もが知っていて当たり前の教養的アーティストとで構成されている。本系統を「教養的Jポップとフェスで楽しむ音楽」と名付ける。

系統8：女性向けコンテンツ

レイヤー1に「NANA」「のだめカンタービレ」がある系統8は、支持するユーザ数12,639人のうち、男性20.93%、女性61.07%、不明18.00%から構成されている。レイヤー1のコミュニティは、ハチクロ同様、恋愛だけではなく青春ストーリーが支持される作品である。

レイヤー2には、「☆僕等がいた☆」や、NANAと同作家の「天使なんかじゃない(矢沢あい)」と、やはりここでも感情移入をして読むことのできる青春定番ものがつながっている。また、「宇多田ヒカル Utada」は音楽系コミュニティで構成される系統7からは離れ本系統にあることから、ユーザら

にとって単なる音楽のひとつではなく、歌詞に自身を重ね別格の扱いをするアーティストとして位置づけられると考えられる。

レイヤー3には、「のだめカンタービレ【実写版】」のキャスト「玉木宏」がいる他、「ニンテンドー3DS & DS」といったライトなゲーム機器とそのソフト「逆転裁判所」「動物の森遊び方研究」が連なり、マンガや映画以外にどのようなコンテンツで楽しんでいるかが浮かび上がってくる。

このように、系統8は、自身を重ねる青春ストーリーマンガと、そこから派生した映画やライト層向けゲーム機器から構成されている。本系統を「休日は青春ストーリーマンガとどうぶつの森」と名付ける。

4-4-5-2 解釈に基づいて命名した各系統

以上、各系統をコミュニティ名に基づき解釈してきた。解釈によって命名された各系統一覧を表7に示す。また、系統名を調整し、左から男性支持が高い順に並べ替えた構造図を図74に示す。

表 7 仮系統名と本系統名

系統ID	仮系統名	本系統名
1	香り	大切にしたい季節と日常
2	本	青春を物語に重ねる
3	芸能	憧れの女性と物語の中の恋人
4	アニメ	教養的アニメ作品群とその周辺事象
5	女性向けライフスタイル	ゆるかわスタイルの手本と実践
6	自分らしさ	メガネ男子とヘタレな私
7	音楽	教養的Jポップとフェスで楽しむ音楽
8	女性向けコンテンツ	休日は青春ストーリーマンガとどうぶつの森

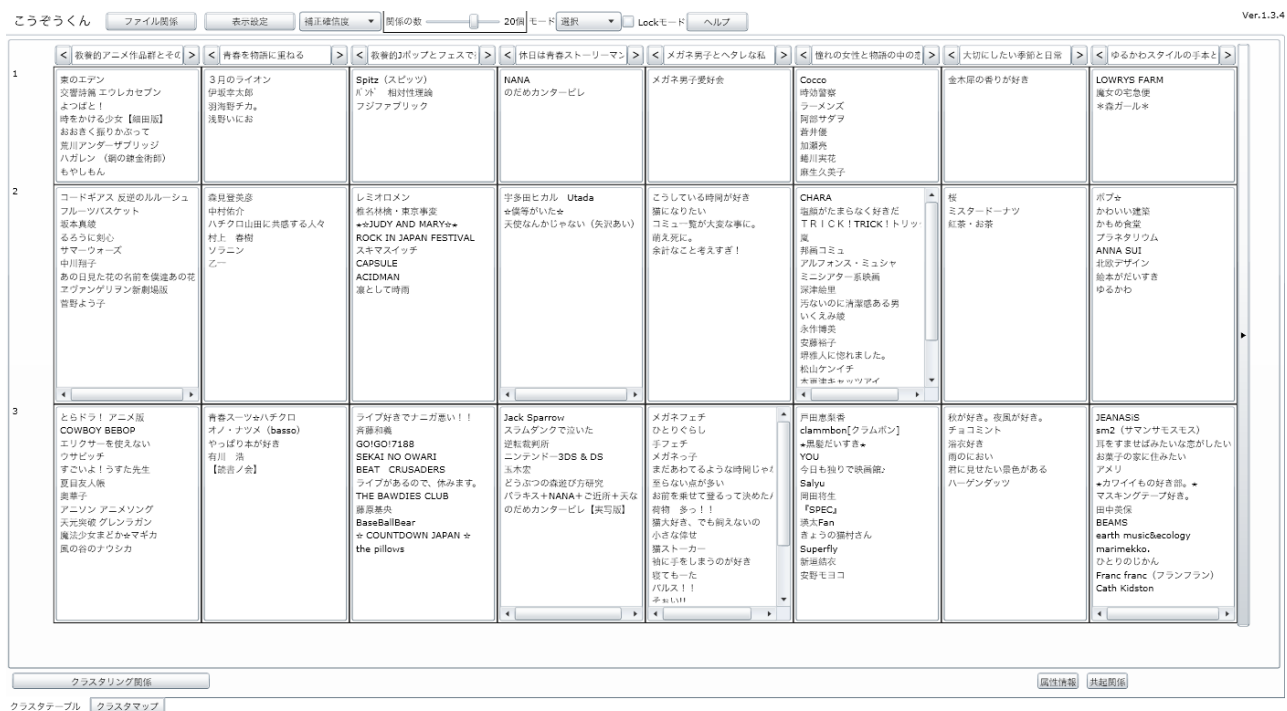


図 74 システムに名前を書き込んだ全体構造図

4-4-6 概念化

これまででは、コミュニティを探索的に構造化し、生成された各システムのコミュニティを解釈することから、ハチクロを支持するユーザらのスタイルを分析してきた。続いて、各システム間の連関を見ることで、ハチクロを支持するスタイルとはどのような概念によって構成されているのか、一段階抽象化した視点から眺める。この分析プロセスを概念化とする。

システム間の連関は、構造図の縦枠を基準として見ていく。したがって本分析では、表 7 に示した 8 つのシステムの連関を見ていくこととなる。

4-4-6-1 連関の調整

まず、デフォルトでの出力である Lift 値 1.15 以上のつながりを図 75 に示す。2 システムが全体とのつながりから分断されてしまっており、解釈が難しい状況にある。

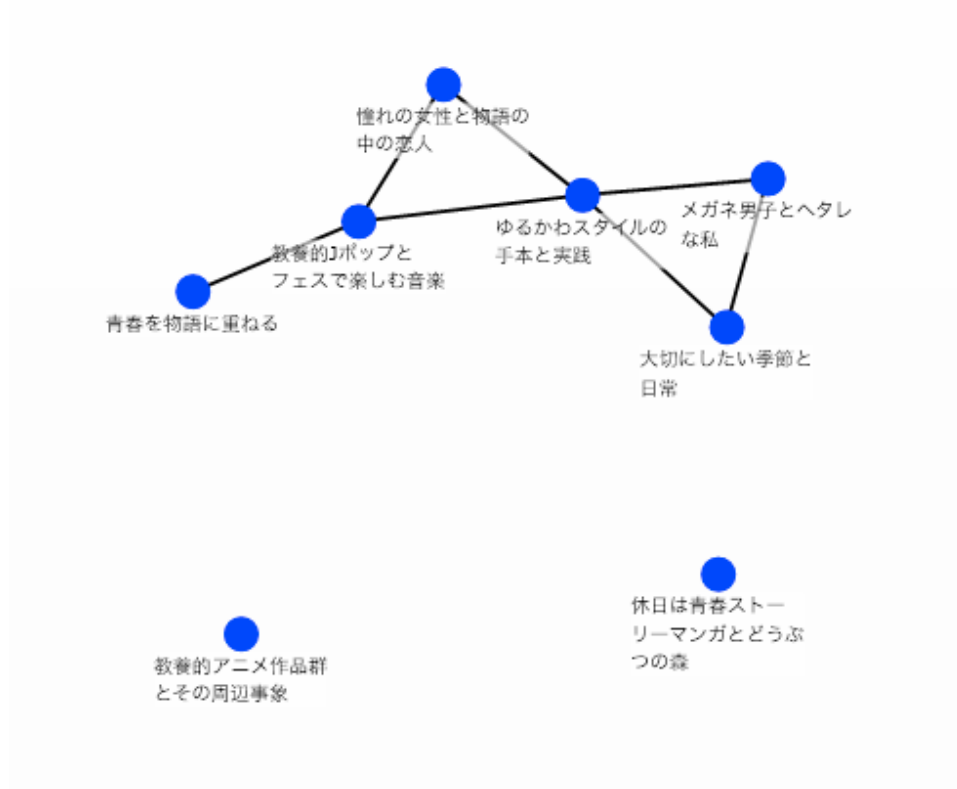


図 75 デフォルト (Lift 値 1.15 以上) でのクラスタ間の関係図

そこで、リフト値を段階的に下げていく。リフト値を 1.13 まで下げた連関を図 76 に示す。

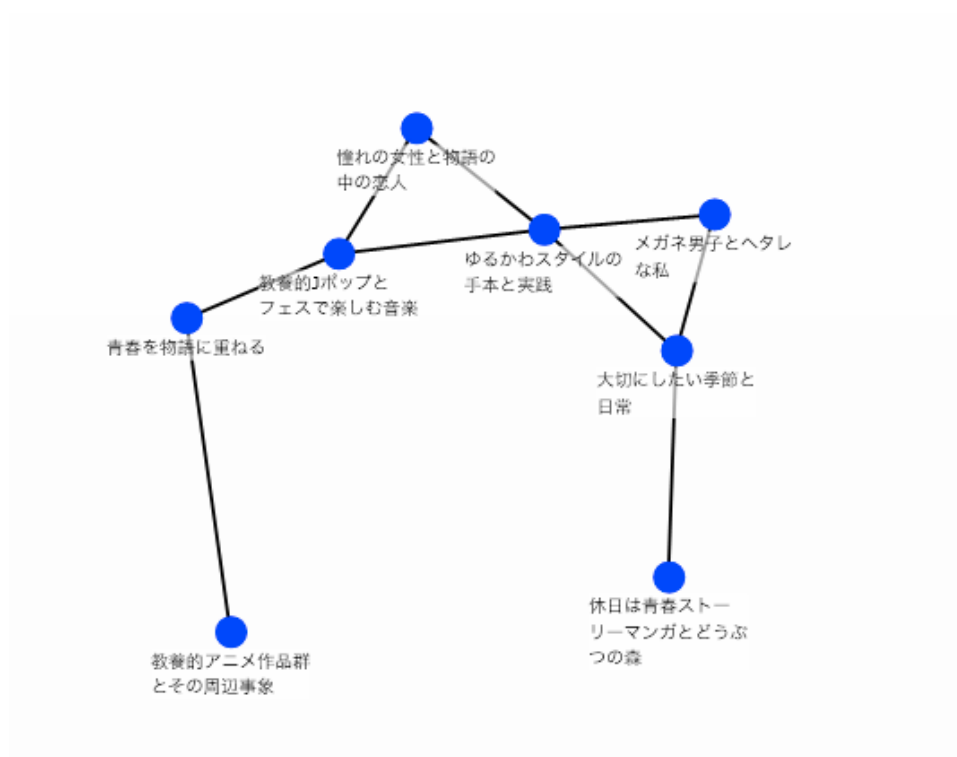


図 76 Lift 値 1.13 以上でのクラスタ間関係図

上記の図 76 の連関を見ると、左側から「教養的アニメ作品群とその周辺事象」と「青春を物語に重ねる」と、アニメと書籍という文化的コンテンツがつながり合い、さらに右側を見ると「休日は青春ストーリーマンガとどうぶつの森」と「大切にしたい季節と日常」の日常系がつながり合っており、解釈的にも妥当であると考えられるため、この調整を採用し、以下解釈していく。

4-4-6-2 連関の解釈

以降では、各系統の連関の解釈を行っていく。系統名は、表 8 のように短縮表記する。

表 8 本系統名と短縮系統名

系統 ID	本系統名	短縮系統名
1	大切にしたい季節と日常	季節日常系
2	青春を物語に重ねる	青春物語重ね系
3	憧れの女性と物語の中の恋人	憧れ女性と恋人系
4	教養的アニメ作品群とその周辺事象	教養アニメ系
5	ゆるかわスタイルの手本と実践	ゆるかわスタイル系
6	メガネ男子とヘタレな私	メガネ男子ヘタレ私系
7	教養的 J ポップとフェスで楽しむ音楽	教養音楽フェス系
8	休日は青春ストーリーマンガとどうぶつの森	休日暇つぶし系

まず、各系統を男女支持率から見た。すると、図 77 のように、図の左側に男性支持の高い系統、図の右側に女性支持の高い系統が位置した。

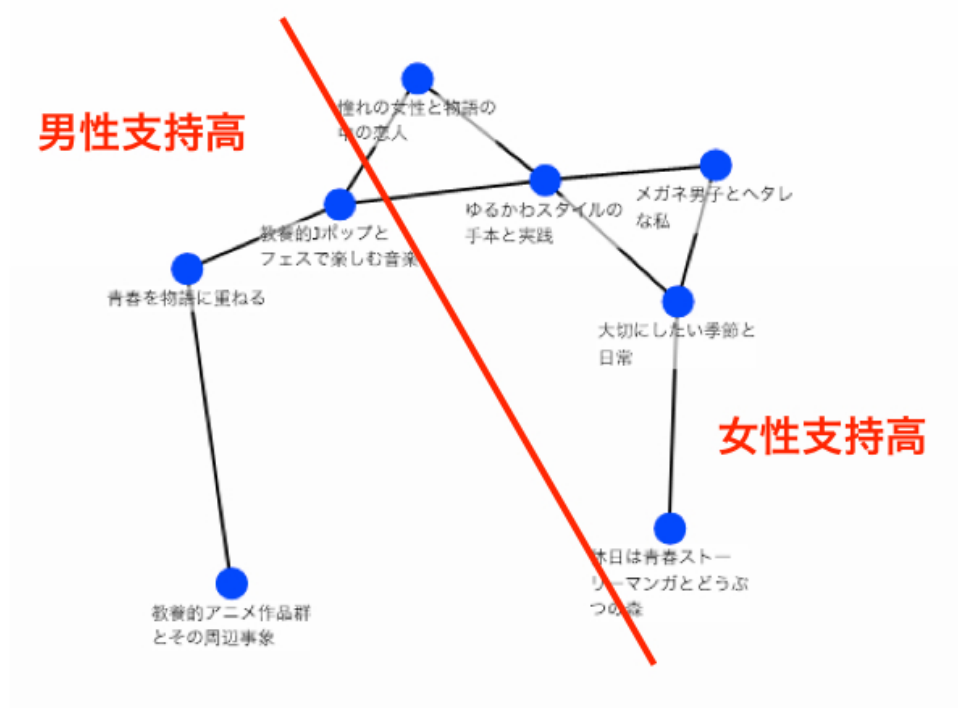


図 77 クラスタ間関係図の解釈 1

男性支持の高い系統に着目すると、図 78 の通り、教養アニメ系から青春物語重ね系へ、そしてそこから教養音楽フェス系へとつながっている。教養アニメ系は、各種アニメ作品名を中心に形成されて

おり、コンテンツへ没入して楽しむ姿が想像できる。そこからつながる青春物語重ね系は、作品名よりも作家名や読書への共感等を中心に成り、青春時代だからこそ誰もが一度は手に取る作家の作品に対して自身の青春を重ね合わせる、コンテンツへの自身の重ね合わせの姿が浮かび上がってくる。さらに連なる教養音楽フェス系は、単に音楽を個人で楽しむということにとどまらず、イベント等で共有し楽しむという、コンテンツとイベントとの密な関係を構築する姿が見てとれる。このように男性支持の高い系統は、コンテンツに没入する系統から、コンテンツに自身を重ね、さらにコンテンツとイベントという、物語世界と現実世界のバランスの濃淡として見ることができる。

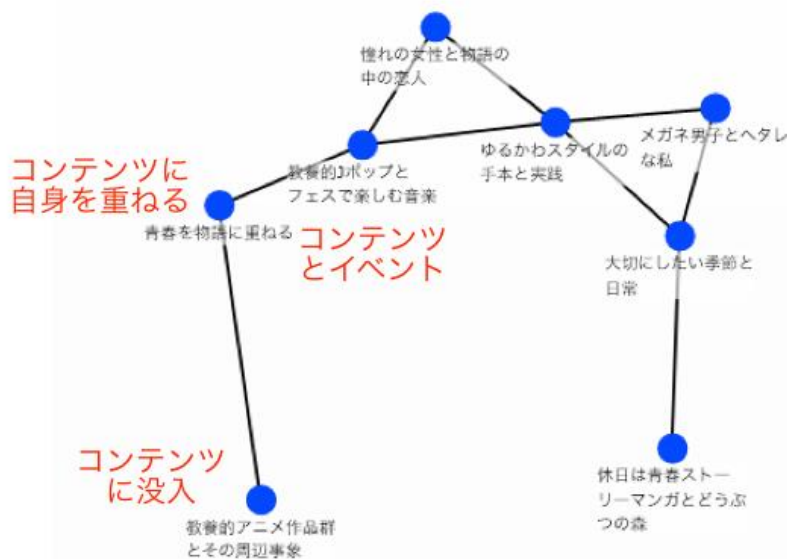


図 78 クラスタ間関係図の解釈 2

さらに、男性支持の高い系統群を、生活の中における位置づけとして再考すると、図 79 のように解釈できる。教養アニメ系は、基本的にはユーザーらは自身のテレビ、あるいは PC やモバイル等の画面で一人で閲覧するだろう。それは、柳田國男の言葉を借りれば、ハレとケの「ケ」、普段の生活に埋め込まれていると想像できる。次いで青春物語重ね系に関しても、基本的には教養アニメ系同様にケに埋め込まれた一部であろう。しかし、自身を重ね合わせるその先は、輝かしくスリルと楽しさにあふれる物語の世界である。コンテンツへの自身の重ね合わせは、その意味において、ハレとケの融合として位置づけられるだろう。最後に教養音楽フェス系に関しては、年に数回のイベントであり、非日常のハレと読み解くことができる。

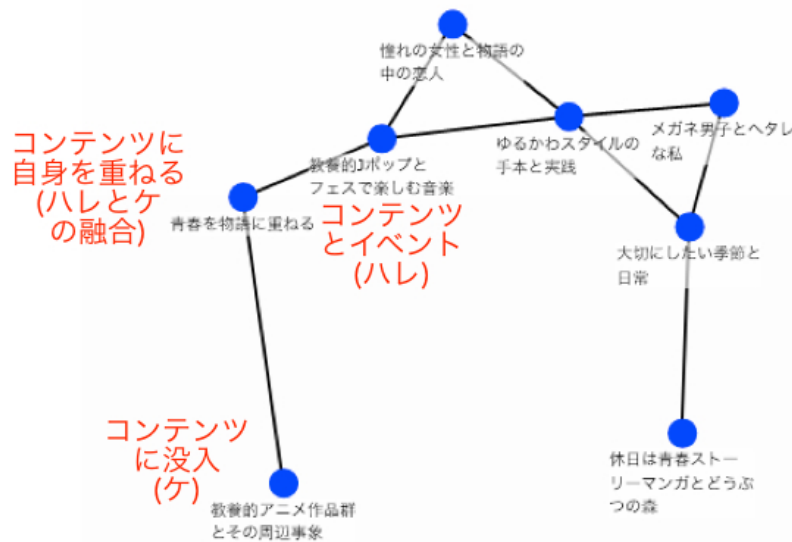


図 79 クラスタ間関係図の解釈 3

女性支持の高い系統群に着目すると、図 80 のように、休日暇つぶし系から季節日常系へ、そこからメガネ男子ヘタレ私系とゆるかわスタイル系へと分岐し、ゆるかわスタイル系からは更に憧れ女性と恋人系と連なっている。休日暇つぶし系は、矢沢あいをはじめとする作品群と任天堂ゲーム機器を中心に構成されており、何度も同じ作品を読み返したり、長く楽しめる育成ゲームで休日の暇をつぶす姿が想像される。そこからつながる季節日常系は、季節の移り変わりや毎日少しだけほっとできる時間など、地味だけれども日々を大切に慈しんでいるだろうことが推測される。メガネ男子ヘタレ私系は、日々の生活ではドジなところもあるネガティブなところもある自分を否定しないが、しかしそんな私も素敵な恋愛をしたい、という膨らむ妄想が想像できる。そこからつながるゆるかわスタイル系は、映画内外で活躍する女優やモデルを具体的な理想像として掲げ、彼女達に少しでも近づけるよう生活空間まで手入れを怠らない、女優やモデルといった手本に基づいた生活スタイルの実践が見てとれる。最後に憧れ女性と恋人系は、映画等の作品内に理想の恋愛を見出す姿として受け取ることができる。このように女性支持の高い系統は、日々をコンテンツで暇つぶししたり、一方で大切に慈しみ、そんな日常に恋愛が欲しいと願ってやまず、コンテンツに憧れの恋愛を見るという、物語をモデルケースとし、モデルと現実とのバランスの濃淡として見ることができる。

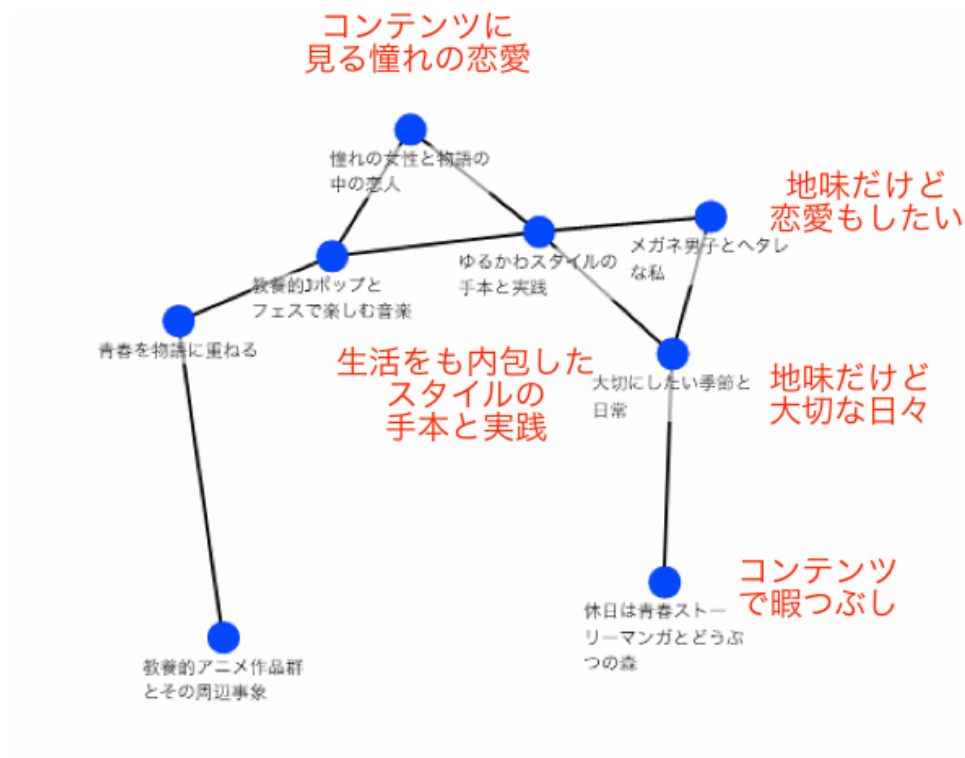


図 80 クラスタ間関係図の解釈 4

さらに、女性支持の高い系統群を、前述の男性支持同様に、生活の中における位置づけとして再考すると、図 81 のように解釈することができる。休日暇つぶし系は、退屈な日常の暇つぶしという意味でケの存在として見る事ができる。続いて季節日常系も同様であり、基本的には地味な日常だからこそ大切に目を向けようとしており、それは日々終わることのないケの世界である。さらにメガネ男子ヘタレ私系も同様に、日常を生きる私はヘタレでパツとしない、代わり映えのないケの世界に生きているからこそ恋愛を夢見ているのだ。しかし、そこからつながるゆるかわスタイル系に関しては、映画や芸能人の生活など普通の人にとっては手の届かないハレと、少しでもそこに近づきたい努力の日常というケの融合として見る事ができる。さらに、憧れ女性と恋人系は、コンテンツ上に構築される憧れの恋愛、ハレそのものと言えるだろう。

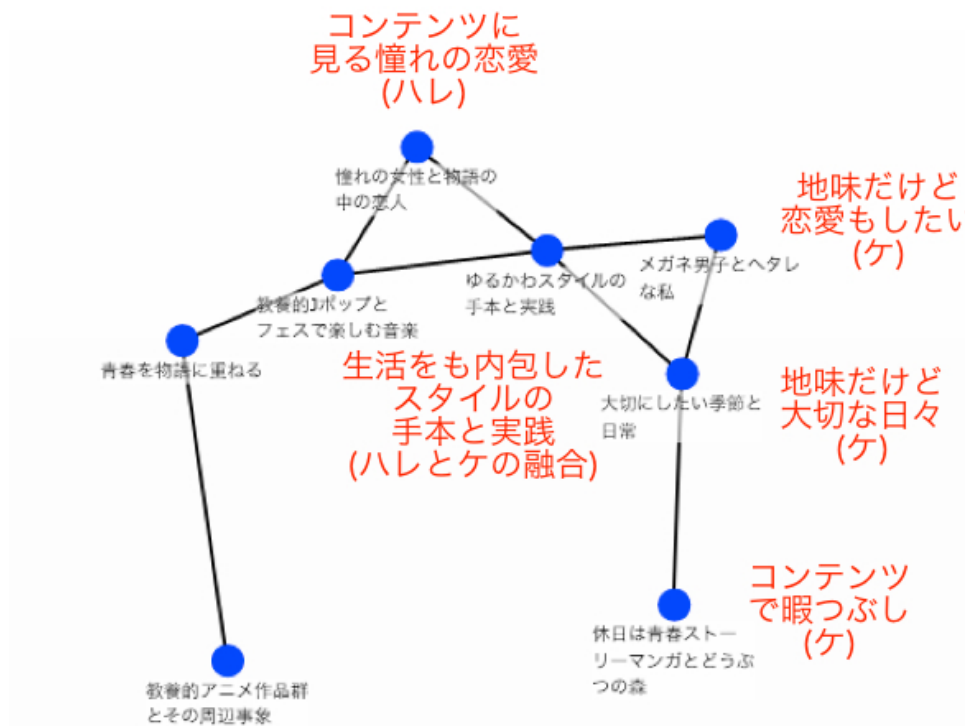


図 81 クラスタ間関係図の解釈 5

4-4-7 結論：普通の若者にとっての「カルチャー」「スタイル」

これまで、ハチクロを支持するユーザらのコミュニティ構造を分析することによって、普通の若者がどのようなライフスタイルを構築しているのかを明らかにしてきた。ハチクロを支持するユーザは、全体のうち男性は 24.26%、女性は 57.87%、不明が 17.87%と女性が過半数を超えており、その意味においてはハチクロを支持する、あるいは SNS 上で支持すると表明する若者は、相対的に女性が多いと推測できる。そのため、本分析で明らかにするカルチャーとスタイルは、女性に特化したものとなることを前提としたい。

全体を 8 系統に分類し、コミュニティ名からの詳細な解釈を通じて系統名を付与し、その後、系統間の連関を見ることで、ハチクロを支持するユーザらのコミュニティ構造の全体像を見てきた。最後に、コミュニティ構造の全体像を再考することで、普通の若者にとっての「カルチャー」と「スタイル」とは何かについて、考察する。

普通の若者にとっての「カルチャー」

コミュニティ構造の 8 系統、および系統間の連関を見る際に特化する性別、ハレとケの解釈を表 9 にまとめる。男性特化の 3 系統に関しては、青春物語重ね系は小説、教養アニメ系はアニメ、教養音楽フェス系は音楽と、小説、アニメ、音楽がカルチャーの 3 本柱になっていることがよくわかる。そして、アニメは日常的に没入し楽しむケの存在として、音楽はフェスという非日常のハレの存在として位置づけられる。それらハレとケをつなぐのは、小説という青春物語への自身の重ね合わせであると言える。

女性特化の系統に関しては、憧れ女性と恋人系は映画・ドラマ、ゆるかわスタイル系は生活雑貨・ファッション、休日暇つぶし系はマンガ・ゲームと、映画・ドラマ、生活雑貨・ファッション、マンガ・ゲームがカルチャーを構成する中心となっていると言える。映画・ドラマはコンテンツに憧れの恋愛を見いだすハレの対象として、マンガ・ゲームは日常の退屈を埋める手段というケの存在として

位置づけられる。そして、ハレとケをつなぐ存在として、生活雑貨・ファッションという、生活をも内包するモデルや女優らによる手本と自らの実践がある。

表 9 系統名とハレとケの解釈

系統ID	本系統名	短縮系統名	特化性別	ハレとケの解釈
1	大切にしたい季節と日常	季節日常系	女性	地味だけど大切な毎日(ケ)
2	青春を物語に重ねる	青春物語重ね系	男性	コンテンツに自身を重ねる(ハレとケの融合)
3	憧れの女性と物語の中の恋人	憧れ女性と恋人系	女性	コンテンツに見る憧れの恋愛(ハレ)
4	教養的アニメ作品群とその周辺事象	教養アニメ系	男性	コンテンツに没入(ケ)
5	ゆるかわスタイルの手本と実践	ゆるかわスタイル系	女性	生活をも内包したスタイルの手本と実践(ハレとケの融合)
6	メガネ男子とヘタレな私	メガネ男子ヘタレ私系	女性	地味だけど恋愛もしたい(ケ)
7	教養的Jポップとフェスで楽しむ音楽	教養音楽フェス系	男性	コンテンツとイベント(ハレ)
8	休日は青春ストーリーマンガとどうぶつの森	休日暇つぶし系	女性	コンテンツで暇つぶし(ケ)

以上、各系統を男女およびハレとケの解釈軸上にプロットした図を図 70 に示す。 図 82 の赤破線内が、若者のカルチャーとして位置づけられる。

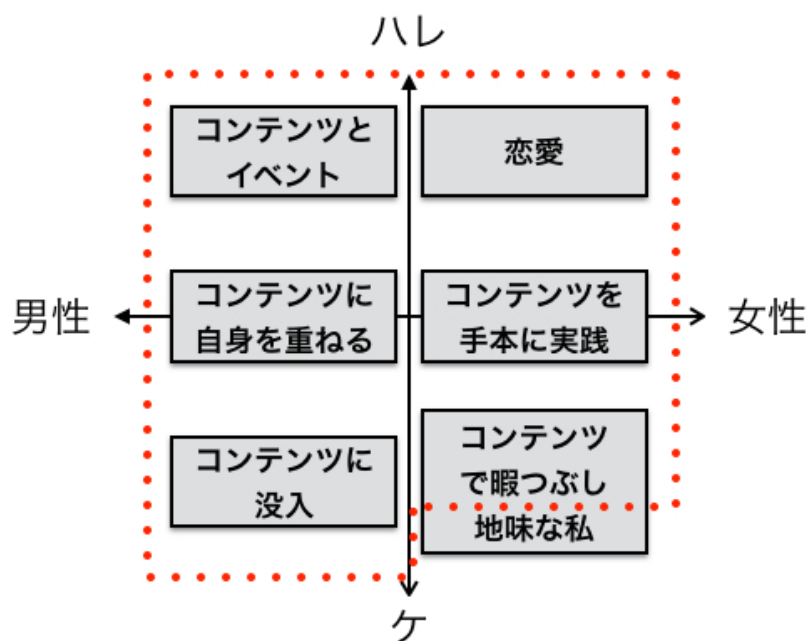


図 82 各系統の位置づけ

では、各系統の中で、カルチャーとして位置づけられない、女性特化のメガネ男子ヘタレ私系はどうか。本系統は、地味だけど恋愛もしたいという、ダメな自己も認める自己主張のコミュニティ群から成っている。この系統こそ、SNS の登場によって、スタイルの一部となった存在である。かつてスタイルとは、ファッションなどの見た目によって規定されていた。それが徐々にカルチャー面を侵食し、さらに今では、心理的な、対面しても見ることが難しい領域にまで広がりを見せているのである。我々は SNS を手に入れるまでは、短くない時間をかけて対人関係を構築し、その中で相手の内面的な特性を理解していくのが通常であった。しかし今や、対面することなく、各自の主張によって内面的な特性をも理解することが可能なのである。その自己主張も含め、各々のスタイルとして我々は違和感無く受け入れているのだ。したがって、図 83 の赤破線内に示すように、ハチクロを支持する若者のスタイルは、コミュニティ構造の全てによって構成されているとすることができるだろう。

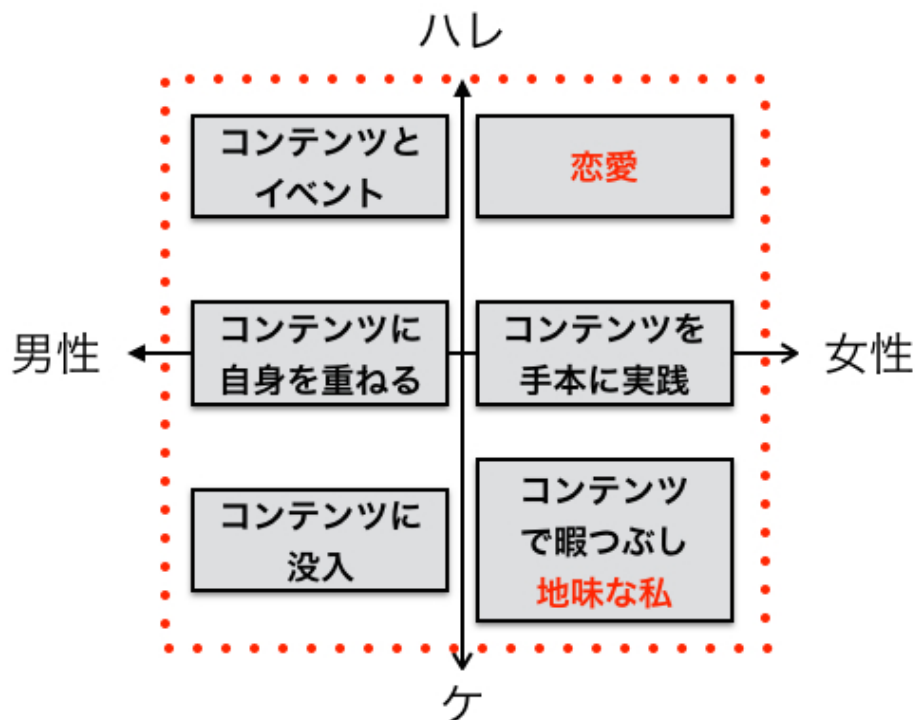


図 83 各系統の位置づけ

4-4-8 実例からみた「こうぞうくん」の価値探索機能

普通の若者のカルチャースタイルについて分析をする過程で、「こうぞうくん」の分析可視化ツールとしての意義を説明してきた。「こうぞうくん」は、ソーシャルメディアという従来の社会調査データとは異なる特性を持つ対象に対して、従来の分析ツールではなし得なかったデータとの対話と解釈を可能とした。ここでは、具体的な成果をまとめて、このツールがどのような意味で価値探索ツールとしての機能を保有しているかを、より一般的な方向で説明する。

「こうぞうくん」の価値探索機能は、表 10 に示す 13 機能である。以下、分析の段階別に機能を説明していく。

表 10 「こうぞうくん」の機能一覧

分析の段階	分析の手法	ツール上の機能
使用の全体	-	セーブ／ロード
構造化	クラスタリング	基準変更
構造化	クラスタリング	統合
構造化	クラスタリング	分割
構造化	相関ルール	ソート
構造化	相関ルール	彩色
構造化	相関ルール	移動
構造化	相関ルール	Lock
構造化	相関ルール	削除
構造化	特化係数	彩色
概念化	クラスタリング	生成
概念化	相関ルール	ソート
概念化	相関ルール	結合

4-4-8-1 使用の全体

使用の全体に関する機能を図 84 に示し、以下、個別の機能について説明する。

使用の全体 ①セーブ／ロード

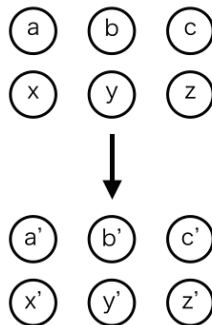


図 84 セーブロード機能の概念

①セーブ／ロード機能

セーブ／ロード機能は、ツールを使用する上で最も重要な機能の 1 つである。なぜなら、「こうぞうくん」ではデータをインプットした後、データとの対話を通じて、後述する緻密な調整によって分析結果を導きだしていくからである。分析者は、分析の段階ごとにデータをセーブし、試行錯誤することが求められる。細かな調整をそのままデータとしてセーブし、次の分析時にストレスなくロードできるように、入力データおよび画面上で可視化されている状態データの両者をまとめて `.clusterTable` の形式で保存している。

4-4-8-2 構造化

構造化に関する機能を、分析手法別に以下で説明する。

クラスタリング

クラスタリングに関する機能は、図 85 に示す通り 3 つある。以下、個別に説明する。

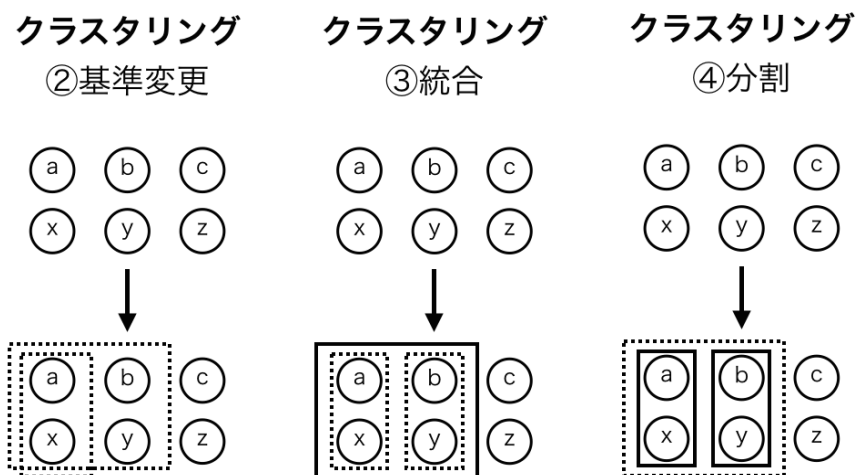


図 85 クラスタリング機能の概念

②基準変更

構造化の第 1 ステップは、データの全体を俯瞰することにある。本分析で対象とした mixi コミュニティは、参加者が各々の思いを持って参加しているという側面においては、極めてたような意味を持っている。言い換えれば、構造化の基準は無数にあり、それらはどれも、ある意味においては正解となる。このような特性を持つ分析対象に関しては、様々な基準で全体を俯瞰し、妥当な基準を採用することが求められる。

評価指標は系統内のつながりの度合い(縦)と全体の散らばりの度合い(横)で評価しており、基本的には大きいほどクラスタリングが適切であると言える。しかし、クラスタリングを開始する際のデータの配置はランダムであるため初期値に依存すること、および、本分析対象のような複数の解が存在するデータに関しては、必ずしも評価指標が最も大きい結果が妥当であるとは言い切れない。そのため、評価指標の上位の結果をいくつか参照し、分析の文脈に沿って妥当だと判断できるものを選択する必要がある。

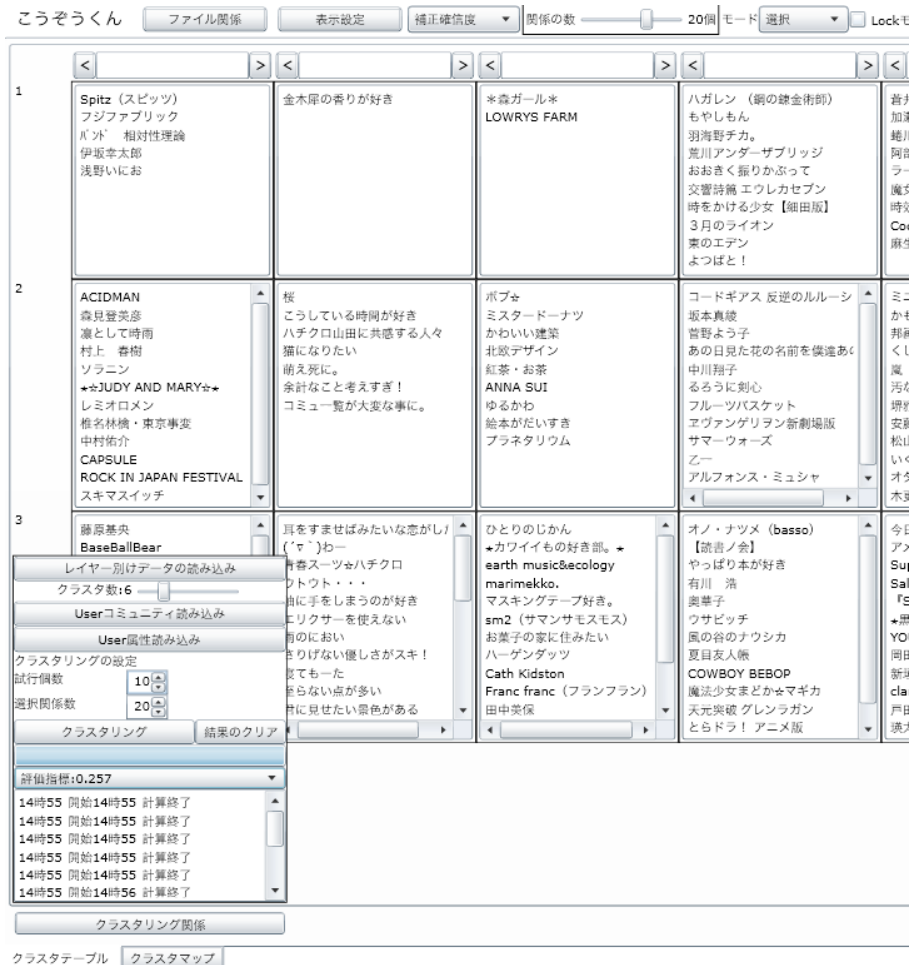


図 86 クラスタリング結果 (評価指標 0.257)

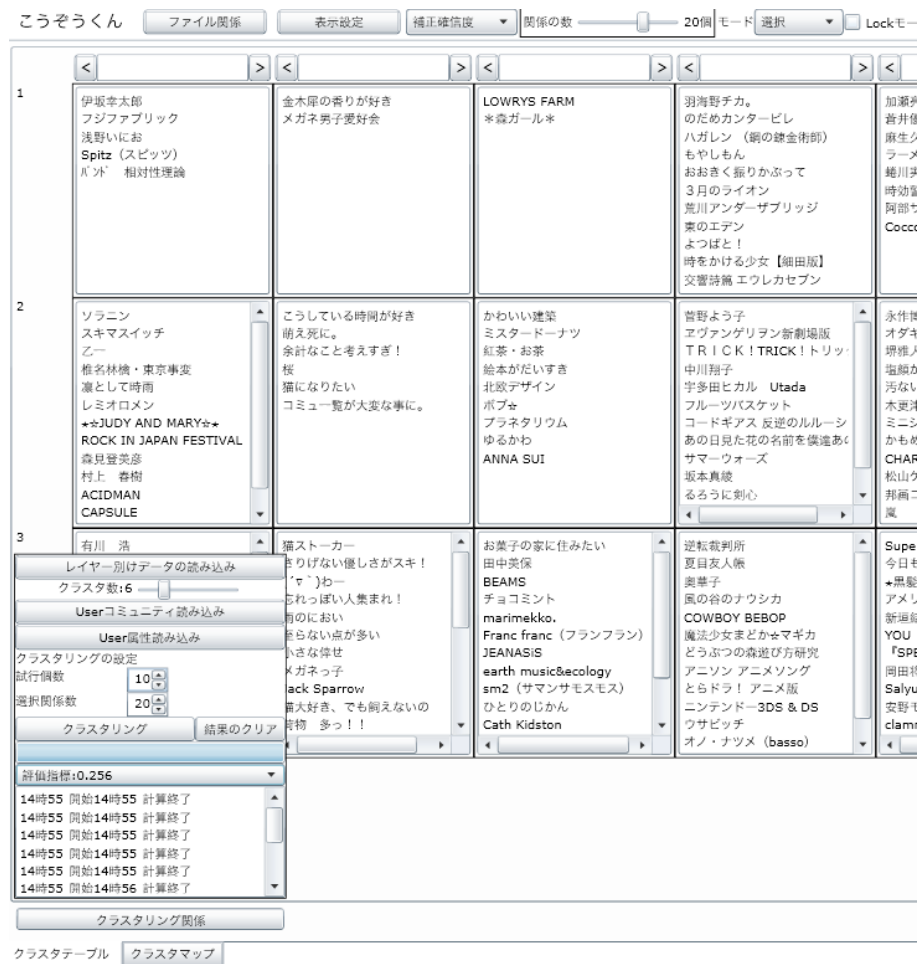


図 87 クラスタリング (評価指標 0.256)

例えば、図 86 は評価指標 0.257、図 87 は評価指標 0.256 のクラスタリング結果である。レイヤー 1 を見ると「金木犀の香りが好き」と「メガネ男子愛好会」が同クラスタのものとそうでないものがあるが、一方でどちらの評価指標でも同クラスタとなっているものもある。結果が妥当であるか見るために基準変更の機能は必須であるが、複数の結果を見比べることで個別のデータ間のつながりの強弱のイメージを掴むことも可能である。

③統合

構造化を行う際に、複数の系統を 1 つに統合することで、全体の見通しが良くなる場合もある。そのような場合には、どれか 1 つの系統に、移動したい系統のデータを全て移動し統合する。

④分割

統合機能とは反対に、各系統内が煩雑で解釈不能な場合には、系統数を増やすことで、各系統を分割する。図 88 は、本分析にて用いたデータで系統数を 6 とした構造図である。レイヤー 1 を見ると、左から 4 番目のクラスタは、データ数が多いだけでなく、「羽海野チカ。」を起点としてデータの共起を見ると、同クラスタ内で共起の高いものと低いものに二分される。このようにレイヤー 1 内でデータのとまとまりが悪いと、系統としての解釈も難しく、全体構造を適切に理解できるとは言いがたい。そのため、図 77 のように、系統数を増やし再クラスタリングを行うことで、系統を分割する。図 89 の分割後を見ると、先ほど起点とした「羽海野チカ。」は左から 3 番目のクラスタにあり、クラスタ内との共起が 6 系統の際よりも強くなり、まとまりが出ていることがわかる。

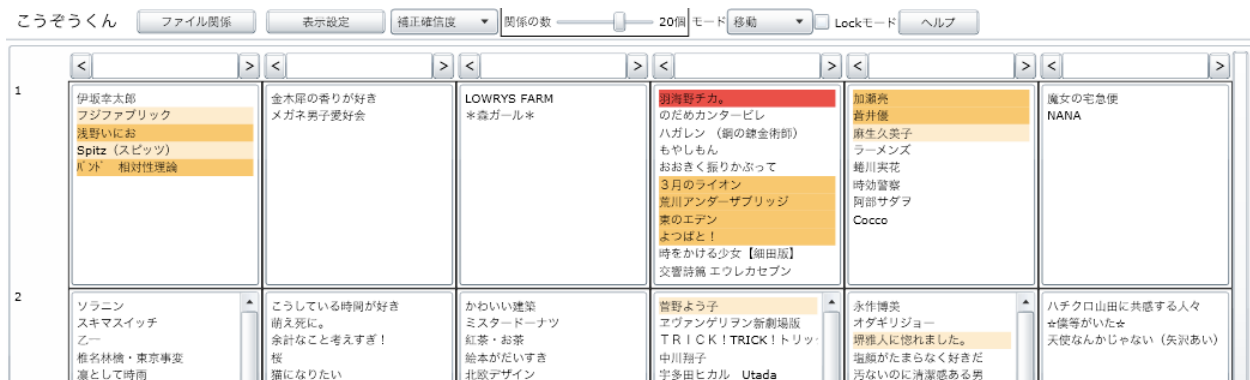


図 88 系統数6で作成したクラスタリング結果

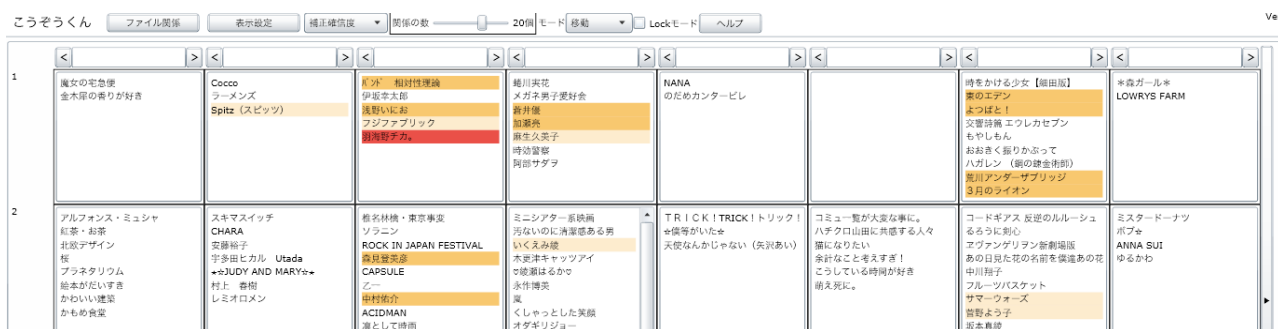


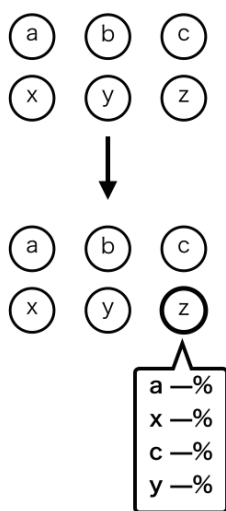
図 89 系統数8で作成したクラスタリング結果

関連ルール

クラスタリングに関係する機能は、図 90 および図 91 に示す通り 5 つある。以下、個別に説明する。

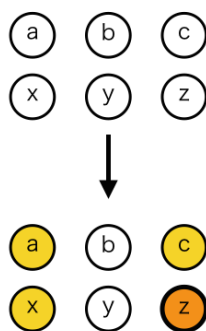
関連ルール

⑤ソート



関連ルール

⑥彩色



関連ルール

⑦移動

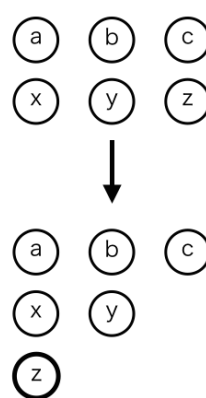


図 90 関連ルールの機能概念 1

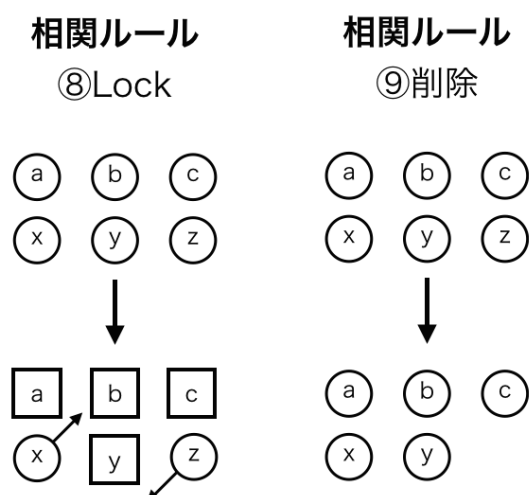


図 91 相関ルールの機能概念 2

⑤ソート

ソート機能は、各データを起点とした時、他データとの相関ルールに関する諸指標を見る際に活用できる。相関ルールに関する諸指標は、実際の共起の規模を示す確信度、およびデータの規模を考慮した補正確信度がある。データを移動・削除等する際は、単なる恣意性によって行うことは分析上望ましくない。そのような場合に、これらの指標を適宜ソートしながら調整を行っていく。

図 92 は、本分析データの例として、「羽海野チカ。」を起点とした際の確信度および補正確信度の一覧を示している。羽海野チカ作品である「3月のライオン」は確信度 0.479、すなわち「羽海野チカ。」コミュニティ参加ユーザの半分弱が「3月のライオン」にも参加していることがよくわかる。それ以降に続くコミュニティ群は、たしかに同じ文化圏のマンガ等が多く表示されているが、データを個別に見た際に「3月のライオン」との関連度合いが極めて高いことも、ソート機能によって理解することが可能である。デフォルトでは補正確信度の降順でソートされており、適宜変更しながら活用する。

度 ▼ 関係の数 20個 モード 移動 Lockモード ヘルプ Ver.1.3.4

	Rank	Name	確信度
バンド 相対性理論	1	3月のライオン	0.479
伊坂幸太郎	2	東のエデン	0.166
浅野いにお	3	浅野いにお	0.099
フジファブリック	4	よつばと！	0.089
羽海野チカ。	5	オノ・ナツメ (basso)	0.058
	6	菅井優	0.134
	7	中村佑介	0.068
椎名林檎・東京事変	8	森見登美彦	0.062
ソラニン	9	加瀬亮	0.108
ROCK IN JAPAN FESTIVAL	10	バンド 相対性理論	0.082
森見登美彦	11	荒川アンダーザブリッジ	0.069
CAPSULE	12	青春スーツ☆ハチクロ	0.055
乙一	13	いくえみ綾	0.055
中村佑介	14	サマーウォーズ	0.060
ACIDMAN	15	麻生久美子	0.082
凜として時雨	16	菅野よう子	0.062

図 92 相関ルールの詳細値の表示

⑥彩色

ソート機能にて、一覧表示された数値を色で可視化した機能が彩色である。個別のデータを起点とした際に関連度合いの高いデータを彩色する。選択しているデータは赤色で、関連度合いが極めて強いデータは濃いオレンジ色で、関連度合いがやや強いデータは薄いオレンジ色で彩色される。

データの彩色は、分析者の解釈に大きく貢献する。ぱっと見た際の視覚的効果は大きいからだ。図 93 は、本分析にて「よつばと！」を起点とした際の彩色例である。レイヤー1 の同クラス内との関連が極めて強く、続いて同系統内での関連も強いことがわかる。すなわち、「よつばと！」は、他の系統へ移動する必要の無い、同系統内でまとまりのあるデータであるということがわかる。もし、起点とするデータを変更し、他系統ばかりと彩色された場合は、データの移動や全体構造の再構成が必要だと考えられる。

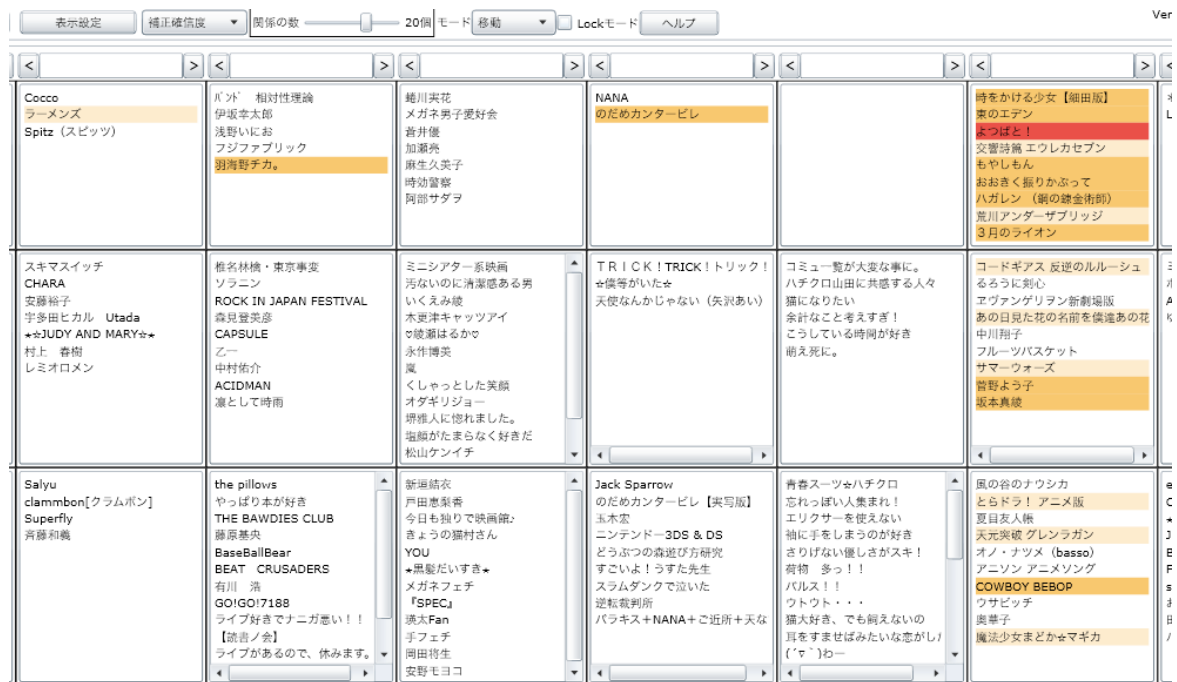


図 93 「よつばと！」を起点とした際の彩色例

⑦移動

⑥にて説明したソート機能に関連し、彩色されたデータを見ながら、データを移動し調整することが求められる。クラスタリングの機能②基準変更にて説明した通り、分析対象とするデータには解が複数あるため、彩色された関連度合いと解釈に基づいて、データを移動させる。

図 94 は、本文内にて説明した、系統 7 から系統 2 への移動である。起点とするデータは他クラスタとの関連が強く、解釈を合わせて移動を行っている。

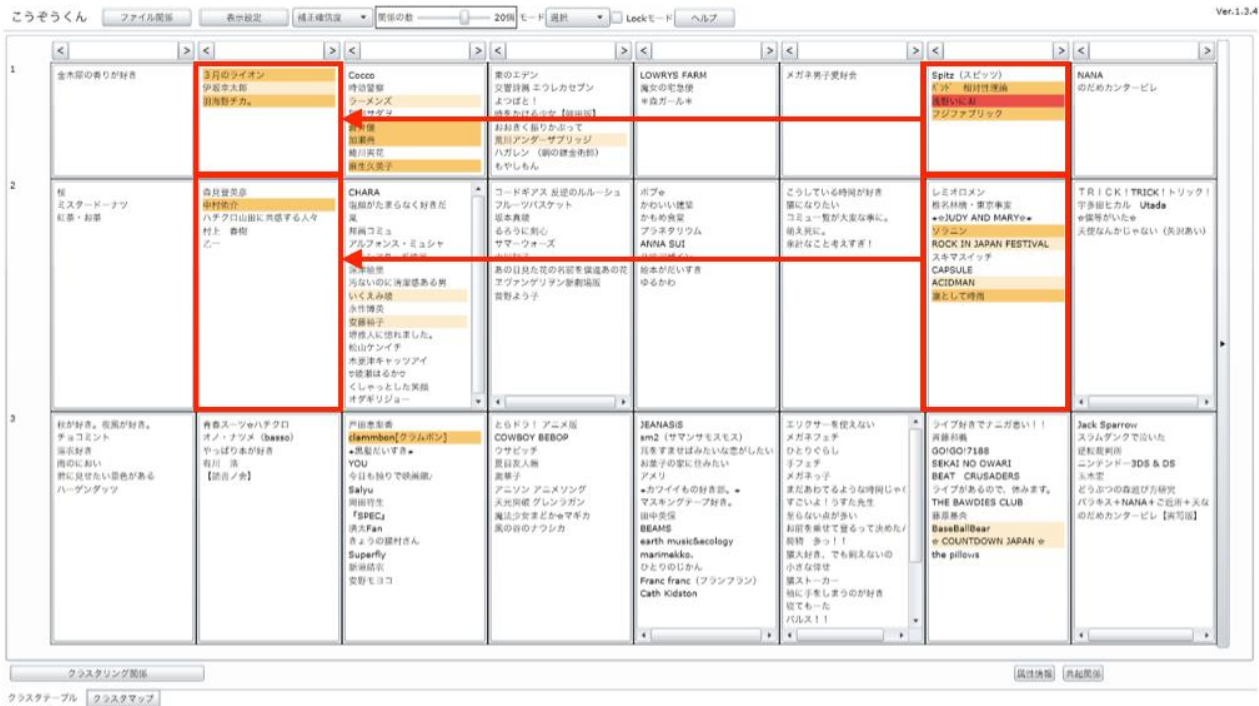


図 94 アイテムの手動による移動

⑧Lock

Lock 機能とは、特定のデータは移動をさせずロックさせる機能である。ロックされた以外のデータを再クラスタリングする際等に使用する。本分析では使用しなかった機能のため、例として、TBS 人気ドラマ「半沢直樹」の公式 Twitter²¹をフォローするユーザーらのポストを対象とした具体例を以下に示す。なお、データ収集は 2013 年 11 月に行った。

図 95 に、Twitter にて半沢直樹アカウントをフォローするユーザーらの発言構造の一部を示す。本ドラマには、ジャニーズグループのメンバーが出演していることもあり、ジャニーズに関する発言が多く見受けられた。それらを全てジャニーズ関連という系統とするとあまりに巨大となり、他の系統との連関を適切に見ることが難しい。当然、系統内の解釈も、ジャニーズ総合として大雑把なものになってしまう。このような場合には、ジャニーズは個別のグループや、メンバーの年齢といった、分析者のメタな視点を取り入れることが有効となる。データの関連だけでなく、つまり機械的には反映されない情報を分析者自身が付与することで、一つの分析軸が出来上がる。図 95 ではジャニーズに関して年齢とグループという情報で分類を行っていくことで、ジャニーズに関する発言の中でも大衆向け〜コアファン向けといった構造が浮かび上がってきている。

< ジャニーズ >	< 嵐 >
もしかしてだけど 有岡大貴 裕翔 知念侑李 山田涼介 ジャニヲタ ジャニショ ジャニ マリウス びんとこな 中島裕翔 手越	相葉 ニノ 翔くん arashi
北斗 中居くん バーナ 中居 鶴戸 亀梨 横アリ ジャニーさん kitty apple なっちゃん	二宮 エイト あいぼん 日産 大野智 家族ゲーム 松潤 大倉 anan 櫻井翔 安田
激おこ フラベチーノ 山下智久 うえーい サマーヌード 昼ドラ つらたん シェアハウス ゴリ 太鼓の達人 ネブリーグ 高田	100均 ヒルナンデス 北川景子

図 95 ロック機能

⑨削除

前述した⑦移動や⑧Lock という機能と関連して、不要なデータは削除するという機能も必要である。本分析のように、ソーシャルなデータを分析対象に据える時、一般的な社会調査の項目と異な

²¹ https://twitter.com/Hanzawa_Naoki

り、全てのデータに意味があるわけではない。そのため、分析の文脈に沿って、不要なデータは削除していくことで、データをよりコンセプチュアルに洗練させていく。

本分析の例では、本文 4-3-6 節にて、自分らしさを表現する系統内の「(▽)`わー」「うとうと・・・」というコミュニティを分析対象外として削除した。これら 2 つのデータは、元ネタが不明であるために何を表出しようとしているのかの判断が難しく、自己の内面性の表現としては内容に乏しいと考えられるため、削除するという判断に至った。(図 96)

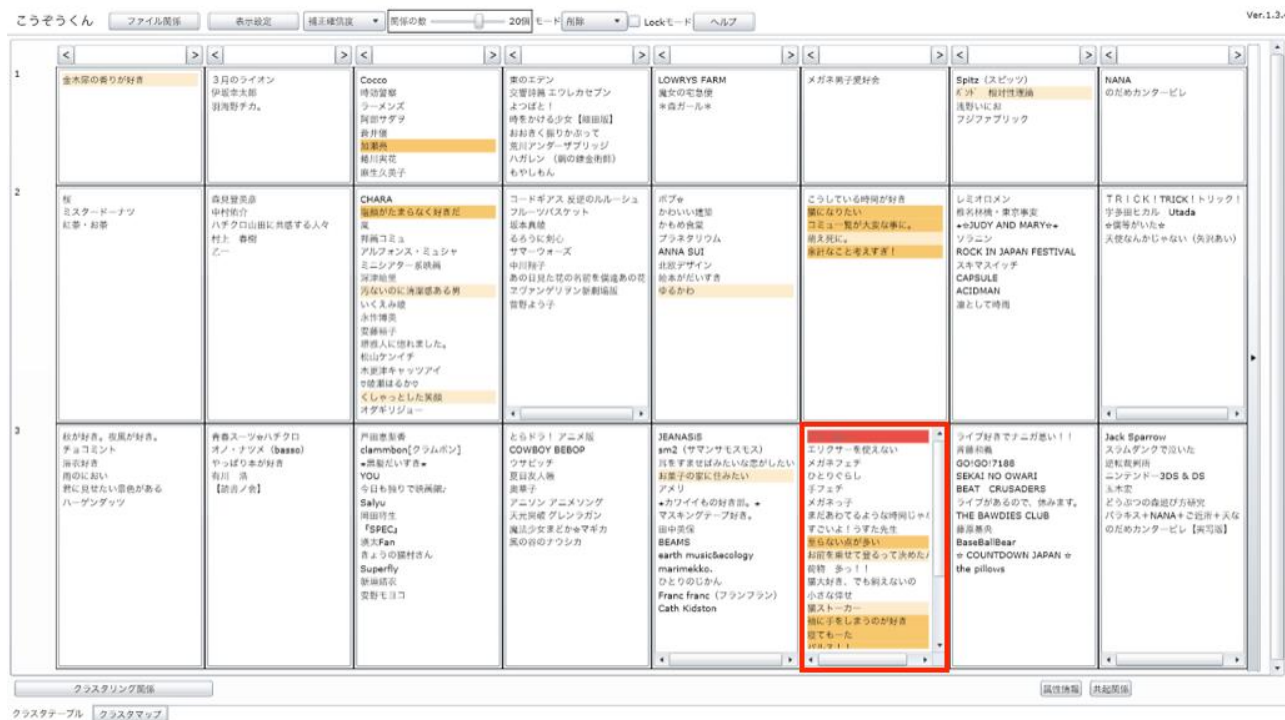


図 96 削除機能

特化係数

特化係数に関する機能は、図 97 に示す 1 つがある。

特化係数

⑩彩色

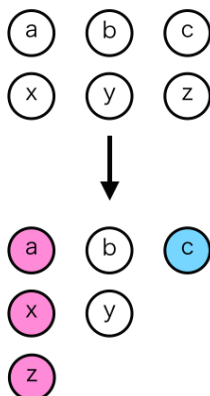


図 97 特化係数の機能の概念

⑩彩色

彩色機能自体は、相関ルールに関する⑥彩色もある。ここでは、ユーザの属性に関する彩色機能について説明する。データとデータの個別の関連度合いを見る際には、相関ルールの⑥彩色機能を活用する。属性に関する⑩彩色機能とは、データ全体を基準とした際の相対的な特化を見るためのものである。本分析では男女の属性を活用し、男女という軸を用いてデータの調整を行っていった。

本分析での例は、図 98 に示す、4-4-2 節の男性特化における調整で⑩彩色機能を活用した。男性属性に特化しているクラスタをピンク色で、非特化クラスタを青色で示している。各系統は解釈する際に一貫性があることが望ましい。そのため、系統内で属性特化にブレがあるデータに関して、⑥彩色機能を併用し、データの調整を検討していった。

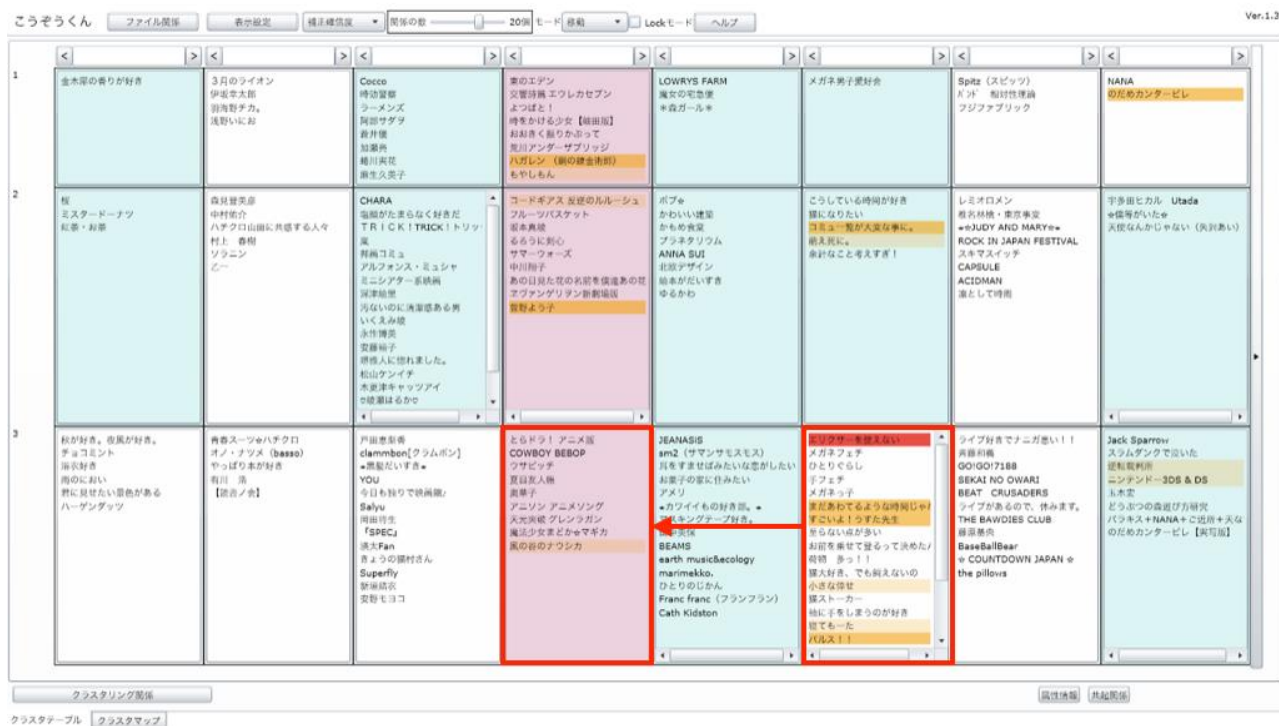


図 98 特化係数による彩色とアイテムの移動

4-4-8-3 概念化

概念化に関する機能を、分析手法別に以下で説明する。

クラスタリング

クラスタリングに関する機能は、図 99 に示す 1 つがある。

クラスタリング

⑪生成

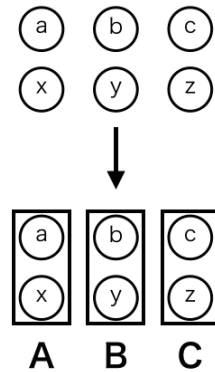


図 99 クラスタリング機能の概念

⑪生成

分析対象の全体構造を把握するための構造化の後は、構造内の系統間の関係を明確化するための概念化が必要である。概念化を通じて、具体的なデータ(本分析ではコミュニティ)から一段階抽象化したレベルで分析対象を理解することが可能となる。一段階の抽象化とは、具体的には系統の明確化、すなわちクラスタリングと個別の調整を通じた各系統の生成となる。通常データマイニングでは、クラスタリングとは探索的にグループを作り出すことだが、作り出されたグループ間の関係やグループ自体の意味を問う機会は多くない。分析対象の理解のために、系統という粒度でデータをまとめあげる生成という機能がここで重要な意味を持つ。

本分析の例では、図 100 に示すように、構造図で各系統として示していたデータをひとまとまりとしてクラスタマップ上に示している。

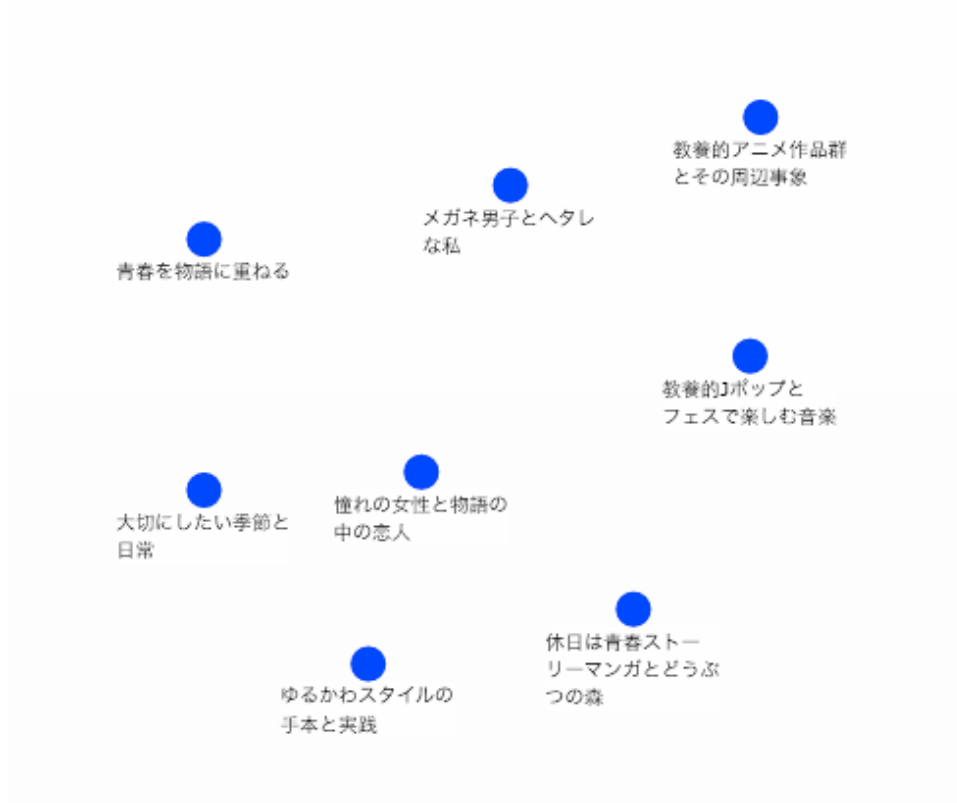


図 100 クラスタマップ

相関ルール

相関ルールに関する機能は、図 101 に示す 2 つがある。以下、個別に説明する。

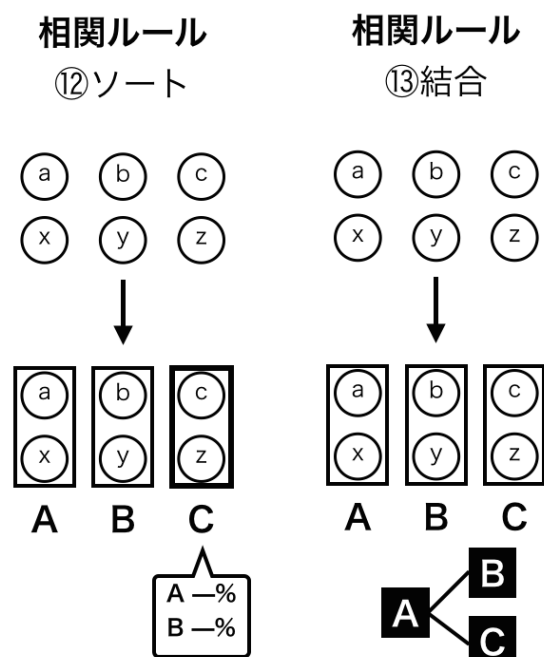


図 101 相関ルールの概念（ソート・結合）

⑫ソート

⑪生成にて系統を作成した後は、系統間の関係を見ていくことが求められる。系統間の関係は、構造化時の相関ルールにて活用する⑤ソート機能と同様に、各系統を起点とした際の関連度合いを示す。⑫ソート機能では、評価指標として Jaccard と信頼度比を採用している。

図 102 に、本分析にて「メガネ男子とヘタレな私」系統を起点としたデータのソート例を示す。

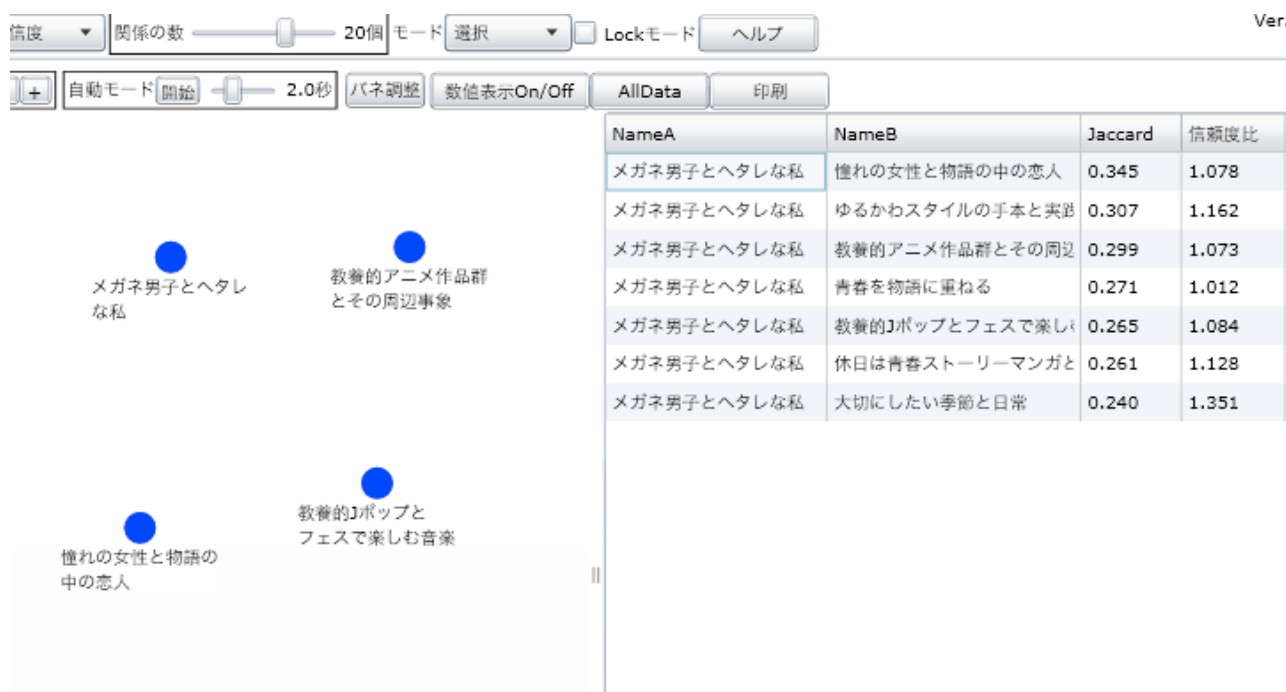


図 102 クラスタマップのデータ詳細例

⑬結合

⑫ソート機能に基づいて、系統間を結合する機能も必要となる。系統間を結合することで、概念化の一通りが終わる。図 103 は、本分析における結合の例である。構造化同様に、ここでも探索的な対話が必要となるため、結合の足切り値やリンク数の調整が可能となっている。⑫ソート機能にてデータの関連度を見ながら、解釈としても妥当な概念構造を導きだす。

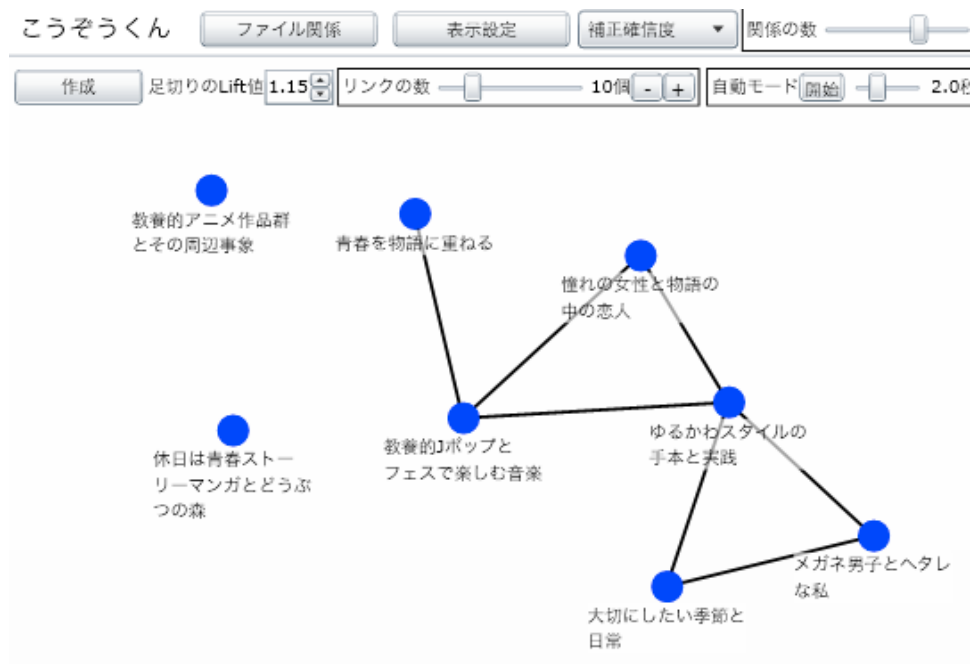


図 103 クラスタマップによるクラスタの結合例

第五章 探索的深層化ツール

ふかぼりくん

第五章 探索的深層化ツールの開発と実査

本章では、探索的社会調査ツールの第3番目のツールである、探索的深層化ツール「ふかぼりくん」の開発と実査について考察する。ここでは、対象となる社会データは、それなりに構造化されている。つまり、その意図は、いわゆるアスキングツールとセットになって、構造化されたデータの深層を深堀することを目的としているからである。前章までは、まったく構造化されていないデータの海から、そこで隠された関係を表出化、構造化することをツール開発の前提としていたが、ここでは、事前のある程度予想された構造をもとに、イメージをも含んだデータを取得するアスキングツールをも独自に開発し、そこで取得されたデータをもとに、その構造を特定化させることで、従来の社会調査では不可能であったレベルまで、深く関係性を模索することを可能にすることを開発目的とした。それが探索的深層化ツール「ふかぼりくん」である。

以下、アスキングツールであるグーカブーカ「GoocaBooca」の開発過程を説明し、その次に、探索的深層化ツールの目的・コンセプトから実装、そして実査について考察する。

5-1 Asking ツール（GoocaBooca）の開発と実装

5-1-1 目的とコンセプト

GoocaBooca は、イメージ調査を行うためのアンケートサイトである（図 104）。ふわふわとした、ぼんやりとした、社会にある雰囲気を感じ取るための調査である。例えば、「雰囲気かわいい」という言葉があったとして、どこらへんが皆が感じている「雰囲気かわいい」なのか？という境界を明らかにする調査である。言葉では、表現しづらい微妙な差異を調査するためには、このような画像について回答させるしかないということである。

画像について、Yes か No を答えるだけというシンプルなアンケートである。最後には、通常のアンケートのように、性別や年齢などの質問をできるようにしている。回答者に携帯電話、スマートフォンを使って、ランダムな画像 40 枚程度に対して、Yes（Good）か No（Bad）を回答してもらうものである。最後には、性別、年齢などの基本的な質問もできるようにしている。



図 104 GoocaBooca 全体像

この GoocaBooca は、ランダムに画像を見せて、回答させることをしている。スマートフォンに対応

する形で、気軽に回答できるようにしている。ランダム故に、すべての用意した画像に対して答えることができない。これは回答者の負担を減らし、そのため、画像間の関係など算出できないものがある。しかし、それ故に「多くの画像を見せて判断させる」ということが可能となる。データマイニングは完全なデータに対する分析ではなく、疎のデータに対する分析として発展してきた。これに対応する形で、データの収集と分析をするツール開発を行った。

5-1-2 実装

実装は、ASP.Net の MVC4 を使って、クラウド上に作成した。回答は、データベースに蓄積するようにしている。回答時、ランダムに画像を表示させているが、完全なランダムではない。完全なランダムでは、期待される画像の種類すべてを回答者に見せることができないためだ。そのため、画像にはカテゴリをつけられており、そのカテゴリの中で何個かを表示させるという仕組みになっている。

質問の登録は、図 105 のような指定の XML ファイルを作り、それを Web にアップロードすることで行えるようにしている。このようにしたのは、入力フォームを作り、入力させるよりも、一覧性が増し、ある程度は、コピーアンドペーストで行うことができるので、複数のアンケートでおなじ質問を行うときに便利だからである。ResearchIdName が同じなら、質問票 XML をアップロードしたとき、既存の質問を書き換えられるようになっている。

```
<?xml version="1.0" encoding="utf-8" ?>
<Research>
  <Base>
    <ResearchName>恋服</ResearchName>
    <ResearchIdName>koifuku2013</ResearchIdName>
    <Description>
      これから表示される写真を見て、好きかそうでないかを答えて下さい。
      好きなら「YES」を、そうでなければ「NO」を選択して下さい。ただし、異性の場合は” 恋人
      が着る服として” 好きかそうでないかを判断基準にして下さい。
      写真は全部で40枚です。
      それではスタート！
    </Description>
    <QuestionText>ようこそ！「恋服（こいふく）」はケータイを使ったリサーチです。ファッション
    の趣向から男女のマッチングを見い出します。</QuestionText>
    <AnswerCount>40</AnswerCount>
    <Hidden>False</Hidden>
    <ResearchType>GooCaBooca</ResearchType>

    <ImageBaseUrl>http://web.sfc.keio.ac.jp/~kiichi/ORF2013/koifuku2013/</ImageBaseUrl>
    <ResearchMainImage>mainImage.png</ResearchMainImage>
    <ExtendAnalyzeResultUrl></ExtendAnalyzeResultUrl>
  </Base>
  <ItemAnswerChoiceList>
    <AnswerString Tag="Key">YES</AnswerString>
    <AnswerString>NO</AnswerString>
  </ItemAnswerChoiceList>
  <Categories>
    <Category CategoryName="キュートガール">
      <Item ImageUrl="cute1.jpg"></Item>
      <Item ImageUrl="cute2.jpg"></Item>
      <Item ImageUrl="cute3.jpg"></Item>
    <!--繰り返し-->
  </Categories>
</Research>
```

```

</Category>
<Category CategoryName="カジュアルガール">
  <Item ImageUrl="ca1.jpg"></Item>
  <Item ImageUrl="ca2.jpg"></Item>
  <Item ImageUrl="ca3.jpg"></Item>
  <!--繰り返し-->
</Category>
<Category CategoryName="クールガール">
  <Item ImageUrl="cool1.jpg"></Item>
  <Item ImageUrl="cool2.jpg"></Item>
  <Item ImageUrl="cool3.jpg"></Item>
  <!--繰り返し-->
</Category>
<Category CategoryName="ナチュラルガール">
  <Item ImageUrl="mori1.jpg"></Item>
  <Item ImageUrl="mori2.jpg"></Item>
  <Item ImageUrl="mori3.jpg"></Item>
  <!--繰り返し-->
</Category>
</Categories>
<Questions>
  <Question>
    <QuestionName>性別</QuestionName>
    <QuestionText>性別は？</QuestionText>
    <QuestionChoices>
      <QuestionChoiceText>男性</QuestionChoiceText>
      <QuestionChoiceText>女性</QuestionChoiceText>
    </QuestionChoices>
  </Question>
  <Question>
    <QuestionName>年齢</QuestionName>
    <QuestionText>年齢は？</QuestionText>
    <QuestionChoices>
      <QuestionChoiceText>18 未満</QuestionChoiceText>
      <QuestionChoiceText>18～20 歳</QuestionChoiceText>
      <QuestionChoiceText>21～24 歳</QuestionChoiceText>
      <QuestionChoiceText>25 歳以上</QuestionChoiceText>
    </QuestionChoices>
  </Question>
  <Question>
    <QuestionName>恋人の有無</QuestionName>
    <QuestionText>恋人はいますか？</QuestionText>
    <QuestionChoices>
      <QuestionChoiceText>いる</QuestionChoiceText>
      <QuestionChoiceText>いない</QuestionChoiceText>
    </QuestionChoices>
  </Question>
  <Question>
    <QuestionName>デート服</QuestionName>
    <QuestionText>デートを想定した時、どちらを意識しますか？</QuestionText>
    <QuestionChoices>

```

```

<QuestionChoiceText>自分好みの服</QuestionChoiceText>
<QuestionChoiceText>モテそうな服</QuestionChoiceText>
</QuestionChoices>
</Question>

<Question>
<QuestionName>フ リ ー</QuestionName>
<QuestionText>どんな服装の女の子がモテると思いますか？</QuestionText>
<QuestionType>FreeText</QuestionType>
</Question>
</Questions>
</Research>

```

図 105 質問設定の XML ファイル

最後の質問項目には、選択肢、ドロップボックス、自由回答文の三種類が扱えるようになっている。Web ページでは、指定に基づきアンケートとして作成される。

5-2 探索的深層化ツールの開発と実査

5-2-1 目的・コンセプト

データ分析において、分析のアルゴリズムも重要であるが、それ以上に重要なのは、データを見ることと、適切なデータにすることである。適切なデータにすることは、データを見て、不要なデータを捨てることや、値の統合や、複数の変数を一つにまとめることによって成される。このようなことは、人間が目的意識を持って、試行錯誤の末にできることなので、これを支援するようなアンケート分析ツールを作成した（図 106）。簡易的に自然言語処理もしてアンケートのフリーアンサー分析もできるようにしている。

Web から使える、ツールとして作成し、データの深堀りができるので、「ふかぼりくん」と命名した。

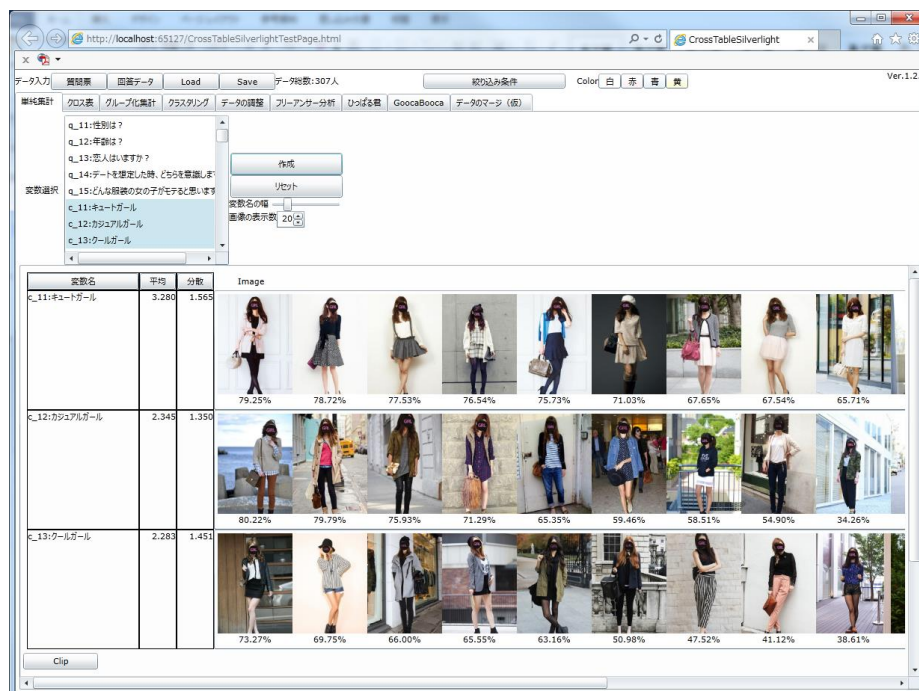


図 106 「ふかぼりくん」 ツール画面

5-2-2 解決する問題

データ分析のために、新たにソフトウェアをインストールするのは、トラブルの元であるので、やりたくないことであり、今は、Windows と Mac の 2 つのマシンが普及しているので、それらの OS のマシンに対応したツールをつくろうとすると、それは Web 上に作るのが最適である。しかし、Web 上にアンケート分析ツールをつくろうとすると、データの入出力に困る。データベースと通信する形もあるが、それは開発コストが高いものになってしまうし、柔軟なデータに対応するのが難しい。また、アンケートデータは、個人情報に係る場合、秘匿性が高いものもあり、利便性のために、安易に Web 上にデータをアップロードしていいものでもない。そのため、Web ブラウザで起動は簡単にできるが、データはローカルに持ってローカルで分析するというツールを開発した。これは、ブラウザプラグインの Silverlight がローカルファイルを読めるという特質を利用したものである。

また、はじめにでも述べたように、データ分析は、どれだけ細かいものが見ることができるのかが重要である。そのため、データの条件を絞り込むことに注力したデータ分析ツールである。

5-2-3 開発の経緯と目標

Web 上でのアンケート分析用のツールが欲しいということで作成した。開発を依頼したところが、データの秘匿を希望したのと、Window,Mac の両方に対応する必要があったため、このような仕様にした。その後、GoocaBooca 用のデータ分析ツールとして使えるということで、それ用のインターフェースを用意した。アンケートには、フリーアンサーもあるということで、自然言語処理も加えることにした。

5-2-3-1 設計

質問データ

アンケート分析のためのツールということで、入力データは、質問データと回答データに分けている。ともに、エンコードが UTF8 の TSV ファイルである。質問データは、質問の型を定義するファイルである。列名は「Key,Text,Type,Answers」それぞれはこのように対応する。

表 11 質問データの列名とその意味と書式

列名	意味	書式
Key	回答データの列名	回答データに対応する列名
Text	質問の Text	自由。画像 URL を {} で囲むと画像表示
Type	質問の Type	離散,順序,文字列,ラベル
Answers	回答と文字列の対応データ	「値:ラベル」でカンマ区切り。なくてもいい
Children	質問の親子関係を示す	Key 名をカンマ区切り

Key は、回答データとの対応を取るためのものであり、回答データの列名と必ず一致させなければならない。Text はその回答の名前である。自由につけていい。ただし、TSV の形式上、Tab は禁止である。画像を表示させるためには、{} でその画像の URL を囲むとその画像が表示される。GoocaBooca は画像を回答するアンケートのため、このような拡張した。

Type は、集計時どのような振る舞いをするのか？ということを示す。平均を算出するにも、男女のような連続値ではないもので、平均をとっても意味が無い。そのため、連続値か離散値かを入力してもらう。連続値の時は、計算時その数値を使うが、離散値の時は、その頻度の集計のみである。

Answers は、回答と文字列との対応関係を記述するところである。素っ気ない数値だけでは、データをイメージ出来ない。そのため、数値とラベルとの対応関係を記述する。

Children は、質問の親子関係を記述する。単純集計時、子の集計結果も表示する。

回答データ

アンケートの回答データも、UTF-8 の TSV データである。列名は先の質問データと対応する。TSV 形式にしたのは、Excel との相性を考えた上である。また、毎回質問データを作らないといけないというのは、不便なので、回答データのみでもこのツールを動くようにした。この時、最適な型などの情報をつかうことができないが、通常の集計なら機能する。

絞り込み機能

探索的分析を支援するために、インタラクティブな絞り込み機能を作成した。質問回答データを選択することにより、可能にする。質問回答データの表示上の名前は変えられるようにする。

作成した機能コンセプト

基本的に、データを見ることに注力を置いている。そのため、単純集計、クロス集計、グループ化集計を用意した。また、クラスタリングによる新たに作られた変数をデータに加えて、集計処理をやれるようにしている。これにより、変数を統合したものからの集計、深堀りができるようになっている。

5-2-3-2 実装

実装は、C# で Silverlight を作成した。Silverlight のため、Web 上から、Windows でも Mac でも使えるようにしている。

Web ブラウザから使えるとはいえ、Silverlight の機能により、ローカルファイルの読み書きができ、入力ファイルは、ローカルファイルを使う。そのため、データをネットに流さない。

質問データを読み込んだ時、質問オブジェクトと、質問回答オブジェクト、質問回答グループオブジェクトが作られる。これは、質問の絞り込み、統合、削除に対応するために、作られる。これらは、親子関係になっている。内部的には、一つ一つの質問回答をオブジェクトとして保管し、その親に、グループ化した質問回答オブジェクト、その親に、質問オブジェクトという構成にしている。一人一人の回答データを読みながら、対応する質問回答オブジェクトを取得し、グループ化された質問回答オブジェクトを返すという形で実現している。絞りこみ情報をソフト全体で共有する形にしたことで、同じ絞り込みで、様々な分析ができるようにしている。以下、諸機能について説明する。

Save & Load

質問データ、回答データ、クラスタリングデータを保存とロードができる。セーブとロードは、ローカルファイルで行い、Web 上にデータを流すことはない。分析の再開が出来る仕組みである。ファイル形式は、機械生成された XML 形式であり、入力した質問データと、回答データを統合したものであり、一度読み込みセーブした後は、このファイルを読み込めば使えるようになっている。

単純集計

指定した質問項目で単純集計をする。データの全体像を把握するのに使う。一つ一つの項目に対して、平均と分散を出力し、画像を含むものに関しては、画像を表示する。

また、質問項目に Children を含む項目に関しては、集計しそのランキングを表示する。この表示数に関しては、「画像の表示数」でコントロールできる。

クロス集計

指定した質問項目でクロス表を作る。2つの変数間の関係を見る。グループ化変数、ターゲット変数を指定すると、クロス表が作成される。相関があるかどうかの検定は、自由度とカイ二乗値を見る。

「Clip」はそのクロス表をテキストとしてクリップボードに転送する。そのため、他のソフトにそのデータを持って行くことができる。

彩色機能



図 107 カラーパレット

上部にある、カラーパレット（図 107）で、色を選択した状態で、クロス表の気になる数字をクリックすると、そこに指定した色がつく。（図 108）

Clip		q_12:年齢は？						
q_11:性別は？		18未満	18～20歳	21～24歳	25歳以上	合計	平均	分散
	男性	3 (1.0%)	70 (23.1%)	172 (56.8%)	58 (19.1%)	303	26.94	0.68
	女性	6 (1.4%)	184 (43.6%)	212 (50.2%)	20 (4.7%)	422	26.58	0.61
	合計	9 (1.2%)	254 (35.0%)	384 (53.0%)	78 (10.8%)	725	26.73	0.66

χ²乗値:56.84 自由度:3

図 108 彩色したクロス表

このように色を分析者自身が付けていくことで、自分がどの変数を注目しているかを可視化する。指定した質問項目でクロス表を作る。2つの変数間の関係を見る。グループ化変数、ターゲット変数をしていると、クロス表が作成される。相関があるかどうかの検定は、自由度とカイ二乗値を見る。

グループ化集計

指定した変数をグループ化して、対象変数を集計する。グループ化する変数は複数指定可能です。そのため、データとして存在するすべての組み合わせを出力する。足切りは、「最小値」のスライダーで行う。そして、各要素は、ソートできるようになっている。偏差値を出力しているので、それを見ることで、全体での位置がわかるようになっている。

クラスタリング

指定した変数に基づいてクラスタリングを行う（図 109）。クラスタリングは、このツールでは、次元圧縮に使う。すなわち、複数の変数からひとつの変数を作る。複数の変数を集約化するので、解釈がやりやすくなる。新しく作った変数を元に、単純集計、クロス集計ができる。

クラスタリングのアルゴリズムはK-Means 法[40]である。そのため、予め、作成するクラスタ数を指定する必要がある。クラスタリングの結果は、レンダーチャートで表示される。変数の情報の表示はチェックボックスで切り替えできる。

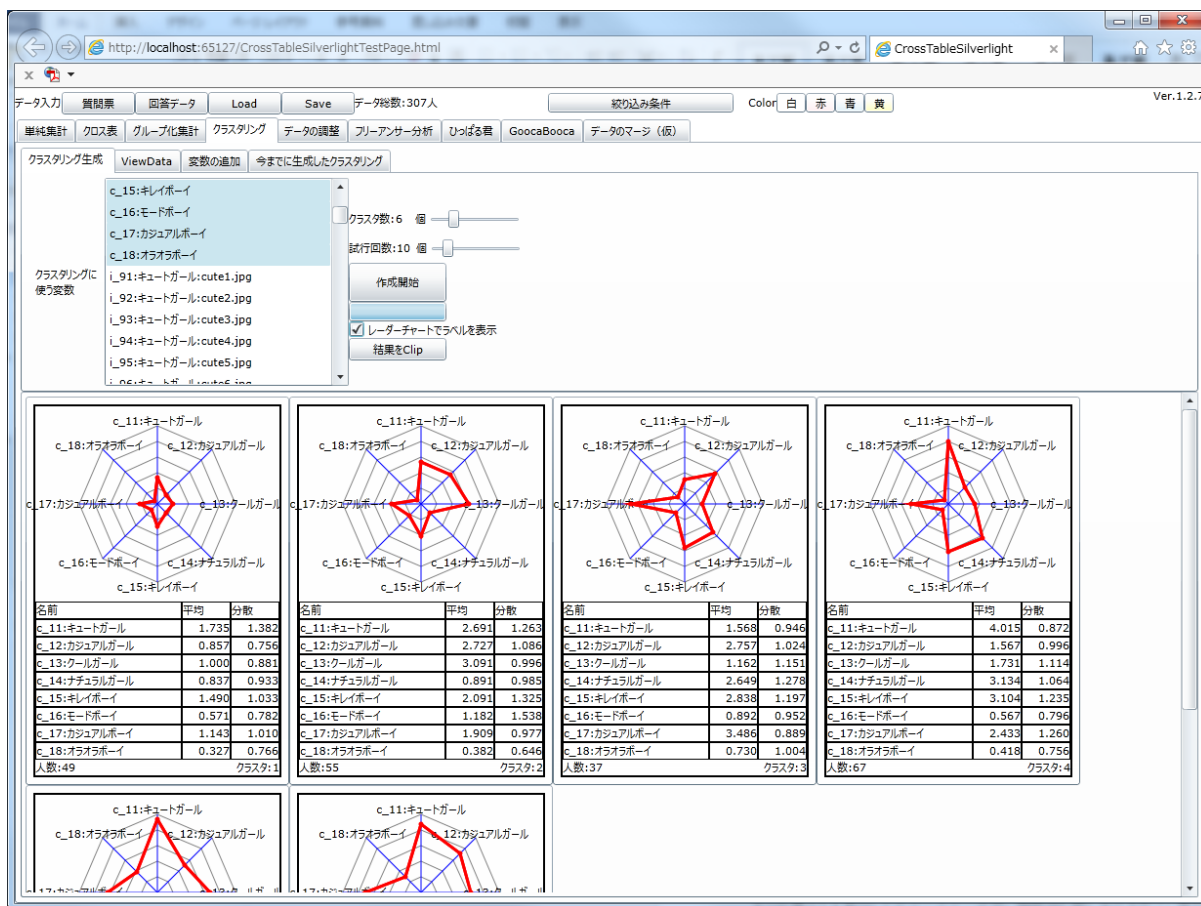


図 109 クラスタリング結果画面

ViewDataにタブを移動後、見たい変数を指定して、クラスタをクリックすると、そのクラスタの変数を見ることができる。

変数の追加では、今作成したクラスタリングの結果を、新たな変数として追加する。

「今までに作成したクラスタリング」では、作成したクラスタリング結果を表示します。これは、「変数の追加」を行ったものだけである。

データの調整

使用するデータの絞り込みを行う（図 110）。ふかぼりくんの深堀たる所以である。

質問項目を選択後、表示される、個々の項目の「+」ボタンを押すと、その項目で絞り込む。おなじ質問では、OR で、異なる質問では、AND となっている。そのため、直感的な絞り込みとなっている。「-」ボタンを押すと、分析の候補から外れる。これは、ノイズデータの削除に使う。

このデータ調整で、分析対象とするデータを指定します。データ総数をみることで、絞りこまれた人数がわかるようになっている。この絞りこまれた人数により、他の機能が使えるようになっている。

絞り込みの条件は、「絞り込み条件」ボタンで、どこからでも見るできるようになっている。

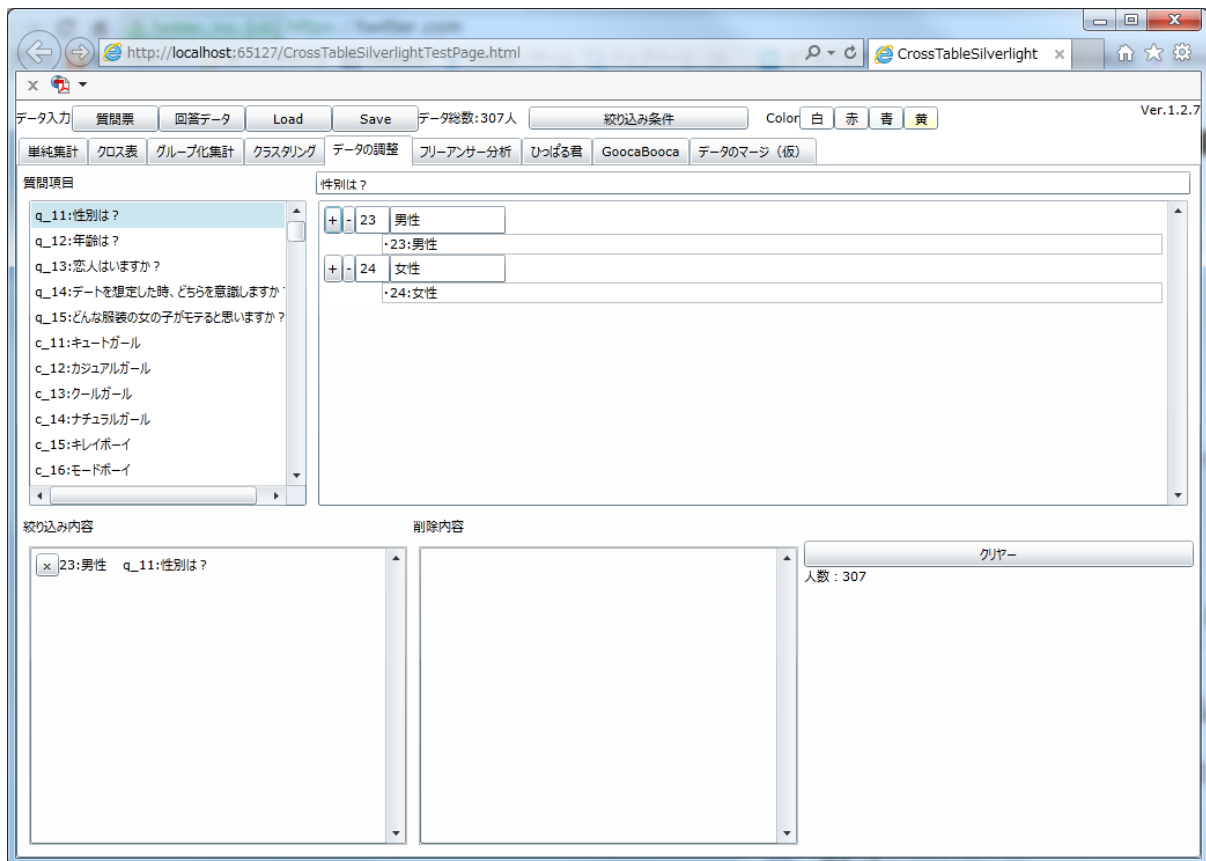


図 110 「データの調整」画面

また、変数のまとめもできる。項目をドラッグドロップすると次のようにデータをまとめることができる。分析もこれに沿って行われる。

このようにして、分析に使う変数の絞り込み、変数の統合を GUI で行うことができるようにした。

フリーアンサー分析

アンケートにつきものの、フリーアンサーの自然言語分析を行う。これは、変数と自由回答分での単語の関係をみるためのものである。。そのため、対象となる自由回答文の設問と指定と、変数の指定が必要となる。

単語は、頻出熟語、未知語も自動的に判別され集計される。それぞれの変数ごとに集計される。特化係数を使うと、その項目だけに特化している単語がわかるようになっている。単語の横のボタンを押すと、その単語を含む文章を見ることができる。

5-2-3-3 開発上の工夫

開発上の工夫は、一つ一つの質問回答をオブジェクトとし、グループ化オブジェクトを挟んで、質問と結びつけたことである。これにより、変数の自由な深堀りを可能とした。初期設計では、グループ化オブジェクトを挟んでいなかったため、そのようなことができなかった。

また、絞り込み情報を共有化したことも工夫である。これにより、他の分析が同じ絞り込み情報を使い行える。絞り込み情報がいつでも見ることができるよう、一番手前に簡易的な絞り込み情報を見ることができるようになっている。

クラスタリングにより新たな変数を追加することが簡単にできるようにしたのも工夫の一つである。これにより、複数の変数をまとめた変数を作ることができる。

マップとカルテのコンセプト通り、全体像と一つ一つの部分を見せるために、クラスタリング後の結果

の詳細データを見ることができるようになっている。これにより、クラスタリング後のデータがどれだけ使えるものになったのかわかるようになっている。

ブラウザ上で分析を行うため、自然言語処理の基本の形態素解析は、定番の Mecab²²を使うことができなかった。Web 上で形態素解析を行うには、Yahoo!社が提供する自然言語処理 API²³を使うという方法もあるが、一日 5 万件の API 制限があり、使い勝手はよくない。そのため、辞書を必要としない TinySegmenter²⁴をベースにユーザ辞書を扱えるように改良したのを作り、頻出連語、未知語に対応した。やり方は、一度、すべての文章を TinySegmenter に通したあと、頻出連語を集計し、高いものをユーザ辞書として加えた上で、もう一度、TinySegmenter にかけて集計した。欲しいものは頻出する言葉の塊であり、それを抽出するためには、ユーザ辞書を動的にやりたかったのが、このような形にした。

Silverlight を使うことで、ブラウザ上から、Windows でも Mac でも使うことができるようにした。ただし、Silverlight は陳腐化した技術なので、これから継続的に使うのは難しいだろう。

5-2-3-4 ソフトウェア開発として特徴的な点

このソフトウェアは、アジャイル的に作られた。今から思うと、適切に設計して作ったほうが良かったと思うが、全体的な設計を行わず、初期バージョンは、手早く 1, 2 週間で作り上げた。設計は特に行わなかったが、MVC モデルを意識して、データは、Model にする、質問も、Model にするというのを念頭に置いて作成した。後々になって、MVVM モデルに開発パターンを切り替えた。

当然だが、このソフトウェアは、データを扱うものであるもので、MVC モデルとの相性は良かった。また、MVVM 的な枠組みも、データと表示を混在させることなく、分けることができたので、アジャイル的な開発での、仕様変更への対応が局所化された。

アジャイル的に開発する上で、Silverlight のブラウザで手軽に使うことができること、と言うのは有利に働いた。最新版の配布は、インストールの手間がいらず、Web サーバにアップするだけで完了し、テストする側も Web サイトにアクセスするだけで済むからだ。しかし、ブラウザ側にキャッシュが残っていると適切に最新版にならないということが、多少手間であったが、ダウンロードしてインストールしないといけないというよりは簡単である。また、データの絞り込みに特化している点である。

5-2-3-5 開発活動を通じて得た知見

インストール不要なことは、授業などでこのソフトウェアを集団で使うとき便利である。インストールは、それぞれの PC の個別環境に依存するので、トラブルがつきものだが、そういったものが少ないのは、嬉しい。Silverlight のプラグインのインストール作業が必要ではあるが、一度インストールすれば問題ない話であり、BtoC といった一般人相手のツールというわけでもないで、そこまでの不都合はない。

5-3 実査 6 女性ファッションスタイル（恋服）の解析

価値探索ツールの 3 番目が、多様性探索ツールの「ふかぼりくん」である。これは、通常の社会調査の解析ツール、たとえば SPSS のようなツールの簡易版としての位置づけをすることも可能ではある。しかしこのツールの本来の価値は、ある程度構造化された社会調査のデータの中に、事前の想定(仮説)を超えた価値ある知識を創発させることである。従来の社会調査の解析ツールが統計解析としての価値を実現することを目的としているのとは異なり、この「ふかぼりくん」は、そのネーミングに示されるように、データの既存構造をもとに、さまざまな条件を付加することで、データの深堀を行い、今まで自明とされた構造内部において、いかに多様な要因とそれらの関係性が潜んでいたかを抉り出すことこ

²² <https://code.google.com/p/mecab/>

²³ <http://developer.yahoo.co.jp/webapi/jlp/>

²⁴ <http://chasen.org/~taku/software/TinySegmenter/>

そ、本来の価値なのである。だからこそ、統計解析ツールの特徴をもちながら、統計解析ではなく、多様な価値探索を実践するツールなのである。

さらに、このツールは、データ入力ツールであるグーカブーカ (GoocaBooca) とセットにすると、そこでは画像データを扱っているの、画像データを含んだ価値探索が可能になり、従来の社会調査ツールとは全く異なった特徴をもつことになる。従来、この種の調査は、よく市場調査において、定性調査と絡んで実施されることが多い。インタビューによる定性調査は、ある意味、価値探索的な調査手法であるから、その意味では、ふかぼりくんに通ずる特性をもつのは、この種の定性調査であるといえよう。ただ、定性調査の場合、インタビューによる深掘はできても、その調査対象者は限定されており、しかもデータの解析についても、かなりの主観的な解釈が重視される。これに対して、ふかぼりくんでは、データを大量に取得することは可能であり、しかもそこで集計されたデータ群と研究者と相互補完的な関係が継続されるという点で、定性調査とは異なった新しい社会調査の可能性をしめすものといえよう。

かくして、ふかぼりくんは、グーカブーカとセットにすることで、従来の社会調査にはないユニークな価値探索を創発するツールである。そこで、以下、実際の社会調査を実例に挙げながら、このツールがどのような意味で、価値探索的であるか、と個別具体的に示す。

ここでは、2013 年の熊坂研究室での卒業制作の 1 つである「恋服－恋するあなたの勝負服とは－」のデータをもとに、「ふかぼりくん」ツールが、具体的に、どのような意味で多様性探索ツールであるかを明確にする。

「恋服」調査は、現代の若者が恋愛関係に入るとき、男女それぞれ、自分の服装のファッションと相手のファッションにはなんからの相性を求めるとすると、その相性には、どのようなヴァリエーションがあるかを探索しようとする、恋愛関連のファッション嗜好を認識しようとするライフスタイル調査である。

5-3-1 調査フレーム

まず、調査フレームを以下に示す。(図 111)

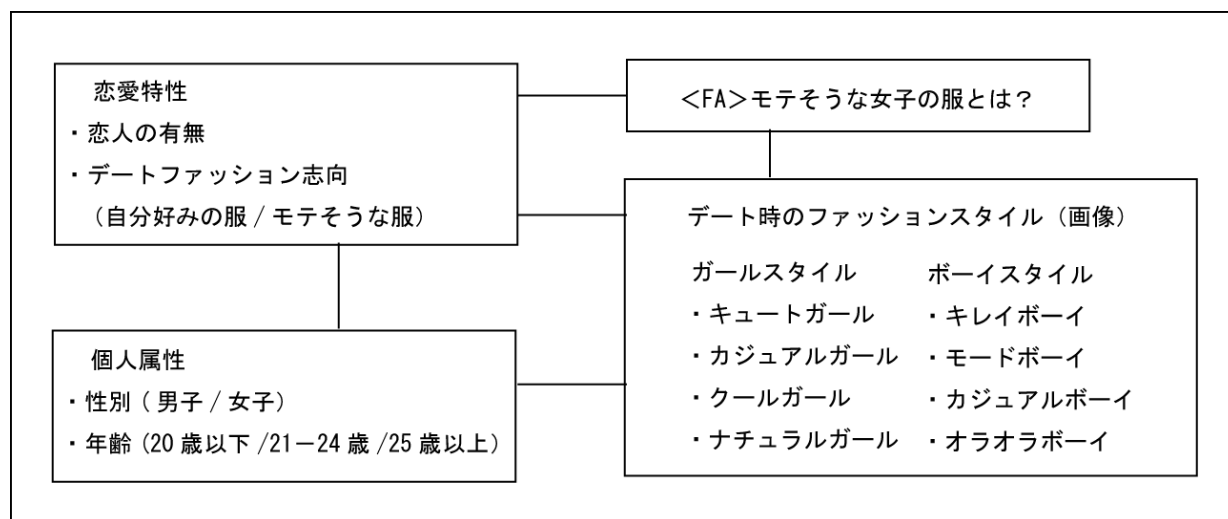


図 111 「恋服」調査の調査フレーム

この調査はスマホを活用した調査なので、非常にシンプルな構造になっている。第 1 に個人属性は、性別と年齢の 2 つだけである。第 2 は、恋愛特性で、2 つの項目からなる。1 つが恋人の有無であり、その有無はデートファッションの選択に影響をもつと思われる。ただ、ここで仮説として設定するほど、明確な関連性を設定できてはいない。もう 1 つの項目はデートする時のファッションとして、自分好みの服を選ぶのか、相手との距離を考えてモテそうな服を選択するか、である。

第 3 は、デート時のファッションスタイルの項目で、ここでは、さらにガールスタイルとボーイスタ

イルについて、それぞれ、女子ならば、カールスタイルについて、自分のデートファッションとして好ましいかを質問し、ボーイスタイルについては、想像上の彼氏のデートファッションとして好ましいかを質問する。男子の場合は、ガールスタイルでは想像上の彼女のファッションスタイルとして好ましいかを聞き、ボーイスタイルでは自分のデートファッションとして好ましいかを聞いている。また、各ガールスタイルとボーイスタイルそれぞれについて4スタイルを想定し、かつそれぞれについて15枚の画像を用意した。したがって用意した画像は、 $2 \times 4 \times 15 = 120$ 枚である。ただ、スマホでの調査なので、120枚すべてについて質問するのは調査設計としては好ましくないので、各スタイルについて、15枚からランダムに選定して5枚を提示して、判断してもらうことにした。したがって、ユーザとしては $2 \times 4 \times 5 = 40$ 枚の画像について判断してもらうことにした。

ここでは、ガールスタイルとしては、つぎの4スタイルを想定(仮説もどき)した。

1) 女子の場合

1. キュートガール：スカートやワンピースを着用し、コンサバ系やふわふわした女の子らしいファッション
2. カジュアルガール：ジーンズやスニーカーなどを着用したラフなファッション
3. クールガール：パンツスタイルやカチッとしたジャケットまたはセクシー系ファッション
4. ナチュラルガール：森ガール系や草食系のファッション

2) 男子の場合

1. キレイボーイ：ジャケットやシャツなどを着用した清潔感のあるファッション
2. カジュアルボーイ：パーカーやキャップなどを着用したスポーティでラフなファッション
3. オラオラボーイ：いわゆるヤンキー系やホスト系が着用するフィット感やギラギラ感のあるファッション
4. モードボーイ：モノトーンや柄と柄を合わせたものなど、ハイファッションと呼ばれるファッション

ここで採用された画像とファッションスタイルとの対応関係は、調査主体がア priori に設定してものであり、その対応関係については、解析の過程で多様な解釈が必要とされる。

第4は、フリーアンサーで、「モテそうな女子のファッションはなにか」である。これは上記3項目が構造化された質問項目であるのに対して、フリーアンサーなので、全く構造化されてはいない。このフリーアンサーをいかに解析過程で扱うか、それが、「ふかぼりくん」の特性の1つである。

5-3-2 価値探索プロセス

以下、どのように価値探索を実施したかについて、そのプロセスを示しながら、そこでの価値探索が具体的にはどのようなことであるかを確定して、このツールの特性を明確にしていく。

5-3-2-1 単純集計

ここでは、ガールスタイルについての単純集計の結果を表示する。図112に示すように、4つのガールスタイルの集計結果が、「統合画面」と「個別画面」が同時に表示される。そのことで、統合画面では、4つのガールスタイルそれぞれの全体像(平均と分散)を明確にし、かつそれらの全体像が容易に「比較」できるようになっている。つまりここでは、「全体と比較」を通して、単純集計の段階で、調査の大枠に関する全体的なイメージの把握が容易になされる仕掛けになっている。これは、価値探索の最初の特徴である。

個別画面では、各ガールスタイルについて、0個から5個の分布状況が実数とパーセント表示で分かるようになっている。グーカブーカのユーザは、各スタイルがもつ15枚の画像からランダムに5枚が選定され、それに「似合う/似合わない」と回答するので、スタイルごとに、「似合う」程度が0から5の段階となり、その集計結果がこの個別画面で明確になるような仕組みになっている。統合画面と同一画面に位置することで、全体と比較の詳細な確認が、この個別画面でできる仕組みになっている。ここには、「全体と個別の相互関連が容易に確認できる」ようになっており、ここでも探索行為を容易にする仕掛けがセットされている。

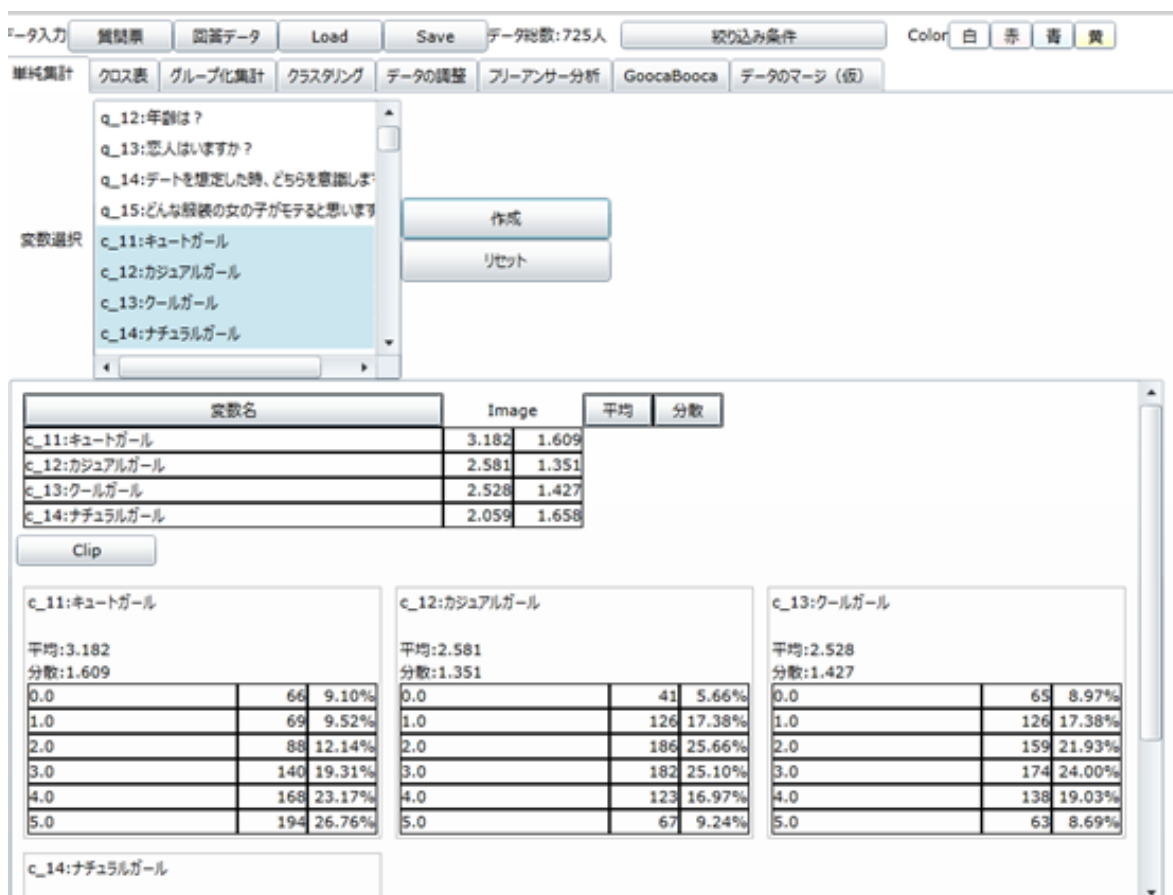


図 112 単純集計の表示結果

さらに、分かりやすさを示すのが、平均と分散のボタンを押すことで、比較が降順と昇順でソートできる仕組みになっており、単純集計の比較分析がさらに探索可能な仕組みになっている。「ソート機能」は、価値探索を容易にする重要な機能である。

図 113 と図 114 に示すように、ガールスタイルとボーイスタイルを全体として比較すると、8 スタイルの比較であるから、図 113 では、比較の意図がなかなか達成されない。要は、比較分析が困難なのである。そこで、図 114 に示すように、平均のボタンを押すことで、ソート機能が発揮される。すると、8 スタイルの平均値が降順でソートされる。その結果、比較の解釈が容易になされる。ソート機能が活きてくる。

その成果として、

1. 8 スタイルでは、ガールスタイルの支持がボーイスタイルよりも相対的に高い。
2. ガールスタイルのキュートガールとボーイスタイルのキレイボーイの支持が高い。

データ入力 質問票 回答データ Load Save データ総数: 725人 絞り込

単純集計 クロス表 グループ化集計 クラスティング データの調整 フリーアンサー分析 GoocaBooca デ

変数選択

- c_11:キュートガール
- c_12:カジュアルガール
- c_13:クールガール
- c_14:ナチュラルガール
- c_15:キレイボーイ
- c_16:モードボーイ
- c_17:カジュアルボーイ
- c_18:オラオラボーイ
- i_91:キュートガール:cute1.jpg

作成

リセット

変数名	Image	平均	分散
c_11:キュートガール	3.182	1.609	
c_12:カジュアルガール	2.581	1.351	
c_13:クールガール	2.528	1.427	
c_14:ナチュラルガール	2.059	1.658	
c_15:キレイボーイ	2.968	1.391	
c_16:モードボーイ	0.912	1.074	
c_17:カジュアルボーイ	2.739	1.454	
c_18:オラオラボーイ	0.683	0.993	

Clip

図 113 ソート機能操作前

変数名	Image	平均	分散
c_11:キュートガール	3.182	1.609	
c_15:キレイボーイ	2.968	1.391	
c_17:カジュアルボーイ	2.739	1.454	
c_12:カジュアルガール	2.581	1.351	
c_13:クールガール	2.528	1.427	
c_14:ナチュラルガール	2.059	1.658	
c_16:モードボーイ	0.912	1.074	
c_18:オラオラボーイ	0.683	0.993	

Clip

図 114 ソート機能操作後

- ボーイスタイルでは、モードボーイとオレオレボーイが特に支持が低い。その意味では、ボーイスタイルの支持は、キレイボーイとカジュアルボーイとの間で、2分している。
- ガールスタイルでは、ナチュラルガールが他の3つと比較して、やや支持が低い。ガールスタイルでは、ナチュラガールが他の3スタイルとの間に差異があると推測される。

このように、ソート機能が画面上で容易に利用できる仕組みは、調査結果の解析と価値探索にとって、重要な貢献をしていることが分かる。

つぎに、8スタイルに関して、性別での比較をするために、「データ調整」機能を活用する。ここで、データ調整のボタンを押し、さらに性別項目を開いて、絞り込み項目として女性を選択する。そうすると、全体の725名から、422名の女性だけを対象とする解析が可能になる(図 115)。

データ入力 質問票 回答データ Load Save データ総数: 422人 絞り込み条件 Color 白 赤 青 黄

単純集計 クロス表 グループ化集計 クラスティング データの調整 フリーアンサー分析 GooCaBooca データのマージ (仮)

質問項目

性別は？

q_11:性別は？

q_12:年齢は？

q_13:恋人はいますか？

q_14:デートを想定した時、どちらを意識しますか？

q_15:どんな服装の女の子がモテると思いますか？

c_11:キュートガール

c_12:カジュアルガール

c_13:クールガール

c_14:ナチュラルガール

c_15:キレイボーイ

c_16:モードボーイ

絞り込み内容

削除内容

× 24:女性 q_11:性別は？

クイヤー

人数: 422

図 115 データの調整 (女性で絞込)

そこで、再度、8スタイルの単純集計をとると、422名の女子と303名の男子が、自分のデートファッションと相手のファッションについて、どのような選択していることが分かる。(図 116/図 117)

その結果、つぎのような傾向が発見できる。

- 1：全体からみて、女子と男子の好みの傾向は非常に似ている。序列をみるかぎり、女子と男子は、ほぼ同じような傾向を示している。これは意外な発見である。女子が自分が着たいファッションと男子が女子に着てほしいファッション、またその逆について、全体的傾向が同一的である、という事実の発見は、重要な価値探索である。
- 2：詳細をみると、女性は自分のクールガールを男子よりも高く支持で、男子は女子にはナチュラルガールをより期待していることが分かる。クールとナチュラルの差異は、男女のファッション嗜好と同時に恋愛スタイルへの期待についても、非常に示唆的な傾向を示しているといえよう。
- 3：オラオラボーイの支持の低さは予想を超えている。1つは、調査対象者の多くが大学生に限定されていることが影響していると解釈できる。マイルドヤンキーが新しい若者のライフスタイルとして台頭している現代事情を考慮すると、このスタイルの支持が低すぎると思える。

変数名	平均	分散
c_11:キュートガール	3.123	1.638
c_15:キレイボーイ	3.000	1.331
c_17:カジュアルボーイ	2.839	1.492
c_12:カジュアルガール	2.758	1.325
c_13:クールガール	2.709	1.378
c_14:ナチュラルガール	1.893	1.657
c_16:モードボーイ	0.872	1.015
c_18:オラオラボーイ	0.578	0.896

図 116 女子：422名

変数名	平均	分散
c_11:キュートガール	3.264	1.566
c_15:キレイボーイ	2.924	1.468
c_17:カジュアルボーイ	2.601	1.389
c_12:カジュアルガール	2.333	1.346
c_14:ナチュラルガール	2.290	1.631
c_13:クールガール	2.277	1.456
c_16:モードボーイ	0.967	1.148
c_18:オラオラボーイ	0.828	1.098

図 117 男子：303 名

このように、データ調整の「絞り込み」機能を活用して、データ比較を単純集計の段階ですることができることは、しかもその操作が非常に簡単にできることは、全体傾向を比較の視点で理解する場合、これ以上のツールはないといえる。

通常の社会調査だと、性別とのクロス集計をとって、8 スタイルを個別に比較する方法しかない。これではスタイル全体のイメージを一挙に把握することができない。それに対して、ここで「絞り込み」を容易にするデータ調整機能を活用することで、従来、困難であったことが非常に容易に可能になる。このことも、データからの発見の探索をより誘発する機能である。

さらに、この絞り込み機能を活用して、画像との関連性で比較することが可能になる。この点を説明しよう。キュートガールに限定して、全体と女子と男子について、どの画像の支持率が高いかを画像で示すと、次のようになる。(図 118：全体 / 図 119：女子 / 図 120：男子) この操作は、表示での「画像の数」を設定する（ここでは全 15 枚の画像）だけで、上位からソートされた画像が提示される。これは、上記の図 115・図 116・図 117 について、画像の数を 0 ではなく、それ以上の数字を設定するだけで、画像が貼り付けられる仕組みになっている。この操作は、従来の社会調査のツールにはありえない機能であり、価値探索の機能として第 1 級の価値をもつ。

この 3 図から、キュートガールといっても、具体的にはどのようなイメージをもったファッションなのか、かつそこでの 15 枚のファッションについて、どの画像がどの程度支持されるのか、それがヴィジュアルとして確認できる仕組みになっている。とすると、通常の数値表現だけではなく、ヴィジュアル表現がもつ具体的なキュートガールの共有理解が可能になり、解釈に深みがでてくるはずである。つまり同じキュートガールというレッテルを貼り、そこで了解された意味の具体性が問われたとき、ヴィジュアルの個別性との対応関係を明確にして、レッテルの意味の確認ができるので、そこでは、多様なキュートガールについての価値の探索が可能になるはずである。このように、「深い解釈」を求める場合には、画像対応による探索機能は重要な機能を発揮する。



図 118 全体：725 名



図 119 女子：422 名



図 120 男子：303 名



図 121 画像の対応関係

上記の画像データの集計結果について、簡単な成果をまとめると、つぎのようになる。(図 121)

- 1: 全体の傾向としては、画像にナンバーを振った順番で、支持率の序列が示されている。Cu-1 は、80% の支持率を獲得しており、Cu-15 は 43% の支持率しかない。同じキュートガールとレッテルを貼っても、その内部での支持率の幅は非常に大きい。また、この 15 枚の画像をキュートガールとしてレッテルを貼ったのは、あくまでもアプリオリの前提であり、それが構造化されているかは 15 枚の相互関連の問題であるので、クラスタリングの章で言及する。ここでは、あくまでも 15 枚の画像をキュートガールとしてレッテルを貼ると、そのかぎりでは、支持率に差異は倍近く存在し、かつ Cu-1 を最上位として、上位の画像であるほど、よりキュートガール的であると解釈するということである。
- 2: さらに性別の視点から、キュートガールの認知的支持率について差異があるかをみると、全体的な傾向として、大きな差異があるわけではない。ほぼ同じような支持率の傾向にある。その意味では、ある程度安定した構造になっていると判定できる。15 枚を上位から 5 枚ずつの 3 階層（上位・中位・下位）に設定すると、そこでの性別での移動状況を判定すると、下の図 122 のようになる。

	上位 (1-5)	中位 (6-10)	下位 (11-15)
女子	1, 2, 4, 3, 7	5, 8, 6, 10, 9	12, 13, 11, 14, 15
男子	3, 1, 6, 4, 2	5, 11, 9, 8, 10	14, 7, 12, 13, 15

※枠内の数字は、全体での序列の数字
図 122 女子、男子での画像の序列

まず全体構造との比較をすると、女子の場合は、上位と中位に 1 つずつのズレがみられる。それが、上位の Cu-7 と中位の Cu-5 である。下位にズレはない。他方、男子の場合には、上位に 1 つ (Cu-6)、中位に 2 つ (Cu-5, Cu-11)、下位に 1 つ (Cu-7) ある。

さらに、性別での比較をすると、Cu-7 が、女子では上位層に帰属するのに、男子では下位層に帰属し、評価がもっとも割れている。反対に、女子よりも男子の階層が高いのが、Cu-6 と Cu-11 である。画像で示すと、



は、より女子志向であり、



は、より男子志向になっている。同じキュートガール

でも、男女の差異は、この程度はある、ということである。このような画像があることで、具体的な差異を深く探索することが可能になっている。

5-3-2-2 クロス集計

クロス集計では、個人属性とその関連のクロス表と 8 スタイル内部のクロス表を作成する。第 1 に、個人属性について、まず性別と年齢のクロス表について検討する。年齢の単純集計 (図 123) をみると、明らかに、18 歳未満が少ないので、解析としては価値を生まないのでは、ここで、データ調整機能を使ってデータ統合 (図 124) を行う。

変数名	Image	平均	分散
q_12:年齢は？	26.732	0.660	

Clip

q_12:年齢は？

平均:26.732
分散:0.660

18未満	9	1.24%
20歳以下	254	35.03%
21～24歳	384	52.97%
25歳以上	78	10.76%

図 123 年齢の単純集計

質問項目	年齢は？
q_11:性別は？	+ - 25 18未満
q_12:年齢は？	+ - 26 20歳以下
q_13:恋人はいますか？	・25:18未満
q_14:デートを想定した時、どちらを意識しますか？	・26:18～20歳
q_15:どんな服装の女の子がモテると思いますか？	+ - 27 21～24歳
c_11:キュートガール	・27:21～24歳
c_12:カジュアルガール	+ - 28 25歳以上
c_13:クールガール	・28:25歳以上

図 124 年齢のデータ調整 (20 歳以下で変数をまとめる)

データ統合をしたあと、性別とのクロス表 (図 125) を作成する。

このデータをみると、男女の比較をする場合には、年齢の相違が有意な影響をもっているのでは、21～24 歳に限定することが必要であると思われる。男性は 25 歳以上が多く、女性の場合は 20 歳以下が多く、したがって、単純に男女比較をすると、男性年齢が相対的に高いということになる。ある程度の年齢上のバイアスがあるので、厳密には、年齢の特定化、つまり 21～24 歳への特定化が必要であろう。

Clip

q_12:年齢は？

q_11:性別は？		20歳以下	21～24歳	25歳以上		合計	平均	分散
	男性	73 (24.1%)	172 (56.8%)	58 (19.1%)		303	26.95	0.66
	女性	190 (45.0%)	212 (50.2%)	20 (4.7%)		422	26.60	0.58
	合計	263 (36.3%)	384 (53.0%)	78 (10.8%)		725	26.74	0.64

χ²乗値:56.72 自由度:2

図 125 データの調整後の年齢のクロス表

このように、データ統合することで、クロス表での相関をきちんと確認できるという意味で、データ調整機能を容易に利用できる価値は、価値探索にとっても重要である。

つぎに、ファッションスタイルの8スタイル（4 ボーイスタイル/4 ガールスタイル）と個人属性（性別と年齢）とのクロス表について検討する。

その前提として、8スタイルは、そのすべてについて、個数のカウントで構成された尺度なので、得点範囲は0から5点である。ただ、この得点分布を利用すると、バラつきが大きいので、解析の精度を高めるために、得点範囲を6段階から3段階に統合・縮約する。つまり0-1と2-3と4-5の3段階（低・中・高）に再編成する。この場合も、データ調整を活用して、データを統合・縮約する（図 126）。

データ入力 質問票 回答データ Load Save データ総数:725人 絞り込み条件

単純集計 クロス表 グループ化集計 クラスタリング データの調整 フリーアンサー分析 GooCaBooca データのマージ（仮）

質問項目

q_14:デートを想定した時、どちらを意識しますか？
q_15:どんな服装の女の子がモテると思いますか？
c_11:キュートガール
c_12:カジュアルガール
c_13:クールガール
c_14:ナチュラルガール
c_15:キレイボーイ
c_16:モードボーイ
c_17:カジュアルボーイ
c_18:オラオラボーイ
i_91:キュートガール:cute1.jpg

オラオラボーイ

+ - 0 0.0
+ - 1 0-1
・1:1
・0:0
+ - 2 2.0
+ - 3 2-3
・3:3
・2:2
+ - 4 4-5
・4:4
・5:5
+ - 5 5.0

絞り込み内容 削除内容

図 126 6段階評価から3段階に統合・縮約

その結果、性別とのクロス表を作成し、カイ二乗のスコアの高い順にソートすると、つぎのようになる。（図 127）

ここでのカイ二乗のスコアから判別できるように、1%水準で有意な違いがみられるスタイルは、8つのスタイル中4つのファッションスタイルがそれに相当する。そのうち、3つがガールスタイルで、キュートガール以外の3つで、男女差が明確である。しかし、その違いについては、カジュアルガールとクールガールについては、女性の支持が高いが、ナチュラルガールについては、反対に男性が多い。男性ファッションで有意なファッションはオラオラボーイで、これは予想できるように、男性の支持が高い。このように、デートファッションについては、それなりに性差が有効であることがわかる。



図 127 データ調整後のクロス表一覧

このように、ここでも、スタイルの得点範囲を簡単に統合・縮約するためのデータ調整機能が重要であることがわかる。クロス表を作成する場合には、クロスをかけるために適切な準備をする必要があり、そのために、データ調整機能がいかに有効であるか、がわかる。

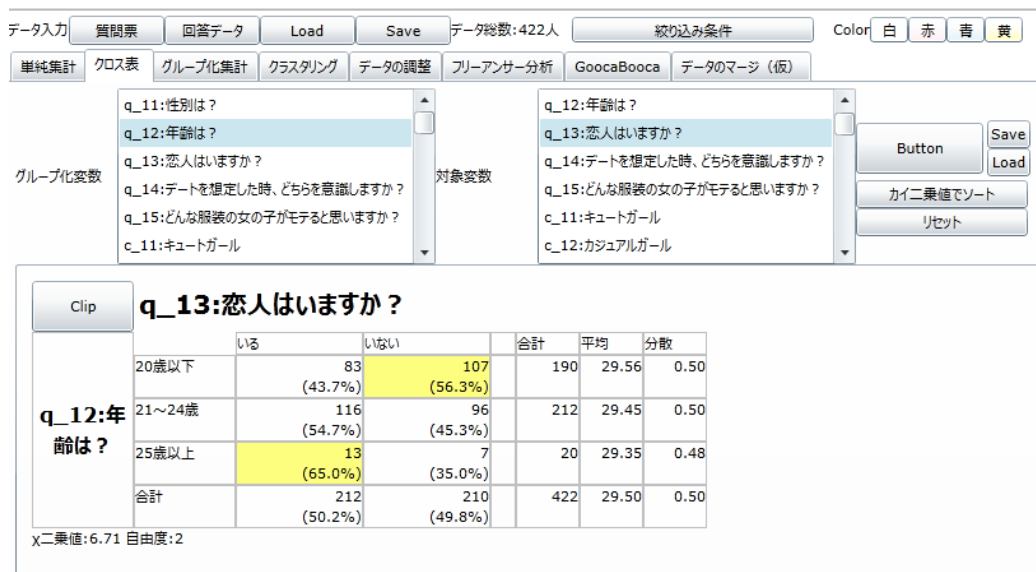


図 128 年齢と恋人の有無のクロス表

さらに、個人属性と恋愛特性とのクロス表（図 128）について言及する。ここでは生別を分けることを前提とするので、まずは、データ調整で、女性を絞り込み、その上で、年齢と恋愛経験との相関をみる。ここでは、その結果として、女性に関しては、年齢が上昇するほど、恋人のいる度合いが上昇している。それは 2.5%水準で有意である。

なお、有意な関係があるとき、図に色を付加することで関係が分かりやすくしている。このような「色つけ」の操作も、単純なことであるが、探索行為をより容易にかつ有効にする手法である。もちろん、ここでも前提条件を設定（データ調整機能による性別データの絞り込み機能）することが探索を有効化している。

第2に、女子スタイルと男子スタイルそれぞれの内部相関を検討する。ここでも、上記のように、各スタイルの得点分布の統合・縮約をして、それぞれの相関をクロス表で確認する。

まず、女子の4スタイルの相関について、個別のクロス表を検討する。すると、クールガールとナチュラルガールとの関連については相関がなく、それ以外の5つの関連性は1%水準で有意な相関を示している。クールガールとナチュラルガールは関係性が弱く、独立した傾向を示すが、それ以外の関係については強い関連性が形成されている。特に、クールガールとカジュアルガールとの相関は非常に高い。それら5つの個別の関係を図に示すと、以下の図 129、図 130、図 131、図 132、図 133 となる。

Clip	c_12:カジュアルガール						
	0-1(低)	2-3(中)	4-5(高)	合計	平均	分散	
c_13:クールガール	0-1(低)	83 (43.5%)	90 (47.1%)	18 (9.4%)	191	2.23	1.11
	2-3(中)	62 (18.6%)	185 (55.6%)	86 (25.8%)	333	2.89	0.99
	4-5(高)	22 (10.9%)	93 (46.3%)	86 (42.8%)	201	3.21	0.91
	合計	167 (23.0%)	368 (50.8%)	190 (26.2%)	725	2.80	1.07

X²乗値:94.59 自由度:4

図 129 女子の4スタイルのクロス表 1

Clip		c_13:クールガール					
c_11: キュートガ ール		0-1(低)	2-3(中)	4-5(高)	合計	平均	分散
	0-1(低)	63 (46.7%)	52 (38.5%)	20 (14.8%)	135	2.21	1.18
	2-3(中)	60 (26.3%)	120 (52.6%)	48 (21.1%)	228	2.68	1.08
	4-5(高)	68 (18.8%)	161 (44.5%)	133 (36.7%)	362	2.99	1.06
	合計	191 (26.3%)	333 (45.9%)	201 (27.7%)	725	2.75	1.13

X²乗値:55.42 自由度:4

図 130 女子の4スタイルのクロス表2

Clip		c_14:ナチュラルガール					
c_12:カ ジュアル ガール		0-1(低)	2-3(中)	4-5(高)	合計	平均	分散
	0-1(低)	101 (60.5%)	42 (25.1%)	24 (14.4%)	167	1.68	1.04
	2-3(中)	156 (42.4%)	123 (33.4%)	89 (24.2%)	368	2.06	1.18
	4-5(高)	62 (32.6%)	64 (33.7%)	64 (33.7%)	190	2.35	1.25
	合計	319 (44.0%)	229 (31.6%)	177 (24.4%)	725	2.05	1.19

X²乗値:32.55 自由度:4

図 131 女子の4スタイルのクロス表3

Clip		c_14:ナチュラルガール					
c_11: キュートガ ール		0-1(低)	2-3(中)	4-5(高)	合計	平均	分散
	0-1(低)	79 (58.5%)	39 (28.9%)	17 (12.6%)	135	1.67	0.99
	2-3(中)	107 (46.9%)	77 (33.8%)	44 (19.3%)	228	1.92	1.11
	4-5(高)	133 (36.7%)	113 (31.2%)	116 (32.0%)	362	2.27	1.25
	合計	319 (44.0%)	229 (31.6%)	177 (24.4%)	725	2.05	1.19

X²乗値:30.72 自由度:4

図 132 女子の4スタイルのクロス表4

Clip

c_12:カジュアルガール

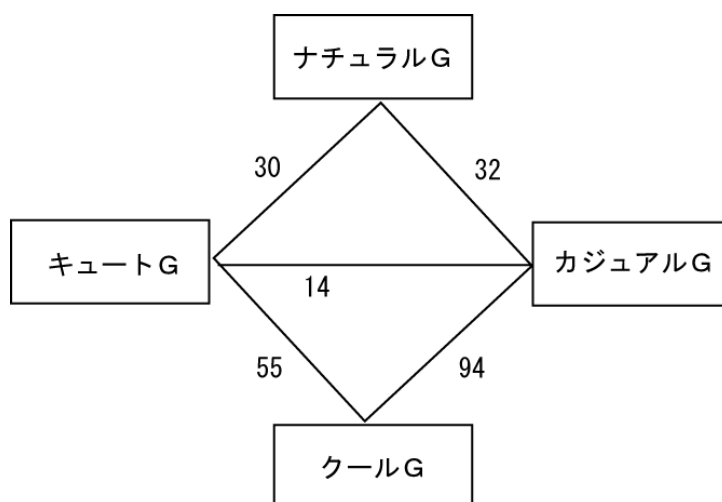
c_11:
キュートガ
ール

	0-1(低)	2-3(中)	4-5(高)		合計	平均	分散
0-1(低)	45 (33.3%)	65 (48.1%)	25 (18.5%)		135	2.52	1.13
2-3(中)	52 (22.8%)	121 (53.1%)	55 (24.1%)		228	2.79	1.05
4-5(高)	70 (19.3%)	182 (50.3%)	110 (30.4%)		362	2.92	1.03
合計	167 (23.0%)	368 (50.8%)	190 (26.2%)		725	2.80	1.07

x²乗値:14.64 自由度:4

図 133 女子の4スタイルのクロス表5

以上の関連性を全体の構造として示すと、図 134 となる。



※数字は、自由度4のカイ二乗値

図 134 女子の4スタイルの関連図

つぎに、男子スタイルについて検討する。4スタイルは、オラオラボーイとモードボーイの関連だけを除く、5つの関係について1%水準で有意な相関がみられた。下記のように、もっとも相関の高いレイボーイとカジュアルボーイの関係（図 135）を示し、男子スタイルの全体構造を図 136 に示す。

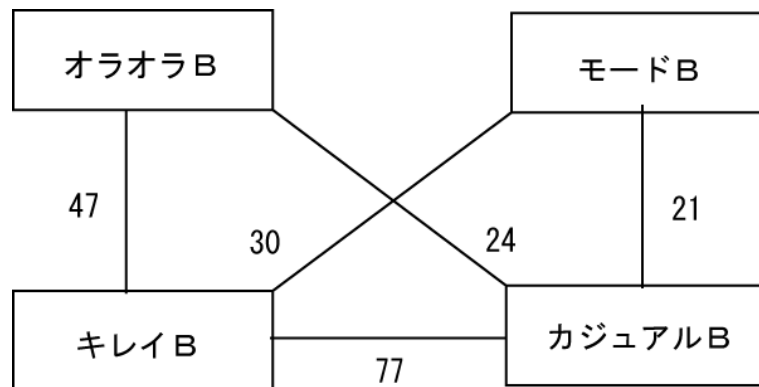
Clip

c_17:カジュアルボーイ

	0-1(低)	2-3(中)	4-5(高)	合計	平均	分散
0-1(低)	55 (43.3%)	43 (33.9%)	29 (22.8%)	127	2.02	1.16
2-3(中)	75 (24.3%)	153 (49.5%)	81 (26.2%)	309	2.28	1.10
4-5(高)	26 (9.0%)	130 (45.0%)	133 (46.0%)	289	2.83	1.11
合計	156 (21.5%)	326 (45.0%)	243 (33.5%)	725	2.46	1.16

x²乗値:77.80 自由度:4

図 135 男子の4スタイルのクロス表



※数字は、自由度 4 で有意なカイ二乗値

図 136 男子の 4 スタイルの関連図

このように、クロス表を作成することで、項目間の関係が明確になるが、そのためには、その準備としてデータ調整機能を活用して、データ統合と縮約をし、またデータの絞り込みをし、またクロス表での色づけをするという操作が重要であることがわかる。従来の社会調査では無視され、したがってクロス表の作成が自在に展開できなかったことが、このツールでは容易にできるような仕組みになっている。この価値は大きい。

5-3-2-3 クラスタリング

クロス表が作成され、女子と男子のファッションスタイルについての全体的なイメージが描けたので、ここからクラスタリングをやって、どのようなクラスが形成されるかを確定する。

まず、女子クラスを作成する。クラスタリング生成を開き、女子スタイルの 4 つを選択し、クラスタリングを始動させる。ここでは、クラス数と試行回数を設定するだけで、クラスタリングの結果が提案される。この結果をみて、研究者が納得できるならば、採用し、気に入らなければ、さらに、作成開始を何度も試行する。その場合、クラス数の変更や、試行回数を変えることも可能である。(図 137)

要は、研究者が、クロス表での解析からイメージされた全体像に適合するクラスが提案されるまで、何度も試行を繰り返す、この試行錯誤がここで一番重要なことである。従来の多変量解析ツールを使用すると、この試行錯誤を繰り返すことが前提とされていない。つまり解析ツールが提案するクラスが絶対で、研究者がそれに対して積極的に関与する仕組みになっていない。これに対して、このツールでは、研究者の思考と思惑がツールの結果と対話をして、何度も条件を変更しながら、価値を発見するまで執拗に探索を繰り返すことが期待されている。だからこそ、このクラスタリングは、価値探索において最も重要な試行過程である。

図 137 女子 4 スタイルのクラスタリング設定画面

なぜ、価値探索ツールと呼び、統計解析ツールとは呼ばないか、それは、データを媒介としてユーザと研究者との間主観的な関係の中で、新しい価値の発見を探索していこうとするからである。そのためには、データをさまざまな視点から探索するプロセスが必要である。そのとき、データの統合・縮約・削除・絞り込みなどのデータ調整を簡便に操作可能であることが不可欠である。その簡便で容易な操作可能性が確保されないかぎり、研究者の発想を豊かにすることはできない。瞬間に発想されたアイデアをすぐに操作可能にすると、新しい価値の創発が実現されるのである。そのための仕組みと仕掛けを多様にすることが、このツールを作成した根拠なのである。

クラスタリングの成果に納得すると、図 138 に示すように、クラスタタイプのネーミングと変数項目のネーミングを確定して、クラスタを登録する。具体的に、女子クラスタがどのような結果になったかを、以下に示す。

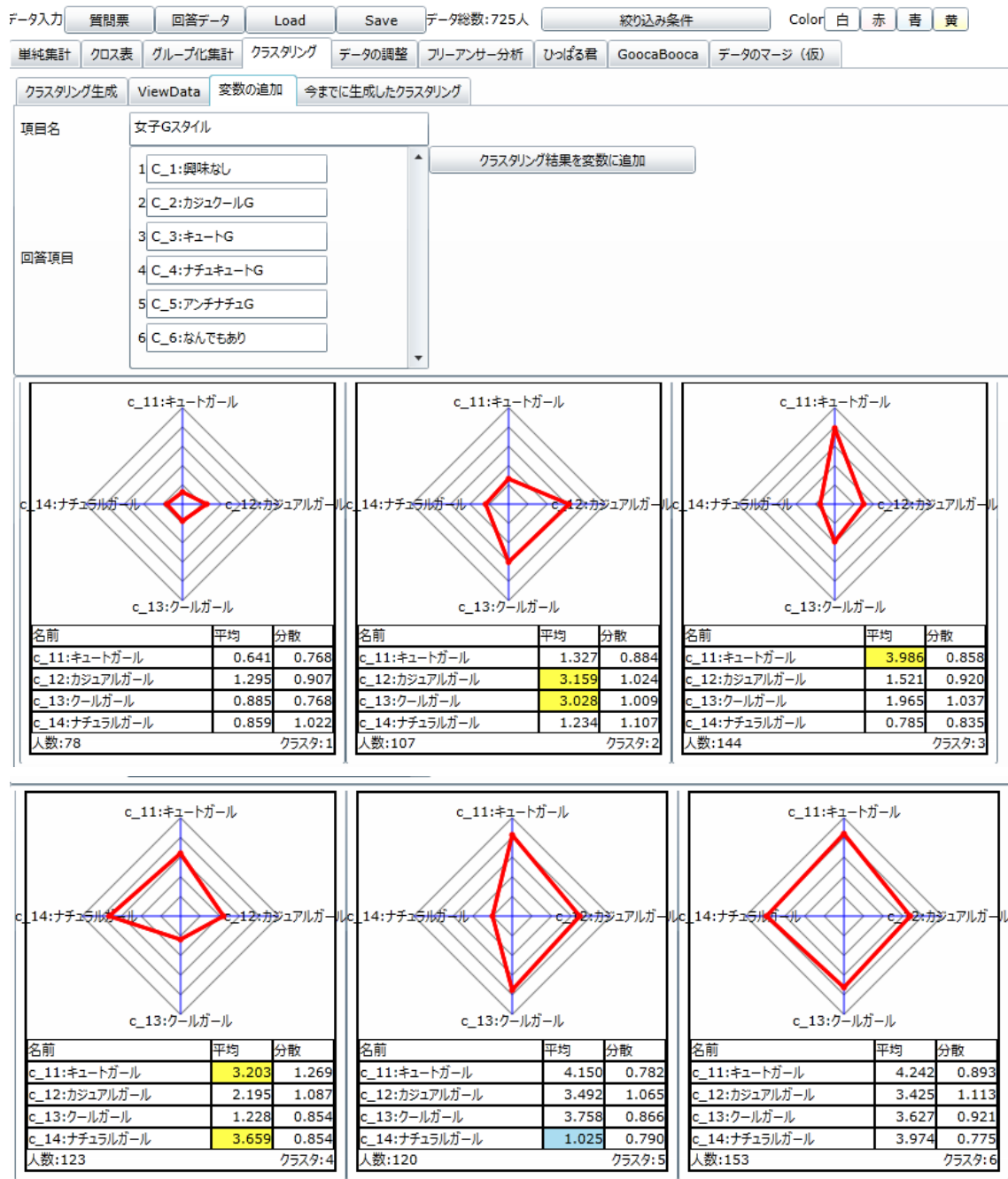


図 138 女子4スタイルによるクラスタリング結果

ここでは、6 クラスを導出した。しかし、ここでは2つのクラスは、クラスとしては価値をもたない。それが、クラス1（興味なし）とクラス6（何でもあり）である。前者は、女子のデートファッションに興味のないクラスで、どの画像にも支持がないクラスである。これは、恋愛ファッションの価値発見の意図からはずれるので、クラスの分析としては排除する。もう1つの後者も、どのようなファッション画像にも支持を示すので、クラスとしての価値はないので、これも排除する。

すると、残るのは次の4クラスである。

1: クラス2は、「カジュクールGスタイル」である。これは、クロス表でもあったように、カジュアルガールとクールガールとの相関が一番高いことを反映したクラスである。この2つのガールファッションが融合したクラスがカジュクールである。

- 2: クラスタ3は「キュートGスタイル」である。これは、キュートガールに強く特化したクラスタで、他のガールファッションにはほとんど関心をもっていない類型である。
- 3: クラスタ4は、「ナチュキュートGスタイル」である。これは、クラスタ3と対照的なポジションにあるクラスタで、ナチュラルガールとキュートガールが融合したクラスタである。キュートガールについては、それ単独で特化するスタイルと、このクラスタのように、ナチュラルガールと融合するクラスタに区分されることがわかる。
- 4: クラスタ5は、「アンチナチュラGスタイル」である。これは、ナチュラルガールには反応しないが、残りの3つのガールファッションについては強く支持を表明するクラスタである。ナチュラルガールだけはダメだ、というアンチでまとまるクラスタである。
- このようにクラスタを確定すると、キュートガールが一番高い支持率をもつ根拠は、単独（クラスタ2）、ナチュラルガールとの融合（クラスタ4）そして、このアンチナチュラルという意味でのクラスタ（クラスタ5）の3類型から構成されていることが分かる。上記のクロス表との関連でクラスタを位置づけると、下の図 139・図 140 となる。

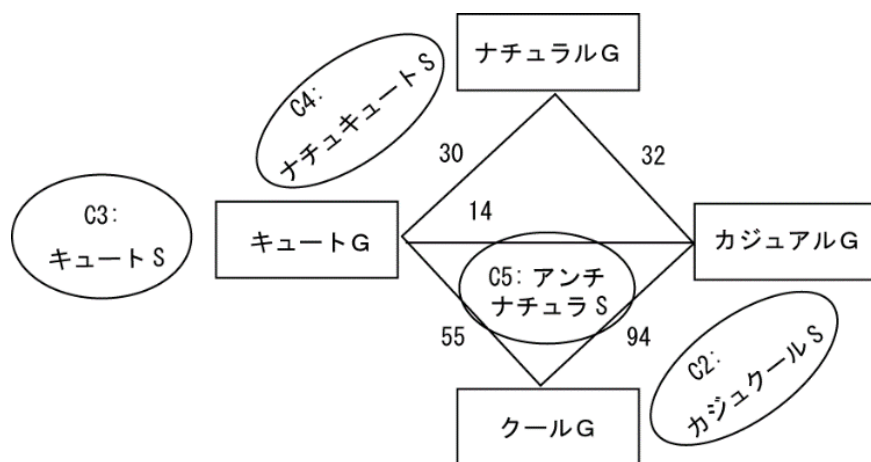


図 139 女子の4スタイルの関連図とクラスタとの関係のプロット

変数名	平均	分散
cluster_5:女子Gスタイル	3.771	1.648

Clip

cluster_5:女子Gスタイル

平均:3.771
分散:1.648

C_1:興味なし	78	10.76%
C_2:カジュアルG	107	14.76%
C_3:キュートG	144	19.86%
C_4:ナチュキュートG	123	16.97%
C_5:アンチナチュG	120	16.55%
C_6:なんでもあり	153	21.10%

図 140 クラスタの人数分布

このように、クラスタリングでは、クラスタの確定過程において強く探索的であり、また、設定後にすぐに他の項目との関係性が確認し、さらに新しい価値の探索を開始できることも、このツールの探索

性を示す特性である。では、このクラスと他の項目との関連性を探索することにする。
男子のクラスタリングの結果を示すと、以下ようになる。(図 141 図 142 図 143)

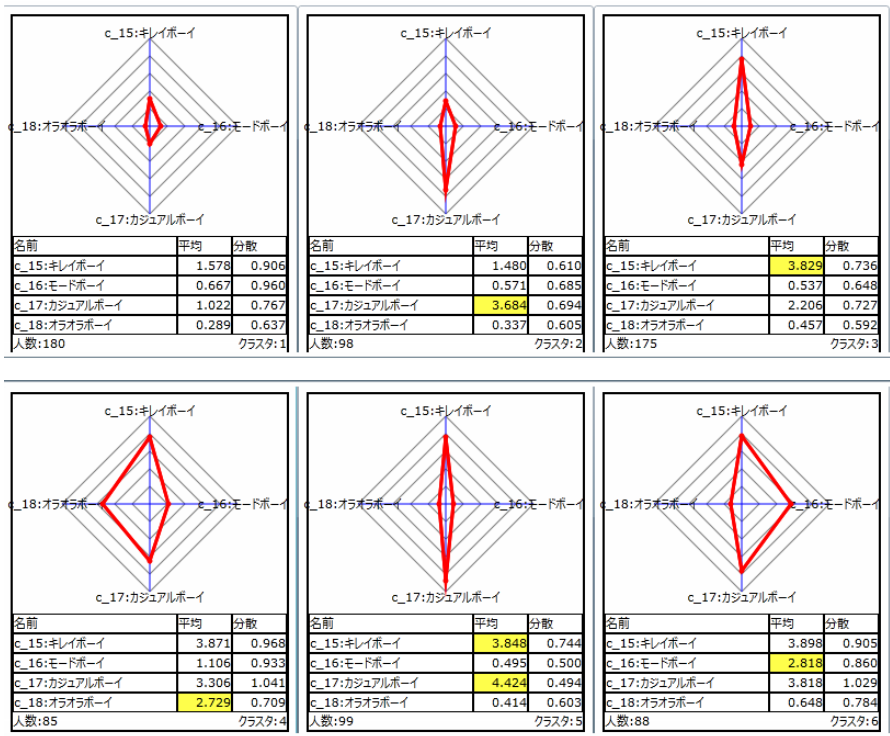


図 141 男子 4 スタイルによるクラスタリング結果



図 142 クラスタリング結果の人数分布

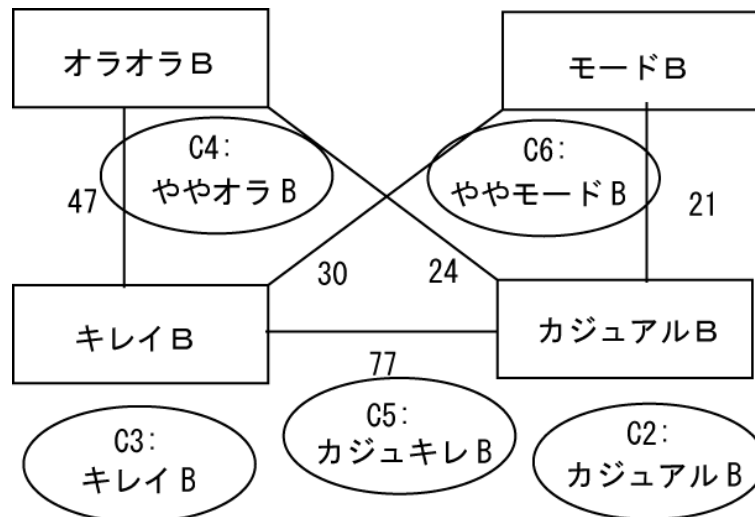


図 143 男子の4スタイルの関連図とクラスタとの関係のプロット

男子クラスタは、興味なしを除くと、意味あるクラスタが5つ作成された。ここでは、クロス表で分かっていたように、支持率において、ケイ B とカジュアル B の支持率は高く、対照的に、Ora Ora B とモード B の支持率は低いという2分化の状態にあったので、クラスタリングも、その状況をふまえた構成になっている。なので、男子スタイルのクラスタは、ケイ志向がカジュアル志向か、を基本として構成されている。

つぎに、クラスタと他項目との相関をみる。

まず、性別との相関をみる。そのためには、年齢分布と揃えることと、無駄なクラスタを排除する。データ調整機能を活用して、絞り込みと削除を図 144 のようにする。

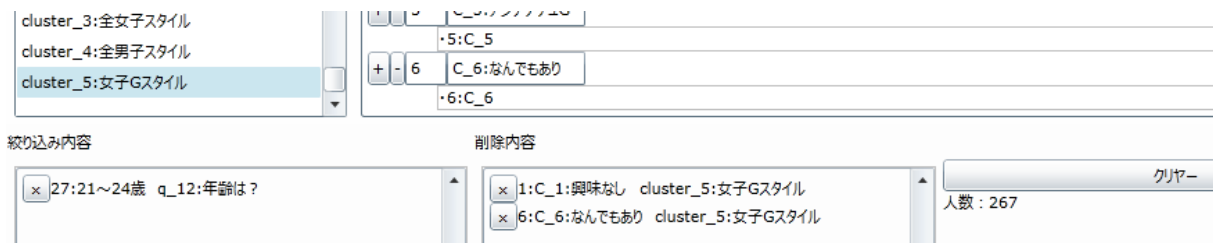


図 144 年齢での絞り込みと、無駄なクラスタの排除

以上の操作をもとに、性別とクラスタの相関をみると、男女の違いが非常に明確になる。

男子 (115 名) はナチュラルキュートに特化し、女性 (152 名) は、カジュクールとアンチナチュラルに特化する傾向が明確になる。20 代前半に限定する限り、男子はナチュラル系の女子のファッションを好み、それに対して、女子はそれを拒否して、カジュクールのクラスタに好みを鮮明にしている。その限りでは、男女にミスマッチが生じているともいえよう。(図 145)

Clip		q_11:性別は？				
cluster_5:女子Gスタイル		男性	女性	合計	平均	分散
	C_2:カジュアル	23 (32.9%)	47 (67.1%)	70	23.67	0.47
	C_3:キュートG	29 (39.7%)	44 (60.3%)	73	23.60	0.49
	C_4:ナチュキュー	42 (68.9%)	19 (31.1%)	61	23.31	0.46
	C_5:アンチナチュ	21 (33.3%)	42 (66.7%)	63	23.67	0.47
	合計	115 (43.1%)	152 (56.9%)	267	23.57	0.50

χ²乗値:22.28 自由度:3

図 145 女子スタイルクラスと性別とのクロス表

同様に、男子スタイルのクラスタについても、性別との相関をみると、つぎのようになる。(図 146) ここでの絞り込み条件は、年齢の統一 (21-24 歳) と男子クラスタ C1 の削除である。よって、293 名を対象とし、男子が 129 名、女子が 164 名である。

Clip		q_11:性別は？				
cluster_4:全男子スタイル		男性	女性	合計	平均	分散
	C_2:カジュアルB	22 (41.5%)	31 (58.5%)	53	23.58	0.49
	C_3:キレイB	35 (38.5%)	56 (61.5%)	91	23.62	0.49
	C_4:ややオラオラ	29 (65.9%)	15 (34.1%)	44	23.34	0.47
	C_5:カジュキレイB	20 (34.5%)	38 (65.5%)	58	23.66	0.48
	C_6:ややモードB	23 (48.9%)	24 (51.1%)	47	23.51	0.50
	合計	129 (44.0%)	164 (56.0%)	293	23.56	0.50

χ²乗値:12.43 自由度:4

図 146 男子スタイルクラスとクロス表

ここでは、女子の場合ほど、高い相関はない (2.5%水準で有意) が、それなりの違いはある。女子は、男性スタイルについての知識が乏しいのか、基本のクラスタであるキレイ B スタイルとカジュアル B スタイルを強く支持し、それに対して、男子はややオレ B とややモード B を支持している。

つぎに、クラスタと年齢との相関を確認する。この場合、年齢について、25 歳以上が少ないので、データ調整機能を活用して、年齢階層の統合を行う (図 147)。

質問項目

q_11:性別は？
q_12:年齢は？
q_13:恋人はいますか？
q_14:デートを想定した時、どちらを意識しますか？
q_15:どんな服装の女の子がモテると思いますか？
c_11:キュートガール
c_12:カジュアルガール
c_13:クールガール
c_14:ナチュラルガール
c_15:キレイボーイ
c_16:モードボーイ

年齢は？

+ - 25 18未満
+ - 26 20歳以下
+ - 27 21歳以上
+ - 28 25歳以上

絞り込み内容

× 24:女性 q_11:性別は？

削除内容

× 1:C_1:興味なし cluster_5:女子Gスタイル
× 6:C_6:なんでもあり cluster_5:女子Gスタイル

人数：292

クリアー

図 147 年齢の統合（20 歳以上、以下の 2 つにする）

その結果、女子にとっての、クラスと年齢との相関は、つぎのようになる（図 148）。

Clip		q_12:年齢は？					
		20歳以下	21歳以上	合計	平均	分散	
cluster_5:女子Gスタイル	C_2:カジュクール	20 (27.8%)	52 (72.2%)	72	26.72	0.45	
	C_3:キュートG	40 (47.1%)	45 (52.9%)	85	26.53	0.50	
	C_4:ナチュキュー	30 (60.0%)	20 (40.0%)	50	26.40	0.49	
	C_5:アンチナチュ	38 (44.7%)	47 (55.3%)	85	26.55	0.50	
	合計	128 (43.8%)	164 (56.2%)	292	26.56	0.50	

X二乗値:13.23 自由度:3

図 148 女子スタイルクラスと年齢のクロス表（女子限定）

年齢を統合してカテゴリを縮約したことで、関係が鮮明になる。ここでは、低年齢（20 歳以下）ほど、ナチュキュートスタイルを志向し、20 歳を超えると、多くがカジュクールクラスを求める、という関連が明確になる（2.5%水準で、有意）。

同様に、男子に限定して、男子スタイルでの関係を求めたが、有意な相関はみられなかった（図 149）。

Clip		q_12:年齢は？						
cluster_4:全男子スタイル		20歳以下	21歳以上	25歳以上		合計	平均	分散
	C_2:カジュアルB	4 (10.3%)	22 (56.4%)	13 (33.3%)		39	27.23	0.62
	C_3:キレイB	20 (29.4%)	35 (51.5%)	13 (19.1%)		68	26.90	0.69
	C_4:ややオラオラ	10 (21.3%)	29 (61.7%)	8 (17.0%)		47	26.96	0.62
	C_5:カジュアルB	8 (25.0%)	20 (62.5%)	4 (12.5%)		32	26.88	0.60
	C_6:ややモードB	8 (22.9%)	23 (65.7%)	4 (11.4%)		35	26.89	0.57
	合計	50 (22.6%)	129 (58.4%)	42 (19.0%)		221	26.96	0.64

X²乗値:11.31 自由度:8

図 149 男子スタイルクラスと年齢のクロス表（男子限定）

以上、クラスタリングに関連した価値探索過程を検討してきた。ここでも、多様な操作が容易に可能なことが、いかにして探索行為を価値創発的に示すことができた。

5-2-3-4 グループ化集計

グループ化集計は、ある項目を説明するには、どのような項目が有効なのかを探索する仕組みである。ここでは、女子クラスターの4スタイルについて、女性に限定すると、各スタイルの選択は、①年齢なのか、②恋人の有無なのか、それとも③デート時の服装選択（自分好みの服か、それともモテそうな服か）の3項目のどの項目が説明項目として、より有効であるか、を探索するツールである。まず、条件を整えるために、データ調整機能を活用する（図 150）。

The screenshot shows a software interface for data adjustment. On the left, a list of clusters is shown: 'cluster_4:全男子スタイル' and 'cluster_5:女子Gスタイル'. The 'cluster_5:女子Gスタイル' is selected. In the center, there are two panels: '絞り込み内容' (Filtering Content) and '削除内容' (Deletion Content). The '絞り込み内容' panel shows a condition: '24:女性 q_11:性別は？'. The '削除内容' panel shows two conditions: '1:C_1:興味なし cluster_5:女子Gスタイル' and '6:C_6:なんでもあり cluster_5:女子Gスタイル'. On the right, there is a 'クリアー' (Clear) button and a display showing '人数: 292' (Number of people: 292).

図 150 絞り込み条件の指定

女子の292名について、グループ集計機能を活用すると、つぎのような画面が出る。その結果、各クラスターごとにソート機能をかけると、各クラスターは何によって説明されているか、が判明する。

カジュアルGスタイルの場合（図 151）

※今回の場合、平均・分散・偏差値の項は無意味なので、無視する。



図 151 年齢、恋人、服でグループ化集計（カジュクールでソート）

ここでは、カジュクールのクラスタに帰属する割合が高い順にソートし、それを説明する3項目の状態をチェックすると、このクラスタの場合には、「自分の好みの服」の項目が上位3までで同じ支持を獲得していることが、わかる。

この説明項目のパターンは12パターン（年齢3、恋人の有無2、デート服2： $3 \times 2 \times 2 = 12$ ）で、そのパターンに共通の項目があれば、それが説明要因として機能しているかもしれない。カジュクールの場合、自分好みの服が共通項目なので、それがこのクラスタの説明には重要であると推測される、ということである。

キュート G スタイルの場合（図 152）

対象: cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
26:20歳以下	29:いる	32:モテそうな服	9	3.889	0.875	53.3	0 (0.00%)	4 (44.44%)	2 (22.22%)	3 (33.33%)
26:20歳以下	29:いる	31:自分好みの服	45	3.578	0.954	50.6	5 (11.11%)	19 (42.22%)	11 (24.44%)	10 (22.22%)
27:21歳以上	29:いる	32:モテそうな服	24	3.417	1.077	49.2	5 (20.83%)	10 (41.67%)	3 (12.50%)	6 (25.00%)
26:20歳以下	30:いない	32:モテそうな服	21	3.762	1.109	52.2	3 (14.29%)	7 (33.33%)	3 (14.29%)	8 (38.10%)

図 152 年齢、恋人、服でグループ化集計（キュート G スタイルでソート）

キュート G スタイルでは、「恋人がいる」項目が有効な説明項目として機能していると思われる。他のクラスタとの比較からすると、キュート G スタイルを志向する女子には、恋人がいるという状況が、そのようなファッションスタイルを選考させているのかもしれない。

ナチュキュート G スタイルの場合（図 153）

対象: cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
26:20歳以下	30:いない	31:自分好みの服	53	3.679	1.146	51.5	12 (22.64%)	10 (18.87%)	14 (26.42%)	17 (32.08%)
26:20歳以下	29:いる	31:自分好みの服	45	3.578	0.954	50.6	5 (11.11%)	19 (42.22%)	11 (24.44%)	10 (22.22%)
26:20歳以下	29:いる	32:モテそうな服	9	3.889	0.875	53.3	0 (0.00%)	4 (44.44%)	2 (22.22%)	3 (33.33%)
27:21歳以上	30:いない	31:自分好みの服	51	3.235	1.165	47.6	19 (37.25%)	12 (23.53%)	9 (17.65%)	11 (21.57%)

図 153 年齢、恋人、服でグループ化集計（ナチュキュート G スタイルでソート）

ナチュキュート G スタイルの場合には、同じキュート系ではあっても、ナチュラル系が強くでているクラスタなので、ここでは年齢の若さ（20歳以下）が有効な項目として機能している。

アンチナチュ G スタイルの場合（図 154）

対象:cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
28:25歳以上	30:いない	32:モテそうな服	2	5.000	0.000	63.0	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (100.00%)
27:21歳以上	30:いない	32:モテそうな服	12	4.000	1.225	54.3	2 (16.67%)	3 (25.00%)	0 (0.00%)	7 (58.33%)
28:25歳以上	30:いない	31:自分好みの服	4	3.750	1.299	52.1	1 (25.00%)	1 (25.00%)	0 (0.00%)	2 (50.00%)
26:20歳以下	30:いない	32:モテそうな服	21	3.762	1.109	52.2	3 (14.29%)	7 (33.33%)	3 (14.29%)	8 (38.10%)

図 154 年齢、恋人、服でグループ化集計（アンチナチュ G スタイルでソート）

アンチナチュ G スタイルの場合には、「恋人がいない」ことが有効な機能を果たしているようだ。これは、同じキュート系のスタイルを共有させているとしても、キュート系に特化したキュート G スタイルの説明項目が「恋人がいる」というのとは対照的である。また、このクラスはアンチナチュラルであるから、ナチュラキュート G スタイルの説明項目が低年齢であるの対照的に、高年齢の特性を共有するかと予想されるが、そのような結果にはなっていない。恋人がいない、という年齢とも相関がありそうな、しかし年齢ではない別の項目が、このクラスを有効に説明している。

このように、グループ集計化は、説明項目を探索する機能をもつ点で重要なボタンである。現象を理解する場合、その発想の原点である「なぜ」に対して、それなりの探索を容易にする、このグループ化集計機能は価値探索にとって重要である。

5-3-3 結論

この調査の目的は、服装ファッションから男女の恋愛パターンを探索することであるので、最後に、女子と男子の相性を発見することにする。

まず、女性に限定し、かつクラスタについては、女子クラスタでは、「興味なし」と「なんでも」の2クラスタを、データ調整機能で削除し、男子クラスタも「興味なし」クラスタは削除した。その結果、下記のような関連性が導出された（図 155）。

Clip		cluster_4:全男子スタイル								
cluster_5:女子Gスタイル		C_2:カジュアルB	C_3:キレイB	C_4:ややオラオラB	C_5:カジュキレB	C_6:ややモードB	合計	平均	分散	
	C_2:カジュクール	21 (36.8%)	14 (24.6%)	2 (3.5%)	11 (19.3%)	9 (15.8%)	57	3.53	1.52	
	C_3:キュートG	10 (16.1%)	34 (54.8%)	8 (12.9%)	10 (16.1%)	0 (0.0%)	62	3.29	0.92	
	C_4:ナチュキユー	6 (15.8%)	10 (26.3%)	7 (18.4%)	10 (26.3%)	5 (13.2%)	38	3.95	1.30	
	C_5:アンチナチュ	7 (9.6%)	22 (30.1%)	8 (11.0%)	13 (17.8%)	23 (31.5%)	73	4.32	1.42	
	合計	44 (19.1%)	80 (34.8%)	25 (10.9%)	44 (19.1%)	37 (16.1%)	230	3.78	1.38	
	x二乗値:50.88 自由度:12									

X²乗値:50.88 自由度:12

図 155 女子スタイルと男子スタイルのクロス表（女性限定）

女性からすると、

- ①カジュクール G スタイルは、カジュアル B スタイルを好む。
- ②キュート G スタイルは、キレイ B スタイルを好む。
- ③ナチュラキュート G は、相対的には、カジュキレ B スタイルを好む。

相対的の意味は、キレイ B スタイルと同じ支持であるが、②と比較すると、カジュキレ B との意味的な相関が重視される、ということである。

④アンチナチュ G スタイルは、ややモード B スタイルを好む。

⑤ややオラ B スタイルに強く共感するガールスタイルのクラスタはない。

このように、女子からすると、女子クラスタから男子クラスタへの関連性は非常に明確である。

これにたいして、男子で判定すると、上記のクラスタ削除条件を同様にしてみると、つぎのような関連性がみられる（図 156）。

Clip		cluster_5:女子Gスタイル				合計	平均	分散
		C_2:カジュールG	C_3:キュートG	C_4:ナチュキュートG	C_5:アンチナチュG			
cluster_4:全男子スタイル	C_2:カジュアルB	3 (9.4%)	10 (31.3%)	12 (37.5%)	7 (21.9%)	32	3.72	0.91
	C_3:キレイB	10 (20.4%)	16 (32.7%)	19 (38.8%)	4 (8.2%)	49	3.35	0.89
	C_4:ややオラオラ	3 (10.7%)	6 (21.4%)	7 (25.0%)	12 (42.9%)	28	4.00	1.04
	C_5:カジュキレB	4 (25.0%)	1 (6.3%)	10 (62.5%)	1 (6.3%)	16	3.50	0.94
	C_6:ややモードB	1 (4.5%)	4 (18.2%)	9 (40.9%)	8 (36.4%)	22	4.09	0.85
	合計	21 (14.3%)	37 (25.2%)	57 (38.8%)	32 (21.8%)	147	3.68	0.97

χ²乗値:26.78 自由度:12

図 156 男子スタイルと女子スタイルのクロス表（男子限定）

ここでは、カイ二乗検定では 1%水準で有意ではあるが、トップの関連性をみると、4 つの男子クラスタがナチュキュート G スタイルに集中しており、クラスタの解析としては期待したものになっていない。そこで、男子にかぎってデータをさらに絞り込むことにする。それは、年齢層の 20 歳以下を削除してみることである。その結果は、図 157 となる。

Clip		cluster_5:女子Gスタイル				合計	平均	分散
		C_2:カジュールG	C_3:キュートG	C_4:ナチュキュートG	C_5:アンチナチュG			
cluster_4:全男子スタイル	C_2:カジュアルB	3 (10.3%)	10 (34.5%)	10 (34.5%)	6 (20.7%)	29	3.66	0.92
	C_3:キレイB	8 (22.9%)	7 (20.0%)	17 (48.6%)	3 (8.6%)	35	3.43	0.93
	C_4:ややオラオラ	1 (4.8%)	5 (23.8%)	5 (23.8%)	10 (47.6%)	21	4.14	0.94
	C_5:カジュキレB	4 (33.3%)	1 (8.3%)	7 (58.3%)	0 (0.0%)	12	3.25	0.92
	C_6:ややモードB	1 (5.3%)	4 (21.1%)	6 (31.6%)	8 (42.1%)	19	4.11	0.91
	合計	17 (14.7%)	27 (23.3%)	45 (38.8%)	27 (23.3%)	116	3.71	0.98

χ²乗値:28.32 自由度:12

図 157 男子スタイルと女子スタイルのクロス表（男子限定、20 歳以上）

ここで、やっと関連性がやや拡散する。関係性を確定すると、以下のようなになる。

①カジュアル B スタイルは、キュート G スタイルを好む。

②キレイ B スタイルとカジュアル B スタイルは、ナチュキュート G スタイルを好む。

③ややオラ B スタイルとややモード B スタイルは、アンチナチュ G スタイルを好む。

④カジュール G スタイルを好む男子クラスタはない。

そこで、上記の 2 図（図 155・図 157）から、黄色の関係を図化すると、つぎの図 158 のようにな

る。

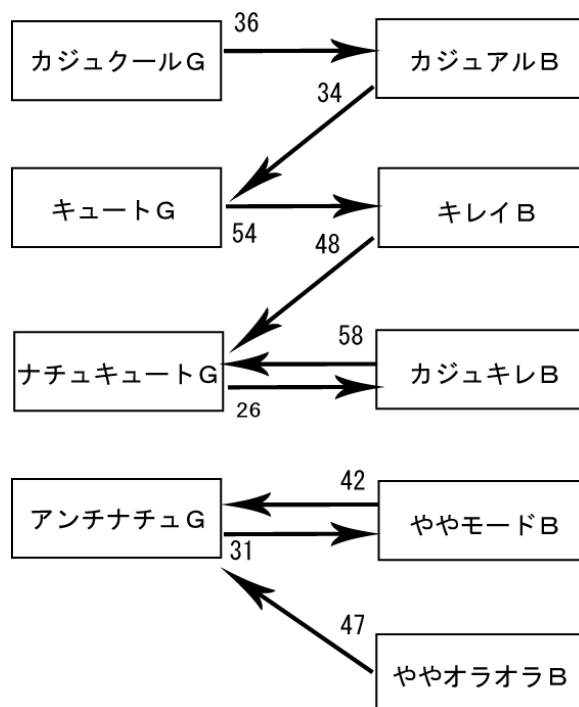


図 158 図 143・図 145 から、黄色の関係を図化

このように、男女のクラスタ間の関連性は、2つのクラスタでの相思相愛のカップルが形成される。それが、ナチュキュートGスタイルとカジュキレイBスタイルの関連性であり、またアンチナチュGスタイルとややモードBスタイルの関連性である。

それ以外のクラスタについては、さらに、完全な片思いのクラスタとすれ違いのクラスタが存在する。片思いクラスタは、女子ならばカジュアルGスタイルであり、男子ならばややオラBスタイルである。後者のすれ違いクラスタは、女子ならばキュートクラスタであり、男子ならばカジュアルBスタイルとキレイBスタイルの2つである。このように、ファッションを媒介として、男女の相性を探索すると、以上のような関連性が探索された。

通常の社会調査では、ここまでの結論で終わるが、ここでは、このような関係性がなぜ成立しているのかを、さらに画像を活用することで、さらに深堀をすることができる。そこで、各クラスタの特性を画像で確認する。以下、女子クラスタに限定して検討する。

①カジュアルGスタイルの場合（21名）

カジュアルGスタイルで、カジュアルBスタイルを好む21名の女子がどのようなファッションを好んでいるかを画像で確認する。絞り込みの検索は、図159になる。



図 159 女性限定で、カジュアルG、カジュアルBで絞り込み
その結果の画像は、つぎのようになる（図160）。

変数名	平均	分散	
c_17:カジュアルボーイ	3.619	0.653	 100.00%  100.00%  100.00%  100.00%  88.89%  81.82%  80.00%
c_13:クールガール	3.429	0.955	 100.00%  100.00%  88.89%  87.50%  83.33%  75.00%  71.43%
c_12:カジュアルガール	3.048	1.090	 85.71%  83.33%  81.82%  80.00%  77.78%  75.00%  75.00%
c_15:キレイボーイ	1.524	0.499	 85.71%  66.67%  66.67%  60.00%  60.00%  60.00%  28.57%
c_11:キートガール	1.190	0.852	 60.00%  57.14%  42.86%  40.00%  33.33%  28.57%  25.00%
c_14:ナチュラルガール	1.000	1.113	 57.14%  40.00%  40.00%  33.33%  33.33%  30.00%  20.00%
c_16:モードボーイ	0.714	0.628	 41.67%  25.00%  25.00%  22.22%  16.67%  14.29%  12.50%
c_18:オラオラボーイ	0.190	0.393	 33.33%  22.22%  16.67%  0.00%  0.00%  0.00%  0.00%

図 160 図 159 の条件に基づく画像の集計

上記の画像が、カジュアル G スタイルの女子 21 名が選考した画像である。各クラスタでの画像の平均をみれば、上位 3 つのスタイル（男子のカジュアルボーイ：3.6、女子のクールガール：3.4、カジ

ュアルガール：3.0）が圧倒的に支持されており、個別の画像をみても、この3つのクラスタについては、最低の支持率がクールガールの7番目の画像の71%である。この数字は、それ以外のクラスタの数字と比較してみると、いかに高い数字であるかが分かる。キレイボーイの1番の画像が85%で、これのみが最低の画像を超えている。このように、カジュークールGスタイルが好む、自分と彼の恋服がどのようなファッションであるのか、がこれによって具体的に提示されている。

以下、クラスタについて、特徴的な画像を提示する。

②キュートGスタイルの場合（34名）（図161）

変数名	平均	分散							
c_11:キュートガール	4.235	0.730							
			100.00%	100.00%	100.00%	100.00%	100.00%	92.86%	92.31%
c_15:キレイボーイ	3.647	0.681							
			100.00%	100.00%	92.31%	91.67%	90.91%	90.00%	86.67%

図161 キュートGスタイルで絞込での集計

キュートGスタイルは、キレイBスタイルを好むので、そこに相当する34名について画像を検索すると、非常に明確な支持率が得られる。平均が高いのは、この2つのクラスタのみであり、個別の支持率も圧倒的に高い。キュートGスタイルが、いかにキュートガールのファッションに特化しているか、それは、上位7位であっても、90%を超える支持率を示していることでもわかる。

③ナチュキュートGスタイルの場合（10名）：（図162）

変数名	平均	分散	
c_17:カジュアルボーイ	4.300	0.458	
c_14:ナチュラルガール	3.800	0.748	
c_15:キレいボーイ	3.500	0.500	
c_11:キュートガール	3.200	1.400	

図 162 ナチュキョート G スタイルで絞り込み集計

データが 10 名なので、不安定であることは承知の上で、上記のような結果が提示されている。ナチュラルガールについての志向がいわゆる森ガールの雰囲気をもつガールファッションであり、それがキュートガールを融合すると、単純平均であつたら序列とはかなり異なったパターンがキュートガールにおいてもみられる。ここに、ナチュキョート G スタイルの個性がみてとれよう。男子スタイルの嗜好についても、カジュアルボーイの序列は、カジュクール G スマイルが好むカジュアルボーイとはやや異質なパターンになっており、同じカジュアルボーイを好むといっても、その意味内容は異なるのであろう。

④アンチナチュラル G スタイルの場合 (23 名) : (図 163)

アンチナチュラル G スタイルは、ややモード B スタイルを好む、という関連性であるが、画像をみればわかるように、なぜ「ややモード」とネーミングしているかの理由が明確である。この女子は、モードスタイルを一番支持しているわけではなく、その意味での一番はカジュアルボーイであり、二番はキレいボーイであり、三番目がモードボーイである。しかし他のクラスとよ比較で、特化しているのがモードボーイである、という意味で、ややモード B スタイルは納得できるものである。

その前提をもとに、画像をみていくと、女子クラスタのアンチナチュラルについても、カジュアル・キュート・クールガールの意味あいも、上記のスタイルとは、画像の選択において、それなりに異質なパターンを示している。

変数名	平均	分散	
c_12:カジュアルガール	4.000	0.933	 100.00%  100.00%  100.00%  100.00%  100.00%  100.00%  100.00%
c_17:カジュアルボーイ	4.000	1.180	 100.00%  100.00%  90.00%  90.00%  88.89%  85.71%  81.82%
c_11:キートガール	3.913	0.830	 100.00%  100.00%  100.00%  100.00%  100.00%  90.91%  90.00%
c_15:キレイボーイ	3.913	0.830	 100.00%  100.00%  100.00%  90.00%  88.89%  87.50%  85.71%
c_13:クールガール	3.826	0.701	 100.00%  100.00%  100.00%  100.00%  100.00%  90.00%  87.50%
c_16:モードボーイ	2.696	0.906	 100.00%  100.00%  88.89%  85.71%  83.33%  66.67%  57.14%

図 163 アンチナチュラル G スタイルで絞り込み集計

以上、画像とクラスタの詳細を検討した。さらに、クラスタとフリーアンサーとの関係も、クラスタの詳細を知るには、重要である。従来の社会調査では、フリーアンサーの扱いは、つねに別扱いであり、本筋の添え物としての役割しか担ってこなかった。しかし、ふかぼりくんでは、データの深堀には、画像と並んで自由回答もまた重要な使命をもっている。そこで、以下、上記の4クラスタとの関連で、どのような自由回答の解釈が可能であるかを検討する。

①カジュクール G スタイル (72 名)

まず、データの取得は、図 164 に示すように、データの調整機能のもとで、絞り込みを実行する。

cluster_5:女子Gスタイル

+ - 6 C_6:なんでもあり

・6:C 6

絞り込み内容

削除内容

人数 : 72

× 24:女性 q_11:性別は?

× 2:C_2:カジュアルG cluster_5:女子Gスタイル

図 164 絞り込み条件（女性限定、カジュアル G）

つぎに、女子のカジュアル G スタイルの 72 名が、どのようなファッションが大切であるかをフリーアンサーで語っている。それが単純集計の成果として示される（図 165）。

変数名	平均	分散
q_15:どんな服装の女の子がモテと思いますか？	0.000	0.000

Clip

q_15:どんな服装の女の子がモテと思いますか？

平均:0.000
分散:0.000

他人にどう見られるか意識した服装。	1	1.39%
自分に似合う服をきこなしてこ	1	1.39%
ふわふわワンピース	1	1.39%
(回答なし)	18	25.00%
小綺麗	1	1.39%
シンプルにさらっとした着こなし。	1	1.39%
ある程度露出がある服と、甘辛な服	1	1.39%
かわいらしい、上品な女性らしいやわらかい印象を与える服装 みゆたん頑張れ～（ミユヅミ）	1	1.39%
膝丈スカート 体のラインみえる デコルテをだすか、足を出す。露出のバランス	1	1.39%
甘すぎず、カジュアルすぎず	1	1.39%
ふわふわ系 ワンピ	1	1.39%

自分に合っている服	1	1.39%
ぶりぶりすぎず、ちょっと女の子っぽい服装	1	1.39%
シンプルで自分に似合っている服をわかる人。	1	1.39%
コンサバ系	1	1.39%
かっこいい自分らしい服	1	1.39%
モテようと思っていない服装	1	1.39%
メスっぽい服。スカートにヒールなど。	1	1.39%
優しい印象を与える 胸元がゆるい	1	1.39%

リア充男子→アンサンブルユニットを着ているJJ系女子/ビビッドな色やら流行を取り入れたり海外セレクトとか好きそうなViVi系女子 サバカル系	1	1.39%
自分らしさが感じられる服装。	1	1.39%
デコルテがあいてデコルテ清潔で 似合う服をきてるこ	1	1.39%
かわいらしい、女の子らしい服装	1	1.39%
カジュアルで清潔感のある服装	1	1.39%
体のラインが出ている服	1	1.39%
シンプル	1	1.39%
カッパ可愛い服。	1	1.39%
シンプル！	1	1.39%
モテるの概念がわからないけど、きめすぎない服装が支持されそーな気がします。	1	1.39%
女の子らしい服装。スカートとか。	1	1.39%
はやりばかりにとらわれないで自分の顔や雰囲気にあう洋服をきている子	1	1.39%
緊張しない服装	1	1.39%
綺麗だけど個性がある。シンプルでも色や形が綺麗な服の子。	1	1.39%
派手すぎないけど女性らしい格好、女性らしいラインや細い部分がアクセントになるコーディネート	1	1.39%
おしゃれ	1	1.39%
JJ系	1	1.39%
エレガント	1	1.39%
シンプル 白黒とか	1	1.39%
だらしくなくて、自分に似合うものが分かってる	1	1.39%

シンプルかつセクシー！	1	1.39%
流行を追うだけでなく、自分の好きな服を選んで着ている子。自分の目指す、あるいは自分に合う雰囲気をはかっている子。でも奇抜すぎ	1	1.39%
自分を主張しすぎない服	1	1.39%
自分の体型に似合うことを分かっている女の子	1	1.39%
きれいめファッション	1	1.39%
スカート履いてる子、ある程度女らしさが出ている服	1	1.39%
自分にあった服をえらべる子だとおもいます。	1	1.39%
かわいいけど、かわいすぎない女らしさがある服装	1	1.39%
	1	1.39%
普通、個性的でない服	1	1.39%
似合うもの	1	1.39%
その人に合ってる服	1	1.39%
ふわふわした感じ	1	1.39%
自分に合った服を着てる人。	1	1.39%
ややコンサバ	1	1.39%
	1	1.39%
自分に合っている服	1	1.39%
ぶりぶりすぎず、ちょっと女の子っぽい服装	1	1.39%
シンプルで自分に似合っている服をわかる人。	1	1.39%
コンサバ系	1	1.39%

図 165 カジュアル G のフリーアンサー

上記の個別のコメントから、カジュアルの個人属性である、年齢層が高い、自分好みの服を着る、という条件での絞り込みの結果も配慮し、かつ他の3つのクラスタとの差別化を意識すると、つぎのようなコメントが重視される。

- ・膝丈スカート、体のラインがみえるデコラテをみせるか、足をだす。露出のばらんす。
- ・体のラインが出ている服
- ・派手すぎないけど女性らしい格好、女性らしいラインや細い部分がアクセントになるコーディネート
- ・JJ系
- ・エレガント
- ・シンプルかつセクシー
- ・コンサバ系
- ・媚びない自分らしい服
- ・メスっぽい服、スカートにヒール。

上記のコメントを表示する画像を、上記の条件のうち、年齢を 25 歳以上に特定した時に選択された上位の画像で示すと、下記ようになる（図 166）。



図 166 カジュアル G の上位の画像（年齢 25 歳以上）

このように、フリーアンサーと画像をリンクすることで、言語表現が具体的な画像として表示される。これこそ、このツールが示す価値探索の成果である。

②キュート G スタイル（85 名）

自分に似合っていて可愛い系の洋服
その人に似合っている服装
ワンピース、ヒール、全体的に暗すぎない服
カジュアルだけどガーリーなところもある
その人に似合った服装
派手すぎず、地味すぎない無難な感じ
女の子らしさを多すぎずに取り入れた服装。
ふわふわした白い服
女の子らしい服装！
スカート
清潔感のある服装 TPOを弁えた服装
シンプルだけど女の子らしい服装の女の子
ワンピース シンプル 大人っぽい
スカート、淡い色、ハンドバック 3 つは身につけていることが前提条件とした服装
凝りすぎていない服
無難にかわいい服装
ほんわかした女の子らしい服を着てる子
自分に似合っている服を着ている子。
可愛さもあるけど、少し大人っぽい服。きちっとすぎない程度
TPOに合った服が着れる人！自分はこのテイスト！みたいなのを決めつけすぎないひと

ワンピース
女の子すぎない シンプルな感じが好きです。
自分に似合う服を着ている
ゆるふわ系天然女子
やせてるこ
ワンピース、スカート
清潔感のあるかんじ
女の子らしいフォルムの服
清楚で標準的な雑誌に載っているような服
清楚な服装
品のいい感じ
清楚だけど、他とはちょっと違う女の子
パステルカラー 自分に合ったサイズ ふわふわ
ふわふわ系
自分にあってる洋服であるかどうかが大事故かな、と思います♪
シンプル
露出が多すぎず、その人に合っているもの。女らしい要素があるもの。
かわいらしい服装
清楚なワンピース
カジュアルで可愛いかんじ
可愛い雰囲気
自分に似合う服を着ている人
大人っぽい(きれいめ)服装
ふんわり
エロすぎなく女の子らしい服
大人っぽい 密着した
自分に合った服を着てる
清楚系
清楚系
清楚な服
清楚 シンプル
女の子らしい定番の服装
清楚すぎない
清楚
自分に似合っていて、なおかつスタイルをよく見せる・際立たせる服！
自分に似合う服を着てる女子。好きな服じゃなくて、似合う服。
ふわふわ CamCam系
個性を主張し過ぎでない服装
華奢な感じ
流行に振り回されてない服装
清楚
CanCam系

図 167 キュート G スタイルのフリーアンサー

キュート G スタイルのコメントで特徴的なものを、図 167 から列記すると、つぎのようになる。

- ・ふわふわした白い服
- ・清潔感のある服装、TPO を弁えた服装
- ・スカート、淡い色、ハンドバックの 3 つは身に着けていることが前提

- ・可愛さもあるが、少し大人っぽい服。きちっとしすぎない程度。
- ・ワンピース
- ・清楚で標準的な雑誌に載っているような服
- ・大人っぽい（きれいめ）服装
- ・清潔シンプル
- ・ふわふわ CamCam 系

以上を表示する画像については、キュート G スタイルで、年齢像を 21－24 歳に限定してみると、以下のような画像が抽出される。これが、このスタイルの具体的なイメージである。（図 168）※すべてキュートガールの画像



図 168 キュート G スタイルの上位画像（キュートガール限定・21－24 歳に限定）

従来ならば、コメントをもとに、プロフィールを文章で作成し、そのライフスタイルを記述するという方法が一般的ではあるが、そこには研究主体の関与があまりにも大きく、かなり恣意的な解釈に入ってしまう。これに対して、このような価値探索ツールを活用すると、さまざま条件でのデータ調整（絞り込みや統合）を実行することで、その条件ごとの結果についてのコミュニケーションが展開され、つねに具体的な方法での探索が可能になり、従来のように、一足飛びに解釈を飛躍させてしまう危険性がない。あくまでもデータの検索の試行錯誤を繰り返しながらの価値探索が可能になる。このように、フリーアンサーと画像をリンクさせ、それを多様な操作によって、納得し合意される成果を発見するまで継続する思考過程がなされる、ということこそが、このツールの新しい価値である。まさに、価値探索から価値を創発するツールであるといえよう。

③ナチュキュート G スタイル（50 名）

つぎに、ナチュキュート G スタイルのコメントを列記する。このスタイルを志向する女子は 50 名とやや少数派である。個人属性としては、他の 3 クラスと比較して若い層が多い（図 169）。

白いワンピースを着た女の子
奇抜でない、スタイルがよく見える、スカートや肌見せなど女性らしさが現れる服装
自分が好きな服を着ている子
短すぎないニットワンピース
かわいい
奇抜すぎず地味すぎずが良いかと
スカートを履いた女の子
ナチュラル 自分に似合う服 ワンピース
思わず触りたくなるかんじ！
きっちりちゃんとしてるけど、どこかに女子らしさのある服装
可愛い、清楚 ワンピースとかスカートとか
清潔感のある服
清楚系
清楚系。あまり露出はー...
可愛い
甘カジ
キレイ系
自分に似合っている服装
キレイめ
フリフリ

可愛い服の女
ふわふわしてるパステルカラーの服
清潔感があり、適度な露出がある服装だと思います。
自分に似合った服装の女の子
本人に似合っている服
自分に合った服
ふわふわした服
ボーイッシュな格好
ふんわり系
ふるふわ
人によると思います
ゆるふわ
フェミニン系の服
自分に似合う服を着てる女の子
なんかびんくとか白っぽい服でつるつるした素材でべらべらした服をきている子
もてるのは露出ミニスカとか出ないかと思う 好かれるのは清楚すぎないシンプルな感じ
普通
ロング丈の服（ワンピースとかロングカ）が着こなせてる子
自由に好きな格好が一番だと思います
清楚系や一般的な服で飛び抜けてない服装。

図 169 ナチュキュート G スタイルのフリーアンサー

ここのコメントを、他のクラスタとの比較でこのクラスタの特徴を列記すると、つぎのようになる。

- ・自分が好きな服を着ている子
- ・かわいい
- ・ナチュラル、自分に似合う服、ワンピース
- ・ふりふり
- ・ふわふわしているパステルカラーの服
- ・ゆるふわ

- ・なんかピンクとか白っぽい服でつるつるした素材でぺらぺらした服をきている子
- ・ロング丈（ワンピースとかロンスカ）が着こなせる子
- ・自由に好きな恰好が一番だと思います。

このようなコメントを、このクラスタの中で、より若い 21-24 歳の層に限定した画像を選択すると、下記の 5 つの画像が検出される（図 170）。



図 170 ナチュキョート G スタイルの上位画像（21-24 歳に限定）

いわゆる森ガールに固有なナチュラルでゆるふわなファッションとキュートで可愛さをもつファッションが共存や融合するファッションスタイルの具体的な画像が上記の 5 画像である。

④アンチナチュ G スタイル（85 名）

最後は、アンチナチュラル G スタイルである（図 171）。

このスタイルは、ネーミングに示されているように、ナチュラルガールのみ否定的で、それ以外の 3 ガールスタイルには強く共感している。その意味では、3 つのガールスタイルが混然一体になっているともいえ、それ固有の意味世界があるのか、それを知るのがフリーアンサーなのかもしれない。

カジュアルだけどおしゃれ
相手の男性が好きそうな雰囲気、自分らしく着こなせる子
ひらひら ふわふわ ゆらゆら すけすけ の法則を守る子
シンプル ふわふわ 森ガールのやつがカジュアルなやつ
綺麗めで、女性っぽさが出る服
Camcan系
シンプル

自分に似合ってる服を堂々と着ている子
清楚なコーディネート。
スカート 露出しすぎない服
いいえ
おしゃれ
シンプル、クール
自分の魅力を引き出す服装の女の子
スカート
似合う服を着る人
程よい露出
ナチュラル系
趣味に走り過ぎず、かつその人の雰囲気にあったもの。フェミニンすぎないもの。
淡い色
ニーハイブーツが合う服装。
女の子らしい服 ゆるニット、ワンピース
きれいめの服
自分に似合う突飛しすぎでない服装をきているおんなのこ
清楚系。
かわいい感じの子？
白とかピンクとかの淡い色ベースの服

シンプルな綺麗目
・やりすぎない肌見せ・オシャレすぎない・男性が理解できないような流行アイテムを使わない
自分が似合ってる服を着ている女の子
甘過ぎないワンピース
クール
ガーリーな可愛い服装
女の子らしい服装、綺麗な服装
ワンピース 淡い色
着飾りすぎず、流行に乗りすぎるよりも自分らしい服！
清潔感のある服装。クセがありすぎず、シンプルすぎない服。万人受けしそうな服。
清楚な服装
ふわふわ
ふわふわしてる女の子らしさではなく、清楚だったり品のある女の子らしい服装
ピンクや白などの可愛いらしい服
おんなのこらしい服
落ち着いていて女の子らしい服
自分にあった服を着ている
TPOにあった服をきている人
シンプル
ワンピース

あまり行き過ぎない服装。最先端な服はそんなに人気がないそう。
ガーリー
ふわふわ、ミニスカート
服装 < 顔！ 気がする！
cancam系の服装
ふわふわ系

少しカジュアルっぽいけど、女の子らしい服装。
可愛くて清楚
目立たない服
その人の雰囲気、スタイルに合った服を着ている子
ぶりぶりすぎず、シンプルな服装の女子！
TPOに合わせて、そのなかで自分にあう服装
ワンピース 女の子らしくするのは良いが女の子らしすぎるのは駄目ほどよくカジュアルなのがモテると思う。
その子に似合っている服であることが大切だと思います。
その子に似合っていれば！
キャンキャン系を少し大人しくした感じ
ラインが出る服
無難、お酒落すぎない、ひらひら
自分に似合う服を来てる子。流行にとらわれないで、同一系統の服を着回せてる人。
カジュアルで気取ってない感じ

図 171 アンチナチュ G スタイルのフリーアンサー

上記コメントから、このクラスに固有のコメントを列記すると、つぎのようになる。

- ・ カジュアルだけどおしゃれ
- ・ 相手の男性が好きそうな雰囲気を、自分らしく着こなせる子
- ・ 綺麗めで、女性っぽさが出る服
- ・ 自分の魅力を引き出す服装の女の子
- ・ 自分に合っている服を堂々と着ている子
- ・ 趣味に走りすぎず、かつその人の雰囲気にあったもの、フェミニンすぎないもの
- ・ 自分に似合う突飛過ぎない服装を着ているおんなの子
- ・ やりすぎない肌みせ、オシャレすぎない、男性が理解できないような流行アイテムを使わない
- ・ 飾りすぎず、流行に乗りすぎるよりも自分らしい服
- ・ 清潔感のある服、クセがありすぎず、シンプルすぎない、万人受けしそうな服
- ・ 落ち着いていて女の子らしい服
- ・ キャンキャン系を少し大人しくした感じ

このようなコメントに強く反応した画像は、下記の 6 点である（図 172）。



図 172 アンチナチュ G スタイルの上位画像

この 6 つの画像は、3 つのスタイルの上位 2 つの画像を選択したものである。この 6 つの画像の支持率が高いことから分かるように、このクラスには、3 つのガールスタイルを統合する何かの感覚が共有されているのだろう。

5-3-4 実例からみた「ふかぼりくん」の価値探索機能

「恋服」の解析をする過程で、そのツールである「ふかぼりくん」を活用することで、従来の社会統計ツールとは、似て非なるツールであることを、多岐にわたって説明してきた。ここでは、そこでの具体的な成果をまとめて、このツールがどのような意味で価値探索ツールとしての機能を保有しているかを、より一般的な方向で説明する。

5-3-4-1 使用の全体

「ふかぼりくん」に固有の価値探索機能は、以下の10機能である（図 173）。

- ①統合機能
- ②絞込機能
- ③削除機能
- ④彩色機能
- ⑤ソート機能
- ⑥スケール機能
- ⑦生成機能
- ⑧付加機能
- ⑨リンク機能
- ⑩セーブ（ロード）機能

①統合機能

統合機能は、項目をまとめることで、項目それ自体としての価値が発揮できるように変換する機能である。たとえば、今回のケースならば、低年齢層の数が少なすぎるので、ある程度の規模が構成されるように、年齢層の統合を行い、その結果、年齢階層が縮約されることで、年齢項目が他の項目との関連で、有効な価値を発揮するようになった、このような統合機能が容易に実行できる仕組みになっていることは、価値探索を促進すると確信できる。

②絞込機能

絞込機能は、ある特定の項目に限定することで、他の項目との関連性を明確にする機能である。データを深堀して、より特定の条件を設定して、そこでの関連性を求めようとする、絞込機能の活用が非常に重要である。この機能を活用することで、テーマの追究が明確になり、だからこそ、そこから新しい価値探索が可能になる。

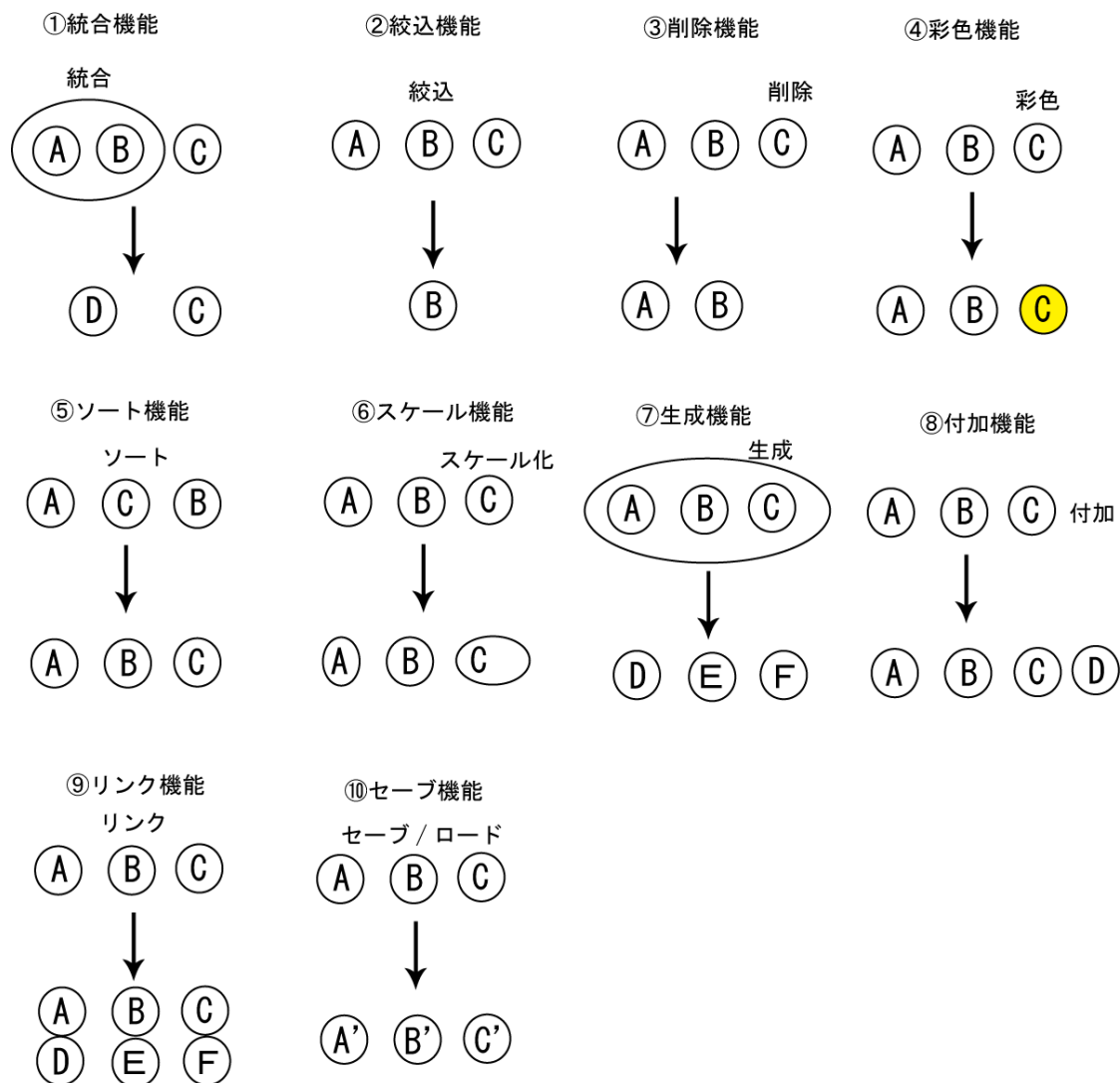


図 173 「ふかぼりくん」の価値探索機能

③削除機能

削除機能は、項目のなかで、解析に必要な変項については、削除することが必要である。成果のより明確な関係を認識するためには、データとしては存在していても、結論を導くためには無用なデータについては、削除する必要がある。そうすることで、成果がより明確になる。もちろん、この削除機能は、特定条件での削除であるので、つねに復活は可能な仕組みにはなっている。だからこそ、容易に削除機能を活用して、多様な関係の確定が可能になる。

④彩色機能

これは、一見、意味のなさそうな機能であるが、実際にデータ解析を実行しているプロセスでは、非常に効果のある機能である。たとえば相関が明確なデータのクロス表などの場合、どの変項がその有効性を誘導しているかを確定するためには、彩色することで、その確定が明確な可視化として表示される。彩色機能は、このようなデータの可視化をより鮮明にするためには重要な機能を果たしている。

⑤ソート機能

ソート機能は、彩色機能と並んで、可視化の機能として重要である。平均や分散などの結果について、そこでの序列を変更する機能が容易に発揮できると、そこでの関連性が非常に容易に理解可能になる。それが、可視化の1つの重要機能である。ソート機能が、単純集計ばかりでなく、グループ化集計にお

いても容易に活用できることは、成果の理解を強化する機能として高い価値をもつ。

⑥スケール機能

スケール機能は、変項の数値の規模をどこまでに設定するか、特に最小値をどこに設定することで、解析をするかを特定化する機能である。たとえば、グループ集計の場合ならば、各クラスタの特性する変項の規模（最小値）をどこに設定するか、を確定することで、価値ある成果を探索しようとする機能である。そこでの規模感をどこに設定するか、が研究者の探索したい成果を直接関係するので、非常に重要な機能である。もちろん、この機能は数値データの場合だけでなく、画像データでもフリーアンサーのテキストデータの場合でも活用できる機能であり、とくに、このような非数値データの場合で重要な機能を果たしている。

⑦生成機能

生成機能は、新しい項目を産出する機能であり、その典型がクラスタリングである。数種の項目から新しい価値をもった項目を誘発させることで、解析段階を一段アップさせる機能である。ここでは、何度も試行錯誤が繰り返させることが可能になっており、最適な、納得できるクラスタが創発させることが可能になっている。この生成機能があることで、既存のデータから、さらなる高度なデータの生成が創発される仕組みになっている。そうすることで、価値探索がさらに促進することが可能になっている。

⑧付加機能

生成機能が発揮されれば、それをすぐに新しい項目として付加して、さらなる多様な解析が可能になる仕組みが必要である。それが付加機能である。これによって、生成機能での成果が次の解析に継続的・連続的に実行可能になる。

⑨リンク機能

リンク機能は、数値データを非数値データ（画像データとテキストデータ）に関連づける機能である。従来の社会解析ツールでは、数値データと非数値データとは独立して解析することになっているが、ここでは、その関連性を明確にすることで、データの個別具体的な解釈を付加させることが可能になり、価値探索のレベルが全く異質なものにもっていくことが可能になっている。これによって、比較分析が、画像とテキストが付加されることで、非常に深い解釈が実現するようになっている。

⑩セーブ（ロード）機能

最後はセーブとロード機能である。上記の諸機能を活用することで、データの構造が変更されるので、それを新しいデータ構造としてセーブすることが必要である。その機能がセーブ機能であり、それによって、何度も解析を繰り返す過程で、次のロード機能が活かしてくる。

以上の機能を駆使するとき、「ふたぼりくん」が価値探索をする過程で可能となる分析は、つぎの2つの分析手法である。

- 1：比較多層分析
- 2：マルチデータ分析

5-3-4-2 比較多層分析

どの社会調査でも、比較分析は当たり前であるが、ここでは、統合・絞込・削除機能を活用することで、深堀を何度も繰り返すことで、表面的な比較分析のレベルから、さらに特定化する過程で、データの統合と絞込を展開し、無用なデータは削除することで、深堀が可能になる。そうすることで、通常の分析では得られない深い階層での解析が可能になり、新しい価値の探索が可能になる。これを、比較多層分析と呼ぶ。

具体的に、女子スタイルの場合で検討する。

レベル1：女子スタイルの単純集計（725 名）

すでに生成機能と付加機能によって作成された女子クラスタについての単純集計を表示する。これがレベル1である（図 174）。

変数名	平均	分散
cluster_5:女子Gスタイル	3.771	1.648

Clip

cluster_5:女子Gスタイル

平均:3.771
分散:1.648

C_1:興味なし	78	10.76%
C_2:カジュアルG	107	14.76%
C_3:キュートG	144	19.86%
C_4:ナチュキュートG	123	16.97%
C_5:アンチナチュG	120	16.55%
C_6:なんでもあり	153	21.10%

図 174 全体での女子 G スタイルの単純集計

レベル2：女子スタイルの性別での比較（女子：422 名 / 男子 303 名）

次に、レベル2では、性別でのクロスによる比較を検討する（図 175、図 176、図 177）。

変数名	平均	分散
cluster_5:女子Gスタイル	3.765	1.665

Clip

cluster_5:女子Gスタイル

平均:3.765
分散:1.665

C_1:興味なし	43	10.19%
C_2:カジュアルG	72	17.06%
C_3:キュートG	85	20.14%
C_4:ナチュキュートG	50	11.85%
C_5:アンチナチュG	85	20.14%
C_6:なんでもあり	87	20.62%

図 175 女子 422 名による女子 G スタイルの単純集計

変数名	平均	分散
cluster_5:女子Gスタイル	3.779	1.623

Clip

cluster_5:女子Gスタイル

平均:3.779
分散:1.623

C_1:興味なし	35	11.55%
C_2:カジュアルG	35	11.55%
C_3:キュートG	59	19.47%
C_4:ナチュキュートG	73	24.09%
C_5:アンチナチュG	35	11.55%
C_6:なんでもあり	66	21.78%

図 176 男子 303 名による女子 G スタイルの単純集計

Clip cluster_5:女子Gスタイル

	C_1:興味なし	C_2:カジュアルG	C_3:キュートG	C_4:ナチュキュートG	C_5:アンチナチュG	C_6:なんでもあり	合計	平均	分散
q_11:性別は？									
男性	35 (11.6%)	35 (11.6%)	59 (19.5%)	73 (24.1%)	35 (11.6%)	66 (21.8%)	303	3.78	1.62
女性	43 (10.2%)	72 (17.1%)	85 (20.1%)	50 (11.8%)	85 (20.1%)	87 (20.6%)	422	3.77	1.66
合計	78 (10.8%)	107 (14.8%)	144 (19.9%)	123 (17.0%)	120 (16.6%)	153 (21.1%)	725	3.77	1.65

X二乗値:27.54 自由度:5

図 177 女子 G スタイルと性別のクロス集計

このクロス表にあるように、性別と女子スタイルとの間には 1%水準で有意な相関がみられる。しかもその関連は、重要な 3 つのクラスにみられる。そこで、クラス 1 (興味なし) とクラス 6 (なんでもあり) のクラスを削除し、そこから、さらに深堀をする。

レベル 3 : 女子 4 クラスの説明項目の探索

2 クラスを削除した 292 名の女子について、そこでの 4 クラスを説明する項目は何かの検討をする。これがレベル 3 である (図 178、図 179、図 180、図 181)。

対象:cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちらが好み？	カウント	平均	分散	偏差値	C_2:カジュアルG	C_3:キュートG	C_4:ナチュキュートG	C_5:アンチナチュG
28:25歳以上	29:いる	31:自分好みの服	6	2.833	1.213	44.1	4 (66.67%)	0 (0.00%)	1 (16.67%)	1 (16.67%)
27:21歳以上	30:いない	31:自分好みの服	51	3.235	1.165	47.6	19 (37.25%)	12 (23.53%)	9 (17.65%)	11 (21.57%)
27:21歳以上	29:いる	31:自分好みの服	65	3.338	1.193	48.5	21 (32.31%)	19 (29.23%)	7 (10.77%)	18 (27.69%)
28:25歳以上	30:いない	31:自分好みの服	4	3.750	1.299	52.1	1 (25.00%)	1 (25.00%)	0 (0.00%)	2 (50.00%)

図 178 グループ化集計 (カジュアルグループでソート)

対象:cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
26:20歳以下	29:いる	32:モテそうな服	9	3.889	0.875	53.3	0 (0.00%)	4 (44.44%)	2 (22.22%)	3 (33.33%)
26:20歳以下	29:いる	31:自分好みの服	45	3.578	0.954	50.6	5 (11.11%)	19 (42.22%)	11 (24.44%)	10 (22.22%)
27:21歳以上	29:いる	32:モテそうな服	24	3.417	1.077	49.2	5 (20.83%)	10 (41.67%)	3 (12.50%)	6 (25.00%)
26:20歳以下	30:いない	32:モテそうな服	21	3.762	1.109	52.2	3 (14.29%)	7 (33.33%)	3 (14.29%)	8 (38.10%)

図 179 グループ化集計（キュート G でソート）

対象:cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
26:20歳以下	30:いない	31:自分好みの服	53	3.679	1.146	51.5	12 (22.64%)	10 (18.87%)	14 (26.42%)	17 (32.08%)
26:20歳以下	29:いる	31:自分好みの服	45	3.578	0.954	50.6	5 (11.11%)	19 (42.22%)	11 (24.44%)	10 (22.22%)
26:20歳以下	29:いる	32:モテそうな服	9	3.889	0.875	53.3	0 (0.00%)	4 (44.44%)	2 (22.22%)	3 (33.33%)
27:21歳以上	30:いない	31:自分好みの服	51	3.235	1.165	47.6	19 (37.25%)	12 (23.53%)	9 (17.65%)	11 (21.57%)

図 180 グループ化集計（ナチュキュート G でソート）

対象:cluster_5:女子Gスタイル

q_12:年齢は？	q_13:恋人はいますか？	q_14:デートを想定した時、どちら	カウント	平均	分散	偏差値	C_2:カジュク	C_3:キュート	C_4:ナチュキ	C_5:アンチナ
28:25歳以上	30:いない	32:モテそうな服	2	5.000	0.000	63.0	0 (0.00%)	0 (0.00%)	0 (0.00%)	2 (100.00%)
27:21歳以上	30:いない	32:モテそうな服	12	4.000	1.225	54.3	2 (16.67%)	3 (25.00%)	0 (0.00%)	7 (58.33%)
28:25歳以上	30:いない	31:自分好みの服	4	3.750	1.299	52.1	1 (25.00%)	1 (25.00%)	0 (0.00%)	2 (50.00%)
26:20歳以下	30:いない	32:モテそうな服	21	3.762	1.109	52.2	3 (14.29%)	7 (33.33%)	3 (14.29%)	8 (38.10%)

図 181 グループ化集計（アンチナチュ G でソート）

このように、カジュクール G スタイル（72 名）、キュート G スタイル（85 名）、ナチュキュート G スタイル（50 名）、アンチナチュ G スタイル（85 名）それぞれが、年齢と恋人有無とデート服の 3 項目の説明項目とどのような関連性があるか、を探索する。ここには、それなりに価値ある探索の成果が発見される。

レベル 4：カジュクール G スタイルと 4 ガールスタイルとの関連（図 182）

変数名	平均	分散	
c_12:カジュアルガール	3.236	1.021	
c_13:クールガール	3.083	0.968	
cluster_5:女子Gスタイル	2.000	0.000	
c_14:ナチュラルガール	1.292	1.160	
c_11:キートガール	1.222	0.931	

図 182 カジュアル G スタイルの支持の高い 4 ガールスタイル画像

さらに、カジュアル G スタイル（72 名）に限定して、4 ガールスタイルの画像について、どのような支持が高いかをソート機能を活用して、表示する（図 182）。

ここで、具体的なイメージが特定され、カジュアル G スタイルがどのようなファッションを志向しているかが明確になる。

レベル 6：カジュアル G スタイルとモテ服の自由回答との関連性

最後に、カジュアル G スタイルを志向する女子の中で、さらに、カジュアルスタイルとクールスタイルの得点が低い層（2 点以下）を削除して、より典型的なカジュアル G スタイルの女子に限定して、その層 35 名が、モテ服について、どのようなフリーアンサーを記入しているか、を表示する。（図 183）

シンプルにさらっとした着こなし。	1	2.86%
膝丈スカート 体のラインみえる デコルテをだすが、足を出す。露出のバランス	1	2.86%
ふわふわ系 ワンピ	1	2.86%
リア充男子→アンサンプルニートを着ているJJ系女子/ビビッドな色やら流行を取り入れたり海外セレブとか好きそうなViVi系女子 サブカル系男	1	2.86%
かわいらしい、女の子らしい服装	1	2.86%
体のラインが出ている服	1	2.86%
シンプル	1	2.86%
カッパ可愛い服。	1	2.86%
シンプル！	1	2.86%
モテるの概念がわからないけど、きめすぎない服装が支持されそーな気がします。	1	2.86%
女の子らしい服装。スカートとか。	1	2.86%
はやりばかりにとらわれないで自分の顔や雰囲気にあう洋服をきている子	1	2.86%
おしゃれ	1	2.86%
エレガント	1	2.86%
シンプル 白黒とか	1	2.86%
流行を追うだけでなく、自分の好きな服を選んで着ている子。自分の目指す、あるいは自分に合う雰囲気をわかっている子。でも奇抜すぎず、	1	2.86%
自分を主張しすぎない服	1	2.86%
自分の体型に似合うことを分かっている女の子	1	2.86%
スカート履いてる子、ある程度女らしさが出ている服	1	2.86%
自分にあった服をえらべる子だとおもいます。	1	2.86%

図 183 カジュアル G スタイルのフリーアンサー（一部）

これによって、さらに、深堀が可能になり、このクラスがどのような特徴をもっているか、非常に深いレベルまで、検討が可能になる。

5-3-4-3 マルチデータ分析

「ふかぼりくん」は、多層な比較分析をするツールであると同時に、もう1つ、非数値データも併用して扱うことができるツールなので、数値データと非数値データとの複合的な分析を可能にしている。それが、画像データとテキストデータと数値データを、リンク機能を活用して解析する手法である。上記のレベル5とレベル6で提示した例は、このマルチデータ分析でもある。

以下、簡単に事例を紹介する。

まず、フリーアンサー分析の事例である。これは、女子 G スタイルの4ファッションについて、女子に限定した場合での、フリーアンサーの比較分析である（図 184）。

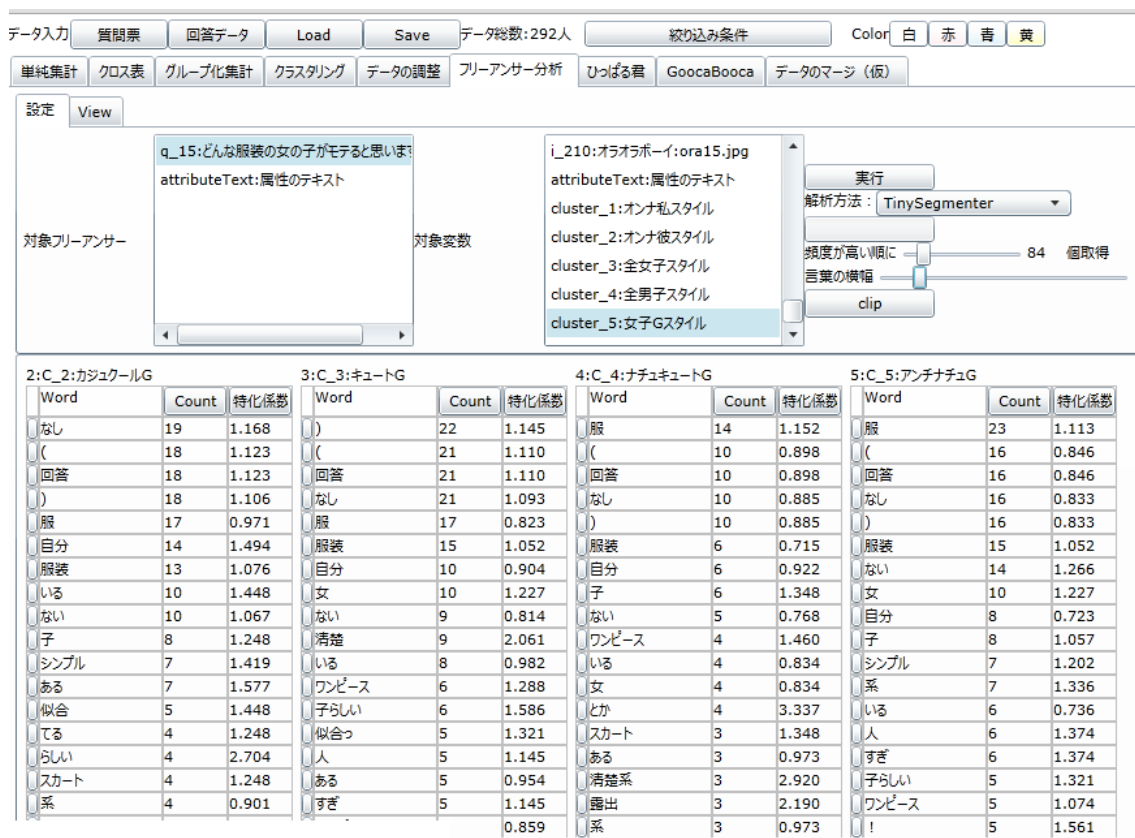


図 184 フリーアンサー集計

この分析は、フリーアンサーの形態素分析の結果を、クラスごとに、頻度と特化係数でソートさせたもので、さらに、特定の言葉を View でチェックすると、その言葉を含んだフリーアンサーが表示される (図 185)。



図 185 フリーアンサー集計 (詳細表示)

ここでは、カジュクール G スタイルの「服」を View でチェックした結果である。すると、このスタイルで服と書いた人のモテ服のコメントが列挙される仕組みになっている。このように、絞込機能などを駆使することで、ターゲットにしたユーザ層のテキストが自動的に分析されて導出される仕組みになっ

ている。このような表示は、通常の数値データでの解析ツールにはないことであり、このツールに固有な価値探索を求めているからこそ実現されたマルチ複合ツールとしての特性である。

しかも、各クラスタでのコメントでの特徴を示す層が、どのような画像を支持しているかを特定すると、下記のようなになる、ここでは、データ調整機能として、図 186 のような絞り込と削除機能を活用した結果、図 187 のような結果が探索された。

絞り込み内容

- ☒ 24:女性 q_11:性別は?
- ☒ 2:C_2:カジュアルG cluster_5:女子Gスタイル
- ☒ 29:いる q_13:恋人はいますか?

削除内容

- ☒ 1:C_1:興味なし cluster_5:女子Gスタイル
- ☒ 6:C_6:なんでもあり cluster_5:女子Gスタイル
- ☒ 0:0.0 c_13:クールガール
- ☒ 1:1.0 c_13:クールガール
- ☒ 2:2.0 c_13:クールガール
- ☒ 0:0.0 c_12:カジュアルガール
- ☒ 1:1.0 c_12:カジュアルガール
- ☒ 2:2.0 c_12:カジュアルガール
- ☒ 26:20歳以下 q_12:年齢は?

人数: 13

図 186 絞り込み条件の設定

変数名	平均	分散	Image
c_12:カジュアルガ	3.846	0.863	    
c_13:クールガール	3.385	0.625	    

図 187 図 174 での絞り込み条件による画像

このように、数値データを前提に、それにテキストデータと画像データを、リンク機能を活用して、連動させることで、上記のような個別具体的なファッションイメージが明確になってくる。従来の調査では、言葉だけですべての処理をしていたのに対して、ここでのマルチなデータの活用によって、ユーザのプロフィールが非常に鮮明に描けることになる。だからこそ、新しい価値の発見や探索が容易になる、といえよう。

以上、「ふかぼりくん」が価値探索ツールとして、どのような特性をもつかを検討した。「恋服」のデータをもとに、多様な解析過程を確認し、その一般的な価値探索機能として、10の機能を確認した。さらに、その10機能の活用によって、従来の社会調査の社会統計ツールとは違って、比較多層分析とマルチデータ分析のツールとして有効であることを確認した。

第六章 結論と今後の課題

第六章 結論と今後の課題

6-1 作成したツールの全体像

インターネット化が進み、ソーシャルメディアを使い、人々はコミュニケーションする時代になってきた。このような時代において、社会の声、社会的現実、すでにソーシャルメディアに蓄積されている。そのための分析する道具として、「Rawler」「ひっぱりくん」「こうぞうくん」「GoocaBooca」「ふかぼりくん」を開発した。

これらの位置づけとしてこのような形になる。(図 188)

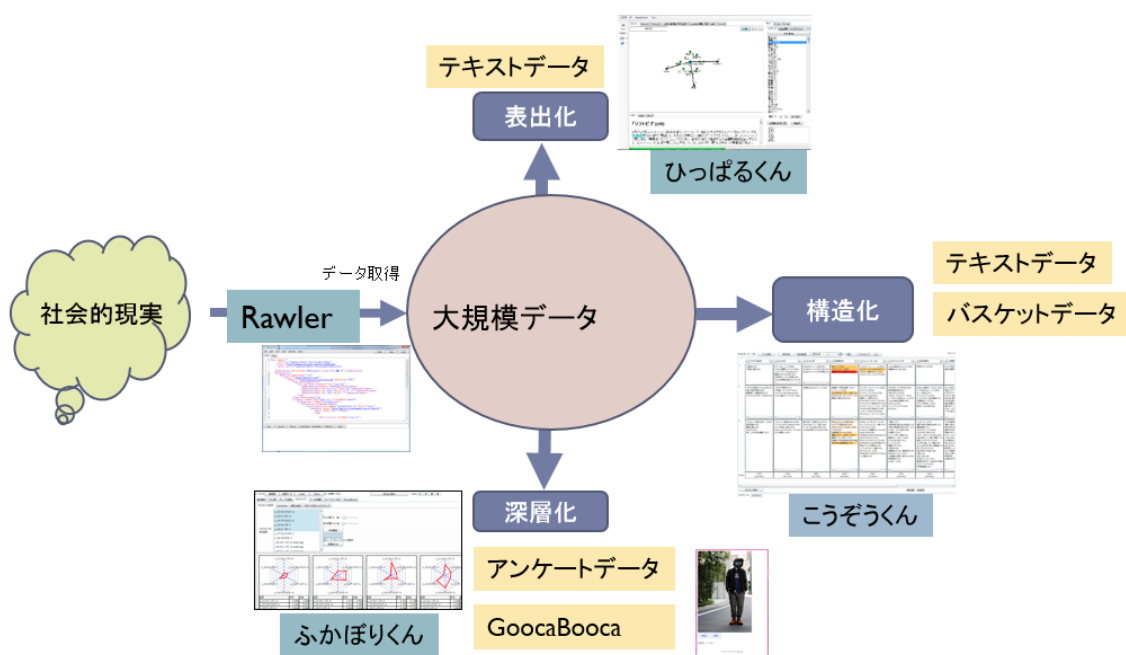


図 188 ツールの全体像

社会的現実ともいえるべき、Web のデータを取得するためのツールが、Rawler であり、これは大規模データとなる。Rawler は、Web にあるデータを取得するためのメタツールであり、前処理ツールである。データ分析には、データ分析それ自体も重要だが、データの取得も重要であり、データ取得のイノベーションを図りたくて作成した。対象とするデータは、様々な形であるので、様々な形に対応するためにプログラミングを行うという形にしている。Rawler は、関数型言語、宣言型言語の特徴を持つため、対象とする Web サイトへのデータ取得の設定のみを記述し、それ以外の手続きについての記述は省くことができるようにした。このようなツールを使うことで、Web から大規模データの取得を容易にし、多様な Web サイトからデータ取得し分析する障壁を低くする。

この大規模データを分析するための方法論を、3 つ用意した。一つ目は、データから、分析者の考えを表出化させる「ひっぱりくん」、これは、テキストデータをおもに対象とし、テキストデータから気になる単語を「ひっぱり」ことで、言葉のネットワークを作るツールである。実査 1,2,3 で示した。「ひっぱりくん」はデータから多様な解釈を可能となるツールであり、データが同じでも結果は一意に決まらない。しかし、インタラクティブに自身の考えを反映するために、思考しながら探索することを要求するツールであり、創発的な発見を誘発させるツールである。

二つ目は、データを構造化させ、その構造を修正、解釈する中で、分析者のデータへの理解を深める

「こうぞうくん」である。これは、購買履歴のようなバスケット形式のデータ、分ち書き済みのテキストデータを対象とし、その共起関係からクラスタリングをして、可視化を行い、構造化を行う。ソフト上では、おおまかなクラスタリングを自動で行い、アイテムの解釈に基づいて、分析者が細かいクラスタの調整を探索的に行えるようにしている。機械的に生成されたクラスタリング結果を絶対視せず、分析者が介入できる仕組みを用意しており、半教師付きクラスタリングを行えるようにしている。このようにしてクラスタを作った上で、クラスタに名前をつけるという行為をさせる。これは、作られたクラスタの解釈を分析者にさせることであり、分析者の背景知識、暗黙知に問いかけるものであり、クラスタを概念として解釈できるものに変換する行為である。そして、その概念間の関係をデータによりグラフとして可視化し、概念の全体像を掴むことができるようなツールになっている。

最後に、データへの条件指定し、絞り込みに特化した深層化ツール「ふかぼりくん」を作成した。これは、アンケートデータの分析用のツールであり、画像アンケートシステムの、GoocaBooca と連携する。画像に対する印象は多様であり、言語化が難しいものである。しかし、例えば服装などは、社会としてイメージは確かに存在し、ある場面での、どのような服装、雰囲気が適切不適切かを答えることは可能であり、漠然とした暗黙知として存在する。そのようなものを分析対象として GoocaBooca は、アンケート回答者にその暗黙知を表出させるようにしている。そして、そのデータの分析に「ふかぼりくん」を使う。「ふかぼりくん」は、データの中に様々な人達がいることを前提とし、探索的に様々な人達の小集団を探して探索的に深堀りをしてデータ分析できるような仕組み支援するツールである。アンケートでの回答項目への絞り込み機能とクラスタリング機能を合わせることで、画像とテキストと回答を組み合わせて考えながら探索する仕組みを提供していることを示した。

「こうぞうくん」「ふかぼりくん」に関しては、Web ブラウザで使えるツールにした。そのため、自身の PC にインストールをすることなく使うことができ、インストールトラブルを防ぐことができるため、授業などの多くの人が同時に使う場面において有効である。また、ソフトウェアのアップデートもそのサイトにアクセスしたときに行われるので、アップデートトラブルも発生せず、常に最新版がつかえることになる。「Rawler」「ひっぱるくん」に関しては、Windows ソフトウェアであるが、ClickOnce という技術を使い、Web ブラウザから起動できるようにしている。これも同様に、最新版に保つことができ、インストールの手間を省き、使いたいときに使えるようなものになっている。

6-2 探索性、創発性

これらのツールに共通するのは、探索的であることをサポートする機能である。データの可視化を行い、まず、データを目に見える形にし、その後、人間が主観を交えてその可視化をコントロールすることを基本的なパターンとしている。分析者自身の関心に基づき、データの可視化を行い、現前化させる。そして、その現前化されたものを人が観察し解釈を加えていく中で、新たな気づきを得る。その気づきから、分析者は新たな関心をもち、さらなる可視化を行う、というものになっている。これは、「メタ認知」プロセス[18]と同型であり、FNS ダイアグラム[19]が示しているプロセスである。なにかしらの現前化をもって、思考は働くのであり、その中で創発性が生まれる。

ある特定のゴールに向かって真っすぐすすむというものではなく、漠然としたゴールはあるとはいえ、手さぐりに進んでいく。手さぐり故に、目的への最短距離を取ることができないため、愚かである、というような価値判断をするかもしれない。しかし、データはあくまで、事実の一部分であり、新たな価値はデータの中にあるというよりも、データの外にある。データの可視化は、形式的にとらえられないものを人間が解釈してとらえることを目的としており（形式的に捉えられるなら、数式的変換して判定した方がいい）、背景知識、暗黙知、インスピレーション、気づきを引き出すものである。

一方、ツールのコンセプトである、「表出化」「構造化」「深層化」は、どのように考えればいいのか？補助線として、KJ 法[22]を使う。

KJ 法は、川喜田二郎氏によって提案された、データをまとめる方法論であり、「創造性開発」に効果があるとされている。KJ 法は、「カードを作る」「カードからグループを作る」「図解化」「叙述化」というプロセスを踏む。

このようなモデルとして捉えると、「ひっぱるくん」の単語をひっぱっていく行為は、データの中から、

「カード」としてふさわしいものを選んでいく行為であり、「表出化」である。そして、表出化された「カード」は、その時点で、可視化し、並べられるので、「カードのグループを作る」「図解化」が行われる。データに基づき、「カード」間の関係が示されるものとなっている。

このように考えると、「ひっぱるくん」は、SECI モデルの知識創造のプロセスとして、「表出化」と「連結化」を担っているといえ、このツールの一番の特徴は、「ひっぱる」行為であり、その意味で、「表出化」を強く意識したツールだといえる。しかし、ツールの機能として、グループ化をして、名前をつけるといった機能は弱く、外部のツールを使って、自身が引っ張っていたことの意味付けを行わなくてはならない。

一方、「こうぞうくん」は、予め表出化されたデータを前提として、その「連結化」を行っているツールであるといえる。KJ 法で言えば、「カードのグループを作る」「図解化」に重きをおいたツールであるといえる。「こうぞうくん」はデータに基づき、クラスタリング（グループ分け）ができるツールであり、自身の考えで、アイテムのグループを移動させることができる。そのため、考えながら「グループの作成」という行為を支援するツールであるといえる。また、「名前をつける行為」は、完全な叙述とはいえないが、カードの塊に対して記述を行うことなので、表出化に相当し叙述化に相当する。

最後に、「深層化」ツールの「ふかぼりくん」は、データの深堀りを行うツールであり、これは、データの塊から部分を取り出すことである。そのため、KJ 法でいう、「カードを作る」行為に相当する。クラスタリング機能は、データの中でのグループ化に相当する。しかし、「ふかぼりくん」は、「図解化」に相当する機能が弱い。全体像との対比としての部分という形での可視化がなく、深堀りをしていった結果を「カード」のようにして、「図解化」するようなことが必要であろう。データに対して自身がつよく認識したところを彩色できることは、「表出化」であるといえる。しかし、「叙述化」までと言えるかといえ、言えないだろう。

以上のことをまとめると表 12 の通りである。

表 12 KJ 法と各種ツールとの対応

KJ 法	ひっぱるくん	こうぞうくん	ふかぼりくん
カードを作る	○	×	○
グループ化	△	○	△
図解化	○	○	×
叙述化	×	△	△

これらのツールとして全体的に苦手なことは、「叙述化」である。可視化したことを文章なりにまとめていく作業へのサポートが弱い。「叙述化」自体は、本論文では実査として、記述したことであり、ツール自体の機能というよりは、ツールを使うワークフローとして、考えたほうがよいと思われる。しかし、「叙述化」を誘発させる、よりシームレスに行うことができるツールはありうるものなので、この点が課題であろう。

KJ 法と「ひっぱるくん」「こうぞうくん」「ふかぼりくん」との比較を行ったが、KJ 法と本論で述べたツール群の大きな違いは、データを扱っているということである。データのない中、ブレインストーミングで、表出化させる、KJ 法と、データの中から表出化させるものを選んでいくという行為の違いである。これには一長一短がある。ブレインストーミングの場合は、データに縛られず、自由な発想ができるということであるし、本論のツールの場合は、データに縛られて、自由な発想を束縛しているともいえる。しかし、データ、事実のない中での発想というより、事実（データ）を手がかりにして表出化を行うことは、社会分析として、事実に基づいている分、有効であるといえる。

連結化についても、データに基づいて行っており、何もない状態から考えながらまとめていくという行為を省いている。しかし、大量のデータがある状態で、一つ一つ行っていくのは大変であり、この省力化は次の大きい発想への手がかりとなっている。このようなデータによる補助の上になりたっているため、データによる補助は必要である。

本論で述べたツール群は、データとの対話を促進するツールである。自身の考えを反映しながらデータに拘束されるという形をとる。このことは、主観的であるが、完全な主観ではなく、データによる客

観性に基づいていることである。

何かしらのデータによる具体物を生成することにことにより、インタラクティブに、探索的に思考させるツールになっている。

6-3 実査から発見された価値探索機能の創発性

本論文の結論の3番目の特性は、3つの社会調査の実査を通して、そこでのオリジナルなツールを活用する結果として、「表出化」と「構造化」と「深層化」という3つの価値探索機能がデータ解析と解釈の過程に対して、それぞれどのような創発特性をもたらすか、を明確にすることである。

6-3-1 「ひっぱりくん」の表出化ツールの価値探索機能

まず、「ひっぱりくん」を活用した表出化を求めた調査では、すでに3章で記述しているように、つぎの6点が表出化による関係生成のフレームワークとして重要な機能をはたしていることが判明している。以下、3章での記述を、再掲する。

- 1) 自由と秩序
- 2) 価値表明と目標達成
- 3) 自明性と創発性
- 4) 目的と手段
- 5) 階層とネットワーク
- 6) 一般と特殊

1) 自由と秩序

「ひっぱりくん」は、基本的には、ツールを活用する主体（解釈者）の自由な選択と探索を重視しており、ほぼ完全なる自由を支持している。この完全なる自由は、ツールが投げかける緑の円環が示唆する特化頻度の序列を無視して、徹底して黒ラインだけを重視して構造化を図ることを可能にしている。ここでは、ささやかに赤ラインが自動生成されて、主体的に選択された用語の関連性を支援するだけである。ただ、この赤ラインの支援がないとネットワーク化ができない、という制約は、もしも解釈者が用語の階層のみで解釈は十分であるという立場ならば、完全自由の実現は維持されるが、そうではなく、ネットワーク化構図には不可欠だとするならば、赤ラインの貢献は不可欠である。

他方、このツールを活用する場合、ルールを確定して、たとえば緑の円環で提示された上位3つの用語のみを選択する、とすると、解釈者の主体的な関与は消失し、単なる引っ張るマシーンになってしまう。これは、「ひっぱりくん」の基本的精神とは異なるが、しかし活用次第で、解釈者をマシーンに変えるものでもある。

要は、「ひっぱりくん」は、主体の自由（選択と探索）とツールの秩序（制約と提示）のバランスをいかにとるか、をケースごとに確定することが不可欠であり、そのバランスを求めることで、生成された構図が、主体とツール（＝もう一人の主体）との間主観的な了解を通して獲得された成果となる。したがって、解釈者は、構図を作成する時に、いかなる秩序を主体に課すか、そのためのルールをしっかりと確定しなければならない。そのルールが提示する「制約と提示」の秩序を前提として、自由な「選択と探索」を繰り返す時、その過程で生成される構図は、間主観的な了解という観点から十分に学術的に正当化される成果の位置を獲得しよう。

2) 価値表明と目標達成

起点ワードは、いかにして設定されるのか。これは、解釈者の価値の表明である。テキスト全体から、1つの用語を選択する基準は、外部から決定されることはない。ランダムに決めるものでないし、客観的に決定するルールは存在しない。ここは、解釈者の主観であり、事前にもつ構図のイメージをシンボリックに表現するであろう用語を選択することである。しかも構図のイメージは、解釈者

が最終的に生成されたいイメージであるから、主体が望ましいと確信する価値そのものである。その価値表明が起点になって、自由と秩序のバランスのもとで、探索が始動する。

しかし価値表明は、ある意味、起点ワードを確定するまでの行為で完結してしまう。それ以降の選択と探索の行為は、価値を意識しながら、その価値実現を目標として、その達成を求めて試行錯誤を繰り返す多難な行為の連続である。その成果が生成された構造図である。この成果は、価値表明（イメージ）との間で評価される。その評価過程は、イメージが不動の基準であり、その基準に沿っているかどうかで、成果が評価される、という一方向的な評価ではない。イメージ（予期された構造図）と成果（生成された構造図）は、成果を追求する過程で、何度も、お互いを修正することが繰り返される。イメージも、成果を追求する過程で、より具体化され、より多様化され、より多層化される。イメージと過程での成果は、その乖離を互いに埋めることで、間主観的（予期する主体と試行する主体）な了解に達するまで、試行錯誤を展開する。その結果、予期する構造図イメージと生成された構造図との間での乖離が縮小し、納得できる評価の妥当性が獲得されると、成果はここで確定し、構造図が完成する。要は、この時点でやっと目標は達成された。

3) 自明性と創発性

価値表明と目標達成の間で、納得いく合意・了解が得られた時、その構造図は生成され成果となって産出されたといえよう。しかしその成果を評価するもう1つの基準がある。それが、自明性の確認と創発性の発見のバランス問題である。もしも、予期した構造図がそのまま成果の構造図になったとしたら、それは、事前の構造イメージがツールによってそのまま反映されたことになる。つまりこれは100%自明性の確認である。これは、確認だけを求めたツール活用ならば、それなりに評価されようが、本来のツールの活用法ではない。このツールを活用する根拠は、事前の予期のイメージが、そこそこに支持されながら、他方、新しい何かの発見が求められている。つまり適度の裏切りが必要なのである。その裏切りが創発性の発見である。とすると、自明性の確認と創発性の発見のバランスが重要である。

単なるネットワークが生成されたのでは意味がなく、そのネットワークの中に、自明な部分と発見の部分の適度のバランスをもって存在するとき、この構造図は価値があるし、目標を達成したと評価されるのである。

4) 目的と手段

構造図は、そのネットワークの中に、なんらかの意味を創発させている。その意味解釈をする場合、ネットワーク全体はいかなる要素（意味集合）から構成されているのか、さらにその要素は目的と手段の関係として理解できるか、が重要である。特に、目的－手段関係が発見できるならば、ネットワークの意味解釈は容易である。

意味解釈の問題は、ネットワークの構造がいくつかの要素に分解でき、さらにその要素間に関係性が発見され、さらにはその関係が目的－手段とか因果関係として理解可能ならば、より良い解釈の方法論が創発されたといえよう。

今回のケースならば、「小さな成長戦略」の目的要素が発見され、さらにそれを説明する手段として、3つの要素（「雇用」「環境」「イノベーション」）が手段として発見された。このような方法論が構造図のネットワークのなかに発見できるならば、解釈は正当に可能になったといえよう。

構造図は、解釈されなければならない。起点ワードをはじめとして探索された用語の集合が単に存在する、というだけでは、解釈にはならない。価値表明との関連で、なぜ構造図は目標を達成するか、というと、それは、構造図のネットワークの中に、目的手段関係が顕在化するからである。それが意味の体系として存在するとき、はじめて構造図は価値を獲得する。

5) 階層とネットワーク

構造図が完成（解釈者が納得する）すると、その用語を階層構造に変換してみるといい。そうすることで、用語の階層が明示され、どの階層レベルで、どの用語を使用し、その同一階層では、どのような用語を探索しているか、が一目で理解可能である。ネットワークのままでは、理解不能な用語のつながりが、ここで整理されることで、全体の意味の体系を一気に理解することが可能になる。しか

も、ここでは、手段と目的の用語がどの階層で使用され、しかもそのような経路をたどりながら、手段から目的への推移がなされたかが容易に判別可能である。しかも下位構造にまで言及すると、基本的用語以外の派生用語がどこで探索されているかも判明し、納得し理解することが容易になる仕組みになっている。こうすることで、構造にポジショニングされた用語が、全体の構造化にとって、どれくらい貢献するのか、またしないのか、という個別用語の評価も容易にできるようになる。対照的に、あまり深いコミットなしに、何気なく選択した用語が、構造化の段階では重要なコンセプトになることもあるはずである。

構造図におけるネットワークは、この階層構造に準拠しながら再検討すると、階層の中の経路として浮かび上がってくるので、当初の解釈とは異なる再解釈も可能となり、新しい視点からの解釈が再構成されることにもなる。こうして、ネットワークと階層が表裏の関係になることで、解釈は深化し、より納得される解釈への変貌するのである。

6) 一般と特殊

これは、用語の抽象度のレベルが高いか低いか、である。より一般的であるほど、他の多くの用語と連動し、より特殊であるほど、他の多くの用語との連動の可能性は低く、ある特定の用語のみとの連動が強まる。この抽象度の高低を活用して、探索行為は、深化したり、多様化したりする。なので、探索行為の遂行において、どの階層での探索なのかを十分に意識して、用語の選択をしなければならない。安易に一般的な用語を選択すると、今までのネットワークを勝手に超えて、意味のない連動に発展し、解釈が不能になる事態を招きかねない。また、ある場合に、特殊すぎる用語を選択すると、その特殊性ゆえに、他の言葉への赤ラインがひかれず、単に孤立するような事態が発生し、解釈の対象としては意味のない事態を引き起こすことになる。

このように、用語の選択と探索においては、つねに用語の抽象度に気遣いながら、実行しなければならない。そうしないかぎり、解釈は困難を極めるばかりになる。だからこそ、用語の選択と探索には、解釈者の熟慮と決断が求められる。解釈者は、その時の用語のネットワークと階層の状況をしっかりと熟慮し、さらにさまざまな試行をしながら、徐々に一手を慎重かつ確実に打つ決断しなければならない。その熟慮と決断の成果が、納得できるネットワーク構造図と階層構造なのである。

上記6点の関連性について、まとめると、つぎのようになろう。

まず、「ひっぱりくん」を使用することで、意識下にある漠然とした前思考（茫漠とした思い）を、上記の操作によって表出化させるには、「自由と秩序」が提示する、表出化のための用語の認知（気づき）をもたらす前提条件についてのバランス感覚を学習することが求められる。つまり、このバランスを学習しないかぎり、「ひっぱりくん」による表出化は不可能である。

と同時に、認知と並行して、「自明性と創発性」が喚起する表出化された構造図についての価値判断ができる能力も学習されなければならない。つまり構造図がもつ自明な側面と、新しい気づきをいわばチャンス発見として価値判断する側面について、その両者を判別できる能力が学習されないかぎり、構造図の理解ができたとはいえない。

この認知と価値判断の基準が学習されると、それによって、「ひっぱりくん」を操作することが可能になる。つまり起点として用語を選択することが、単なる偶然の操作ではなく、しっかりとした価値判断に準拠して行われ、さらに続く一連の操作についても、なんの目的もなく操作するのではなく、十分な思惑をもってすすめられ、その結果として、目的達成する時点まで、操作が進行しつづける。これが、「価値判断と目的達成」のフレームである。このような表出化の操作によって、暗黙知としてしか想定されていなかった前思考が、構造図として現前化し、新しい創発を伴って、新しい価値を誘発させる。

この構造図を生成させるには、次の2つの操作が必要とされる。それが、「目的と手段」と「一般と特殊」である。前者は、ひっぱり操作によって表出化された用語の関係性について、もっとも基本的な意味連関のルールである目的手段関係を当てはめることで、関係性の解釈を容易にさせる。また後者の「一般と特殊」は、表出された用語の抽象度を確定することで、関係性の理解において、包摂関

係を通して、解釈の一般性や多様化と個別性と特定化の切り口から、構造図の解釈を多次的に操作可能にする。

最後の「構造化とネットワーク化」は、この2つの操作の総合化であり、これによって、構造図から構造への変換によって全体の可視化とネットワークのパスが明確になることで、再解釈の繰り返しが可能になる。しかもここでは、目的手段関係と一般特殊関係を自在に駆使することで、解釈がより自在化し、それによって、表出化した構造が新しい創発を呼ぶことが十分に可能になる仕掛けになっている。

このように「ひっぱりくん」を活用することで、暗黙知の表出化が可能になり、そこでは新しい関係生成のフレームワークがみえることになる。

6-3-2 「こうぞうくん」の構造化ツールの価値探索機能

つぎに、「こうぞうくん」を活用すると、第4章で述べたように、つぎのような構造生成のフレームが発見されている。

表 13 「こうぞうくん」の機能一覧

分析の段階	分析の手法	ツール上の機能
使用の全体	－	セーブ／ロード
構造化	クラスタリング	基準変更
構造化	クラスタリング	統合
構造化	クラスタリング	分割
構造化	相関ルール	ソート
構造化	相関ルール	彩色
構造化	相関ルール	移動
構造化	相関ルール	Lock
構造化	相関ルール	削除
構造化	特化係数	彩色
概念化	クラスタリング	生成
概念化	相関ルール	ソート
概念化	相関ルール	結合

ここでは、構造化の段階と概念化の段階に区分して、価値探索機能の創発性についてのフレームが要約されている。ただし構造化と概念化は、ある意味、コインの両面のような関係であり、構造化された構図をネットワーク的に再構成するのが概念化であり、前述の表出化の6特性のうちで、構造化とネットワーク化に相当するのが、この段階区分である。

ここでの構造化は、一口でいえば、構造を構成する系統（クラスタ）に関して、系統内と系統間との最適なバランスを維持している状況のことである。したがって、系統を構成する個々の要素（アイテム）の視点からすると、そのアイテム間の関連性が意味的に十分に高く維持されることで、系統内での関連性が高く、しかも系統間では、その差異が計量的にも意味関連的にも十分に明確にみられる、という状態を示すものである。

そのような構造化を目指して、さまざまな操作が仕組まれているのが、この構造化ツールである「こうぞうくん」である。ここでは、この操作のもっとも基盤として位置づけられているのは「基準変更」である。これは、構造化のスタートであり、同時に最終のゴールである。その意味では、表出化ツールにおける「価値判断と目的達成」のバランスをとる機能をはたす操作である。

この変更基準を基盤に、一方では、クラスタリングの系列として、「統合」と「分割」の操作が重視され、他方、相関ルールの系列では、「移動」「ロック」「削除」の操作とその背操作の次のステップとして「ソート」と「彩色」が重視されている。これらの多様な操作を装備することで、研究主体の自由と選択の意思とマシンの合理的なルールと秩序とのバランスがここで図られる。このバランスは、表出化ツールの場合と同様であり、「自由と秩序」という認識と「自明と創発」という価値判断と

のバランスに対応している。ただ、表出化ツールとの違いは、自由と秩序とのバランスでは、あきらかにマシンの合理的・計算的な秩序要請の度合いが高く、研究主体の自由度はその分制約される。ただ、価値判断については、研究主体の関与は非常に重要であり、その意味では、解析の起点と終点についての価値判断は研究主体に保有されている。

このように構造化が達成されると、それは系統間のネットワーク化という観点から再解釈されることになる。その時、重視される操作が、系統の概念化であり、各系統は、それぞれ他の系統との差異と関連性（「結合」の操作）を意識してネーミングされ、それによって、系統間の構造とそのネットワークの意味が生成される。

このようにして、「こうぞうくん」という構造化ツールについても、個々の多様な操作を通じて、構造化に対して創発的な貢献を果たし、従来の客観的な調査ツールとは異なり、多様な価値探索ツールとしての創発特性をみせている。これこそ、このツールが社会調査の実査において貢献するオリジナルな価値である。

6-3-3 「ふかぼりくん」の深層化ツールの価値探索機能

第3の価値探索ツールである「ふかぼりくん」については、どうであろうか。このツールは、「こうぞうくん」よりも、さらに、確定された構造を前提として、そこでの深層化が志向されるツールであるので、前述した表出化ツールで提示された6点からなるフレームワークとの関連性では、あくまでも表面的には、より秩序志向のフレームワークになっている。

まず、「ふかぼりくん」ツールを活用した実査から得られた創発特性をもつ価値探索機能を列举すると、つぎの10機能となる(図 189)。

- ①統合機能
- ②絞込機能
- ③削除機能
- ④彩色機能
- ⑤ソート機能
- ⑥スケール機能
- ⑦生成機能
- ⑧付加機能
- ⑨リンク機能
- ⑩セーブ（ロード）機能

以下、簡単に説明する。

①統合機能

統合機能は、項目をまとめることで、項目それ自体としての価値が発揮できるように変換する機能である。

②絞込機能

絞込機能は、ある特定の項目に限定することで、他の項目との関連性を明確にする機能である。データを深堀して、より特定の条件を設定して、そこでの関連性を求めようとする、絞込機能の活用が非常に重要である。この機能を活用することで、テーマの追究が明確になり、だからこそ、そこから新しい価値探索が可能になる。

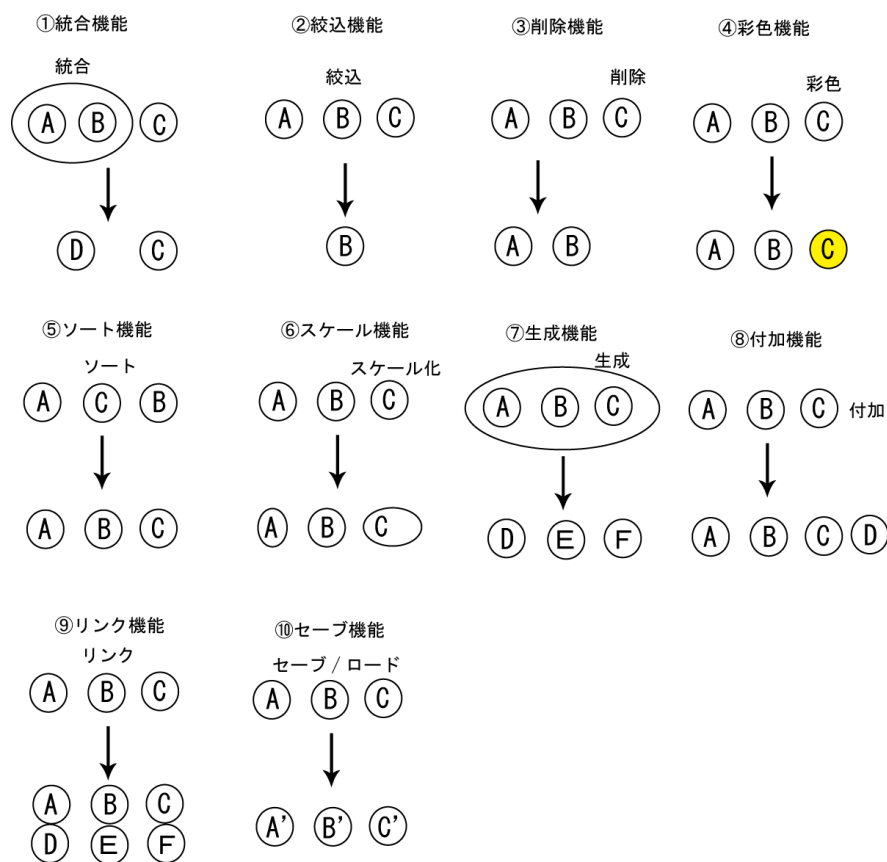


図 189 「ふかぼりくん」の価値探索機能概念図

③削除機能

削除機能は、項目のなかで、解析に必要なのない変項については、削除することが必要である。成果のより明確な関係を認識するためには、データとしては存在していても、結論を導くためには無用なデータについては、削除する必要がある。そうすることで、成果がより明確になる。もちろん、この削除機能は、特定条件での削除であるので、つねに復活は可能な仕組みにはなっている。だからこそ、容易に削除機能を活用して、多様な関係の確定が可能になる。

④彩色機能

これは、一見、意味のなさそうな機能であるが、実際にデータ解析を実行しているプロセスでは、非常に効果のある機能である。たとえば相関が明確なデータのクロス表などの場合、どの変項がその有効性を誘導しているかを確定するためには、彩色することで、その確定が明確な可視化として表示される。彩色機能は、このようなデータの可視化をより鮮明にするためには重要な機能を果たしている。

⑤ソート機能

ソート機能は、彩色機能と並んで、可視化の機能として重要である。平均や分散などの結果について、そこでの序列を変更する機能が容易に発揮できると、そこでの関連性が非常に容易に理解可能になる。それが、可視化の1つの重要機能である。ソート機能が、単純集計ばかりでなく、グループ化集計においても容易に活用できることは、成果の理解を強化する機能として高い価値をもつ。

⑥スケール機能

スケール機能は、変項の数値の規模をどこまでに設定するか、特に最小値をどこに設定することで、解析をするかを特定化する機能である。たとえば、グループ集計の場合ならば、各クラスタの特

性する変項の規模（最小値）をどこに設定するか、を確定することで、価値ある成果を探索しようとする機能である。そこでの規模感をどこに設定するか、が研究者の探索したい成果を直接関係するので、非常に重要な機能である。もちろん、この機能は数値データの場合だけでなく、画像データでもフリーアンサーのテキストデータの場合でも活用できる機能であり、とくに、このような非数値データの場合で重要な機能を果たしている。

⑦生成機能

生成機能は、新しい項目を産出する機能であり、その典型がクラスタリングである。数種の項目から新しい価値をもった項目を誘発させることで、解析段階を一段アップさせる機能である。ここでは、何度も試行錯誤が繰り返させることが可能になっており、最適な、納得できるクラスタが創発させることが可能になっている。この生成機能があることで、既存のデータから、さらなる高度なデータの生成が創発される仕組みになっている。そうすることで、価値探索がさらに促進することが可能になっている。

⑧付加機能

生成機能が発揮されれば、それをすぐに新しい項目として付加して、さらなる多様な解析が可能になる仕組みが必要である。それが付加機能である。これによって、生成機能での成果が次の解析に継続的・連続的に実行可能になる。

⑨リンク機能

リンク機能は、数値データを非数値データ（画像データとテキストデータ）に関連づける機能である。従来の社会解析ツールでは、数値データと非数値データとは独立して解析することになっているが、ここでは、その関連性を明確にすることで、データの個別具体的な解釈を付加させることが可能になり、価値探索のレベルが全く異質なものにもっていくことが可能になっている。これによって、比較分析が、画像とテキストが付加されることで、非常に深い解釈が実現するようになっている。

⑩セーブ（ロード）機能

最後はセーブとロード機能である。上記の諸機能を活用することで、データの構造が変更されるので、それを新しいデータ構造としてセーブすることが必要である。その機能がセーブ機能であり、それによって、何度も解析を繰り返す過程で、次のロード機能が活きてくる。

以上の機能を駆使するとき、「ふたぼりくん」が価値探索をする過程で可能となる分析は、つぎの比較多層分析とマルチデータ分析の2つの分析手法である。

1：比較多層分析

どの社会調査でも、比較分析は当たり前であるが、ここでは、統合・絞込・削除機能を活用することで、深堀を何度も繰り返すことで、表面的な比較分析のレベルから、さらに特定化する過程で、データの統合と絞込を展開し、無用なデータは削除することで、深堀が可能になる。そうすることで、通常の社会分析では得られない深い階層での解析が可能になり、新しい価値の探索が可能になる。これを、比較多層分析と呼ぶ。

2：マルチデータ分析

「ふかぼりくん」は、多層な比較分析をするツールであると同時に、もう1つ、非数値データも併用して扱うことができるツールなので、数値データと非数値データとの複合的な分析を可能にしている。それが、画像データとテキストデータと数値データを、リンク機能を活用して解析する手法である。

この2つの分析手法は、従来の計量的な社会調査では、ほぼ不可能とされ、そこは、いわゆる定性調査の領域とされ、計量的には踏み込まない領域であった。しかし「ふかぼりくん」を活用すると、定性調査の領域にまで踏み込むことが可能となり、従来の社会調査の棲み分けはその境界を失い、シームレスに連続する形で位置づけることが可能になる。これは、従来の社会調査からすれば活気的なことである。つまり「ふかぼりくん」は、定量調査と定性調査の領域を跨ぐことで、両方の調査手法を媒介する機能を果たす。これによって、今まで、社会調査として、まったく異質な社会調査手法と想定されたものが、一気に統合される可能性が、「ふかぼりくん」によって発見されたといえよう。

このような新しい社会調査の地平を想像することができるとすると、「ふかぼりくん」はその起点において、強く構造に規定されることで、秩序志向が前提となり、自由志向は排除された手法として位置づけられるが、しかし、そこでの手法をフルに活用し、比較多層かつマルチデータ分析の段階にま

で深層化を探索すると、そこには、既存構造の象徴的本質ともいえる理念型を探索することになり、そのことによって、逆に、既存の自明な構造から自由に飛躍して、新しい価値を創発させる可能性をもつ時点にまで到達するにいたるのである。これが定性調査の使命だとすると、ふかぼりくんの深層化の究極には、定性調査そのものに合致するフレームワークに到達するという、新しい社会調査の地平が、計量分析の極地として想定可能になる。これこそ、第三の価値探索ツールである「ふかぼりくん」が求めていた社会調査への多大な貢献なのである。

6-4 今後の課題

本論文のツールは、まだ至らないところが多々あり、本格的に使われていくためには、いくつかのハードルがある。1つは、ドキュメントの整備である。ツールの使い方、考え方をしめすドキュメントが開発中ということもあって、圧倒的に不足している。多くの人が使うためには、ツールのドキュメント化と使用事例のドキュメント化が必要であろう。特に、Rawler は、使う人が限られるものなので、英語ドキュメントの拡充が必要であろう。

2つには、分析者の能力に依存するツールであるので、使い方のフレームを作る必要がある。例えば、「ひっぱりくん」の実査において、ルールを定めたようなことが必要である。使い方のフレームを作るためにも、様々なデータに対しての分析事例を増やしていく必要があるであろう。

3つには、先に述べたように、「叙述化」の機能の弱さの補強が必要であろう。すでに、彩色機能、命名機能があるが、もっと分析者自身がデータに対して気づきの目印をつける仕組みを強化した方がいい。現在は、外部のツールを使い、「叙述化」を行うこととなっているが、ツールとして内部的にもったほうがよく、それをさらに分析に使える仕組みができれば理想的である。また、「叙述化」をサポートする意味で、分析者の意識の流れに対するログを取っていく、状態の変化を記録していく仕組みも必要であろう。「ひっぱりくん」には、一部その機能があるが、その強化版が必要であろう。

4つには、比較の機能の拡充である。ものの理解は差異から生まれるのであり、比較のコンセプトを様々なところで仕込む必要があるだろう。「ふかぼりくん」の成功は、深掘りによる比較を可能としたことであり、このことを他のツールにも取り入れるべきである。例えば、「こうぞうくん」において、「ふかぼりくん」のように、データの調整を行いある特定の条件での結果を比較できるようにすべきであり、それを含めた仕組みにすべきである。

5つには、各ツールの融合である。それぞれのツールは今のところそれぞれは関連していない。「ひっぱりくん」と「こうぞうくん」は、同じ共起関係を使った分析の主観的、構造的といった視点を変えて行っているものである。そのため、二つを融合することは、二つの視点を持つこととなり、有用である。また、「ふかぼりくん」のアンケート分析における、自然言語分析にも、「ひっぱりくん」と「こうぞうくん」の仕組みは有用であり、組み込みたいと思う。また、それぞれのツールにおいて、柔軟なデータの inputs を可能とするためにも、「Rawler」から inputs をできるという仕組みにすることにも意味があるだろう。

6つには、共同作業的な仕組みである。Web で使えるデータ分析ツールとはいえ、一人が使うことを前提としている。Web を使っているのだから、共有しながら分析の作業をする仕組みというものがあると便利であり、分析者の主観と主観の混じり合いは、新たな創発を生むことになるだろう。ツールの性質として、一つの解に収斂していくものというよりも、複数の解釈に発散していくものなので、継承、分岐を行える、多様な解釈をサポートするための共同作業をおこなうツールであることが望ましい。これは、共有する場所の提供や、共同作業時の動作の決定、データのマージなどを考えると膨大な作業となるので、あくまで理想でしかない。

以上、課題としては、ドキュメントの整備、使い方のフレームワークづくり、叙述化機能の強化、比較機能の深化、それぞれのツールの融合があり、願望的なものとして、共同作業的な仕組みがある。

謝辞

本論文をまとめるにあたり、多くの方々にご助力いただきましたことを、心より感謝申し上げます。博士課程への進学および研究全般にわたる多大なご支援、ご指導を賜りました指導教授、熊坂賢次先生には、どれほど言葉をつくしても感謝の気持ちを十分に表すことはできないほど、お世話になりました。ありがとうございました。引きこもりがちで、どうしようもない私を根気強く導き、私のバグだらけのソフトウェアの初めのユーザとして、さまざまな助言をいただき、半学半教の精神をもって、学生からも学ぼうとする、研究者としてのあるべき姿についてもご教授くださったことに重ねて感謝いたします。

古川康一先生、諏訪正樹先生、井庭崇先生には、異分野であるにも関わらず快く受け入れてくださったことに、感謝いたします。特に、諏訪先生の身体的メタ認知の考え方に本論文は強く影響を受けるものであり、諏訪先生には研究会等にお誘いいただき、良い刺激を受けました。感謝いたします。それぞれ3人の先生方は、自分自身で研究分野を切り開いており、このことには強く影響を受け、尊敬をしています。様々な形でアドバイスをや励ましをしてくださり、深い感謝をいたします。SFC という場のすごさを感じるところであります。

研究室の先輩であり、現、産能大学准教授である、小野田哲弥先生にもお世話になりました。小野田先生が、柔らかい構造化手法を切り開いていったおかげで、僕の手法があります。研究室の後輩であり、現、産能大学講師の白土由佳先生にもお世話になりました。「こうぞうくん」は、白土先生の博士論文にある可視化がベースであり、その改良へのアドバイスをいただきありがとうございます。「Rawler」と名付けたのも彼女です。本論文についても、構成についてのアドバイスをたくさんいただきました。重ね重ねありがとうございます。

Rawlerを作るきっかけや、Rawlerでのメタプログラミング化をするきっかけとなる問題を持ってきた、研究室の後輩の中村圭佐さんにも、感謝いたします。彼の叱咤激励なしには、作ることはありませんでした。

私の所属する、熊坂研では、毎年、SFC オープンリサーチフォーラム (ORF) への研究成果の発表のため、学部生の4年生を中心に自分たちの研究したいことを研究させるということをしています。この場は、社会分析のためのツールを作るという私の研究の実験場でもありました。毎年毎年、それぞれの学部生が様々なテーマで研究を行ったので私もとても刺激になり勉強になりました。本論文の研究は、この ORF にむけて、研究室の学部生とともに作り上げていったものでもあります。重ね重ね、熊坂研究室のみなさまには感謝いたします。特に、本論文で使用した事例研究を行った、「ひっぱるくん」を使い、自身の味覚の探索を行った、花房真理子さん。「こうぞうくん」を使い、ドラマ「半沢直樹」の分析を行った、八木梨花子さん、熊澤大吉さん、内山貴博さん。「GooCaBooca」を使い「モテる服装」の調査を行った、山田幸さん。事例対象の選び方のセンスから、私のツールを十二分までに使っていただいたことに感謝します。名前を挙げた、彼ら彼女ら以外の研究室の卒業生・現役生にも感謝いたします。そして、このような場を作ってくださった熊坂先生には感謝いたします。

最後に、これまで自分の思う道を進むことに対し、温かく見守りそして辛抱強く支援してくださった両親に対しては深い感謝の意を表して謝辞と致します。

参考文献

- [1] 鈴木謙介・電通消費者研究センター(2007)『わたしたち消費—カーニヴァル化する社会の巨大ビジネス』幻冬舎.
- [2] 宮台真司・大塚明子・石原英樹(1993)『サブカルチャー神話解体—少女・音楽・マンガ・性の30年とコミュニケーションの現在』PARCO出版.
- [3] リースマン, D. 加藤秀俊訳(1964)『孤独な群集』みすず書房.
- [4] 熊坂賢次・山崎由佳(2011)「おしゃべりなロングテールの時代--東京ガールズのネットコミュニティ解析」『法学研究』, pp.530-501.
- [5] ブルデュー, P. 石井洋二郎訳(1990)『ディスタンクシオン <1> 社会的判断力批判』藤原書店.
- [6] 宮台真司(1994)『制服少女たちの選択』講談社.
- [7] ラパポート, S. 電通ソーシャルメディアラボ訳(2012)『リッスン・ファースト! ソーシャルリスニングの教科書』翔泳社.
- [8] 萩原雅之(2011)『次世代マーケティングリサーチ』ソフトバンククリエイティブ.
- [9] ジェームズ・スロウィッキー, 小高尚子訳 (2006)『「みんなの意見」は案外正しい』, 角川書店
- [10] 東浩紀(2011)『一般意志 2.0』講談社
- [11] 小野田哲弥(2007)「柔かい構造化モデル—ロングテール事象のカテゴリーマネジメント支援—」『慶應義塾大学博士学位論文』, pp.1-354.
- [12] 妹尾紗恵(2008)「日本の家庭料理の構造と創発」『慶應義塾大学博士学位論文』, pp.1-235.
- [13] 山崎由佳(2012)「ソーシャルメディアを活用した「おしゃべりなロングテール」をめぐる方法論的展開と実態分析」『慶應義塾大学博士学位論文』, pp.1-351.
- [14] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth, (1996), From data mining to knowledge discovery in databases, AI magazine 17.3.
- [15] 古川康一・植野研・尾崎知伸(2001)『帰納論理プログラミング』共立出版.
- [16] McCarthy, J., Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence". Machine Intelligence 4: 463–502.
- [17] Polanyi Michael(1983), The Tacit Dimension, Gloucester, Mass, Peter Smith (佐藤敬三訳(1998)『暗黙知の次元』紀伊国屋書店)
- [18] 諏訪正樹(2005)「身体知獲得のツールとしてのメタ認知的言語化」『人工知能学会誌』, 20(5), pp. 525-532.
- [19] 中島秀之・諏訪正樹・藤井晴行(2008)「構成的情報学の方法論からみたイノベーション」『情報処理学会論文誌』, 49(4), pp.1508-1514.
- [20] 大澤幸生(2003)『チャンス発見の情報技術』東京電機大学出版局.
- [21] 野中郁次郎(1990)『知識創造の経営—日本企業のエピステモロジー』日本経済新聞社.
- [22] 川喜田二郎(1967)『発想法—創造性開発のために』中公文庫.
- [23] 岡部正幸・山田誠二(2013)「知的インタラクティブシステムにおけるインタラクションデザインとは何か」『人工知能学会全国大会論文集』27.
- [24] Philip Kotler, Kevin Lane Keller, 恩蔵 直人 (監修), 月谷 真紀 (翻訳)(2014)「コトラー&ケラーのマーケティング・マネジメント 第12版」, 丸善出版
- [25] 砂山渡・高間康史・西原陽子・徳永秀和・串間宗夫・阿部秀尚・梶並知記「テキストデータマイニングのための統合環境 TETDM の開発」『人工知能学会論文誌』, 28(1), pp.1-12.
- [26] Gancarz, M. 芳尾桂訳(2001)『UNIXという考え方—その設計思想と哲学』オーム社.
- [27] Johnston, Wesley M.; J.R. Paul Hanna; Richard J. Millar (March 2004). "Advances in Dataflow Programming Languages". ACM Computing Surveys 36: 1–34.
- [28] Hughes, J. (1990), Why Functional Programming Matters, In D. Turner, editor, Research Topics in Functional Programming, Addison Wesley.
- [29] Gul Agha. (1985), Actors: A Model of Concurrent Computation in Distributed Systems. MIT Press

- [30] Anderson, C. 星睦訳(2007)『エッセンシャル WPF:Windows Presentation Foundation (Programmer's SELECTION—Microsoft .net Development Series)』翔泳社.
- [31] Hazzard, K. Bock, J. 長尾高弘訳(2013)『メタプログラミング.NET』アスキー・メディアワークス.
- [32] Anderson, C. (2004), The Long tail, Wired.
- [33] 増田直紀・今野紀雄(2005)『複雑ネットワークの科学』産業図書.
- [34] Zipf, G.K. (1949), Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology, Addison-Wesley.
- [35] Barabási, A. L. (2002), Linked: The New Science Of Networks, Basic Books.
- [36] 山崎由佳・熊坂賢次(2012)「共有化と生活化から生成される 2 つの”かわいい”: 4 ファッションスタイルをめぐるネットコミュニティ分析」『ファッションビジネス学会論文誌』17, pp.25-42.
- [37] Wagstaff, K. and Rogers, S. (2001), Constrained k-means clustering with background knowledge, Proceedings of the Eighteenth International Conference on Machine Learning, 1, pp.577–584, .
- [38] Kohonen T. Self-organizing formation of topologically correct feature maps. Biol Cybern 43:59-69, 1982.
- [39] Breiman, L. Friedman, J.H. Olshen, R.A. and Stone, C.J. (1984), Classification and Regression Trees, Wadsworth.
- [40] MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).