

内積に基づくクラスタリングアルゴリズムの
非線形化とその応用に関する研究

システム情報工学研究科

筑波大学

2006年3月

水谷清隆

目次

第1章	序論	1
1.1	研究の背景	1
1.2	関連研究	5
1.3	研究の意義	6
1.3.1	内積に基づく類似度	6
1.3.2	クラスタリングアルゴリズムの非線形化の必要性	6
1.3.3	本研究の目的	7
1.4	本論文の構成	8
第2章	c -平均クラスタリングとその拡張手法	10
2.1	クラスタリングと類似性測度	10
2.2	コサイン相関	11
2.3	クリスプ c -平均法	13
2.4	ファジィ c -平均法	16
2.4.1	標準的ファジィ c -平均法	16
2.4.2	エントロピー正則化ファジィ c -平均法	18
2.5	ファジィ分類関数	21
2.6	クラスターサイズ調整変数	24
2.6.1	クラスターサイズ調整変数付ファジィ c -平均法	24
2.6.2	クラスターサイズ調整変数付クリスプ c -平均法	26
第3章	クラスタリングアルゴリズムの非線形化	29
3.1	はじめに	29
3.2	カーネル法	30
3.2.1	カーネル関数	30
3.2.2	カーネル・トリック	32
3.3	カーネル関数を利用したクラスタリング	33
3.3.1	カーネル関数を利用したクリスプ c -平均法	33
3.3.2	カーネル関数を利用したファジィ c -平均法	35

3.3.3	カーネル関数を利用したクラスターサイズ調整変数付クリスプ c -平均法	37
3.3.4	カーネル関数を利用したクラスターサイズ調整変数付ファジィ c -平均法	37
3.4	数値例	39
3.5	まとめ	42
第 4 章	競合学習モデルに基づくクラスタリング	43
4.1	はじめに	43
4.2	競合学習クラスタリング	43
4.3	ファジィ競合学習クラスタリング	45
4.4	競合学習クラスタリングの非線形化	47
4.4.1	カーネル・競合学習クラスタリング	47
4.4.2	カーネル・ファジィ競合学習クラスタリング	48
4.5	自己組織化マップによる高次元特徴空間におけるデータの可視化	51
4.5.1	自己組織化マップ	51
4.5.2	カーネル・自己組織化マップ	52
4.6	数値例	54
4.7	まとめ	66
第 5 章	ファジィマルチ集合に基づく文書情報クラスタリング	67
5.1	はじめに	67
5.2	ファジィマルチ集合	67
5.2.1	マルチ集合の性質と基本演算	68
5.2.2	ファジィマルチ集合の性質と基本演算	69
5.3	ファジィマルチ集合のノルムと類似度	71
5.3.1	ファジィマルチ集合のノルム	71
5.3.2	ファジィマルチ集合の類似度	72
5.4	ファジィマルチ集合による文書情報のモデル化	72
5.5	文書情報クラスタリング	75
5.5.1	数値例 1	75
5.5.2	数値例 2	79
5.6	まとめ	92
第 6 章	クラスタリングによる日本周辺の軍事情勢の分析	93
6.1	はじめに	93
6.2	クラスタリングによる分析	94
6.2.1	クラスタリングに用いるデータ	95

6.2.2	クラスタリングアルゴリズム	96
6.3	冷戦終焉期における世界とアジア地域の軍事情勢	101
6.3.1	クラスタリング結果	101
6.3.2	世界の軍事情勢の分析	106
6.3.3	東アジア地域の軍事情勢の分析	108
6.4	現在の世界とアジア地域の軍事情勢	111
6.4.1	クラスタリング結果	112
6.4.2	世界の軍事情勢の分析	117
6.4.3	東アジア地域の軍事情勢の分析	118
6.5	日本をとりまく軍事的リスク	122
6.5.1	東アジア地域における主要国の軍事情勢の変化	122
6.5.2	日本に対する軍事的脅威	127
6.6	クラスタリング結果の妥当性評価	129
6.6.1	主成分分析による妥当性評価	130
6.6.2	核兵器および弾道ミサイルの拡散・保有状況の変化による妥当性評価	133
6.7	まとめ	135
第7章	結論	138
7.1	まとめ	138
7.2	今後の課題と展望	141
	謝辞	143
	参考文献	144
	研究業績	153
付録A	文書情報クラスタリングに用いた文献のタイトル(第5章)	156
A.1	数値例1に用いた文献のタイトル	156
A.2	数値例2に用いた文献のタイトル	158
付録B	クラスタリングによる日本周辺の軍事情勢の分析: 参考資料(第6章)	164
B.1	クラスタリングに用いたデータ	164
B.1.1	1985年	164
B.1.2	2003年	176
B.2	クラスタリング結果($c=2 \sim c=4$): 1985年	188
B.3	クラスタリング結果($c=2 \sim c=4$): 2003年	191

B.4 各変数における上位 10ヶ国	194
B.4.1 1985 年	194
B.4.2 2003 年	196

目次

1.1	階層的クラスタリングのイメージ図	4
1.2	非階層的クラスタリングのイメージ図	4
1.3	ボロノイ図	6
2.1	内積の図解 1	12
2.2	内積の図解 2	12
2.3	クリस्प c -平均法により 2 つのクラスターに分類した例	15
2.4	ファジィ c -平均法により 2 つのクラスターに分類した例	20
2.5	標準的ファジィ c -平均法により求められたファジィ分類関数	23
2.6	エントロピー正則化ファジィ c -平均法により求められたファジィ分類関数	23
2.7	クラスターサイズが異なるデータ	24
2.8	ファジィ c -平均法によりクラスターサイズの異なるデータを 2 つのクラスターに分類した例	28
2.9	クラスターサイズ調整変数付ファジィ c -平均法によりクラスターサイズの異なるデータを 2 つのクラスターに分類した例	28
3.1	数値例：人工データ 1	30
3.2	数値例：人工データ 2	30
3.3	数値例：人工データ 1 — eFCM による分類結果	40
3.4	数値例：人工データ 1 — K-FCM による分類結果	40
3.5	数値例：人工データ 2 — eFCM による分類結果	41
3.6	数値例：人工データ 2 — K-FCM による分類結果	41
4.1	クラスタリングのための競合学習モデル	44
4.2	数値例：人工データ 1 — CCL による分類結果	55
4.3	数値例：人工データ 1 — K-CCL による分類結果	55
4.4	数値例：人工データ 1 — K-SOM により生成されたマップ	55
4.5	数値例：人工データ 2 — FCCL による分類結果	56
4.6	数値例：人工データ 2 — K-FCCL による分類結果	57
4.7	数値例：人工データ 2 — K-FCCL によるファジィ分類関数	57

4.8	数値例：人工データ 2 — ユークリッド距離の 2 乗を用いた K-FCM による分類結果	58
4.9	数値例：人工データ 2 — K-SOM により生成されたマップ	58
4.10	数値例：人工データ 3 — CCL, FCCL による分類結果	59
4.11	数値例：人工データ 3 — K-CCL, K-FCCL による分類結果	60
4.12	数値例：人工データ 3 — K-SOM により生成されたマップ	60
4.13	数値例：人工データ 4 — CCL による分類結果	61
4.14	数値例：人工データ 4 — K-CCL による分類結果	61
4.15	数値例：Iris データの 2 次元散布図	62
4.16	数値例：Iris データ — SOM により生成されたマップ	65
4.17	数値例：Iris データ — K-SOM により生成されたマップ	65
4.18	数値例：Iris データ — SOM (ユークリッド距離の 2 乗) により生成されたマップ	65
4.19	数値例：Iris データ — K-SOM (ユークリッド距離の 2 乗) により生成されたマップ	65
5.1	数値例 — 分類結果	75
5.2	数値例 1 — eFCM による分類結果	77
5.3	数値例 1 — K-FCM による分類結果	77
5.4	数値例 1 — FCMA による分類結果	78
5.5	数値例 1 — K-FCMA による分類結果	78
5.6	数値例 1 — FCCL による分類結果	78
5.7	数値例 1 — K-FCCL による分類結果	78
5.8	数値例 2 — 分類結果 ($c = 2$)	82
5.9	数値例 2 — 分類結果 ($c = 3$)	82
5.10	数値例 2 — 分類結果 ($c = 4$)	83
5.11	数値例 2 — 分類結果 ($c = 5$)	84
5.12	数値例 2 — 分類結果 ($c = 6$)	85
5.13	数値例 2 — 各クラスに属する文献の数	87
5.14	数値例 2 — CCM による分類結果	90
5.15	数値例 2 — CCL による分類結果	90
5.16	数値例 2 — FCCL による分類結果	91
5.17	数値例 2 — K-FCCL による分類結果	91
6.1	クラスタリング結果 (1985 年)	102
6.2	クラスタリング結果 (1985 年) — 世界地図	103

6.3	国防費の比較（1985年）	105
6.4	現役兵力の比較（1985年）	105
6.5	主要陸上戦力の比較（1985年）	105
6.6	主要海上戦力の比較（1985年）	105
6.7	主要航空戦力の比較（1985年）	106
6.8	NATO 諸国とソ連の軍事能力の比較	107
6.9	東アジア地域の国防費の比較（1985年）	109
6.10	東アジア地域における兵力の比較（1985年）	109
6.11	東アジア地域における陸上戦力の比較（1985年）	110
6.12	東アジア地域における海上戦力の比較（1985年）	110
6.13	東アジア地域における航空戦力の比較（1985年）	111
6.14	クラスタリング結果（2003年）	113
6.15	クラスタリング結果（2003年）— 世界地図	114
6.16	国防費の比較（2003年）	115
6.17	現役兵力の比較（2003年）	115
6.18	主要陸上戦力の比較（2003年）	116
6.19	主要海上戦力の比較（2003年）	116
6.20	主要航空戦力の比較（2003年）	116
6.21	東アジア地域の国防費の比較（2003年）	118
6.22	東アジア地域における兵力の比較（2003年）	119
6.23	東アジア地域における陸上戦力の比較（2003年）	120
6.24	東アジア地域における海上戦力の比較（2003年）	120
6.25	東アジア地域における航空戦力の比較（2003年）	121
6.26	東アジア地域における主要国の国防費の比較	122
6.27	東アジア地域における主要国の一人当たりの国防費の比較	123
6.28	東アジア地域における主要国の GDP に占める割合の比較	123
6.29	北朝鮮における GDP の推移	124
6.30	東アジア地域における総兵力の比較	125
6.31	東アジア地域における主要国の人口に占める総兵力の比較	125
6.32	東アジア地域における主要国の陸上戦力の増減	126
6.33	東アジア地域における主要国の海上戦力の増減	126
6.34	東アジア地域における主要国の航空戦力の増減	127
6.35	主成分分析の結果 — 固有値と累積寄与率	131
6.36	主成分分析の結果 — 主成分負荷量	131
6.37	主成分分析の結果 — 主成分得点の分布	132

6.38	主成分分析の結果とクラスタリング結果を融合した図	134
6.39	主成分分析の結果とクラスタリング結果を融合した図 — 拡大図	134
B.1	1985年のクラスタリング結果 ($c = 2$)	188
B.2	1985年のクラスタリング結果 ($c = 2$) — 世界地図	188
B.3	1985年のクラスタリング結果 ($c = 3$)	189
B.4	1985年のクラスタリング結果 ($c = 3$) — 世界地図	189
B.5	1985年のクラスタリング結果 ($c = 4$)	190
B.6	1985年のクラスタリング結果 ($c = 4$) — 世界地図	190
B.7	2003年のクラスタリング結果 ($c = 2$)	191
B.8	2003年のクラスタリング結果 ($c = 2$) — 世界地図	191
B.9	2003年のクラスタリング結果 ($c = 3$)	192
B.10	2003年のクラスタリング結果 ($c = 3$) — 世界地図	192
B.11	2003年のクラスタリング結果 ($c = 4$)	193
B.12	2003年のクラスタリング結果 ($c = 4$) — 世界地図	193

表目次

4.1	数値例：Iris データを用いた実験結果	64
4.2	数値例：Iris データを用いた実験結果 (ユークリッド距離の 2 乗を用いた場合)	64
5.1	数値例 2 — クラスタリング結果	81
5.2	数値例 2 — クラスタリング結果 ($c = 2$)	82
5.3	数値例 2 — クラスタリング結果 ($c = 3$)	82
5.4	数値例 2 — クラスタリング結果 ($c = 4$)	83
5.5	数値例 2 — クラスタリング結果 ($c = 5$)	84
5.6	数値例 2 — クラスタリング結果 ($c = 6$)	85
5.7	数値例 2 — 各クラスに属する文献の数	87
5.8	数値例 2 — CCM と CCL の比較	88
5.9	数値例 2 — CCL と FCCL の比較	88
5.10	数値例 2 — FCCL と K-FCCL の比較	89
5.11	数値例 2 — CCM による分類結果	90
5.12	数値例 2 — CCL による分類結果	90
5.13	数値例 2 — FCCL による分類結果	91
5.14	数値例 2 — K-FCCL による分類結果	91
6.1	ユークリッド距離の 2 乗を用いた場合のクラスタリング結果 (2003 年のデータ)	97
6.2	2003 年における国防支出費の上位 10 カ国	97
6.3	コサイン相関を用いた場合のクラスタリング結果 (2003 年のデータ)	98
6.4	日本における 2003 年の軍事データの一部と正規化した値	99
6.5	クラスタリングに使用した各種パラメータ	100
6.6	クラスタリング結果 (1985 年)	103
6.7	クラスタリング結果 (2003 年)	112
6.8	世界の核兵器保有状況とクラスタリング結果	135

第1章 序論

1.1 研究の背景

近年のコンピュータの高性能化およびそれに伴うネットワーク化によって、高度情報化社会が形成され、その中で多くのデータ化された情報が、国家や企業あるいは大学や研究機関などのデータベースに絶え間なく蓄積され続けている。このような膨大なデータを効率よく取り扱うためには、コンピュータによるデータ解析は非常に有効な手段である。特に近年、データマイニング (data mining)¹ と呼ばれる技術に注目が集められており、様々な研究がなされている [1, 30, 111]。データマイニングとは、統計学やパターン認識、人工知能等のデータ解析の諸技法を大量のデータに網羅的に適用することで知識を取り出す技術のことで、通常のデータの扱い方からは想像が及びにくい、発見的 (heuristic) な知識獲得を目的として使用される。

データマイニングに注目が集められるようになった背景には前述したように、安価で大容量の記憶装置が開発され、ネットワーク環境が整備されたことによって、大量のデータ蓄積が可能となったことが直接的に関係している。デジタル形式でのデータの収集は、コンピュータを用いてデータ解析をすることを念頭に置いて 1960 年代には既に行われつつあったのだが、1990 年代に至りデータ量は爆発的に増大したため、従来の解析方法では追いつかなくなってしまった。これに伴い、データベースにおける大量データを処理するための手法として、1995 年頃にデータマイニングの概念が現れ、統計的手法や人工知能分野での検索技術等が応用されるようになったのである [1, 111]。

データマイニングはデータの中から価値ある情報を見つけ出すための技術であり、様々な手法が用いられている。一般的によく知られているニューラルネットワークなどはパターン認識の分野における代表的なデータマイニング手法であり、多変量解析に用いられている主成分分析や因子分析などはデータ解析の分野における代表的な手法である。つまり、データマイニングとは全く新しい技術ではなく、様々な研究の融合した領域であると言える。

クラスター分析 (cluster analysis) はこのようなデータマイニングの諸技法の一つであり、クラスタリング (clustering) とも呼ばれている。クラスタリングは、解析対象となるデータを外的基準なしに自動的に分類を行うという特徴を持つ [34, 76, 90]。外的基準がない、すな

¹海外では Knowledge - Discovery in Databases (データベースからの知識発見) の頭文字をとって KDD とも呼ばれている。詳細については [26, 27] を参照されたい。

わち教師無し分類手法であるため、分類すべきデータ間に定義された類似性や距離のみに基づいてグループ分けが行われる。大規模データベースに蓄積され続ける膨大なデータを分類することは、迅速な知識発見作業に繋がるため、現在、データマイニングの分野において最も注目が集められている手法の一つでもある。

クラスタリングの諸技法は、20 世紀後半から盛んに研究がされるようになり、その応用は広範囲に渡る [2, 92]。例えば、医学の分野では、患者を治療する際に明確な診断をするために症候群別に分類したり [25, 38, 109]、考古学の分野では、発掘された陶器のかけらや石器等を類似のものからなるグループに分けるのに用いられている [39, 40, 41]。また、マーケティングの分野では、都市の商業的、人口学的な変数からなるデータを用いて都市のグループ化が行われたり [29, 33]、企業の顧客データベースから顧客情報を取り出す際にもクラスタリングが用いられている [8, 15]。さらに、文字認識や画像認識、情報検索などの研究にも応用されており [71, 94, 96, 97]、最近ではクラスタリングを用いた情報検索システムなども構築されている [16]。

さて、このようなクラスタリング技法には、階層的クラスタリング (hierarchical clustering) と非階層的クラスタリング (non-hierarchical clustering) の 2 種類がある。

- 階層的クラスタリング

類似度の大きいデータを 1 組ずつ結合していき、小さなクラスターから次第に大きなクラスターを生成していく手法。出力結果はデータの樹形図²で表される。

- 非階層的クラスタリング

クラスターの数をあらかじめ指定し、データ間の距離などを分類の基準としてクラスターを生成していく手法

階層的クラスタリングは、図 1.1 に示されているような樹形図により結果を表示するので、単に対象をいくつかのクラスターに分類するだけでなく、どのようにクラスターが結合されていくかの過程までも見ることができる。そのため、非常に直感的なアウトプットが得られるという特徴を持っており、様々な分野において用いられている手法である [91]。しかしながら、分類する対象が非常に多い場合には計算量が膨大な量となり実行不可能となったり、結果が不安定になってしまうという欠点がある。このような場合、つまり大量のデータを取り扱う場合には、非階層的なクラスタリング手法の方が有効であるということが知られている [92]。本研究では、この非階層的なクラスタリング手法を中心に取り扱う。

非階層的クラスタリングの代表的な手法に c -平均法 (c -means) がある。 c -平均法は、クラスター中心と各データとの距離などを基準としてクラスタリングを行う手法である。また、この c -平均法に分割のあいまいさであるファジィ理論 [113] を導入したのとして、ファジィ c -

²デンドログラムとも呼ばれる。

平均法 (fuzzy c -means) がある。ファジィ理論を導入することにより解析対象となるデータにあいまいな分類を許すことが可能となり、より柔軟にデータを取り扱うことができるようになった。そのため、ファジィクラスタリングは、現在、クラスタリングの分野において最も研究が盛んに行われている手法でもある [22, 31, 42]。

ファジィ c -平均法は、Bezdek らによって提案された手法 [6] (本研究では標準的ファジィ c -平均法と呼ぶ) が最も良く知られており、多くの研究者によってその応用研究が行われている [7, 17, 36, 63, 108]。また、宮本らによって提案されたエントロピー正則化法によるファジィ c -平均法 [74, 75] は、エントロピー関数を利用してファジィ化するものであり、様々な点において拡張性が高いことが知られている [44, 77, 100]。

さて、先にも述べたように、クラスタリングを行う際には、データ間に定義された類似性や距離に基づいてグループ分けが行われる。この類似性や距離は関連性の尺度とも呼ばれており、この尺度を表すものとして類似度 (similarity) あるいは非類似度 (dissimilarity) と呼ばれるものがある [76]。類似度とは、その値が大きければ大きいほど個体同士はより似ていると判定される測度であり、非類似度とは、その値が小さければ小さいほど個体同士はより似ていると判定される測度である。また、これらの値はそれ自身との値が最も大きく (小さく)³ かつ対称である必要があるとされている [76]。

Bezdek らによる標準的ファジィ c -平均法 [6] では、ユークリッド空間が仮定され、その上で最適化問題が考察されたので、使用する非類似度は自然に定まった。しかしながら、ファジィ c -平均法の変形 [43, 44, 74, 75, 100] では、非類似度の選択は自明ではなく、様々な尺度を用いることが可能である。本研究では関連性の尺度として、データクラスタリングの分野においてあまり研究がされていない、内積に基づく類似度に焦点を当てている点が特徴である。

³類似度を用いた場合はそれ自身との値は最も大きくなり、非類似度を用いた場合はそれ自身との値は最も小さくなる。

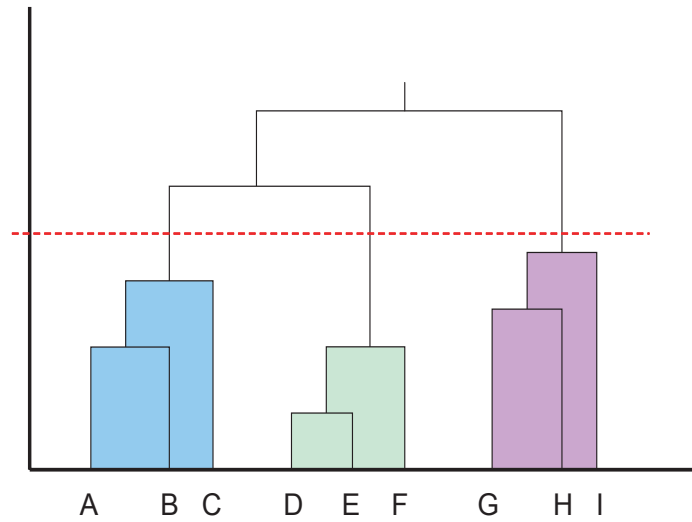


図 1.1: 階層的クラスタリングのイメージ図．縦軸は結合のレベルを表しており，ここでは，結合のレベルを赤の点線でカットしたとき，青，緑，紫の 3 つのクラスターが生成される様子が示されている．

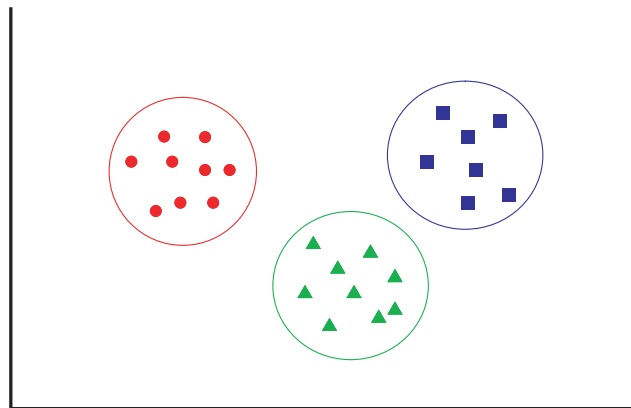


図 1.2: 非階層的クラスタリングのイメージ図．ここでは，各個体がそれぞれ ， ， の 3 つのクラスターに分類されている様子が示されている．

1.2 関連研究

内積に基づく類似度を利用したクラスタリングについては、情報検索の分野においてよく研究されている。特に、コサイン相関と呼ばれる類似度は、文書と用語あるいは文書間同士の類似尺度を計算する際によく用いられている [59, 60, 94, 96, 97]。しかしながら、データクラスタリングの分野では、分類性能において内積は距離に劣るとされているため、あまり用いられていない。なぜならば、原点からの同一直線上のデータは全て同じであるとみなされるからである。

また、従来の c -平均クラスタリング手法では分類境界面が必ず線形になり、しかも、その分類境界面は各クラスター中心への最近隣分類規則（最近中心分類）に基づいて形成されるため、各クラスター中心を結ぶ垂直二等分線から構成されるボロノイ境界を形成する（図 1.3 参照）。そのため、クラスターサイズが異なるようなデータあるいは線形分離不可能なデータに対しては最適な分類を行うことができなかった。

そこで、宮本らはクラスターのサイズを調整する変数を導入した新たなアルゴリズムを提案し、クラスター領域が異なる場合においても適切な分類境界を得ることに成功した [77]。また、非線形サポートベクトルマシン [106, 107] に用いられているカーネル法を利用したクラスタリングアルゴリズムを提案することにより、非線形な分類境界を求めることに成功している [81, 82]。しかしながら、これらの提案手法においては、データ間の非類似度としてユークリッド距離の 2 乗が用いられているため、内積に基づく類似度を利用した際には式の導出方法が異なり、クラスタリングアルゴリズムに用いられている目的関数や最適解の再定式化が必要となる。

また、井口ら [46, 47] は、競合学習というアプローチからクラスタリングを考え、カーネル関数を利用した学習ベクトル量子化によるクラスタリングアルゴリズムを提案することで、 c -平均法で問題となっていたカーネル法を利用した際の計算量の増加を抑えることに成功している。さらに、自己組織化マップのアルゴリズムをカーネル化することにより、高次元特徴空間におけるデータの可視化にも成功している。しかしながら、ここでもデータ間の非類似度としてユークリッド距離の 2 乗が用いられているため、内積を利用した際の更新式の導出方法が異なり、クラスタリングアルゴリズムに用いられている更新式の再定式化が必要となる。

さらに、宮本 [80] は文書情報あるいは Web 情報などの情報要素をクラスタリングする手法として、ファジィマルチ集合 [110] と呼ばれる理論を用いたクラスタリングアルゴリズムを提案している。ここでは、クラスタリングアルゴリズムとして、標準的ファジィ c -平均法およびエントロピー正則化法が用いられているが、クラスターサイズを調整する変数を導入した手法や競合学習モデルを用いた手法はまだ考察されていない。また、カーネル化したアルゴリズムについても考察されていないため、実世界におけるクラスタリング性能の有効性を確かめるためには、このような応用研究も必要であると言える。

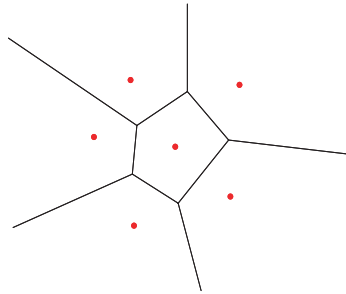


図 1.3: 各点を結ぶ垂直 2 等分線で形成されるボロノイ境界を表した図

1.3 研究の意義

1.3.1 内積に基づく類似度

データクラスタリングの研究においては，データ間の類似性測度としてユークリッド距離が用いられるのが一般的であり，内積に基づく類似度を利用したクラスタリングの研究はあまりなされていない．なぜならば，一般に多変量解析の方法は多次元ユークリッド空間に基づいて定義されており，クラスタリングにおいてもユークリッド空間に基づく数学モデルによって個体間の類似度あるいは関連度を定めることが多いからである．したがって，ユークリッド空間においてデータを分類するということを考える場合には，内積に基づく類似度を用いるよりも距離を用いるほうがより一般的であるとされている．

しかしながら，情報検索におけるクラスター型検索モデル [59, 94] などを考慮する場合には，コサイン相関のような内積に基づく類似度を用いることのほうが妥当であるとされており [60]，データクラスタリングにおいても距離だけでなく内積に基づく類似度を利用したアルゴリズムについて研究を行う必要がある．

さらに，多次元のデータ，特に各変数の単位や性質が異なったデータを取り扱う際，距離を利用する場合にはこれらの変数を比較可能にするためあらかじめデータを尺度化しておかなければならない [34, 90]．しかしながら，内積に基づく類似度を利用したアルゴリズムでは，各変数間の単位が異なっても同じクラスタリング結果を得ることが可能である．なぜならば，内積に基づく類似度を利用したクラスタリングは各個体の持つ値の大きさに依存することなく，個体間の相対的な分布に対してクラスタリングを行うことができるからである．

1.3.2 クラスタリングアルゴリズムの非線形化の必要性

実世界においては線形分離可能なデータよりもむしろ，線形分離不可能なデータの方が多く存在しており，非線形分離可能な手法が必要となる．パターン認識の分野においては，非線形な分類を行う方法としてニューラルネットワークやサポートベクトルマシンなどを応用

した手法が提案されているが、これらはいずれも教師付き分類手法である。そのため、教師無し分類手法であるクラスタリングアルゴリズムの非線形分離可能なアルゴリズムへの拡張が必要となる。本研究では、このような非線形分離可能なアルゴリズムへの拡張のことをアルゴリズムの非線形化と呼ぶことにする。

さらに、競合学習モデルに基づくクラスタリングアルゴリズムはオンライン型のクラスタリング手法とも呼ばれ、大量なデータを取り扱う際に有効な手法であることが知られている [20]。この手法は、 c -平均法のような目的関数の最適化を行うのではなく、入力データとクラスター中心との位置関係だけで分類が行われるのが特徴である。この考えを利用することにより、カーネル法を利用することによるクラスタリングアルゴリズムの非線形化の際、内積に基づく類似度を用いた場合においても計算量の増加を防ぐことが期待できる。

1.3.3 本研究の目的

以上の観点から、本研究では内積に基づく類似度を利用したクラスタリングアルゴリズムをカーネル法を利用することにより非線形分離可能なアルゴリズムに拡張することで、従来の手法よりも分類性能において優れた効果を発揮する可能性について明らかにする。

また、クラスタリングの応用研究として、ファジィマルチ集合という新たな情報検索モデルを用いた文書情報をクラスタリングする手法について示し、従来の手法と分類結果を比較することで、非線形化されたアルゴリズムの有効性を示す。さらに、もう一つの応用研究として、クラスタリングにより日本周辺の軍事情勢の分析を行う。ここでは、非線形化されたアルゴリズムが実際にどのような場面で利用可能なのかを示し、世界の实情および多変量解析などの手法を用いてクラスタリング結果の妥当性を評価していく。

1.4 本論文の構成

本節では、本論文を構成する第 2 章以下の各章の概要について述べる。

第 2 章 c -平均クラスタリングとその拡張手法

非階層的クラスタリングの代表的な手法である c -平均クラスタリングのための類似度としてコサイン相関を定義し、その性質について示す。

また、 c -平均法をファジィ化したファジィ c -平均法について、標準的ファジィ c -平均法とエントロピー正則化ファジィ c -平均法について紹介し、コサイン相関を用いた場合の目的関数やアルゴリズムの変化について示す。このとき、ファジィクラスタリングの特徴でもある、ファジィ分類関数について示し、両手法の持つ性質の違いについて述べる。

さらに、エントロピー正則化ファジィ c -平均法の拡張手法として知られている、クラスターサイズ調整変数を導入したアルゴリズムと、この手法をクリस्प化する方法について述べる。

第 3 章 クラスタリングアルゴリズムの非線形化

従来の c -平均法やファジィ c -平均法では、分類境界が各クラスター中心を結ぶ垂直 2 等分線を形成するように求められ、また、内積を類似度とした場合には、分類境界が各クラスター中心間の角度を 2 等分するように形成されるという特徴を持っている。そのため、非線形な分類境界を持つようなデータに対しては最適な分類を行うことができず、特に、コサイン相関を類似度とした場合、同一直線上の個体はすべて同じであるとみなされてしまうため、線形な分類境界を持つようなデータでさえ最適な分類を行うことができない。

本章では、エントロピー正則化ファジィ c -平均法に基づくクラスタリングアルゴリズムにカーネル法を利用することでアルゴリズムを非線形化し、上記の問題を解決する方法について述べる。また、提案手法の効果、有効性を人工データによる数値実験により視覚的に示す。

第 4 章 競合学習モデルに基づくクラスタリング

c -平均法とは異なるアプローチである学習という概念からクラスタリングを考察することにより、カーネル法を利用する際の問題点である計算量の増加を、内積を利用したクラスタリングアルゴリズムにおいても防げることを示す。また、より複雑な形状をした人工データを用いて数値実験を行い、競合学習モデルの有効性を示す。

さらに、自己組織化マップのアルゴリズムのカーネル化により、高次元特徴空間に写像されたデータの位置関係の可視化を行う。

第 5 章 ファジィマルチ集合に基づく文書情報クラスタリング

クラスタリングを用いた情報検索への応用として，ファジィマルチ集合という新たな数理モデルを用いた，文書情報をクラスタリングする方法について述べる．

ファジィマルチ集合の特徴は，要素の数と帰属度を同時に表現する構造を持っていることである．そのため，同一の用語が異なる度合いを持って複数回出現しているような Web 上の情報あるいは文献情報などを適切に表現することが可能であると考えられている．

本章では，ファジィマルチ集合をモデルとした文書情報クラスタリングの方法について述べ，実際の文書データを用いた実験により，非線形化されたクラスタリングアルゴリズムの有効性を示す．

第 6 章 クラスタリングによる日本周辺の軍事情勢の分析

リスク解析におけるクラスタリングの利用方法の一例として，クラスタリングによる日本周辺の軍事情勢の分析を行う．各国の軍事データを使用してクラスタリングを行い，世界の国々をクラスタリングによりグループ化し，そこから得られる情報を利用して，我が国を取り巻く軍事的リスク（脅威）を分析する．

また，本章で行ったクラスタリング結果の妥当性を，主成分分析により得られた結果と，世界の核兵器および弾道ミサイルの拡散・保有状況の変化から評価する．

第 7 章 結論

本研究の全体を総括し，本研究で得られた成果を要約するとともに，今後の課題および展望について述べる．

第2章 c -平均クラスタリングとその拡張手法

2.1 クラスタリングと類似性測度

クラスタリングとは、データ解析の諸技法の中で、外的基準なしに自動的に分類を行う方法、いいかえれば、データ以外にあらかじめ基準を設定することなく、データの集まりをいくつかのグループに分ける方法のことである。そのため、パターン認識の分野では、教師無し自動分類 (unsupervised automatic classification) と呼ばれており、外的基準がないため、分類すべき個体 (データ) 間に定義された類似性や距離に基づいてグループ分けが行われる。宮本によるとクラスタリングは次のように規定されている [76]。

“分類すべき個体の集まりにおいて、任意の2つの個体間に類似性あるいは非類似性を表す測度が与えられていると仮定する。クラスタリングとは、個体の集まりをいくつかのクラス (部分集合) に分割し、それぞれのクラスの中では個体どうしの類似度が大きく、異なるクラスについては類似度が小さくなるようにすることである。”

ここで、類似性測度 (similarity measure) あるいは類似度とは、個体を2つ与えたとき、その対に対して決まる実数であり、その値が大きければ大きいほど2つの個体は互いに類似しているとされる。逆に、非類似性測度 (dissimilarity measure) あるいは非類似度とは、個体を2つ与えたとき、その対に対して決まる実数であり、その値が小さければ小さいほど2つの個体は互いに類似しているとされる。

類似性を表す測度が与えられた場合には、上記で示されている規定のとおりであるが、非類似度を表す測度が与えられた場合については、同じクラスにおいては非類似度が小さく、異なるクラスについては非類似度が大きくなるようにしなければならない。一般には、非類似度という言葉の代わりに我々が使い慣れた距離 (distance) という用語が用いられることも多い。

クラスタリングではよくクラスター (cluster) という用語が用いられるが、これはクラスタリングによって得られたクラス、すなわち個体のグループのことを指している。分類すべき対象を p 次元ユークリッド空間 R^p 上の n 個の点とすると、データの各個体は $x_k = (x_k^1, \dots, x_k^p)^\top$, $k = 1, \dots, n$ と表される¹。また、これらの個体を c 個のクラスター G_i , $i = 1, \dots, c$ に分類することを考えると、各クラスター G_i の代表点は $v_i = (v_i^1, \dots, v_i^p)^\top$, $i = 1, \dots, c$ と表される。

¹ \top は転置を示す記号であり、したがって x_k は列ベクトルである。

個体 x_k が i 番目のクラスター G_i に属するか否かを二値変数 u_{ik} で定義する．すなわち x_k が i 番目のクラスターに属するならば $u_{ik} = 1$ ，そうでなければ $u_{ik} = 0$ であるとし，1 つの個体は 1 つのクラスターのみ属することができるものとする．各クラスター G_i の代表点 v_i をクラスター中心，個体 x_k が i 番目のクラスター G_i に属するか否かを表す二値変数 u_{ik} を帰属度（メンバーシップ）と呼ぶ．

2.2 コサイン相関

コサイン相関 (cosine correlation) あるいはコサイン係数 (cosine coefficient) とも呼ばれる類似度は，ベクトル間の角度を用いた類似性測度であり，情報検索の分野におけるベクトル空間型モデルにおいてよく利用されている [59, 60, 94, 96, 97]．

X を有限個の点からなる R^p の部分集合とする．このとき， $x, y \in X$ 間の内積（もしくはスカラー積）を次のように定義する．

$$\langle x, y \rangle := x^\top y = \sum_{j=1}^p x^j y^j$$

統計学では，この量は x と y の交差積の和として知られている．ここで， R^p 上のユークリッドノルムを $\|\cdot\|$ とすると， x と y の内積はさらに次のように表される．

$$\langle x, y \rangle = \|x\| \|y\| \cos \theta \tag{2.1}$$

ここで， θ は x と y の間の角度であり，また，スカラー積とノルムの関係は $\|x\|^2 = \langle x, x \rangle$ であることに注意されたい．図 2.1 はこの関係を 2 次元空間上に示したものである． O （原点）から A までの距離は，よく知られているように $\|x\| \cos \theta$ である．この量は， y 上への x の正射影でもあり，したがって内積とは y の長さ $\|y\|$ と y 上への x の正射影の長さとの積となる． $\cos \theta$ について式 (2.1) を解くと，

$$\cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} = \frac{\sum_{j=1}^p x^j y^j}{\sqrt{\left\{ \sum_{j=1}^p (x^j)^2 \right\} \left\{ \sum_{j=1}^p (y^j)^2 \right\}}}$$

となる．この $\cos \theta$ を x と y の類似度

$$s(x, y) = \cos \theta = \frac{\langle x, y \rangle}{\|x\| \|y\|} \tag{2.2}$$

と考えることができる．この類似度 $s(x, y)$ をコサイン相関と呼ぶ．

$s(x, y)$ は、2つのベクトルが平行に近づけば近づくほど類似度が大きくなる。また、この類似度はベクトルの長さとは無関係になり、符号が変わることはありうるが、一様な乗法尺度に対して不変である。この関係を幾何学的に説明すると図 2.2 のようになる。すなわち、 $s(x, y)$ は多対一の変換となり、ベクトル間の相対的な大きさは事実上無視されるという特徴を持つ [92]。

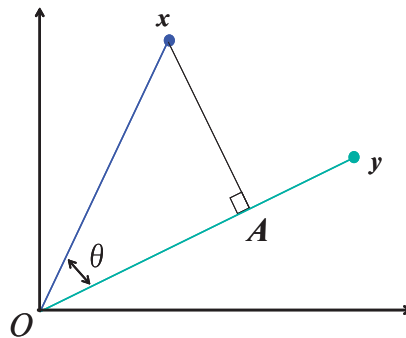


図 2.1: x と y のなす角を θ とし、 y 上へ x を正射影した点を A とする。原点 O から A までの距離は初等幾何学で知られているように $\|x\| \cos \theta$ である。したがって、内積は y の長さと y 上への x の射影の長さとの積となる。

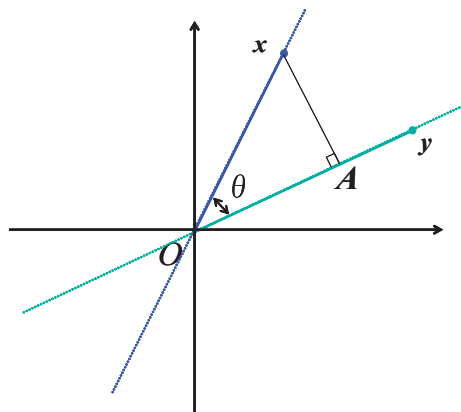


図 2.2: $\cos \theta$ の不変性の幾何学的説明を示した図。直線 Ox およびその両方向への射影（ただし原点 O そのものは除く）上の点は、すべてこの測度のもとでは x と同値となる。この場合、 $\cos \theta$ は正であり、同一象限内にある（2直線上の）点の任意の組についても正である。ただし、点が異なる象限にある場合には負となる。

2.3 クリस्प c -平均法

クリस्प c -平均法とは、非階層的クラスタリングの代表的な手法である c -平均法のことであり、ファジィ c -平均法と区別するために、このクリस्पという用語がよく用いられる。

c -平均法 (c -means) はもともと MacQueen によって K -means と呼ばれていた。MacQueen によれば、以下の方法を K -means プロセスと呼び、 K -means によるクラスタリングとは区別している [66]。

“ K -means の手続きをインフォーマルに述べれば、次のようになる。 K 個ランダムに個体を選び、それぞれのグループの代表とする。ほかの個体を一つずつ選び、最も近い平均値をもつグループに割り当てる。割り当てられたグループについて、平均値を更新する。各ステージにおいて、 K 個の平均値がグループを代表する。したがって、 K -means と呼ぶ。”

実際、MacQueen が K -means クラスタリングのために開発したプログラムには、クラスター数を変更するパラメータなどが含まれており、一般的に使われている c -平均法のアルゴリズムより複雑な手続きをとっている。ここでの c -平均法は、Anderberg [2] において述べられている Forgy [28] のアルゴリズムを指すものとする。

クリस्प c -平均法では分類すべき個体が与えられているときに、 c 個のクラスター中心と、各個体がクラスターに所属するか否かを表す二値変数 (帰属度) を含む目的関数 (評価関数とも呼ばれる) を最適化することによりクラスタリングを行う。最適化はクラスター中心と帰属度について交互最適化を用いて行われる。

ここで、帰属度を $c \times n$ 行列として $U = (u_{ik})$ とする。また、クラスター中心をまとめて $V = (v_1, \dots, v_c)$ とすると、これら 2 つの変数を持つ目的関数 J_{ccm} を以下のように定式化することができる²。

$$J_{ccm}(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_{ik} \quad (2.3)$$

ここで、 s_{ik} は個体 x_k とクラスター中心 v_i との類似度を表す関数で、以下の式で定義される。

$$s_{ik} = \frac{\langle x_k, v_i \rangle}{\|x_k\| \|v_i\|} \quad (2.4)$$

²目的関数 J_{ccm} の添え字の ccm は「crisp c -means」を表している。

この目的関数 J_{ccm} の値は、各クラスターに属する個体とそのクラスター中心との内積の総和である。したがって、この目的関数が最大となるような U, V の値が、最適な解であると言える。また、帰属度 U に関しては次の制約条件が定義されている。

$$M_c = \left\{ (u_{ik}) : u_{ik} \in \{0, 1\}, \sum_{i=1}^c u_{ik} = 1, \forall k \right\} \quad (2.5)$$

この制約条件は、各個体は必ずどれか一つのクラスターに属さなければならないということを示している。以下に、クリस्प c -平均法におけるアルゴリズム CCM (Crisp c -Means) を示す。

アルゴリズム CCM.

CCM1. \bar{U} と \bar{V} の初期値を定める。

CCM2. \bar{V} を固定して

$$\max_{U \in M_c} J_{ccm}(U, \bar{V})$$

を解き、最適解を \bar{U} とする。

CCM3. \bar{U} を固定して

$$\max_V J_{ccm}(\bar{U}, V)$$

を解き、最適解を \bar{V} とする。

CCM4. 解 (\bar{U}, \bar{V}) が収束すれば終了。そうでなければ CCM2 に戻る。

CCM2 と CCM3 の最適化は、実際には次のように行われる。CCM2 では、

$$u_{ik} = 1 \iff v_i = \arg \max_{1 \leq j \leq c} s_{jk} \quad (2.6)$$

$$u_{ik} = 0 \iff j \neq i \quad (2.7)$$

で最適化しており、 s_{jk} が最大となるクラスターに対する帰属度 (メンバーシップ値) u_{ik} を 1 とし、他のクラスターに対する帰属度を 0 にしている。言い換えると、 s_{jk} が最大となるクラスターに個体 x_k を割り当てている。

また、CCM3 における最適解 \bar{V} については、ラグランジュ乗数法を用いて求めることができる。まず、制約条件 $\|v_i\| = 1$ を仮定する。ここで、ラグランジュ乗数 ξ_i ($i = 1, \dots, c$) を

用いて、ラグランジュ関数を

$$L = J_{ccm} + \sum_{i=1}^c \xi_i (\|v_i\|^2 - 1)$$

と定義し、 $\frac{\partial L}{\partial v_i} = 0$ を計算すれば、

$$\sum_{k=1}^n u_{ik} x_k + \xi_i v_i = 0$$

となる．このとき $\|v_i\| = 1$ を考慮して ξ_i を消去すれば、 v_i の最適解は、

$$v_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\left\| \sum_{k=1}^n u_{ik} x_k \right\|} \quad (2.8)$$

($i = 1, \dots, c$) となる．

アルゴリズム CCM は、同一の目的関数を 2 つの変数 U, V によって交互最適化を行っている．したがって、目的関数の値は単調に増加し、アルゴリズムは有限回で終了する．ここで注意すべきことは、このアルゴリズムは初期値に依存するという点である．よって、局所解に陥ることを防ぐために初期値をランダムにとり、クラスタリングを何回か試行して、最も目的関数の値が大きいものを最適解とする必要がある．

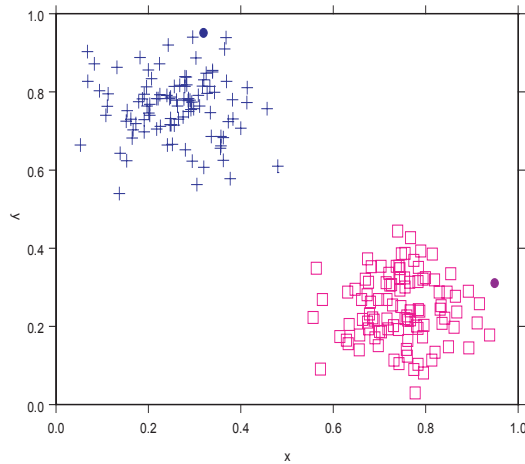


図 2.3: クリスプ c -平均法により 2 つのクラスターに分類した例．各クラスターはそれぞれと + で表されており、図中の ● は各クラスターのクラスター中心を表している．

2.4 ファジィ c -平均法

ファジィ理論は Zadeh [113] によって創始されたもので、物事におけるあいまい性を境界のはっきりしない集合によって取り扱う理論である。ファジィクラスタリングは、個体が各クラスターに所属する度合い、つまり帰属度 u_{ik} に 0 から 1 までの区間の値を連続的にとることを許すことにより、帰属度にあいまいさを認めるという考えに基づいて提案された。

前節の c -平均法では、制約条件 M_c から、1 つの個体は 1 つのクラスターにしか属することができないと決められていた。すなわち、帰属度 u_{ik} の値は属するか属さないかを表す二値変数となるように定義されていた。そこで、 c -平均法にファジィネス³を導入するために、帰属度 u_{ik} に 0 から 1 までの連続的な値をとることを許し、その値を個体 x_k のクラスター G_i への帰属の度合いとみなすことにする。このとき、帰属度の制約条件は次のように緩和される。

$$M_f = \left\{ (u_{ik}) : u_{ik} \in [0, 1], \sum_{i=1}^c u_{ik} = 1, \forall k \right\} \quad (2.9)$$

この制約条件は、一つの個体は複数のクラスターに所属することができるけれども、帰属度をすべて加えると 1 になるということを示している。

しかし、制約条件 M_f のもとで目的関数 J_{ccm} を最適化しても、実際にはアルゴリズム CCM2 における最適解は 0 か 1 のいずれかになってしまう。なぜなら、目的関数 J_{ccm} は帰属度 u_{ik} について線形であり、制約条件 M_f も線形であるので、この問題は線形計画問題となり、最適解は制約条件が形成する多角形の端点、つまりクリスプな解となるからである。

2.4.1 標準的ファジィ c -平均法

Dunn [23, 24] と Bezdek [6] は、 c -平均法をファジィ化するために目的関数にパラメータ m を導入し、次のような目的関数を定義した。

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|^2$$

ここで、 m は $m > 1$ を満たすファジィ化パラメータである。ただし、この目的関数には非類似度であるユークリッド距離の 2 乗が用いられているため、内積に基づく類似度、すなわちコサイン相関を用いる場合には、次のような目的関数 J_{sfcm} になることに注意しなければならない⁴。

$$J_{sfcm}(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m (1 - s_{ik}) \quad (2.10)$$

³fuzziness - ファジィである性質のこと。

⁴目的関数 J_{sfcm} の添え字の $sfcm$ は「standard fuzzy c -means」を表している。

この目的関数を以下のアルゴリズム sFCM (standard Fuzzy c -Means) により最適化を行うと、0 から 1 までの区間の値をもつファジィな帰属度が得られる。

アルゴリズム sFCM.

sFCM1. \bar{U} と \bar{V} の初期値を定める。

sFCM2. \bar{V} を固定して

$$\min_{U \in M_f} J_{sfcM}(U, \bar{V})$$

を解き、最適解を \bar{U} とする。

sFCM3. \bar{U} を固定して

$$\min_V J_{sfcM}(\bar{U}, V)$$

を解き、最適解を \bar{V} とする。

sFCM4. 解 (\bar{U}, \bar{V}) が収束すれば終了。そうでなければ sFCM2 に戻る。

sFCM2 および sFCM3 の最適解はそれぞれラグランジュ乗数法により導出される。

- U の最適解

ラグランジュ乗数 ν_k ($k = 1, \dots, n$) を用いて、ラグランジュ関数を

$$L = J_{sfcM} + \sum_{k=1}^n \nu_k \left(\sum_{i=1}^c u_{ik} - 1 \right)$$

と定義し、最適性の条件から $\frac{\partial L}{\partial u_{ik}} = 0$ を計算すれば、

$$m(u_{ik})^{m-1}(1 - s_{ik}) + \nu_k = 0$$

となる。このとき $\sum_{j=1}^c u_{jk} = 1$ を考慮して ν_k を消去すれば、 u_{ik} の最適解は

$$u_{ik} = \left\{ \sum_{j=1}^c \left(\frac{1 - s_{ik}}{1 - s_{jk}} \right)^{\frac{1}{m-1}} \right\}^{-1} \quad (2.11)$$

となる。なお、この帰属度の式では、 $x_k = Cv_i$ (C はある定数) となる x_k が存在したとき、 $s_{ik} = 1$ となり解が不定となる。そのため、 $x_k = Cv_i$ となる x_k が存在する場合には、 $u_{ik} = 1$ 、 $u_{jk} = 0$ ($i \neq j$) となる。

• V の最適解

まず，制約条件 $\|v_i\| = 1$ を仮定する．ラグランジュ乗数 ξ_i ($i = 1, \dots, c$) を用いて，ラグランジュ関数を

$$L = J_{sfcm} + \sum_{i=1}^c \xi_i (\|v_i\|^2 - 1)$$

と定義し，最適性の条件から $\frac{\partial L}{\partial v_i} = 0$ を計算すれば，

$$\sum_{k=1}^n (u_{ik})^m x_k + \xi_i v_i = 0$$

となる．このとき $\|v_i\| = 1$ を考慮して ξ_i を消去すれば， v_i の最適解は

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\left\| \sum_{k=1}^n (u_{ik})^m x_k \right\|} \quad (2.12)$$

となる．

2.4.2 エントロピー正則化ファジィ c -平均法

工学の諸問題には，真の解が存在しない問題や，観測データが少し変化すると解が大きく変化するような不連続性を持つ問題がしばしば含まれる．このような本質的に解きにくい問題は，適切でない問題あるいは不良設定問題 (ill-posed problem) と呼ばれる．不良設定問題の解き方としてよく知られている考え方に，正則化 (regularization) がある．正則化とは，もとの問題が特異なときにそれを何らかの方法で正則にして近似解を求めるという考え方である．

宮本らは，ファジィ c -平均法は c -平均法の正則化ととらえることはできないかと考え，正則化のための関数としてエントロピー関数を導入したエントロピー正則化ファジィ c -平均法を提案した [74, 75]．このとき，正則化のための関数をエントロピー

$$K = \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik}$$

とするので，正則化された目的関数 J_{efcm} は

$$J_{efcm}(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_{ik} - \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \quad (2.13)$$

となる⁵．

⁵目的関数 J_{efcm} の添え字の $efcm$ は「entropy-based fuzzy c -means」を表している．

ここでもさきほどの標準的なファジィ c -平均法と同じく、目的関数 J_{efcm} には、2 つの変数、帰属度 U とクラスター中心 V が含まれているので、エントロピー正則化ファジィ c -平均法においても、2 段階の交互最適化アルゴリズムによりクラスタリングを行う。以下にエントロピー正則化ファジィ c -平均法のアルゴリズム eFCM (entropy-based Fuzzy c -Means) を示す。

アルゴリズム eFCM.

eFCM1. \bar{U} と \bar{V} の初期値を定める。

eFCM2. \bar{V} を固定して

$$\max_{U \in M_f} J_{efcm}(U, \bar{V})$$

を解き、最適解を \bar{U} とする。

eFCM3. \bar{U} を固定して

$$\max_V J_{efcm}(\bar{U}, V)$$

を解き、最適解を \bar{V} とする。

eFCM4. 解 (\bar{U}, \bar{V}) が収束すれば終了。そうでなければ eFCM2 に戻る。

eFCM2 および eFCM3 の最適解はそれぞれラグランジュ乗数法により導出される。

- U の最適解

ラグランジュ乗数 ν_k ($k = 1, \dots, n$) を用いて、ラグランジュ関数を

$$L = J_{efcm} + \sum_{k=1}^n \nu_k \left(\sum_{i=1}^c u_{ik} - 1 \right)$$

と定義し、最適性の条件から $\frac{\partial L}{\partial u_{ik}} = 0$ を計算すれば、

$$s_{ik} - \lambda^{-1}(1 + \log u_{ik}) + \nu_k = 0$$

となる。このとき $\sum_{j=1}^c u_{jk} = 1$ を考慮して ν_k を消去すれば、 u_{ik} の最適解は

$$u_{ik} = \frac{\exp(\lambda s_{ik})}{\sum_{j=1}^c \exp(\lambda s_{jk})} \quad (2.14)$$

となる。

• V の最適解

まず，制約条件 $\|v_i\| = 1$ を仮定する．ラグランジュ乗数 ξ_i ($i = 1, \dots, c$) を用いて，ラグランジュ関数を

$$L = J_{efcm} + \sum_{i=1}^c \xi_i (\|v_i\|^2 - 1)$$

と定義し，最適性の条件から $\frac{\partial L}{\partial v_i} = 0$ を計算すれば，

$$\sum_{k=1}^n u_{ik} x_k + \xi_i v_i = 0$$

となる．このとき $\|v_i\| = 1$ を考慮して ξ_i を消去すれば， v_i の最適解は

$$v_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\left\| \sum_{k=1}^n u_{ik} x_k \right\|} \quad (2.15)$$

となる．

本研究においては，拡張性の良さから様々な拡張手法が提案されている [43, 44, 77, 100, 112] エントロピー正則化ファジィ c -平均法を中心に扱う．よって，本稿におけるファジィ c -平均法とは，エントロピー正則化ファジィ c -平均法のことを指すことに注意されたい．

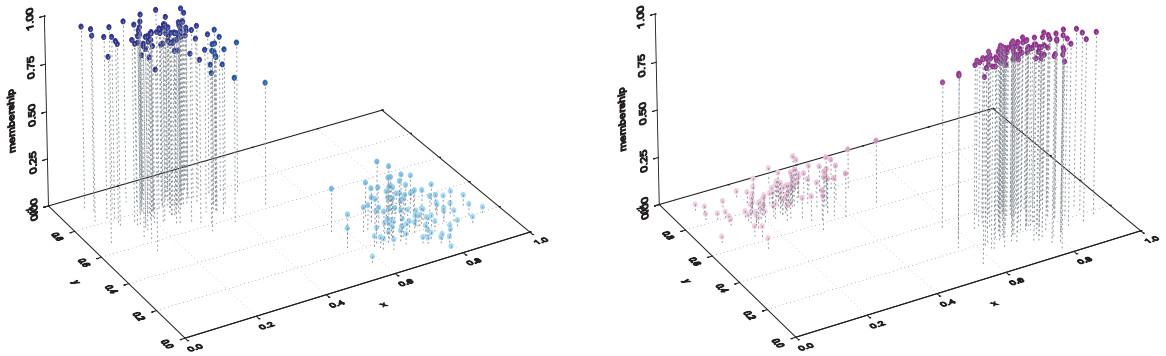


図 2.4: ファジィ c -平均法により 2 つのクラスターに分類した例．縦軸は各クラスターへの帰属度 (メンバーシップ) を表している．ファジィ c -平均法により，個体の各クラスターへの帰属度が 0 から 1 までの値を持つように緩和されたことが分かる．

2.5 ファジィ分類関数

クリスプ c -平均法では，最も類似度が大きいクラスター中心に個体を割り当てるという決定的な分類規則が用いられていた．ファジィ c -平均法においても，データ空間 R^p 内の点 x の各クラスターに対する帰属度を与える関数を求めることができる．このようなファジィな分類規則を与える関数のことを，ファジィ分類関数と呼ぶ [75, 76] ．

このファジィ分類関数はデータ空間全体で定義されているので，分類結果についてのより良い見通しを立てることができ，また，この関数の性質を調べることにより，その手法が持つ特性をより深く分析することができるという特徴を持っている．

ファジィ分類関数は，帰属度 u_{ik} の最適解の式の x_k を x に置き換えたものになる．よって，標準的ファジィ c -平均法のファジィ分類関数は次のようになる⁶ ．

$$U_i^s(x) = \left\{ \sum_{j=1}^c \left(\frac{1 - S_i(x)}{1 - S_j(x)} \right)^{\frac{1}{m-1}} \right\}^{-1} \quad (2.16)$$

ただし， $S_i(x) = 1$ のときは $U_i^s(x) = 1$ である．

また，エントロピー正則化ファジィ c -平均法のファジィ分類関数は次のようになる⁷ ．

$$U_i^e(x) = \frac{\exp(\lambda S_i(x))}{\sum_{j=1}^c \exp(\lambda S_j(x))} \quad (2.17)$$

ここで，

$$S_i(x) = \frac{\langle x, v_i \rangle}{\|x\| \|v_i\|} \quad (2.18)$$

であり， R^p 内の任意の点 x とクラスター中心 v_i との類似度を表している．

宮本によると，ファジィ分類関数の実際的な意義が次のように述べられている [76] ．

- クラスタリングを行った後，新たな個体のデータ y が得られたとする．先のクラスタリングの結果を正しい分類とみなして，新たなデータをどのクラスに入れるかは教師付き分類問題となる．このとき， $U_i^s(y)$, $U_i^e(y)$ などを用いてファジィ分類を行うことができる．
- 分類関数自体の性質を調べれば，クラスタリング自体がもつ特性が明らかになり，クラスタリングに対する見方が深まる．また，分類結果についての見通しが良くなる． u_{ik} からは個体上の値しかわからないが， u_{ik} を内挿 (interpolate) する分類関数は，全空間で定義されているので，性質が調べやすい．

⁶上付きの添え字の s は，標準的ファジィ c -平均法のファジィ分類関数であることを示している．

⁷上付きの添え字の e は，エントロピー正則化ファジィ c -平均法のファジィ分類関数であることを示している．

図 2.5 は、標準的ファジィ c -平均法により個体を 2 つのクラスターに分類した結果得られたファジィ分類関数を表したものであり、図 2.6 は、同じくエントロピー正則化ファジィ c -平均法により個体を 2 つのクラスターに分類した結果得られたファジィ分類関数を表したものである。両者を比較すると、それぞれの手法の特性が異なることが分かる。これにはそれぞれの手法における以下のような性質が関係している。

- 標準的ファジィ c -平均法

標準的ファジィ c -平均法においては、個体 x をクラスター中心 v_i に近づけた場合の個体の帰属度 u_{ik} は 1 となる。

$$\lim_{\|x\| \rightarrow v_i} U_i^s(x) = 1$$

一方、個体 x を空間内で無限に遠ざけた場合の個体の帰属度 u_{ik} は $\frac{1}{c}$ となる。

$$\lim_{\|x\| \rightarrow \infty} U_i^s(x) = \frac{1}{c}$$

- エントロピー正則化ファジィ c -平均法

エントロピー正則化ファジィ c -平均法においては、個体 x をクラスター中心 v_i に近づけた場合の個体の帰属度 u_{ik} は 1 とはならない。

$$\lim_{\|x\| \rightarrow v_i} U_i^e(x) = \frac{1}{1 + E} < 1$$

$$E = \sum_{j=1, j \neq i}^c \exp(\lambda S_j(x)) > 0$$

個体 x を空間内で無限に遠ざけた場合の個体 x の帰属度 u_{ik} は、個体 x がクラスター G_i の領域内であるか、領域外であるかによって値が異なる。すなわち、個体 x がクラスター G_i の領域内である場合

$$\lim_{\|x\| \rightarrow \infty, x \in G_i} U_i^e(x) = 1$$

となり、クラスター G_i の領域外である場合

$$\lim_{\|x\| \rightarrow \infty, x \notin G_i} U_i^e(x) = 0$$

となる。

これらの性質は、図 2.5 や図 2.6 に示されているファジィ分類関数からも読み取ることができる。

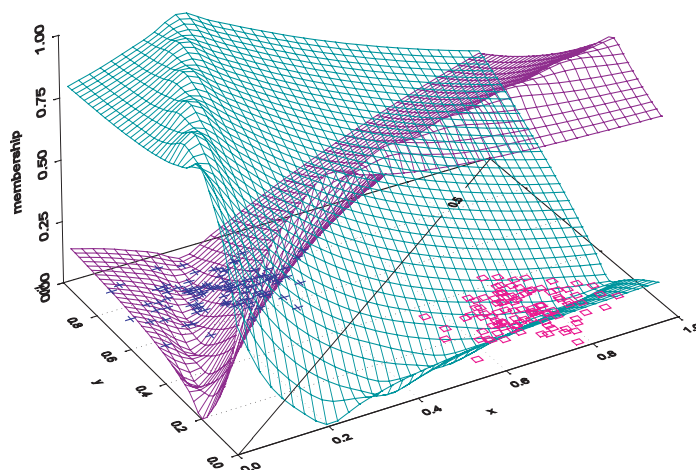


図 2.5: 標準的ファジィ c -平均法により求められたファジィ分類関数．縦軸は各クラスターへの帰属度を表している．また，直線は帰属度 0.5 のクラスターの分類境界を示している．

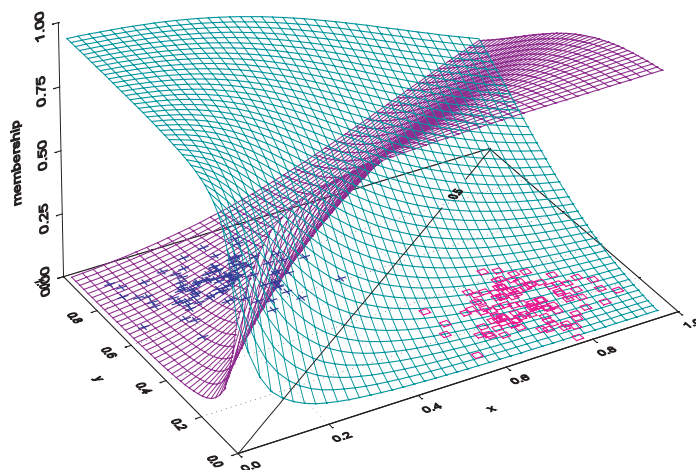


図 2.6: エントロピー正則化ファジィ c -平均法により求められたファジィ分類関数．縦軸は各クラスターへの帰属度を表している．また，直線は帰属度 0.5 のクラスターの分類境界を示している．

2.6 クラスタサイズ調整変数

c -平均法は、クラスタサイズが均等になるように分類される傾向がある。そのため、図 2.7 に示されているようなクラスタサイズの異なるデータが存在した場合においては、サイズの大きいクラスタの外縁部の個体を近接する小さいクラスタが飲み込んでしまうことがある。そこで、個体数が少ないクラスタはそのクラスタの領域も狭くなるという考え方の元、クラスタサイズに偏りがある場合でも効果的に分類することができるように、クラスタのサイズを調整するための変数が導入されたアルゴリズムが宮本によって提案された [77]。これにより目的関数は非線形化され、より最適な分類境界を得ることが可能となった。

本節では、ファジィ c -平均法の拡張手法として、まずエントロピー正則化ファジィ c -平均法にクラスタサイズを調整する変数 α を導入したアルゴリズムについて述べる。続いて、この考えを利用して、ハード技法であるクリस्प c -平均法にクラスタサイズ調整変数を導入したアルゴリズムについて述べる。

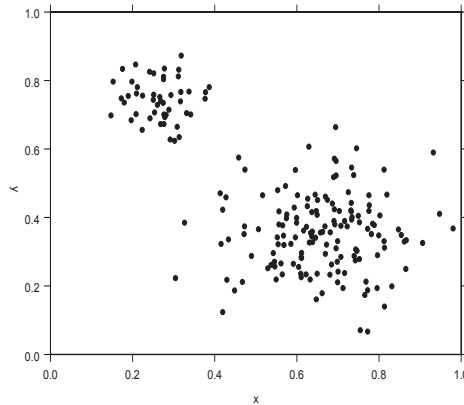


図 2.7: クラスタサイズが異なるデータ

2.6.1 クラスタサイズ調整変数付ファジィ c -平均法

エントロピー正則化ファジィ c -平均法における目的関数 J_{efcm} において、変数 U, V の他にクラスタサイズ調整変数 $\alpha = (\alpha_1, \dots, \alpha_c)$ をもつ次の目的関数 J_{efcma} を定義する。

$$J_{efcma}(U, V, \alpha) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_{ik} - \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \frac{u_{ik}}{\alpha_i} \quad (2.19)$$

ここで, α の制約条件は次のように定義されており,

$$\mathcal{A} = \left\{ \alpha : \sum_{i=1}^c \alpha_i = 1, \quad \alpha_i \geq 0, \quad i = 1, \dots, c \right\} \quad (2.20)$$

α_i はクラスター G_i のクラスターサイズを表す.

目的関数 J_{efcma} には, 3 つの変数 U, V, α が含まれているので, 次のような 3 段階の交互最適化アルゴリズム FCMA によりクラスタリングを行う.

アルゴリズム FCMA.

FCMA1. $\bar{U}, \bar{V}, \bar{\alpha}$ の初期値を定める.

FCMA2. $\bar{V}, \bar{\alpha}$ を固定して

$$\max_{U \in M_f} J_{efcma}(U, \bar{V}, \bar{\alpha})$$

を解き, 最適解を \bar{U} とする.

FCMA3. $\bar{U}, \bar{\alpha}$ を固定して

$$\max_V J_{efcma}(\bar{U}, V, \bar{\alpha})$$

を解き, 最適解を \bar{V} とする.

FCMA4. \bar{U}, \bar{V} を固定して

$$\max_{\alpha \in \mathcal{A}} J_{efcma}(\bar{U}, \bar{V}, \alpha)$$

を解き, 最適解を $\bar{\alpha}$ とする.

FCMA5. 解 $(\bar{U}, \bar{V}, \bar{\alpha})$ が収束すれば終了. そうでなければ FCMA2 に戻る.

各ステップにおける最適解については, ファジィ c -平均法と同様にラグランジュ乗数法を用いて求めることができ, 以下ようになる.

- U の最適解

$$u_{ik} = \frac{\alpha_i \exp(\lambda s_{ik})}{\sum_{j=1}^c \alpha_j \exp(\lambda s_{jk})} \quad (2.21)$$

- V の最適解

$$v_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\left\| \sum_{k=1}^n u_{ik} x_k \right\|} \quad (2.22)$$

- α の最適解

$$\alpha_i = \frac{1}{n} \sum_{k=1}^n u_{ik} \quad (2.23)$$

2.6.2 クラスタサイズ調整変数付クリスプ c -平均法

クラスタサイズ調整変数を導入したクリスプ c -平均法においては，目的関数 J_{efcma} のエントロピー項から $u_{ik} \log u_{ik}$ の項を取り除いた，次の目的関数 J_{ccma} を用いる．

$$J_{ccma}(U, V, \alpha) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_{ik} + \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \alpha_i \quad (2.24)$$

この考え方は，エントロピー正則化ファジィ c -平均法が，エントロピー項を付加することにより目的関数を帰属度 u_{ik} に関してファジィ化しているということから発展したものである．

これは，ファジィ技法からハード技法へと変換を行う，通常の間考え方とは異なる逆の発想であり，これ以外にも様々な研究が行われている [112]．このような変換が可能なのは，エントロピー正則化法の持つ特徴の一つでもある．

この場合においても，目的関数は 3 つの変数 U, V, α が含まれているので，クラスタサイズ調整変数付ファジィ c -平均法と同様の 3 段階の交互最適化アルゴリズムによりクラスタリングを行う．このとき， U の制約条件は M_c であることに注意されたい．

アルゴリズム CCMA.

CCMA1. $\bar{U}, \bar{V}, \bar{\alpha}$ の初期値を定める .

CCMA2. $\bar{V}, \bar{\alpha}$ を固定して

$$\max_{U \in M_c} J_{ccma}(U, \bar{V}, \bar{\alpha})$$

を解き , 最適解を \bar{U} とする .

CCMA3. $\bar{U}, \bar{\alpha}$ を固定して

$$\max_V J_{ccma}(\bar{U}, V, \bar{\alpha})$$

を解き , 最適解を \bar{V} とする .

CCMA4. \bar{U}, \bar{V} を固定して

$$\max_{\alpha \in \mathcal{A}} J_{ccma}(\bar{U}, \bar{V}, \alpha)$$

を解き , 最適解を $\bar{\alpha}$ とする .

CCMA5. 解 $(\bar{U}, \bar{V}, \bar{\alpha})$ が収束すれば終了 . そうでなければ FCMA2 に戻る .

各ステップにおける最適解は次のようになる . CCMA2 における最適解は

$$u_{ik} = 1 \iff i = \arg \max_{1 \leq j \leq c} s_{jk} \quad (2.25)$$

$$u_{ik} = 0 \iff j \neq i \quad (2.26)$$

で求められる . また , CCMA3 および CCMA4 における最適解はそれぞれ , 式 (2.22) および式 (2.23) によって求められる . すなわち ,

$$v_i = \frac{\sum_{k=1}^n u_{ik} x_k}{\left\| \sum_{k=1}^n u_{ik} x_k \right\|}, \quad \alpha_i = \frac{1}{n} \sum_{k=1}^n u_{ik}$$

である .

数値例

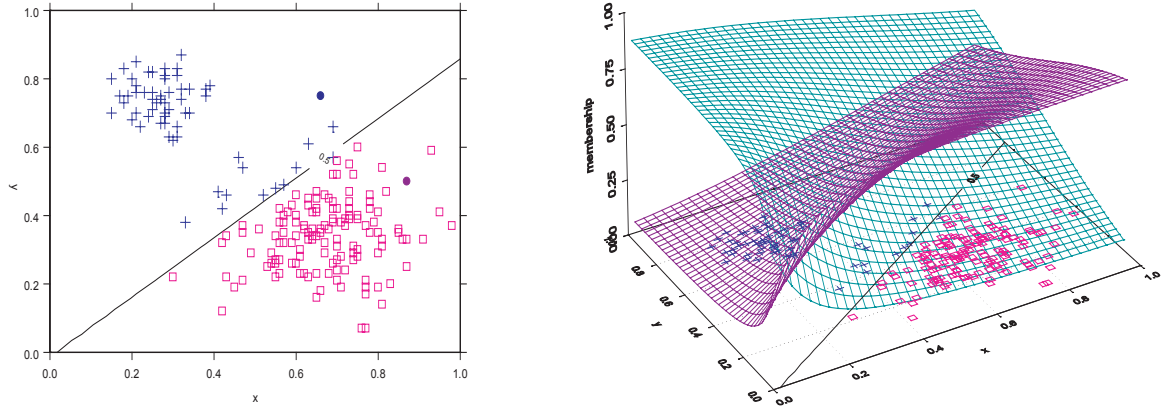


図 2.8: ファジィ c -平均法によりクラスターサイズの異なるデータを 2 つのクラスターに分類した結果と各クラスターのファジィ分類関数が示されている。クラスターサイズの大きいクラスターの外縁部の個体が、近接する小さいクラスターに分類されている様子が分かる。

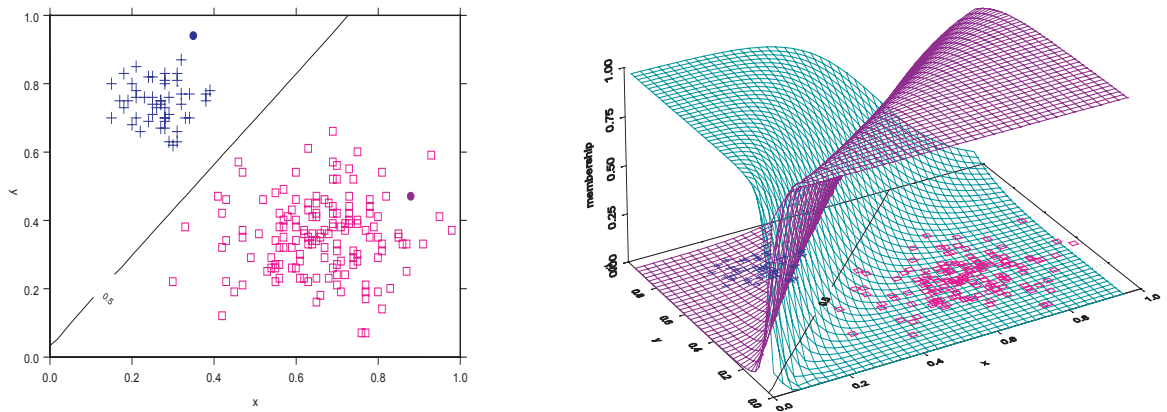


図 2.9: クラスターサイズ調整変数付ファジィ c -平均法によりクラスターサイズの異なるデータを 2 つのクラスターに分類した結果と各クラスターのファジィ分類関数が示されている。クラスターのサイズに関係なく、最適な分類境界が得られていることが分かる。ちなみに、各クラスターのサイズについては、+ のクラスターは 0.2、 \square のクラスターは 0.8 である。

第3章 クラスタリングアルゴリズムの非線形化

3.1 はじめに

従来の c -平均法やファジィ c -平均法では、一般的に非類似度にユークリッド空間での各個体とクラスター中心との距離、すなわちユークリッド距離の2乗が用いられているので、分類境界が各クラスター中心を結ぶ垂直2等分線を形成するように求められる。また、内積を類似度とした場合には、分類境界が各クラスター中心間の角度を2等分するように形成される。そのため、従来の手法では図 3.2 のような非線形な分類境界を持つようなデータに対しては、最適な分類を行うことができないという問題点が指摘されている。特に、コサイン相関を類似度とした場合、原点からの同一直線上の個体はすべて同じであるとみなされてしまうため、図 3.1 のような線形に分類可能なデータにおいても、最適な分類を行うことができないという問題点がある。

これらの問題点を解決する方法として、高次元空間へ個体を写像してクラスタリングを行うということが考えられる。近年、SVM (Support Vector Machines) [106, 107] と呼ばれる非線形な境界を求めることができる教師付き分類が多くの研究者によって注目されている [5, 14, 32]。SVM は 1960 年代に Vapnik によって超平面による特徴空間の分割法として「線形 SVM」が提案され、さらに、1990 年代になりカーネル関数と呼ばれる非線形関数を導入した「非線形 SVM」が提案された。この非線形 SVM における手法の特徴の一つに、無限次元空間を含む高次元特徴空間へ個体を写像して、そこでの内積を表すカーネル関数を利用してデータを分類するという点が挙げられる。この手法はカーネル法とも呼ばれ、最近では様々な分野に応用されている [3]。

宮本らは、このアイデアをクラスタリングに取り入れ、高次元空間への写像を利用したクラスタリングアルゴリズムを提案し、従来の手法では求めることができなかった非線形な分類境界を求めることに成功している [81, 82]。ただし、これらの研究ではデータ間の非類似度としてユークリッド距離の2乗が用いられており、内積に基づく類似度を用いた場合については考察されていない。

本章では、第2章で述べたコサイン相関を利用した4つのクラスタリングアルゴリズムに、カーネル法を利用することでアルゴリズムを非線形化する方法について述べる。また、人工データによる数値例により、コサイン相関を用いた場合でも高次元空間への写像を利用したクラスタリングにより、最適な分類境界を得られることについて示す。

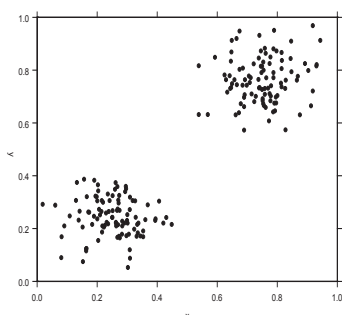


図 3.1: 線形に 2 つのクラスに分類できるのに、線形な分類境界を求めることができない例

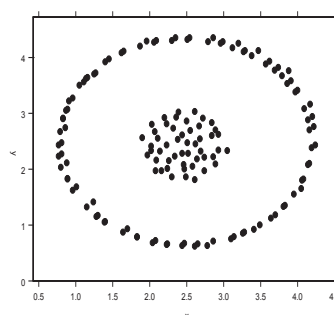


図 3.2: 線形に 2 つのクラスに分類できない例．この場合、最適な分類境界は非線形になるはずである．

3.2 カーネル法

近年、カーネル法と呼ばれる一連の機械学習の手法が提案されてきた [3, 87]．その代表的な手法としては、サポートベクトルマシン [14, 32, 106, 107]，カーネル判別分析 [70]，カーネル主成分分析 [98, 45]，カーネル部分空間法 [67, 88, 103] などが挙げられる．また、カーネル法の応用分野は、物体認識 [9, 13, 18, 64, 95]，テキストマイニング [19, 21, 53]，時系列予測 [69, 85, 86]，DNA やたんぱく質の解析 [37, 52, 114] など多岐にわたる．このように様々な分野に応用されているのは、カーネル法を利用することにより、線形モデルで表される手法をエレガントな方法で非線形モデルに拡張することが可能であるからである．

3.2.1 カーネル関数

高次元特徴空間を H とし、写像を $\Phi : \mathbf{R}^p \rightarrow H$ とする¹．このとき写像を行う非線形関数は、

$$\Phi(x) = (\phi_1(x), \phi_2(x), \dots)$$

と表すことができ、 $\Phi(x)$ が H 上での個体の値となる． x は p 次元ベクトルだが、 $\Phi(x)$ は無限次元を含む高次元特徴ベクトルになる．ここで

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle \quad (3.1)$$

とおき、この H 上での内積を $x, y \in \mathbf{R}^p$ で表現した関数をカーネル関数という．カーネル関数は、2 つの対象 $x, y \in \mathbf{R}^p$ に対して定義される関数であり、次の定理を満たす必要があるとされている [106, 107]．

¹ H は無限次元空間を含む、Hilbert 空間である．

Mercer の定理 [106, 107]

$x, y \in \mathbf{R}^p$ について,

$$K(x, y) = \sum_{j=1}^{\infty} \gamma_j \psi_j(x) \psi_j(y), \quad \gamma_j \geq 0 \quad (3.2)$$

を満たす必要十分条件は, K が対称関数

$$K(x, y) = K(y, x) \quad (3.3)$$

であり, かつ, 半正定値であることである. すなわち, 任意の関数 f について,

$$\int_{\mathcal{X}} \int_{\mathcal{X}} K(x, y) f(x) f(y) dx dy \geq 0 \quad (3.4)$$

を満たすことである. ただし, $\psi_j(x)$ は $K(x, y)$ の固有関数で,

$$\int_{\mathcal{X}} K(x, y) \psi_j(x) dx = \gamma_j \psi_j(x) \quad (3.5)$$

を満たす関数である.

この Mercer の定理を満たすカーネル関数は, Mercer カーネルとも呼ばれており, よく使われるものとして次のものがある.

Mercer カーネル

- 多項式カーネル

$$K(x, y) = (1 + \langle x, y \rangle)^d$$

- ガウシアンカーネル

$$K(x, y) = \exp(-c \text{nst} \|x - y\|^2)$$

- シグモイドカーネル

$$K(x, y) = \tanh(\gamma \langle x, y \rangle + \theta)$$

その他に，入力空間が離散値の場合に用いられる Convolution カーネル [37]，確率モデルから派生した Fisher カーネル [51, 37] や TOP カーネル [104] などが提案されている．本研究では，クラスタリングの分野において性質・性能等に優れているガウシアンカーネル²を主に使用する．

ガウシアンカーネルは，すべての $x \in \mathbf{R}^p$ について $K(x, x) = 1$ のとき，写像された個体は単位長となり， $\|\Phi(x)\| = 1$ となる．すなわち，データを規格化する作用を持っている．そのため，データを高次元特徴空間に写像してもデータ空間におけるデータの順序性というものは保持される．また，すべての $x, y \in \mathbf{R}^p$ について $K(x, y) \geq 0$ のとき，すべての点は特徴空間上の同じ象限内に存在するため，写像された個体が特徴空間の限られた領域に存在するという特徴も持っている．

3.2.2 カーネル・トリック

先程にも述べたように，カーネル法は線形モデルで表される手法を容易に非線形に拡張する方法である．ここでは，その考え方について簡単に述べる．

$x \in \mathbf{R}^p$ とし，以下のような線形モデル（パラメトリックモデル）を考える．

$$f(x, \theta) = w^\top x + b, \quad \theta = \{w, b\}, \quad w \in \mathbf{R}^p, \quad b \in \mathbf{R}$$

このとき，線形モデルの重み w を x との線形結合で表すと，

$$w = \sum_{i=1}^n \gamma_i x_i \tag{3.6}$$

となる．これにより線形モデル $f(x)$ は， x と x_i との内積で表される．

$$f(x) = \sum_{i=1}^n \gamma_i \langle x, x_i \rangle \tag{3.7}$$

このモデルを高次元特徴空間で適用すると，

$$f(\Phi(x)) = \sum_{i=1}^n \gamma_i \langle \Phi(x), \Phi(x_i) \rangle \tag{3.8}$$

となる．このとき， Φ に対応するカーネル K

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

が存在すると仮定すると，高次元特徴空間での線形モデルはカーネル K のみによって置き換えることができ，

$$f(\Phi(x)) = \sum_{i=1}^n \gamma_i K(x, x_i) \tag{3.9}$$

²ガウシアンカーネルの $cnst$ は $cnst > 0$ を満たす，データの分散を調整するパラメータである．

となり，写像 Φ の計算を回避できるのである．このため，写像関数 $\Phi(x)$ の具体的な形が分からなくても，カーネル関数により H 上での内積の値を得ることができ，高次元特徴空間上への写像計算を低次元で行うことができる．このことをカーネル・トリック，またはカーネル法と呼んでいる．

c -平均法を代表とするクラスタリング手法は，非線形分離不可能な線形モデルであるため，このようなカーネル法を利用することにより，非線形分離可能なアルゴリズムに拡張することが期待できる．

3.3 カーネル関数を利用したクラスタリング

3.3.1 カーネル関数を利用したクリスプ c -平均法

ここでは，クリスプ c -平均法にカーネル関数を利用するアルゴリズムについて述べる．

まず，高次元特徴空間における類似度として次の式を定義する．

$$s_H(\Phi(x_k), w_i) = \frac{\langle \Phi(x_k), w_i \rangle}{\|\Phi(x_k)\|_H \|w_i\|_H} \quad (3.10)$$

ここで， w_i は高次元特徴空間におけるクラスター中心であり，まとめて $W = (w_1, \dots, w_c)$ とする．

クリスプ c -平均法にカーネル関数を利用する場合，式 (2.3) の目的関数 J_{ccm} における類似度を，高次元特徴空間における類似度に置き換えたものとなる．

$$J_{kccm}(U, W) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_H(\Phi(x_k), w_i) \quad (3.11)$$

最適解については，アルゴリズム CCM における V を W に置き換えて適用すれば，CCM2 における最適解は次のように求められる．

$$u_{ik} = 1 \iff i = \arg \max_{1 \leq j \leq c} s_H(\Phi(x_k), w_j) \quad (3.12)$$

$$u_{ik} = 0 \iff j \neq i \quad (3.13)$$

また，CCM3 における最適解はラグランジュ乗数法により次のように解くことができる．まず，制約条件 $\|w_i\| = 1$ を仮定する．ここで，ラグランジュ乗数 ξ_i ($i = 1, \dots, c$) を用いて，ラグランジュ関数を

$$L = J_{kccm} + \sum_{i=1}^c \xi_i (\|w_i\|^2 - 1)$$

と定義し, $\frac{\partial L}{\partial w_i} = 0$ を計算すれば,

$$\sum_{k=1}^n u_{ik} \Phi(x_k) + \xi_i w_i = 0$$

となる. このとき $\|w_i\| = 1$ を考慮して ξ_i を消去すれば, w_i の最適解は,

$$w_i = \frac{\sum_{k=1}^n u_{ik} \Phi(x_k)}{\left\| \sum_{k=1}^n u_{ik} \Phi(x_k) \right\|} \quad (3.14)$$

($i = 1, \dots, c$) となる. しかし, $\Phi(x_k)$ の値が不明であるので, これを直接求めることができない. そこで, 次の変形を行う.

まず, 式 (3.14) を類似度の式 (3.10) に代入し, カーネル関数を

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

とおく. さらに,

$$K_{k\ell} = \langle \Phi(x_k), \Phi(x_\ell) \rangle$$

とする. $K_{k\ell}$ は明示的に用いられるので, これを利用して計算すると,

$$\begin{aligned} s_H(\Phi(x_k), w_i) &= \frac{\langle \Phi(x_k), w_i \rangle}{\|\Phi(x_k)\|_H \|w_i\|_H} \\ &= \frac{\sum_{\ell=1}^n u_{i\ell} K_{k\ell}}{\sqrt{K_{kk} \sum_{j=1}^n \sum_{\ell=1}^n u_{ij} u_{i\ell} K_{j\ell}}} \end{aligned} \quad (3.15)$$

となり, カーネル関数 K と u_{ik} だけの式で表すことができる. よって, この式 (3.15) によって類似度の更新を行い, 式 (3.12), 式 (3.13) による規則により各個体を類似度が最も大きいクラスターに再配置することを繰り返せば, カーネル関数を利用したクリस्प c -平均法によるクラスタリングを行うことができる.

カーネル関数を利用したクリस्प c -平均法のアルゴリズム K-CCM (Kernel-based Crisp c -Means) を以下に示す.

アルゴリズム K-CCM.

K-CCM1. 類似度 s_H の初期値を設定する .

K-CCM2. 式 (3.12), 式 (3.13) により, $U = (u_{ik})$ を最適解とする .

K-CCM3. 式 (3.15) により類似度 s_H を更新する .

K-CCM4. $U = (u_{ik})$ が収束すれば終了 . そうでなければ K-CCM2 に戻る .

アルゴリズム K-CCM では w_i を明示的に用いるのではなく, 式 (3.12), 式 (3.13) および式 (3.15) を繰り返すことで最適解を求めている . 従って, K-CCM4 における解の収束性は U のみによって判断される .

クリスプ c -平均法によるクラスタリングでは, クラスタ中心の位置が問題になることがあるが, 高次元特徴空間への写像を利用した場合, w_i は特徴空間における点であり, Φ^{-1} を計算しないので, クラスタ中心を直接用いずにクラスタリングを行うことができる .

3.3.2 カーネル関数を利用したファジィ c -平均法

カーネル関数を利用したファジィ c -平均法についても, クリスプ c -平均法の場合と同様に考えることができる .

この場合の目的関数は, 式 (2.13) の目的関数 J_{efcm} における類似度を, 高次元特徴空間における類似度に置き換えたもので, 次のようになる .

$$J_{kefcm}(U, W) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_H(\Phi(x_k), w_i) - \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \quad (3.16)$$

最適解については, アルゴリズム eFCM における V を W に置き換えて適用すれば, eFCM2 の最適解は次の式で計算できる .

$$u_{ik} = \frac{\exp(\lambda s_H(\Phi(x_k), w_i))}{\sum_{j=1}^c \exp(\lambda s_H(\Phi(x_k), w_j))} \quad (3.17)$$

また, eFCM3 の最適解については, ラグランジュ乗数法により解くことができ, 式 (3.14) を得る . しかし, ここでも $\Phi(x_k)$ の値が不明であるので, これを直接求めることができないので, 類似度の更新式 (3.15) を利用した以下のアルゴリズム K-FCM (Kernel-based Fuzzy c -Means) により, カーネル関数を利用したファジィ c -平均法のクラスタリングを行う .

アルゴリズム K-FCM.

K-FCM1. 類似度 s_H の初期値を設定する .

K-FCM2. 式 (3.17) により , $U = (u_{ik})$ を最適解とする .

K-FCM3. 式 (3.15) により , 類似度 s_H を更新する .

K-FCM4. $U = (u_{ik})$ が収束すれば終了 . そうでなければ K-FCM2 に戻る .

アルゴリズム K-FCM においても , w_i を明示的に用いるのではなく , 式 (3.17) , 式 (3.15) を繰り返す 2 段階の交互最適化により最適解を求めている . 従って , ここでも K-FCM4 における解の収束性は U のみによって判断される .

ファジィ分類関数

ファジィ c -平均法ではクラスター中心の位置を問題にすることがあるが , この場合は w_i は高次元特徴空間上の点であり , 写像 Φ^{-1} は明示的に求まらないので , 高次元特徴空間上でのクラスター中心 w_i を直接求めることはできない . しかし , クラスタ中心 w_i が求まらなくてもファジィ分類関数は次の方法で計算することができる .

$$U_i^c(x) = \frac{\exp(\lambda S_i(x))}{\sum_{j=1}^c \exp(\lambda S_j(x))} \quad (3.18)$$

ただし , 収束後の u_{ik} を用いて

$$S_i(x) = \frac{\sum_{\ell=1}^n u_{i\ell} K(x, x_\ell)}{\sqrt{K(x, x) \sum_{j=1}^n \sum_{\ell=1}^n u_{ij} u_{i\ell} K_{j\ell}}} \quad (3.19)$$

を計算することに注意する . また , カーネル関数 K についてはガウシアンカーネルを用いているので ,

$$K(x, x_\ell) = \exp(-c \text{nst} \|x - x_\ell\|^2)$$

により求められる .

3.3.3 カーネル関数を利用したクラスターサイズ調整変数付クリスプ c -平均法

カーネル関数を利用したクラスターサイズ調整変数付クリスプ c -平均法についても、同様に考えることができる。

この場合の目的関数は、式 (2.24) の目的関数 J_{ccma} における類似度を、高次元特徴空間における類似度に置き換えたもので、次のようになる。

$$J_{kccma}(U, W, \alpha) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_H(\Phi(x_k), w_i) + \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \alpha_i \quad (3.20)$$

最適解については、アルゴリズム CCMA における V を W に置き換えて適用すれば、CCMA2 の最適解は式 (3.12)、式 (3.13) と同じになる。また、CCMA3 の最適解については式 (3.14) となるので、類似度の更新式は式 (3.15) と同じになる。

以上のことから、カーネル関数を利用したクラスターサイズ調整変数付クリスプ c -平均法のアルゴリズム K-CCMA は以下のようなになる。

アルゴリズム K-CCMA.

K-CCMA1. 類似度 s_H と α_i の初期値を設定する。

K-CCMA2. 式 (3.12)、式 (3.13) により、 $U = (u_{ik})$ を最適解とする。

K-CCMA3. 式 (2.23) により、 α_i を更新する。

K-CCMA4 式 (3.15) により、類似度 s_H を更新する。

K-CCMA5. $U = (u_{ik})$ が収束すれば終了。そうでなければ K-CCMA2 に戻る。

アルゴリズム K-CCMA においても、 w_i を明示的に用いるのではなく、式 (3.12)、式 (3.13)、式 (2.23)、式 (3.15) を繰り返す 3 段階の交互最適化によりクラスタリングを行っている。したがって、ここでも K-CCMA5 における解の収束性は U のみによって判断される。

3.3.4 カーネル関数を利用したクラスターサイズ調整変数付ファジィ c -平均法

カーネル関数を利用したクラスターサイズ調整変数付ファジィ c -平均法についても、同様に考えることができる。

この場合の目的関数は、式 (2.19) の目的関数 J_{efcma} における類似度を、高次元特徴空間に

おける類似度に置き換えたもので，次のようになる．

$$J_{kefcma}(U, W, \alpha) = \sum_{i=1}^c \sum_{k=1}^n u_{ik} s_H(\Phi(x_k), w_i) - \lambda^{-1} \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \frac{u_{ik}}{\alpha_i} \quad (3.21)$$

最適解については，アルゴリズム FCMA における V を W に置き換えて適用すれば，FCMA2 の最適解は次の式で計算できる．

$$u_{ik} = \frac{\alpha_i \exp(\lambda s_H(\Phi(x_k), w_i))}{\sum_{j=1}^c \alpha_j \exp(\lambda s_H(\Phi(x_k), w_j))} \quad (3.22)$$

また，FCMA3 の最適解については式 (3.14) となるので，類似度の更新式は式 (3.15) と同じになる．以上のことから，カーネル関数を利用したクラスターサイズ調整変数付ファジィ c -平均法のアルゴリズム K-FCMA は以下のようなになる．

アルゴリズム K-FCMA.

K-FCMA1. 類似度 s_H と α_i の初期値を設定する．

K-FCMA2. 式 (3.22) により， $U = (u_{ik})$ を最適解とする．

K-FCMA3. 式 (2.23) により， α_i を最適解とする．

K-FCMA4. 式 (3.15) により，類似度 s_H を更新する．

K-FCMA5. $U = (u_{ik})$ が収束すれば終了．そうでなければ K-FCMA2 に戻る．

アルゴリズム K-FCMA においても， w_i を明示的に用いるのではなく，式 (3.22)，式 (2.23)，式 (3.15) を繰り返す 3 段階の交互最適化によりクラスタリングを行っている．したがって，ここでも K-FCMA5 における解の収束性は U のみによって判断される．

また，ファジィ分類関数については，クラスターサイズ調整変数付ファジィ c -平均法と同様に考えることができ，ここでは以下の式になる．

$$U_i^{efcma}(x) = \frac{\alpha_i \exp(\lambda S_i(x))}{\sum_{j=1}^c \alpha_j \exp(\lambda S_j(x))} \quad (3.23)$$

ここでも，収束後の u_{ik} と α_i を用いて求める．また， $S_i(x)$ については式 (3.19) と同様である．

3.4 数値例

ここでは、非線形化されたクラスタリングアルゴリズムの効果を、人工データを用いた数値例により示す。従来の手法による分類結果と比較することで、提案手法の有効性を視覚的に確認する。

• 人工データ 1

図 3.3 はファジィ c -平均法 (アルゴリズム eFCM) により、図 3.1 のデータを 2 つのクラスターに分類した結果である。この結果から分かるように、コサイン相関を用いたアルゴリズムでは、線形に分類できるようなデータに対しても最適な分類境界を求めることができないことが分かる。一方、図 3.4 はカーネル関数を利用したファジィ c -平均法 (アルゴリズム K-FCM) により 2 つのクラスターに分類した結果であり、このデータに対して最適な分類境界が得られたことが確認できる。

• 人工データ 2

図 3.5 はファジィ c -平均法 (アルゴリズム eFCM) により、図 3.2 のデータを 2 つのクラスターに分類した結果である。この結果から分かるように、ファジィ c -平均法のアルゴリズムでは、線形分離不可能なデータに対して最適な分類境界を求めることができないことが分かる。一方、図 3.6 はカーネル関数を利用したファジィ c -平均法 (アルゴリズム K-FCM) により 2 つのクラスターに分類した結果であり、最適な分類ができていることが確認できる。この結果から、カーネル法を利用することにより、非線形な分類境界を得ることができるアルゴリズムに拡張できたことが確認できる。

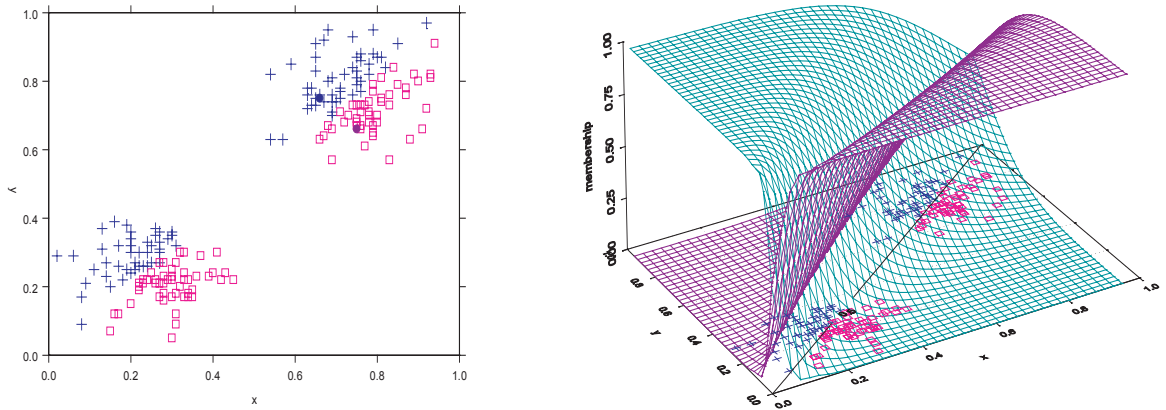


図 3.3: 左の図はアルゴリズム eFCM ($\lambda = 100$) により 2 つのクラスターに分類した結果である。各クラスターは $+$ と \square で表示されており、図中の \bullet は各クラスターのクラスター中心を表している。また、右の図は各クラスターのファジィ分類関数を表しており、帰属度 0.5 の境界線が示されている。

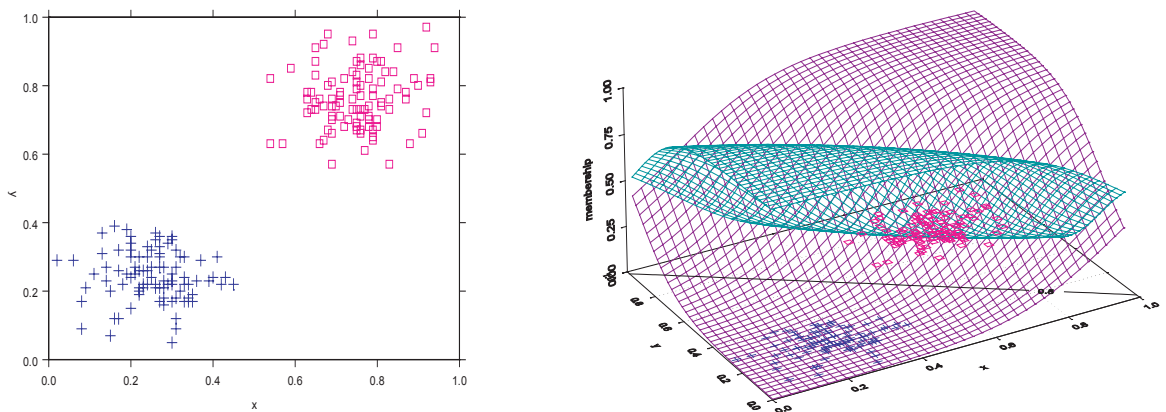


図 3.4: 左の図はアルゴリズム K-FCM ($\lambda = 10, cnst = 1.0$) により 2 つのクラスターに分類した結果である。各クラスターは $+$ と \square で表示されている。また、右の図は各クラスターのファジィ分類関数を表しており、帰属度 0.5 の境界線が示されている。

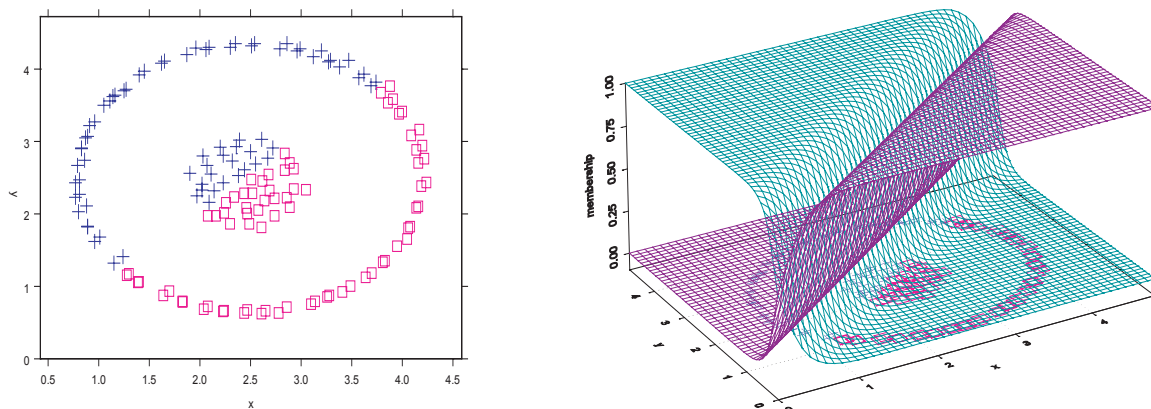


図 3.5: 左の図はアルゴリズム eFCM ($\lambda = 10$) により 2 つのクラスターに分類した結果である。各クラスターはそれぞれ \times と \square で表示されており、線形に 2 つのクラスターに分類されている様子が分かる。また、右の図は各クラスターのファジィ分類関数を表している。

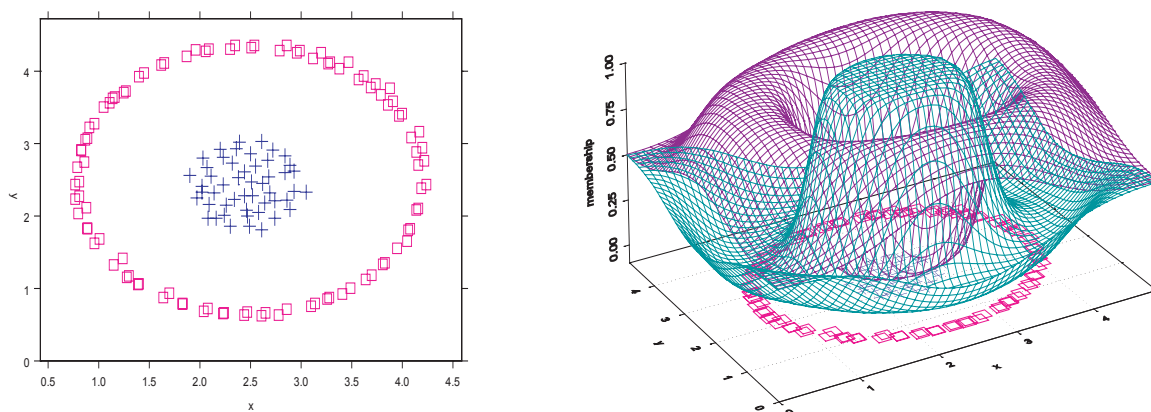


図 3.6: 左の図はアルゴリズム K-FCM ($\lambda = 20, cnst = 10$) により 2 つのクラスターに分類した結果である。各クラスターはそれぞれ \times と \square で表示されており、非線形に 2 つのクラスターに分類されている様子が分かる。また、右の図は各クラスターのファジィ分類関数を表している。

3.5 まとめ

本章では、 c -平均法を中心としたクラスタリングアルゴリズムをカーネル法を利用することで非線形化し、従来の方法では求めることができなかった最適な分類境界を求めることを目的として、その拡張手法について述べた。

カーネル法をクラスタリングに利用する際、高次元特徴空間上のクラスター中心を直接求めることができないという問題を、類似度の更新を行うことによって解決した。その結果、高次元特徴空間を利用したクラスタリングが可能となった。なぜならば、類似度の更新式は帰属度とカーネル関数のみを用いて計算することができるため、高次元空間上のクラスター中心を直接用いなくてもクラスタリングができるからである。

また、カーネル関数を利用したファジィクラスタリングにおいても、ファジィ分類関数の式を導出し、カーネル法を利用した場合、従来の方法とは全く異なる分類規則が存在していることが、人工データを用いた数値例により確認できた。

ただし、カーネル法を利用する際にはカーネル関数のパラメータをどのように決定するのかが重要な問題となる。現在のところ、カーネル関数のパラメータは経験的あるいは実験的に決定しており、厳密な意味でクラスタリングを行っているとは言えない。よって、最適なパラメータの決定方法について今後は考察していかなければならない。また、カーネル法を利用した際の計算量の増加をどのように抑えていくかについても重要なポイントである。なぜならば、 c -平均クラスタリングにカーネル法を利用すると、類似度の更新式の計算量は $O(n^3)$ となり、大量のデータを効率よく解析できるという非階層的クラスタリングの利点が失われてしまうからである。そのため、カーネル法を利用しても計算量が増加しない、もしくは計算量が増加したとしてもできるだけその増加を抑えられるアルゴリズムを考察していかなければならない。さらに、より複雑なデータあるいは実データに対して提案したアルゴリズムを適用していくことで、このアルゴリズムが持つ性能を評価・検討していく必要がある。

第4章 競合学習モデルに基づくクラスタリング

4.1 はじめに

これまで述べてきたクリスプ c -平均法やファジィ c -平均法は、与えられたクラスター数、あるいはクラスター数を仮定した場合の全体評価関数の最適化を実施するものであり、教師無し分類問題を明確に定式化できるものの、期待した結果が常に得られるとは限らないという問題点があった。これは、評価関数（目的関数）のわずかな変化に対してクラスター構造が敏感に変化し、この敏感さがクラスタリングの妥当性の評価を困難にしていると考えられているからである。さらに、評価関数の最適化を行う際には、全てのデータの情報が必要となり、大量のデータを扱う場合には計算量が非常に多くかかってしまうという問題点も指摘されている。

これらの問題点を解決する方法として、競合学習モデルに基づくクラスタリングアルゴリズムが提案されている [20]。この手法は、オンライン型のクラスタリング手法とも呼ばれており、データを逐次的に処理していく手法なので、大量のデータを取り扱う場合に有効であることが知られている。

井口ら [46, 47] は競合学習の考え方をを用いた学習ベクトル量子化¹によるクラスタリングを、カーネル法を適用することで非線形分離可能なアルゴリズムに拡張し、さらに、バッチ型²の c -平均法のカーネル化の際に問題となっている計算量の増加を防ぐことに成功している。また、この応用として自己組織化マップのアルゴリズムをカーネル化することにより、高次元特徴空間におけるデータの可視化にも成功している。

本章では、 c -平均法とは異なるアプローチである競合学習モデルに基づくクラスタリングアルゴリズムを、カーネル法を利用することで非線形化する方法について述べる。

4.2 競合学習クラスタリング

競合学習とは、ニューラルネットワークの分野の自己組織化過程における代表的な学習であり、図 4.1 に示されているように入力層と出力層の2層のネットワークを構成する。その基本的な考え方は、すべてのニューロンの中で入力に対して最も適合するニューロンのみを更

¹LVQ - Learning Vector Quantization と呼ばれる。

²すべてのデータの情報を利用し一括処理する方法

新するという「Winner Takes All - WTA」と呼ばれる規則に基づいている。このため、学習による調整がその時点で与えられるパターンに最も類似度の高いクラスターに限定されるので、現時点でのパターンとは関連の無い、すでに見出されたクラスターの特徴がそのパターンにより乱されることは無いという特徴を持っている。

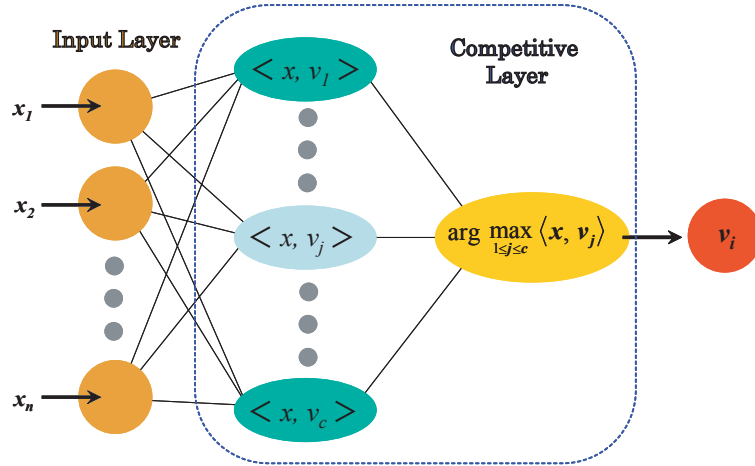


図 4.1: クラスタリングのための競合学習モデル

ここで、時間を表す変数を t とし、 $t = 1, 2, \dots$ と離散的な値をとるものとし、各個体は p 次元ベクトル空間におけるベクトル $x_k = (x_k^1, \dots, x_k^p) \in \mathbf{R}^p, k = 1, \dots, n$ で表されるものとする。また、ここでもクラスター数を c とし、各クラスターを G_i とすると、クラスター中心は $v_i, i = 1, \dots, c$ で表されるものとする。

時刻 t において入力された個体 x_k は各クラスター中心との内積が計算され、最も内積が大きいクラスター中心 v_i を持つクラスター G_i に割り当てられる。

$$i = \arg \max_{1 \leq j \leq c} \langle x_k, v_j \rangle$$

このとき、最も内積が大きいクラスター中心 v_i のみが以下の競合学習則に基づいて、入力ベクトルに近づくように更新される。

$$v_i \leftarrow v_i + \eta(t) x_k$$

$$v_i \leftarrow v_i / \|v_i\|$$

ここで、 $\eta(t)$ は以下の条件を満たす学習率係数と呼ばれるパラメータで、時間と共に減少させる。

$$\sum_{t=1}^{\infty} \eta(t) = \infty, \sum_{t=1}^{\infty} \eta^2(t) < \infty, t = 1, 2, \dots \quad (4.1)$$

ここで、競合学習モデルに基づくクラスタリングアルゴリズム CCL (Clustering by Competitive Learning) を以下に示す。

アルゴリズム CCL.

CCL1. $v_i, i = 1, \dots, c$ の初期値を設定し, $x_k \leftarrow x_k / \|x_k\|$ とする. $t = 1, 2, \dots$ について収束するまで CCL2 と CCL3 を繰り返す.

CCL2.

$$i = \arg \max_{1 \leq j \leq c} \langle x_k, v_j \rangle$$

を求め, x_k をクラスター G_i に割り当てる.

CCL3. v_i を更新する.

$$v_i \leftarrow v_i + \eta(t) x_k$$

$$v_i \leftarrow v_i / \|v_i\|$$

アルゴリズム CCL においては、データが時間と共に一つずつ入力されていき、各クラスター中心との内積が最も大きいクラスターに割り当てられていく。つまり、入力されたデータと各クラスター中心との位置関係だけでクラスタリングが行われていることになる。この手法のことをオンライン型クラスタリングと呼ぶ。なお、アルゴリズムの収束性に関しては、最大学習回数を設定することで保障される。

4.3 ファジィ競合学習クラスタリング

競合学習モデルに基づくクラスタリングアルゴリズムのファジィ化については、ファジィ c -平均法に用いられているメンバーシップ関数（帰属度の式）の利用が考えられる [57, 58]。ファジィクラスタリングの特徴の一つに、ファジィ分類関数がある。ファジィ分類関数の詳細については第 2 章で述べたが、この関数を利用することによりクリスプな分類よりもより多くの情報を得ることができるということが知られている [75, 76]。

ファジィ c -平均法において、個体 x_k がクラスター G_i に所属する帰属度については $c \times n$ 行列 $U = (u_{ik})$ で表される。この時、帰属度 U の制約条件は (2.9) と同様の

$$M_f = \left\{ (u_{ik}) : u_{ik} \in [0, 1], \sum_{i=1}^c u_{ik} = 1, \forall k \right\}$$

となる．帰属度を計算する式として以下のようなものが利用可能である．

$$u_{ik} = \left\{ \sum_{j=1}^c \left(\frac{1 - \langle x_k, v_i \rangle}{1 - \langle x_k, v_j \rangle} \right)^{\frac{1}{m-1}} \right\}^{-1} \quad (m > 1) \quad (4.2)$$

$$u_{ik} = \frac{\exp(\lambda \langle x_k, v_i \rangle)}{\sum_{j=1}^c \exp(\lambda \langle x_k, v_j \rangle)} \quad (\lambda > 0) \quad (4.3)$$

ここで，式 (4.2) は Bezdek らによって提案された標準的ファジィ c -平均法における帰属度の式に基づくものであり，式 (4.3) は宮本らによって提案されたエントロピー正則化ファジィ c -平均法における帰属度の式に基づくものである．本研究では帰属度の計算に式 (4.3) を使用する．

ファジィ化された競合学習モデルに基づくクラスタリングアルゴリズム FCCL (Fuzzy Clustering by Competitive Learning) を以下に示す．

アルゴリズム FCCL.

FCCL1. $v_i, i = 1, \dots, c$ の初期値を設定し， $x_k \leftarrow x_k / \|x_k\|$ とする． $t = 1, 2, \dots$ について収束するまで FCCL2 と FCCL3 を繰り返す．

FCCL2.

$$i = \arg \max_{1 \leq j \leq c} \langle x_k, v_j \rangle$$

を求め，個体 x_k の帰属度を式 (4.3) により計算する．

FCCL3. v_i を更新する．

$$v_i \leftarrow v_i + \eta(t) \sum_{k=1}^n u_{ik} x_k$$

$$v_i \leftarrow v_i / \|v_i\|$$

ファジィ分類関数

競合学習モデルを用いたクラスタリングにおいても，ファジィ分類関数を次の式を用いて計算することができる．本研究では，エントロピー正則化ファジィ c -平均法の帰属度の式を用いているので，この場合のファジィ分類関数は以下ようになる．

$$U_i^c(x) = \frac{\exp(\lambda S_i(x))}{\sum_{j=1}^c \exp(\lambda S_j(x))}$$

ただし，収束後の v_i を用いて

$$S_i(x) = \frac{\langle x, v_i \rangle}{\|x\| \|v_i\|}$$

を計算することに注意する．

4.4 競合学習クラスタリングの非線形化

競合学習モデルに基づくクラスタリングも c -平均法と同様に，非線形分離不可能な線形モデルである．そこで本節では，第 3 章でも述べたようにカーネル法を利用することでアルゴリズムを非線形化する方法について述べる．

4.4.1 カーネル・競合学習クラスタリング

高次元特徴空間上の時刻 t における類似度として次の式を考える．

$$p_{ik}(t) = \langle y_k, v_i \rangle \quad (4.4)$$

ここで， $y_k \leftarrow \Phi(x_k) / \|\Phi(x_k)\|$ である．

カーネル法を利用する場合，特徴空間におけるクラスター中心 v_i を明示的に用いないので，特徴空間における類似度 p_{ik} やカーネル関数のみを用いて計算を進める必要がある．そこで， $p_{ik}(t+1)$ の更新式を次のように求める．アルゴリズム CCL におけるクラスター中心の更新式は

$$\begin{aligned} v_i &\leftarrow v_i + \eta(t)x_k \\ v_i &\leftarrow v_i / \|v_i\| \end{aligned}$$

であるので，

$$\begin{aligned} p_{ik}(t+1) &= \langle y_k, v_i(t+1) \rangle \\ &= \left\langle y_k, \frac{v_i + \eta(t)y_j}{\|v_i + \eta(t)y_j\|} \right\rangle \\ &= \frac{\langle y_k, v_i \rangle + \eta(t)\langle y_k, y_j \rangle}{\|v_i + \eta(t)y_j\|} \end{aligned}$$

となる．ここで， $V_i(t) = \|v_i\|$ とおき，

$$\langle y_j, y_k \rangle = \frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}}$$

$$K_{jk} = K(x_j, x_k)$$

であることに注意すると

$$p_{ik}(t+1) = \frac{p_{ik}(t) + \eta(t) \frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}}}{V_i(t+1)} \quad (4.5)$$

$$V_i^2(t+1) = V_i^2(t) + 2\eta(t)p_{ij}(t) + \eta^2(t)K_{jj} \quad (4.6)$$

となる．これらの式を用いて p_{ik} の更新を行うことにより，カーネル法を利用した競合学習クラスタリングを行うことができる．以下に，カーネル法を利用した競合学習クラスタリングアルゴリズム K-CCL (Kernel-based Clustering by Competitive Learning) を示す．

アルゴリズム K-CCL

K-CCL1. $p_{ik}, i = 1, \dots, c, k = 1, \dots, n$ の初期値を設定し， $t = 1, 2, \dots$ について収束するまで K-CCL2 と K-CCL3 を繰り返す．

K-CCL2.

$$i = \arg \max_{1 \leq j \leq c} p_{jk}(t)$$

を求め，個体 x_k をクラスター G_i に割り当てる．

K-CCL3. p_{ik} を式 (4.5) と式 (4.6) を用いて更新する．

4.4.2 カーネル・ファジィ競合学習クラスタリング

カーネル法を利用したファジィ競合学習クラスタリングも前述と同様に考えることができる．帰属度の計算については，高次元特徴空間における類似度の式 (4.4) を用いて，

$$u_{ik} = \frac{\exp(\lambda p_{ik})}{\sum_{j=1}^c \exp(\lambda p_{jk})} \quad (4.7)$$

で計算される .

アルゴリズム FCCL におけるクラスター中心の更新式は

$$v_i \leftarrow v_i + \eta(t) \sum_{k=1}^n u_{ik} x_k$$

$$v_i \leftarrow v_i / \|v_i\|$$

であるので ,

$$p_{ik}(t+1) = \langle y_k, v_i(t+1) \rangle$$

$$= \left\langle y_k, \frac{v_i + \eta(t) \sum_j u_{ij} y_j}{\|v_i + \eta(t) \sum_j u_{ij} y_j\|} \right\rangle$$

$$= \frac{\langle y_k, v_i \rangle + \eta(t) \sum_j u_{ij} \langle y_k, y_j \rangle}{\|v_i + \eta(t) \sum_j u_{ij} y_j\|}$$

となる . ここで , $V_i(t) = \|v_i\|$ とおき ,

$$\langle y_j, y_k \rangle = \frac{K_{jk}}{\sqrt{K_{jj} K_{kk}}}$$

$$K_{jk} = K(x_j, x_k)$$

であることに注意すると

$$p_{ik}(t+1) = \frac{p_{ik}(t) + \eta(t) \sum_{j=1}^n u_{ij} \frac{K_{jk}}{\sqrt{K_{jj} K_{kk}}}}{V_i(t+1)} \quad (4.8)$$

$$V_i^2(t+1) = V_i^2(t) + 2\eta(t) \sum_{j=1}^n u_{ij} p_{jk}(t) + \eta^2(t) \sum_{j=1}^n \sum_{\ell=1}^n u_{ij} u_{i\ell} K_{j\ell} \quad (4.9)$$

となる . これらの式を用いて p_{ik} の更新を行い , 各個体の帰属度を計算することを繰り返せば , カーネル法を利用したファジィ競合学習クラスタリングを行うことができる . 以下に , カーネル法を利用したファジィ競合学習クラスタリングアルゴリズム K-FCCL (Kernel-based Fuzzy Clustering by Competitive Learning) を示す .

アルゴリズム K-FCCL

K-FCCL1. $p_{ik}, i = 1, \dots, c, k = 1, \dots, n$ の初期値を設定し, $t = 1, 2, \dots$ について収束するまで K-FCCL2 と K-FCCL3 を繰り返す .

K-FCCL2.

$$i = \arg \max_{1 \leq j \leq c} p_{jk}(t)$$

を求め, 個体 x_k の帰属度を式 (4.7) を用いて計算する .

K-FCCL3. p_{ik} を式 (4.8) と式 (4.9) を用いて更新する.

ファジィ分類関数

第 3 章でも述べたが, カーネル法を利用する場合, 写像 Φ^{-1} は明示的に求まらないので, 高次元特徴空間上でのクラスター中心を直接求めることはできない . しかし, クラスター中心が求まらなくても, ファジィ分類関数を次の方法で計算することができる .

$$U_i^e(x) = \frac{\exp(\lambda P_i(x))}{\sum_{j=1}^c \exp(\lambda P_j(x))}$$

ただし, 収束後の u_{ik}, p_{ij} を用いて

$$P_i(x) = \frac{\sum_j u_{ij} K(x, x_j)}{\sqrt{1 + 2 \sum_j u_{ij} p_{ij} + \sum_j \sum_{\ell} u_{ij} u_{i\ell} K_{j\ell}}}$$

を計算することに注意する . また, 学習率係数 η については, クラスタリングが行われた後は微小数になっているので, 本研究においてはファジィ分類関数を計算する際に省略しているということに注意されたい .

さらに, カーネル関数 K についてはガウシアンカーネルを用いているので,

$$K(x, x_j) = \exp(-c \text{nst} \|x - x_j\|^2)$$

により求められる .

4.5 自己組織化マップによる高次元特徴空間におけるデータの可視化

自己組織化マップ³とは、競合学習に基づいた位相マップ形成の代表的なモデルであり、複雑かつ高次元のデータを相互間の位置関係を保持したまま、低次元のマップに写像を行う手法である [62] .

高次元空間への写像を利用したクラスタリングにおいては、カーネル法を利用することで非線形な分類境界を求められることが確認できているが、実際に高次元特徴空間上でどのようにデータが分離されているのかが見ることができない。それは、写像関数が陽に現れないため、どのように非線形写像が行われているのかが分からないからである。しかしながら、高次元特徴空間上でデータの位置関係を知ることは、多次元のデータを分析する際には、非常に有効な手段となり得るであろう。

井口らは、アルゴリズム SOM にカーネル法を利用したカーネル SOM を提案し、高次元特徴空間におけるデータの位置関係の可視化に成功している [46, 47] . これにより、概念的に考えられていた高次元特徴空間上におけるデータの線形構造が視覚的に確認可能になった。

本節では、内積を類似度とした自己組織化マップにカーネル法を利用することで、高次元特徴空間上のデータの位置関係を可視化する方法について述べる。

4.5.1 自己組織化マップ

SOM は入力層と出力層の 2 層のネットワーク構造を持ち、基本的に競合過程、協調過程の 2 つの過程から成る。競合過程では、WTA の原則に基づいた競合が行われる。ここで、 K 個のニューロンが格子状に配置されているとする。SOM ではニューロンのことをノードと呼ぶ。この過程では、以下のような競合則により最もデータに適合する勝者ノードが選ばれる。

$$m_\ell(t) = \arg \max_{1 \leq i \leq K} \langle x(t), m_i(t) \rangle \quad (4.10)$$

ここで、 $m_i \in \mathbf{R}^p, i = 1, \dots, K$ は、各ノードに含まれている参照ベクトルと呼ばれるものである。

SOM においては WTA の原則で競合過程が行われるが、これは局所的な動作であり、大局的な動作、すなわち位相情報の保持を実現するためには協調過程が必要である。勝者ノードを中心とした近傍ノードも、勝者ノードの動きに同調するように次の学習則に従う。

$$m_\ell(t+1) := m_\ell(t) + h_{c_i}(t)x(t) \quad (4.11)$$

位相近傍の定義はノードの格子の形によるが、一般的な 2 次元マップのグリッドの配列としては正方形格子と六角形格子がある。データ構造の視覚化を意図する本研究では、ノード間

³SOM - Self-Organizing Maps とも呼ばれる。

の距離が等しい六角形格子を用いる。一般に、近傍領域 N_c は時間に関する単調減少関数であって、 N_c の半径は時間とともに小さくなる⁴。

$$h_{ci}(t) = \begin{cases} \eta(t), & i \in N_c \\ 0, & i \notin N_c \end{cases} \quad (4.12)$$

また、以下のようなより滑らかな近傍関数を用いることも可能である。

$$h_{ci} = \eta(t) \cdot \exp\left(\frac{\|r_c - r_i\|^2}{2\delta^2(t)}\right) \quad (4.13)$$

ここで、 r_i はノードの位置、 δ は近傍半径を表す。

アルゴリズム SOM

SOM1. 初期ノード $m_i(1), i = 1, \dots, K$ を定め、 $x_k \leftarrow x_k / \|x_k\|$ とする。 $t = 1, 2, \dots$ について収束するまで SOM2 と SOM3 を繰り返す。

SOM2.

$$\ell = \arg \max_{1 \leq i \leq K} \langle x_k, m_i \rangle$$

を求める。

SOM3. 勝者ノード m_ℓ の近傍 N_c に含まれるすべてのノードを更新する。

$$m_i \leftarrow m_i + \eta(t)x_k$$

$$m_i \leftarrow m_i / \|m_i\|$$

競合学習モデルに基づくクラスタリングでは勝者ベクトルのみを学習させていたのに対し、SOM では勝者ノードの近傍領域 N_c に含まれる全てのノードが学習されるのが特徴である。また、 K はノードの数を表すものであり、クラスター数を表すものではないことに注意されたい。

4.5.2 カーネル・自己組織化マップ

カーネル法を利用した SOM は、競合学習モデルに基づくクラスタリングと同様に導くことができる。

⁴ここで用いられている N_c の下付の c はクラスター数を表す記号ではなく、近傍領域のノードの位置を表す記号であることに注意されたい。

まず，高次元特徴空間上の時刻 t における類似度として次の式を考える．

$$p_{ik}(t) = \langle y_k, m_i \rangle \quad (4.14)$$

ここで， $y_k \leftarrow \Phi(x_k)/\|\Phi(x_k)\|$ である．

アルゴリズム SOM における参照ベクトルの更新式は

$$\begin{aligned} m_i &\leftarrow m_i + \eta(t)x_k \\ m_i &\leftarrow m_i/\|m_i\| \end{aligned}$$

であるので，

$$\begin{aligned} p_{ik}(t+1) &= \langle y_k, m_i(t+1) \rangle \\ &= \left\langle y_k, \frac{m_i + \eta(t)y_j}{\|m_i + \eta(t)y_j\|} \right\rangle \\ &= \frac{\langle y_k, m_i \rangle + \eta(t)\langle y_k, y_j \rangle}{\|m_i + \eta(t)y_j\|} \end{aligned}$$

となる．ここで， $M_i(t) = \|m_i\|$ とおき，

$$\begin{aligned} \langle y_j, y_k \rangle &= \frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}} \\ K_{jk} &= K(x_j, x_k) \end{aligned}$$

であることに注意すると

$$p_{ik}(t+1) = \frac{p_{ik}(t) + \eta(t)\frac{K_{jk}}{\sqrt{K_{jj}K_{kk}}}}{M_i(t+1)} \quad (4.15)$$

$$M_i^2(t+1) = M_i^2(t) + 2\eta(t)p_{ij}(t) + \eta^2(t)K_{jj} \quad (4.16)$$

となる．これらの式を用いて p_{ik} の更新を行うことにより，カーネル法を利用した SOM のアルゴリズムを実行することができる．以下に，カーネル法を利用した自己組織化マップのアルゴリズム K-SOM (Kernel-based Self Organizing Maps) を示す．

アルゴリズム K-SOM

K-SOM1. $p_{ik}, i = 1, \dots, K, k = 1, \dots, n$ の初期値を設定する . $t = 1, 2, \dots$ について収束するまで K-SOM2 と K-SOM3 を繰り返す .

K-SOM2.

$$\ell = \arg \max_{1 \leq i \leq K} p_{ik}(t) \quad (4.17)$$

を求める .

K-SOM3. 勝者ノード m_ℓ の近傍 N_c に含まれるすべてのノード i に対する p_{ik} を式 (4.15) と式 (4.16) によって更新する .

アルゴリズム K-SOM においては m_i の値を明示的に用いないことに注意する . 用いるのはグリッド上にある m_i の位置だけである .

4.6 数値例

競合学習モデルに基づくクラスタリングアルゴリズムにカーネル法を利用することで、非線形な分類境界を得ることができるアルゴリズムに拡張できたことを、人工データおよび実データを用いた数値例により示す .

• 人工データ 1

ここでは、個体数 150 の 3 次元人工データに対して、アルゴリズム CCL と K-CCL により 2 つのクラスターに分類を行う . さらに、アルゴリズム K-SOM によりマップを生成し、高次元特徴空間上に写像されたデータの位置関係の視覚化を行う .

図 4.2 はアルゴリズム CCL により、2 つのクラスターに分類した結果であり、内積を用いたアルゴリズムでは、原点を通る超平面により 2 つのクラスターに分類されている様子が確認できる . 一方、図 4.3 はアルゴリズム K-CCL により 2 つのクラスターに分類した結果であり、中心の塊とその周りのリングの 2 つのクラスターに分類されている様子が分かる . すなわち、このデータに対して最適な分類が行われたということが確認できる .

また、図 4.4 はアルゴリズム K-SOM により生成されたマップである . この図から、データ空間における外側のリングと内側の塊が、高次元特徴空間上では境界のはっきりした線形分離可能な状態になっていることが確認できる . このように、カーネル法を利用した自己組

織化マップにより，内積を用いた場合においても，高次元特徴空間上のデータの位置関係を可視化することに成功したことが分かる．

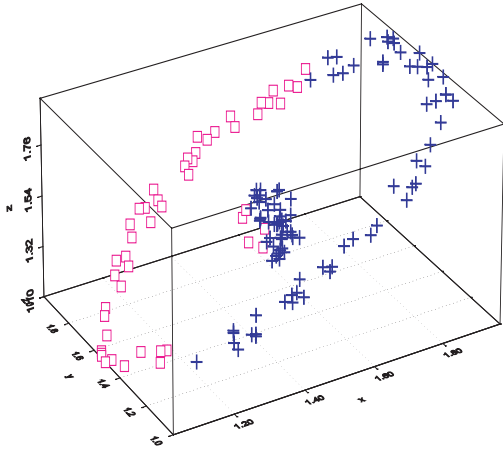


図 4.2: アルゴリズム CCL ($\eta(0) = 0.5$) により 2つのクラスターに分類した結果．各クラスターについては，それぞれ と+で表されており，原点を通る超平面により線形に 2つのクラスに分類されている様子が確認できる．

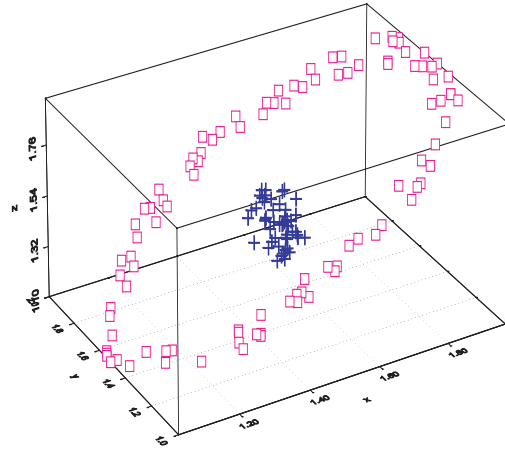


図 4.3: アルゴリズム K-CCL ($\eta(0) = 0.05, cnst = 20$) により 2つのクラスターに分類した結果．各クラスターについては，それぞれ と+で表されており，非線形に分類されている様子が確認できる．

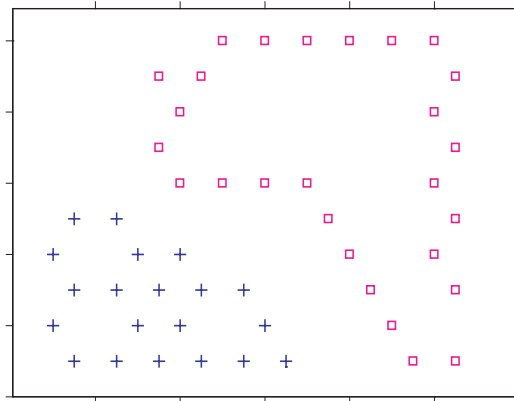


図 4.4: アルゴリズム K-SOM ($\eta(0) = 0.05, cnst = 20$) により生成されたマップ．各クラスターについては，図 4.3 と同じ と+で表されており，高次元特徴空間上では線形分離可能な状態になっていることが確認できる．

- 人工データ 2

ここでは、個体数 150 の 2 次元人工データに対して、アルゴリズム FCCL と K-FCCL により 3 つのクラスターに分類を行った例について示す。

図 4.5 はアルゴリズム FCCL により 3 つのクラスターに分類した結果であり、線形に 3 つのクラスターに分類されている様子が確認できる。一方、図 4.6 はアルゴリズム K-FCCL により 3 つのクラスターに分類した結果であり、真ん中の塊とその周りの 2 つのリングに分類されている様子が分かる。すなわち、非線形に最適な分類ができていることが確認できる。また、図 4.7 はそれぞれのクラスターのファジィ分類関数を示しており、非線形な分類境界が得られている様子が分かる。これらの結果から、カーネル法を利用することにより、より複雑なデータに対しても期待通りの分類を行うことができるアルゴリズムに拡張できたことが確認できる。ちなみに、このデータをバッチ型のアルゴリズム K-FCM で 3 つのクラスターに分類しても、図 4.6 と同様の結果を得られることが確認できている。しかしながら、ユークリッド距離の 2 乗を用いたアルゴリズムでは、カーネル法を利用した手法であってもこのデータを期待通りに分類することはできていない（図 4.8 参照）

図 4.9 はアルゴリズム K-SOM により生成されたマップである。この図から、データ空間における 3 つのクラスターが、高次元特徴空間上では 3 つのクラスターが線形分離可能な状態であることが確認できる。

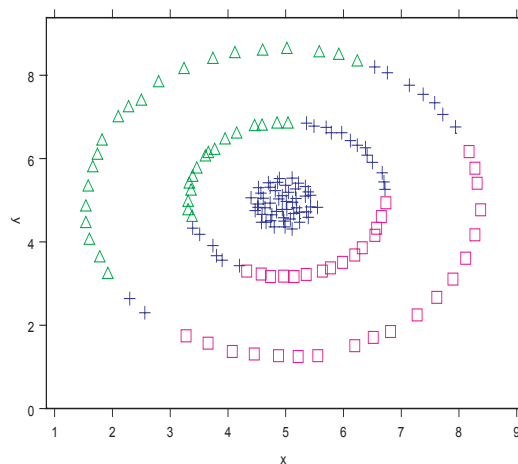


図 4.5: アルゴリズム FCCL ($\lambda = 10, \eta(0) = 0.5$) により 3 つのクラスターに分類した結果である。各クラスターは Δ , $+$, \square で表示されており、線形に 3 つのクラスターに分類されている様子が確認できる。

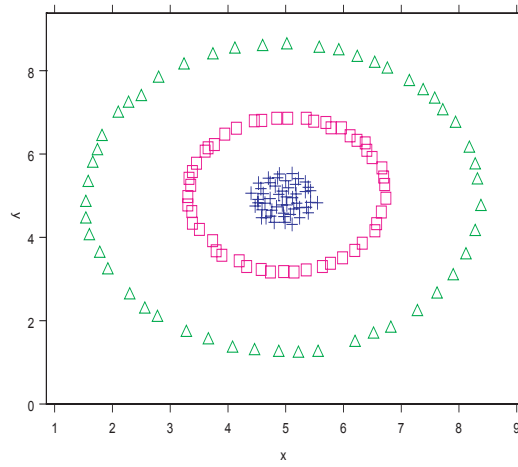


図 4.6: アルゴリズム K-FCCL ($\lambda = 10, cnst = 1.5, \eta(0) = 0.05$) により 3つのクラスターに分類した結果である。各クラスターは \square , $+$, \triangle で表示されており、非線形に 3つのクラスターに分類されている様子が確認できる。

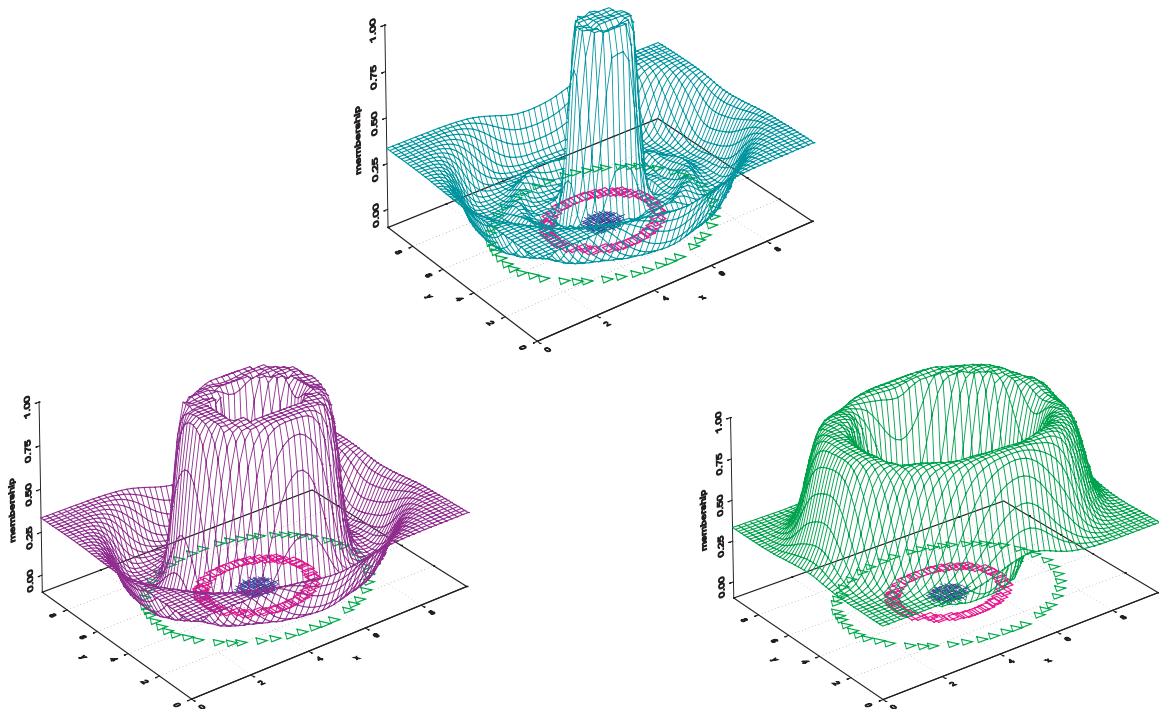


図 4.7: アルゴリズム K-FCCL により 3つのクラスターに分類した結果得られた各クラスターのファジィ分類関数。縦軸は各クラスターに対する帰属度（メンバーシップ）を表している。

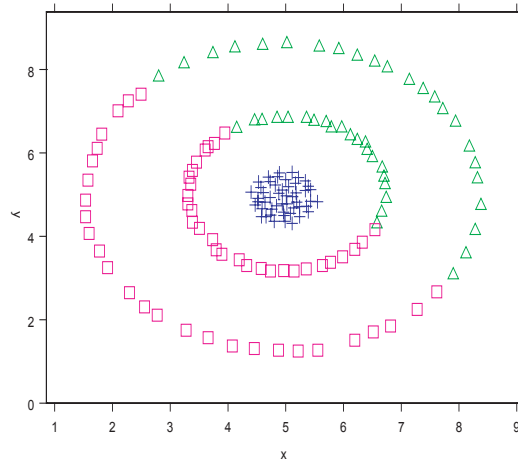


図 4.8: ユークリッド距離の 2 乗を用いたカーネル法を利用したファジィ c -平均法 ($\lambda = 10, cnst = 1$) により 3 つのクラスターに分類した結果である。各クラスターは \square , $+$, \triangle で表示されており, 真ん中のクラスターのみが非線形に分類されたが, 周りのリングについては非線形に分類されていない様子が確認できる。

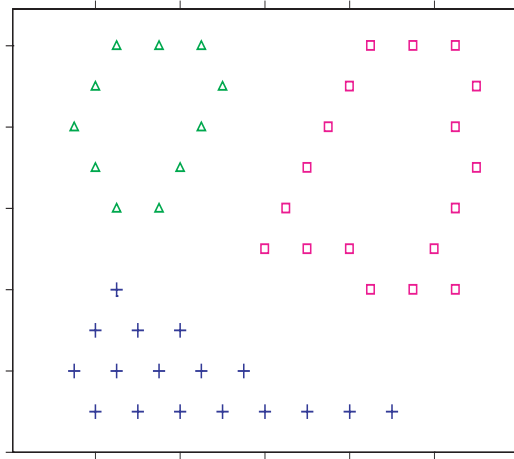


図 4.9: アルゴリズム K-SOM ($\eta(0) = 0.05, cnst = 1$) により生成されたマップ。各クラスターは図 4.6 と同じ \square , $+$, \triangle で表示されている。データ空間における内側の塊と外側の 2 つのリングが, 高次元特徴空間上では境界のはっきりした線形分離可能な状態になっていることが確認できる。

● 人工データ 3

ここでは、個体数 300 の 3 次元人工データに対して各手法により、球面状に分布しているデータ中にもう一つの球面状の分布を持つデータを 2 つのクラスターに分類を行った例について示す。

図 4.10 はアルゴリズム CCL およびアルゴリズム FCCL により、2 つのクラスターに分類した結果であり、超平面によって線形に 2 つのクラスターに分類されている様子が確認できる。すなわち、カーネル化していないクラスタリングアルゴリズムでは分類境界は線形になっている様子が確認できる。一方、図 4.11 はアルゴリズム K-CCL およびアルゴリズム K-FCCL により 2 つのクラスターに分類した結果であり、内側の球と外側の球の 2 つのクラスターに最適な分類ができていることが確認できる。これらの結果から、カーネル法を利用することにより、3 次元においてより複雑な形状をしたデータに対しても、最適な分類を行うことができるアルゴリズムに拡張できたことが確認できる。

また、図 4.12 は、アルゴリズム K-SOM により生成したマップであるが、データ空間における内側の球と外側の球が、高次元特徴空間上では境界のはっきりした線形分離可能な状態になっていることが確認できる。

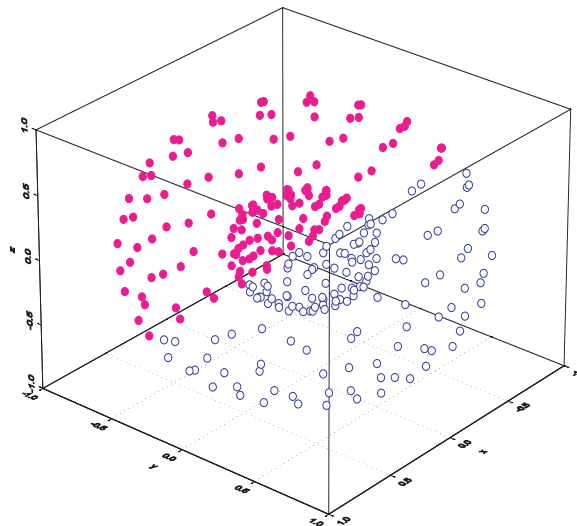


図 4.10: アルゴリズム CCL ($\eta(0) = 0.5$) および FCCL ($\lambda = 10, \eta(0) = 0.5$) により 2 つのクラスターに分類した結果である。各クラスターは●, ○で表示されており、線形に 2 つのクラスターに分類されている様子が確認できる。

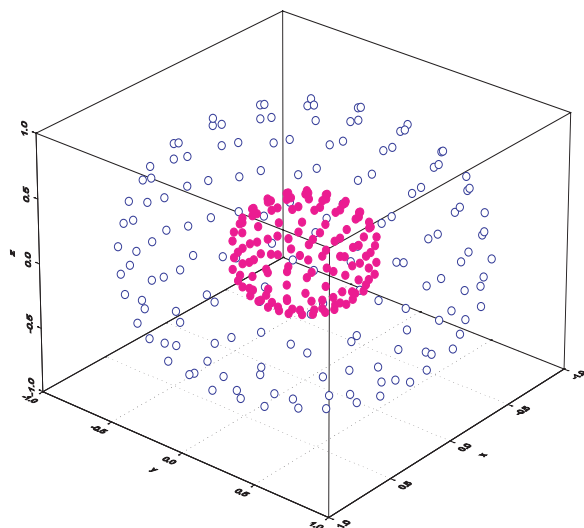


図 4.11: アルゴリズム K-CCL ($cnst = 10, \eta(0) = 0.05$) および K-FCCL ($\lambda = 10, cnst = 10, \eta(0) = 0.05$) により 2 つのクラスターに分類した結果である。各クラスターは●, ○で表示されており, 内側の球と外側の球に非線形に 2 つのクラスターに分類されている様子が確認できる。

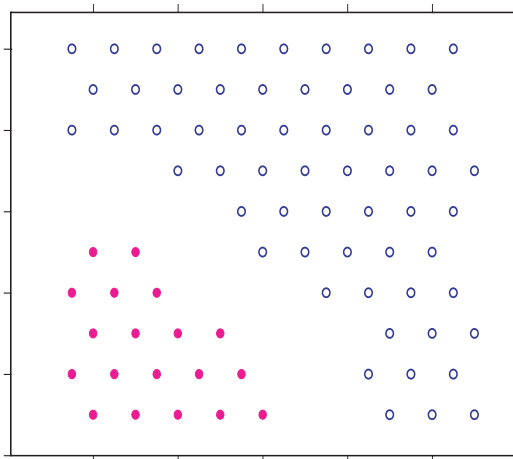


図 4.12: アルゴリズム K-SOM ($\eta(0) = 0.06, cnst = 7.8$) により生成されたマップ。各クラスターは図 4.11 と同じ●, ○で表示されており, 内側の球のデータと外側の球のデータの間にはっきりとした境界ができている様子が確認できる。すなわち, 高次元特徴空間上では線形分離可能な状態になっているということである。

● 人工データ 4

ここでは、個体数 200 の 3 次元人工データに対して、アルゴリズム CCL と K-CCL により螺旋状のデータを 2 つのクラスターに分類を行った例について示す。

図 4.13 はアルゴリズム CCL による分類結果を示しており、螺旋状のデータは原点を通る超平面により 2 つのクラスターに線形に分類されている様子が分かる。一方、図 4.14 はアルゴリズム K-CCL による分類結果を示しており、螺旋状のデータは所望の 2 つのクラスターに分類されている様子が分かる。これらの結果からカーネル法を利用したアルゴリズムにより、より複雑なデータに対しても、期待通りの分類結果を得ることが可能であることが確認できる。

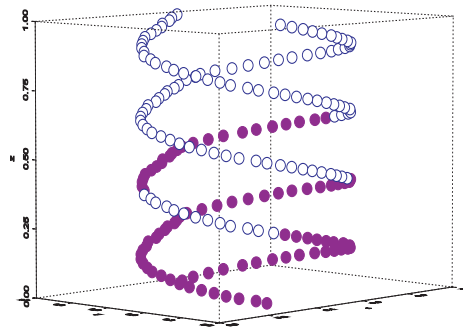


図 4.13: アルゴリズム CCL ($\eta(0) = 0.5$) により 2 つのクラスターに分類した結果である。各クラスターは●, ○で表示されており、超平面により 2 つのクラスターに線形に分類されている様子が確認できる。

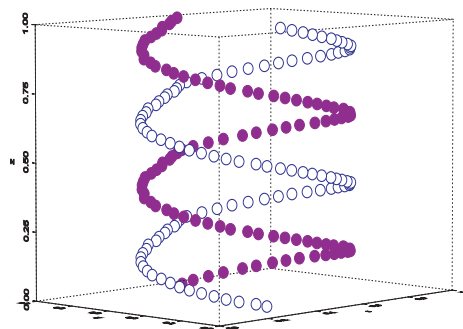


図 4.14: アルゴリズム K-CCL ($cnst = 200, \eta = 0.5$) により 2 つのクラスターに分類した結果である。各クラスターは●, ○で表示されており、2 つの螺旋状の形に分類されている様子が確認できる。

● 実データ

ここでは、実データとして Iris データを用いた実験について示す。Iris データ [105] とは、統計学におけるデータ解析の例としてよく用いられるデータセットであり、以下のようなアヤメ科の 3 品種の寸法を表す個体数 150 の 4 次元データである。

Iris データ

個体数 : 150

属性 (次元) 数 : 4 (各属性は実数値)

ガクの長さ (sepal length)

ガクの幅 (sepal width)

花弁の長さ (petal length)

花弁の幅 (petal width)

クラス数 : 3 (アヤメの品種)

Iris-setosa (50)

Iris-versicolor (50)

Iris-virginica (50)

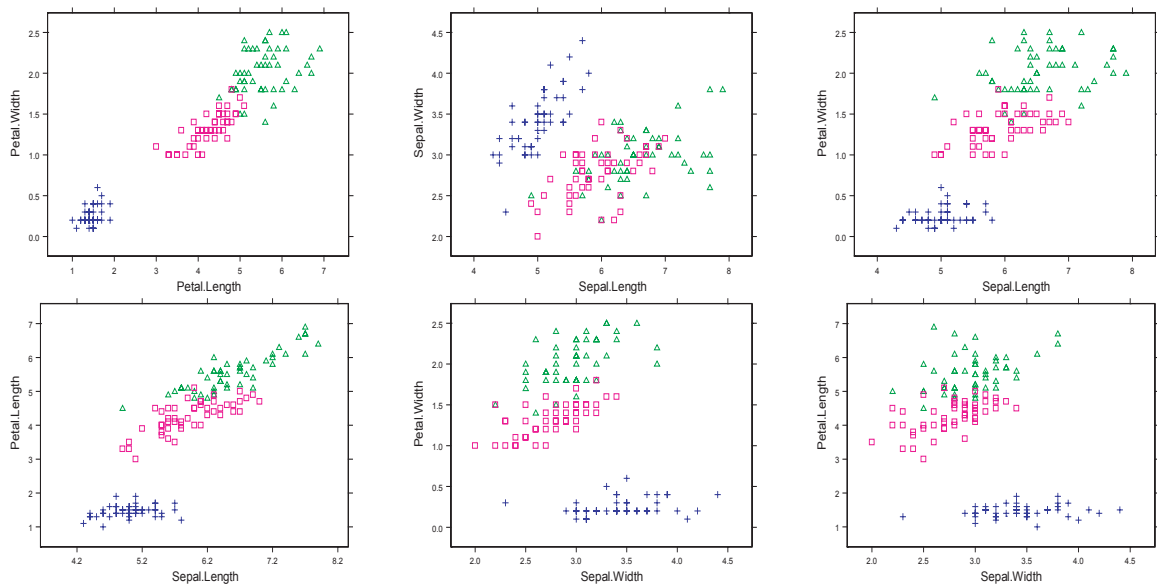


図 4.15: Iris データの 2 次元散布図。+ は *Iris-setosa* , □ は *Iris-versicolor* , △ は *Iris-virginica* の各品種を表している。

図 4.15 に各属性ごとの 2 次元散布図を示す。これらの図が示すように、このデータは 3 つの品種のうち *Iris-setosa* は他の 2 品種と区別できるが、*Iris-versicolor* と *Iris-verginica* の 2 品種はデータの一部が混在しているため線形に分離できないという特徴を持っている。また、数値モデルという観点から見ると、Iris データは、花の各品種の特徴を各代表点で表すことのできる集中型の数値モデルとして見なすことができる。

ここでは、各クラスタリングアルゴリズムによりアヤメ科の 3 品種である 3 つのクラスに分類を行い、誤分類数および計算時間を比較することにより実データにおける提案手法の有効性を示す。

表 4.1 は、それぞれのアルゴリズムにおける誤分類数、収束するまでの繰り返し回数および収束までに要した時間を示している。なお、収束回数および計算時間については、異なる初期値についてそれぞれ 100 回ずつアルゴリズムを適用した際の平均をとっている。

これらの結果から、まずバッチ型のアルゴリズム (CCM, FCM, K-CCM, K-FCM) においてはカーネル化することにより誤分類数が増加してしまっていることが分かる。その逆に、競合学習モデルを用いた場合には、カーネル化により誤分類数が減少していることが分かる。誤分類数のみで結果を比較すると、アルゴリズム CCM および FCM が最も優れているようであるが、今回の実験に用いた Iris データにおいては、初期値によって大きく結果が異なってしまう。

すなわち、実際に今回の実験で得られた結果 (誤分類数: 4) は目的関数値が最も大きくなったときのものであるが、100 回の試行中 9 回しか得られていない。したがって、アルゴリズムの安定性の面からみると、必ずしも優れたアルゴリズムであるとは言えない。ただし、このことは競合学習モデルを用いた場合も同様のことが言える。すなわち、アルゴリズム CCL および FCCL においても、初期値によって結果が大きく異なるという結果を得ている。

一方、カーネル法を利用したアルゴリズムにおいては、初期値によって大きく結果が異なるということがなく、100 回の試行中における誤分類数の差はほとんどない。このことから、カーネル法を利用することにより、初期値依存性を減らすことが可能であることが確認できた。

また、計算時間に注目してみると、カーネル法を適用したバッチ型のアルゴリズム (K-CCM, K-FCM) においては、カーネル法を適用しないアルゴリズムと比較して約 900 倍から 1100 倍もの増加が見られるが、競合学習モデルを用いたクラスタリングでは、カーネル化の際の計算時間の増加が約 22 倍から 200 倍で抑えられていることが確認できる。

ちなみに、表 4.2 はユークリッド距離の 2 乗を用いた場合の分類結果であるが、これらの誤分類数を見てみると、内積に基づく類似度を使用したほうが分類性能が良いことが分かる。

表 4.1: 各手法における誤分類数, 収束・学習回数および計算時間の比較

アルゴリズム	誤分類数	収束・学習回数	計算時間 (秒)
CCM	4	4	9.0×10^{-4}
FCM	4	8.5	4.5×10^{-3}
CCL	9	5000	1.7×10^{-2}
FCCL	9	1200	1.3×10^{-2}
K-CCM	10	7.7	1.03
K-FCM	9	20.6	4.2
K-CCL	5	5000	3.9×10^{-1}
K-FCCL	5	3150	2.5

Computer performance:

CPU: Athlon 1.73 GHz, Memory: 256 MB

OS: Microsoft Windows XP

表 4.2: 各手法における誤分類数, 収束・学習回数および計算時間の比較 (ユークリッド距離の 2 乗を用いた場合)

アルゴリズム	誤分類数	収束・学習回数	計算時間 (秒)
CCM	17	2.9	1.6×10^{-3}
FCM	16	14.7	7.8×10^{-3}
CCL	14	5000	1.4×10^{-3}
K-CCM	15	11	1.7
K-FCM	10	10	4.6
K-CCL	9	5000	2.6×10^{-2}

Computer performance:

CPU: Athlon 1.73 GHz, Memory: 256 MB

OS: Microsoft Windows XP

また，図 4.16～図 4.19 は，自己組織化マップのそれぞれの手法により 2 次元上にマップ化したものである．すなわち，図 4.16 は内積を用いたアルゴリズム SOM により生成されたマップ，図 4.17 は内積を用いたカーネル法を利用したアルゴリズム K-SOM により生成されたマップ，図 4.18 はユークリッド距離の 2 乗を用いたアルゴリズム SOM により生成されたマップ，図 4.19 はユークリッド距離の 2 乗を用いたカーネル法を利用したアルゴリズム K-SOM により生成されたマップを表している．図中の + は *Iris-setosa*，△ は *Iris-veisicolor*，□ は *Iris-verginica* の各品種を表している．これらの結果から，Iris データに関しては，ユークリッド距離を用いるよりも内積に基づく類似度を用いて分類した方がより精度の良い結果が得られることが分かる．なぜならば，本研究で提案したアルゴリズム K-SOM によりマップを形成した結果が，各品種間の境界が最もはっきりと現れているからである．

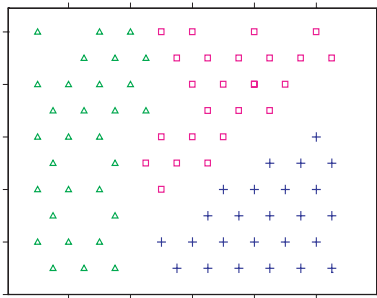


図 4.16: アルゴリズム SOM ($\eta(0) = 0.5$) により生成されたマップ．*Iris-setosa* と他の 2 品種の境界がはっきりと出ていることが確認できる．

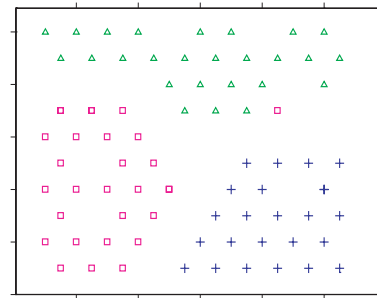


図 4.17: アルゴリズム K-SOM ($\eta(0) = 0.4, cnst = 20$) により生成されたマップ．各品種の境界がはっきりと出ていることが確認できる．

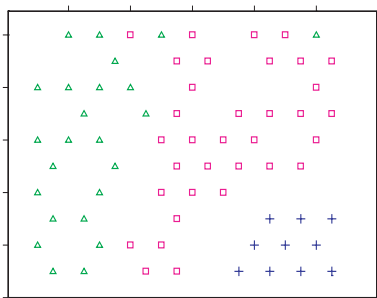


図 4.18: ユークリッド距離の 2 乗を用いたアルゴリズム SOM ($\eta(0) = 0.5$) により生成されたマップ．

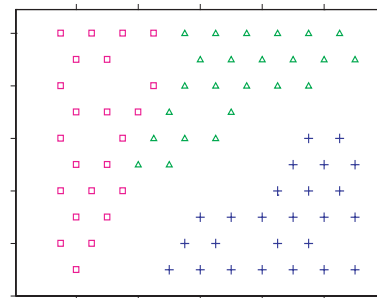


図 4.19: ユークリッド距離の 2 乗を用いたアルゴリズム K-SOM ($\eta(0) = 0.05, cnst = 20$) により生成されたマップ．

4.7 まとめ

本章では、 c -平均法とは異なるアプローチである競合学習モデルを利用したクラスタリングアルゴリズムを、カーネル法を利用することで非線形化し、従来の方法では求めることができなかった最適な分類境界を求めることを目的として、その拡張手法について述べた。あわせて、競合学習の考え方をういた自己組織化マップにカーネル法を利用することで、高次元特徴空間上に写像されたデータを可視化する方法についても述べた。

4つの人工データを用いた実験により、競合学習モデルとカーネル法を組み合わせることで、従来の方法では求めることができなかった最適な分類境界を求めることができるアルゴリズムに拡張できたことを示した。また、実データによる実験として、統計学におけるデータ解析の例としてよく用いられている Iris データを用いた実験においては、競合学習モデルの利点であるカーネル化の際の計算量の増加を、内積を用いた場合においても抑えることが可能であることを示した。さらに、ユークリッド距離の2乗を用いた分類結果と比較して、Iris データにおいては内積に基づく類似度を利用した方が、距離を用いた場合よりも優れた性能を発揮することを示した。

ただし、競合学習モデルを用いた場合においても、カーネル法を利用する際にはカーネル関数のパラメータをどのように決定するのが重要な問題となるため、今後は最適なパラメータの決定方法について考察していく必要がある。また、ファジィ競合学習クラスタリングにおいては、理論的にはカーネル法を利用した際の計算量の増加を防ぐことができていない。この問題を解決するヒントとして、井口ら [48] によって提案されたバッチ型のファジィLVQをカーネル化しても計算時間が増加しない方法を利用することが考えられる。そのため、内積を利用した場合においても、どのようにして競合学習クラスタリングの特徴であるオンライン型の特徴を活かして計算量を防いでいくのかということについて考察していく必要がある。さらに、より多くの実データに対してこのアルゴリズムを適用することで、どこまでの性能を発揮するのかを検討していく必要があり、その際の結果の評価の仕方などについても検討していく必要があると考えられる。

第5章 ファジィマルチ集合に基づく文書情報クラスタリング

5.1 はじめに

計算機システムの高性能化とそれに伴うネットワーク化によって、我々の身のまわりには大量の電子化された情報が存在するようになってきた。これらの情報を効率よく取り扱うための方法として、情報検索と呼ばれる技術がある [59, 60]。

この情報検索において情報アクセスを支援する手段の1つとして、文書その内容に応じて分類・整理するということが考えられる。大規模な文書集合を分類・整理するためにはコンピュータによる自動的な処理が必要となるわけであるが、このための技術の一つとしてクラスタリングが利用可能である [71, 94, 96, 97]。しかしながら、文書情報をクラスタリングするためには、文書に関する情報をなんらかの数理モデルを用いて表現する必要がある。そこで、本研究では文書集合を数値化するための数理モデルとして、ファジィマルチ集合 (fuzzy multisets) を用いることにする。

ファジィバッグ (fuzzy bags) とも呼ばれるファジィマルチ集合は、Yager [110] によって提案された。また、宮本 [72] によってその基本演算の性質が論じられ、ファジィマルチ集合をクラスタリングする手法も提案されている [79]。ファジィマルチ集合の特徴は要素の数と帰属度を同時に表現する構造を持つため、同一の用語が異なる度合いを持って複数回出現しているような Web 上の情報あるいは文献情報などを適切に表現することができると考えられている [80, 84]。本章では、ファジィマルチ集合をモデルとした文書情報クラスタリングの方法について述べる。

5.2 ファジィマルチ集合

通常の集合ではある要素は一つしか存在しないが、要素の重複が許される集合が考察されている [10, 61, 68]。このような集合はマルチ集合あるいはバッグと呼ばれており、通常の集合はマルチ集合の一種とみなすことができる。また、マルチ集合のファジィ化が Yager [110] によって提案され、宮本 [65, 72, 73, 78] によってその基本演算の性質が論じられた。本節ではマルチ集合とファジィマルチ集合の特徴及びそれらの基本演算について述べていく。

5.2.1 マルチ集合の性質と基本演算

マルチ集合とカウント関数

X を通常有限集合, $N = \{0, 1, 2, \dots\}$ を自然数の集合とする. 有限集合 $X = \{x_1, \dots, x_n\}$ 上にマルチ集合が定義されるものと仮定する. クリस्पなマルチ集合 M は,

$$M = \{x_i, \dots, x_i, x_j, \dots, x_j\}$$

のように要素の重複を許す集合であり, 各 $x \in X$ に自然数を対応づける *Count* 関数

$$Count: X \rightarrow \{0, 1, 2, \dots\} \quad (5.1)$$

で特徴付けられる. この関数は要素の重複度, すなわちマルチ集合 M における要素 x の出現回数を表しており, $Count_M(x_1) = 3$ ならば, マルチ集合 M には要素 x_1 が 3 個存在することを意味している.

マルチ集合の基本演算

2 つのマルチ集合 M, N について, マルチ集合の基本演算及びその関係は, *Count* 関数を用いて次のように定義される [65, 72].

マルチ集合の基本演算

1. 包含関係

$$M \subseteq N \iff Count_M(x) \leq Count_N(x), \quad \forall x \in X$$

2. 相等関係

$$M = N \iff Count_M(x) = Count_N(x), \quad \forall x \in X$$

3. 合併集合

$$Count_{M \cup N}(x) = Count_M(x) \vee Count_N(x)$$

4. 共通集合

$$Count_{M \cap N}(x) = Count_M(x) \wedge Count_N(x)$$

5. 和

$$Count_{M \oplus N}(x) = Count_M(x) + Count_N(x)$$

なお, \vee および \wedge はそれぞれ \max と \min を表す. 合併集合と共通集合は, 交換法則, 結合法則, 分配法則を満たすことが知られている [68, 110]. マルチ集合が有限であると仮定すると, これらの演算及び関係は通常の集合におけるそれらと同様に定義されることが分かる.

例: $X = \{a, b, c, d\}$ を全体集合とし, $M = \{3/a, 1/b, 3/c\}$ と $N = \{2/a, 3/b, 2/d\}$ が X のマルチ集合であるとする. その合併集合, 共通集合, 和演算は次のように計算される.

$$\begin{aligned} M \cup N &= \{3/a, 3/b, 3/c, 2/d\} \\ &= \{a, a, a, b, b, b, c, c, c, d, d\} \\ M \cap N &= \{2/a, 1/b\} \\ &= \{a, a, b\} \\ M \oplus N &= \{5/a, 4/b, 3/c, 2/d\} \\ &= \{a, a, a, a, a, b, b, b, b, c, c, c, d, d\} \end{aligned}$$

5.2.2 ファジィマルチ集合の性質と基本演算

ファジィマルチ集合とカウント関数

ファジィマルチ集合は各要素の帰属度をマルチ集合に導入したものである. Yager [110] の定義によれば, 通常の有限集合 X 上のファジィマルチ集合 A は, 単位区間 $I = [0, 1]$ のマルチ集合に値をとる *Count* 関数で特徴付けられている. すなわち,

$$Count_A(x) = \{\nu, \nu', \dots, \nu''\}, \quad \nu, \nu', \dots, \nu'' \in I. \quad (5.2)$$

このように定義された X 上のファジィマルチ集合の全体を $\mathcal{FM}(X)$ と書くことにする.

ファジィマルチ集合の基本演算の定義にはメンバーシップ列, すなわち $Count_A(x)$ を降順にソートした列を用いる必要がある [72]. そこで, 上記の数列を降順にソートすることができると仮定する.

このようにして定義されたメンバーシップ列を

$$Count_A(x) = \{\mu_A^1(x), \mu_A^2(x), \dots, \mu_A^m(x)\}$$

と表す. $\mu_A^1(x) \geq \mu_A^2(x) \geq \dots \geq \mu_A^m(x)$ であることに注意する. X を全体集合すると上記のファジィマルチ集合は次のように表すことができる.

$$A = \{(\mu_A^1(x), \mu_A^2(x), \dots, \mu_A^m(x)) / x : x \in X\}$$

ここで, $\{(\mu_A^1(x), \mu_A^2(x), \dots, \mu_A^m(x))/x\}$ は要素 x がファジィマルチ集合 A に m 個存在し, 帰属度はそれぞれ $\mu_A^1(x), \mu_A^2(x), \dots, \mu_A^m(x)$ であることを表している. 一般に $\mu_A^j(x)/x$ ($j = 1, 2, \dots, m$) において, $\mu_A^j(x) = 0$ ならば要素が存在しないことを表している. 従って, あるファジィマルチ集合には $\mu_A^j(x) = 0$ となる要素 $\mu_A^j(x)/x$ を任意個追加しても, 省略してもファジィマルチ集合は変わらないという性質を持っている.

ファジィマルチ集合の基本演算

2つのファジィマルチ集合 A, B の基本演算とそれらの関係は, メンバーシップ列の演算と関係により次のように定義されている [65, 72].

ファジィマルチ集合の基本演算

1. 包含関係

$$A \subseteq B \iff \mu_A^j(x) \leq \mu_B^j(x), \quad j = 1, \dots, m, \quad \forall x \in X$$

2. 相等関係

$$A = B \iff \mu_A^j(x) = \mu_B^j(x), \quad j = 1, \dots, m, \quad \forall x \in X$$

3. 合併集合

$$\mu_{A \cup B}^j(x) = \mu_A^j(x) \vee \mu_B^j(x), \quad j = 1, \dots, m, \quad \forall x \in X$$

4. 共通集合

$$\mu_{A \cap B}^j(x) = \mu_A^j(x) \wedge \mu_B^j(x), \quad j = 1, \dots, m, \quad \forall x \in X$$

5. 和

$$Count_A(x) = \{\nu_1, \nu_2, \dots\}, \quad Count_B(x) = \{\nu'_1, \nu'_2, \dots\}$$

のとき,

$$Count_{A \oplus B}(x) = \{\nu_1, \nu'_1, \nu_2, \nu'_2, \dots\}$$

6. 積

$$Count_A(x) = \{\nu_1, \nu_2, \dots, \nu_\ell\}, \quad Count_B(x) = \{\nu'_1, \nu'_2, \dots, \nu'_\ell\}, \quad \ell \leq \ell'$$

のとき,

$$Count_{A \cdot B}(x) = \{\nu_1 \nu'_1, \nu_2 \nu'_2, \dots, \nu_\ell \nu'_\ell, 0, \dots\}$$

ファジィマルチ集合の演算では、要素の帰属度を並び替えて順序付けをしている点が重要である。このようなファジィマルチ集合はメンバーシップ列で特徴付けられるが、ベクトル値のメンバーシップを持っているわけではない。すなわち、メンバーシップ列の長さが異なるような場合に、帰属度 $\mu_A^j(x) = 0$ の要素を任意個付加して、メンバーシップ列の長さを同じにすることができる。一般に、有限個のファジィマルチ集合を取り扱うのであれば、同様の操作により各ファジィマルチ集合について x におけるメンバーシップ列を構成し、それらの列の長さをすべての $x \in X$ とすべてのファジィマルチ集合について一定に取ることができる。なぜならば、対象とする有限個のファジィマルチ集合と有限集合内の要素についてそれぞれの長さを計算し、それらの最大値を取ればよいからである。

例： $X = \{a, b, c, d\}$ を全体集合とし、

$$A = \{(0.7, 0.4, 0.1)/a, (0.5, 0.3)/b, (0.9, 0.2)/c\}$$

$$B = \{(0.6, 0.5)/a, (0.7)/b, (0.6)/c\}$$

が X のファジィマルチ集合であるとする、その合併集合、共通集合、和演算、積演算は次のように計算される。

$$A \cup B = \{(0.7, 0.5, 0.1)/a, (0.7, 0.3)/b, (0.9, 0.2)/c\}$$

$$A \cap B = \{(0.6, 0.4)/a, (0.5)/b, (0.6)/c\}$$

$$A \oplus B = \{(0.7, 0.6, 0.5, 0.4, 0.1)/a, (0.7, 0.5, 0.3)/b, (0.9, 0.6, 0.2)/c\}$$

$$A \times B = \{(0.42, 0.2)/a, (0.35)/b, (0.54)/c\}$$

5.3 ファジィマルチ集合のノルムと類似度

5.3.1 ファジィマルチ集合のノルム

ファジィマルチ集合 A の長さを

$$L(x; A) = \max\{j : \mu_A^j(x) > 0\}$$

と定義する。明らかに、次の式が成り立つ。

$$L(x; A \oplus B) = L(x; A) + L(x; B)$$

$$L(x; A \cup B) = L(x; A) \vee L(x; B)$$

$$L(x; A \cap B) = L(x; A) \wedge L(x; B)$$

ここで、ファジィマルチ集合 A のノルムを次式で定義する

$$\|A\| = \sqrt{\sum_{x \in X} \sum_{j=1}^{L(x;A)} \{\mu_A^j(x)\}^2} \quad (5.3)$$

5.3.2 ファジィマルチ集合の類似度

クラスタリングのために、ファジィマルチ集合間の類似度として次のような式を定義する。ファジィマルチ集合 A, B に対して

$$s(A, B) = \frac{\sum_{x \in X} \sum_{j=1}^{L(x;A) \wedge L(x;B)} \mu_A^j(x) \mu_B^j(x)}{\|A\| \|B\|} \quad (5.4)$$

と定義する。 $s(A, B)$ は次の性質を満たす。

- (i) $s(A, B) \geq 0$
- (ii) $s(A, B) = s(B, A)$

なお、(i) で $s(A, B) = 1$ が満たされるのは、ある定数 C に対して

$$\mu_A^j(x) = C \mu_B^j(x), \quad j = 1, \dots, m, \quad \forall x \in X$$

のときかつそのときに限る。

この類似度は情報検索におけるベクトル空間モデルのコサイン相関を、ファジィマルチ集合間の類似度として一般化したものである。コサイン相関あるいはコサイン係数と呼ばれている類似度は、文献の類似検索やクラスタリングにおいてよく利用されている [59, 60, 94, 96, 97]。

5.4 ファジィマルチ集合による文書情報のモデル化

情報検索では、用語の関連性解析や用語あるいは文献クラスタリングのため、用語-文献行列 (Term-Document Matrix) [96, 97] が用いられる。これは次のように表される。用語の集合を $T = \{t_1, \dots, t_p\}$ とし、文献の集合を $D = \{d_1, \dots, d_q\}$ とする。用語-文献行列 $F = (f_{ij})$ は、 $p \times q$ 行列で、用語 t_i ($i = 1, \dots, p$) が文献 d_j ($j = 1, \dots, q$) に f_{ij} 回出現することを表す。

いま、用語 t_i が文献 d_j に異なる度合いをもって複数回現れるとし、そのメンバーシップが ν_1, \dots, ν_s であるとする。このとき、整数 f_{ij} の代わりにファジィマルチ集合

$$f_{ij} = \{\nu_1, \dots, \nu_s\}$$

を用いる。

クラスタリングに関していえば、用語のクラスタリングと文献のクラスタリングのいずれもが可能である。用語をクラスタリングする場合、 t_k に対応するファジィマルチ集合を A_k とし、 $X = D$ と置くと、

$$\text{Count}_{A_i}(d_j) = f_{ij} = \{\nu_1, \dots, \nu_s\}$$

であり、また、文献をクラスタリングする場合、 d_k に対応するファジィマルチ集合を A_k とし、 $X = T$ と置くと、

$$\text{Count}_{A_j}(t_i) = f_{ij} = \{\nu_1, \dots, \nu_s\}$$

であることに注意すれば、文書情報をファジィマルチ集合を用いてモデル化することができる。

数値例

ここでは、ごく簡単な数値例について示す。用語の集合を $T = \{t_1, t_2, t_3\}$ 、文献の集合を $D = \{d_1, d_2, d_3, d_4, d_5\}$ とし、これらの文献に対応するファジィマルチ集合を A_k ($k = 1, \dots, 5$) とする。

このとき、用語帰属度が、

1. 用語が文献のタイトルにあれば、帰属度 1.0 とし、なければ帰属度 0 とする
2. 用語が文献のキーワードにあれば、帰属度 0.5 とし、なければ帰属度 0 とする
3. 用語が文献のアブストラクトにあれば、帰属度 0.2 とし、なければ帰属度 0 とする

というルールに従って与えられ、文献に対応するファジィマルチ集合が以下のように与えられたとする。

$$A_1 = \{(1.0, 0.5)/t_1, (0.5, 0.2)/t_2, (0)/t_3\}$$

$$A_2 = \{(0.5)/t_1, (1.0, 0.5, 0.2)/t_2, (0.2)/t_3\}$$

$$A_3 = \{(1.0, 0.2)/t_1, (1.0, 0.2)/t_2, (0.5)/t_3\}$$

$$A_4 = \{(0.2)/t_1, (0.5)/t_2, (1.0, 0.5, 0.2)/t_3\}$$

$$A_5 = \{(0)/t_1, (0.5, 0.2)/t_2, (1.0)/t_3\}$$

このとき用語-文献行列は，

$$\begin{array}{c}
 \\
 \\
 \\
 \\
 \\
 \end{array}
 \begin{array}{c}
 t_1 \\
 t_2 \\
 t_3
 \end{array}
 \begin{array}{c}
 (1.0, 0.5) \\
 (0.5) \\
 (1.0, 0.2) \\
 (0.2) \\
 (0)
 \end{array}
 \begin{array}{c}
 (0.5, 0.2) \\
 (1.0, 0.5, 0.2) \\
 (1.0, 0.2) \\
 (0.5) \\
 (0.5, 0.2)
 \end{array}
 \begin{array}{c}
 (0) \\
 (0.2) \\
 (0.5) \\
 (1.0, 0.5, 0.2) \\
 (1.0)
 \end{array}$$

となる．この時，文献 d_1 と d_2 の類似度，すなわちファジィマルチ集合 A_1 と A_2 の類似度は

$$\begin{aligned}
 s(A_1, A_2) &= \frac{1.0 \times 0.5 + 0.5 \times 1.0 + 0.2 \times 0.5}{\|1.54\| \|1.58\|} \\
 &= 0.82
 \end{aligned}$$

と計算される．

これらの文献をファジィ c -平均法のアルゴリズム（ただし $\lambda = 2.0$ ）を用い，クラスター数 $c=2$ で分類すると，クラスタリング結果は以下ようになる．

$$(u_{1A_1}, u_{2A_1}) = (0.74, 0.26)$$

$$(u_{1A_2}, u_{2A_2}) = (0.68, 0.32)$$

$$(u_{1A_3}, u_{2A_3}) = (0.85, 0.15)$$

$$(u_{1A_4}, u_{2A_4}) = (0.20, 0.80)$$

$$(u_{1A_5}, u_{2A_5}) = (0.22, 0.78)$$

$$v_1 = \{(0.1)/t_1, (0.5, 0.1)/t_2, (0.8, 0.2, 0.1)/t_3\}$$

$$v_2 = \{(0.7, 0.2)/t_1, (0.7, 0.2)/t_2, (0.2)/t_3\}$$

ここで， u_{iA_1} から u_{iA_5} ($i = 1, 2$) はそれぞれのクラスに属する帰属度を表しており， v_1 ， v_2 はそれぞれクラスター中心を表している．

また，図 5.1 は分類結果を示しており，グラフの横軸は文献番号，縦軸はそれぞれのクラスに属する帰属度を表している．すなわち，文献 d_1 についてみると，この文献はクラス 1 に 0.74，クラス 2 に 0.26 の割合で属していることになる．

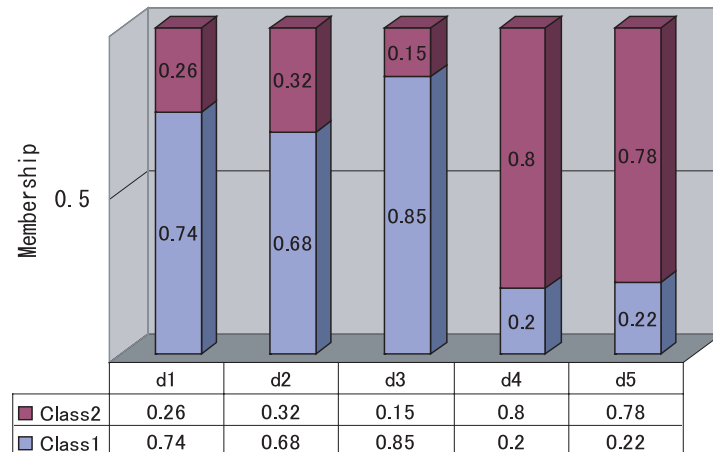


図 5.1: 分類結果

5.5 文書情報クラスタリング

5.5.1 数値例 1

データの説明

ここでは、30 件の文献に対してファジィマルチ集合のモデルを適用し、文書に関するクラスタリングを行った例について示す。なお、今回選択した文献、用語、用語-帰属度の設定ルールについて以下に示す。

1. 文献：クラスタリングアルゴリズムをテーマとする文献（21 件）とクラスタリングには関連しないアルゴリズムをテーマとする文献（9 件）の 2 種類を選択（付録：A.1 参照）
2. 用語：
 - t_1 ：クラスタリング
 - t_2 ：ファジィ
 - t_3 ：アルゴリズム
3. 用語-帰属度を定めるルール
 - (a) 用語が文献のタイトルにあれば、帰属度 1.0 とし、なければ帰属度 0 とする。
 - (b) 用語が文献のキーワードにあれば、帰属度 0.5 とし、なければ帰属度 0 とする。
 - (c) 用語が文献のabstractにあれば、帰属度 0.2 とし、なければ帰属度 0 とする。

これにより得られたファジィマルチ集合を提案した手法により 2 つのクラスに分類する。

分類結果

図 5.2 はエントロピー正則化ファジィ c -平均法 (eFCM, $\lambda = 15$) による分類結果, 図 5.3 はカーネル関数を利用したエントロピー正則化ファジィ c -平均法 (K-FCM, $\lambda = 10, cnst = 1$) による分類結果, 図 5.4 はクラスターサイズ調整変数付エントロピー正則化ファジィ c -平均法 (FCMA, $\lambda = 15$) による分類結果, 図 5.5 はカーネル関数を利用したクラスターサイズ調整変数付エントロピー正則化ファジィ c -平均法 (K-FCMA, $\lambda = 10, cnst = 1$) による分類結果, 図 5.6 はファジィ競合学習クラスタリング (FCCL, $\lambda = 10$) による分類結果, 図 5.7 はカーネル・ファジィ競合学習クラスタリング (K-FCCL, $\lambda = 10, cnst = 1.6$) による分類結果である。

図の横軸はファジィマルチ集合のノルム, 縦軸は文献がクラスタリングをテーマとする文献のクラスに属する帰属度を示している。帰属度の高さ 0.5 のラインを基準に, ラインの上はクラスタリングをテーマとする文献のクラスで, ラインの下はクラスタリングとは関連しないアルゴリズムをテーマとする文献のクラスである。

分類に用いた文献数は 30 件であり, このうち, クラスタリングをテーマとする文献 (21 件) は図中の青の ‘ \circ ’ で表示し, その他のアルゴリズムをテーマとする文献 (9 件) は図中の赤の ‘ $+$ ’ で表示している。

eFCM と K-FCM の両者の分類結果を比較してみると, eFCM による分類結果 (図 5.2) では, クラスタリングに関係がある文献が 4 件, クラスタリング以外のアルゴリズムに関係がある文献 3 件が誤分類されていることが分かる。これは, ファジィマルチ集合を構成する際の帰属度の設定において, 文献のタイトルにその用語が含まれれば高い帰属度を与えるというルールによるものであると考えられる。一方, K-FCM による分類結果 (図 5.3) では, クラスタリングに関係がある文献 2 件が誤分類される結果となった。カーネル関数を利用することで, 誤分類された文献数は 7 件から 2 件に減少した。しかし, この誤分類された 2 件の文献に注目してみると, カーネル関数を利用することで若干ではあるがさらに帰属度が低くなっていることがわかる。これは, 得られたファジィマルチ集合の性質に大きく関係しており, これらはクラスタリングに関係がある文献であるが, タイトル, あるいはキーワードにクラスタリングという言葉が含まれておらず, 逆にキーワードやアブストラクトにアルゴリズムという言葉が含まれていたからであると考えられる。しかし, その他の文献に関しては, クラスタリングに関する文献は帰属度がより高く, 逆にアルゴリズムに関する文献は帰属度がより低くなった。

FCMA と K-FCMA の両者の分類結果を比較してみると, FCMA による分類結果 (図 5.4) では, クラスタリングをテーマとする文献 2 件, その他のアルゴリズムをテーマとする文献 3 件が誤分類されていることが分かる。理由については先程と同様のことが考えられる。一方, K-FCMA による分類結果 (図 5.5) では, クラスタリングをテーマとする文献が 1 件, その他のアルゴリズムをテーマとする文献 3 件が誤分類となる結果となり, 誤分類数について

は 1 件しか減らすことができなかった。

さらに、FCCL と K-FCCL の両者の分類結果を比較してみると、FCCL による分類結果 (図 5.6) では、クラスタリングをテーマとする文献 5 件、その他のアルゴリズムをテーマとする文献 1 件が誤分類されているのに対して、K-FCCL による分類結果 (図 5.7) では、クラスタリングをテーマとする文献の 2 件のみが誤分類となる結果となり、競合学習モデルにおいても、カーネル化したアルゴリズムによって誤分類数を大幅に減らすことができた。

分類結果を全面的に考察すると、ファジィマルチ集合のモデルを用いたクラスタリングにより、文書情報を有効に分類できることが分かった。また、カーネル関数を利用した非線形クラスタリングアルゴリズムにより、文献の誤分類を減らすことができる、すなわち分類性能を向上させることができるということも確認できた。

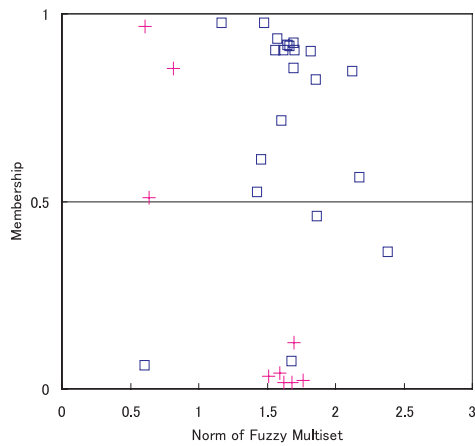


図 5.2: アルゴリズム eFCM による分類結果

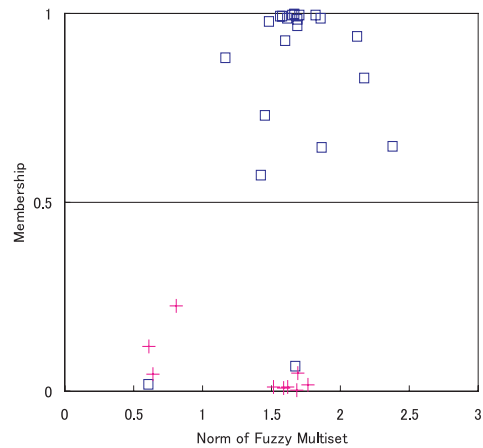


図 5.3: アルゴリズム K-FCM による分類結果

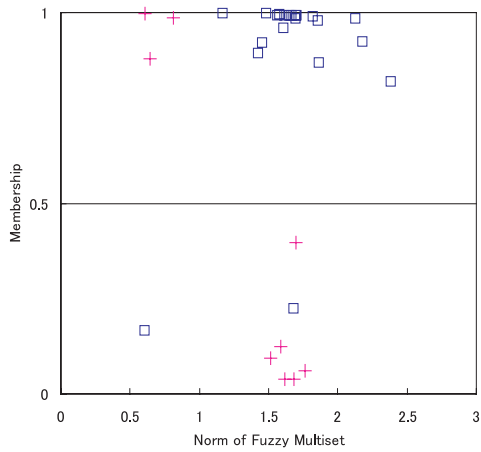


図 5.4: アルゴリズム FCMA による分類結果

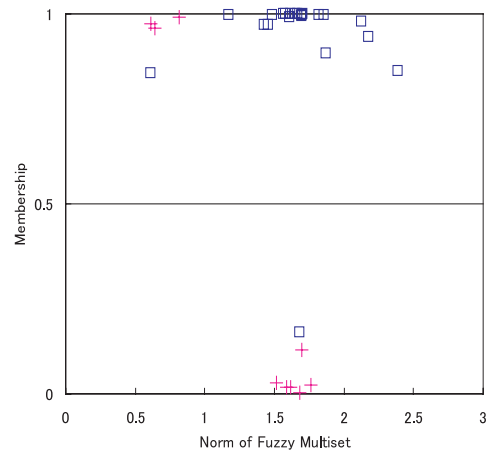


図 5.5: アルゴリズム K-FCMA による分類結果

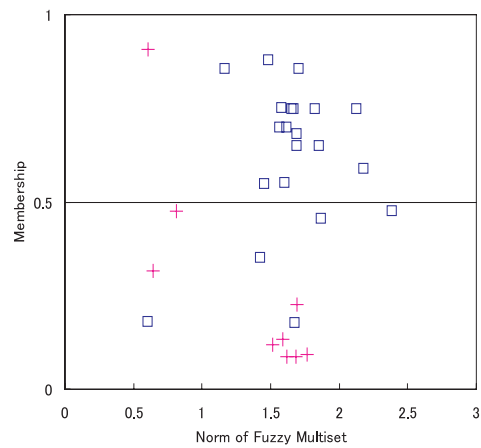


図 5.6: アルゴリズム FCCL による分類結果

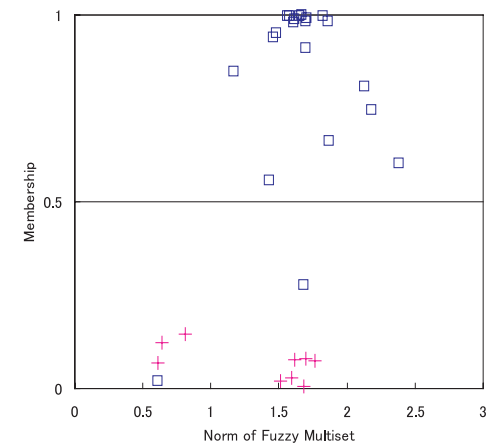


図 5.7: アルゴリズム K-FCCL による分類結果

5.5.2 数値例 2

リスクとは、リスク学事典 [89] によると、「人間の生命や経済活動にとって、望ましくない事象の発生の不確実さの程度およびその結果の大きさの程度」、「危険をもたらす確率と事象によってもたらされた損失や障害の大きさとの積、ないし和集合」とされている。一般的には、危険量を定量的に示したもので、被害規模と起きる発生確率の積で表現され、その事象が顕在化すると社会に好ましくない影響が発生するというものである。

我々は、昔から自然災害や伝染病などの様々なリスクにさらされ、それらを克服してきた。しかしながら、豊かさや快適性を追求している現代社会においては、巨大技術に伴う事故や大都市災害、大気汚染などの様々な人工リスクが発生している。さらに、近年の情報化の進展を背景にサイバーリスクなどの問題が社会的に注目を集めている。このように、新規のリスクが顕在化する背景として、潜在的に存在していたリスクに対して、早い段階でリスクの回避・低減のための適切な対応が取られなかったことが考えられる。新規のリスクを管理するためには早い段階で発見し、早期警告を発することが必要であることが最近では認められるようになってきており、様々なリスク解析のための研究が行われるようになってきた [4, 54]。

このような、リスクへの早期発見を支援するための一つの手段として、リスクに関する情報を分類するということが考えられる。存在する様々なリスク情報を分類することは、それらの情報が持つ特徴を抽出するだけでなく、その情報が持つ他の情報との関連性を見出すことも可能であり、リスク解析を容易にするための有効手段と言える。これはソフトコンピューティングと呼ばれている技術の一つであり、存在する様々なリスク情報をグループ化することで、それらのデータあるいは情報が持つ特徴を抽出することを目的として行うものである。抽出した情報を用いてリスク解析を行うことは、個々に存在する情報を直接用いて解析を行うことに比べて、より容易にかつより詳細な分析を可能にすることができるであろう。このことから、クラスタリングはリスク解析のための方法論として基礎的な研究であると考えることができる。

序論でも述べたように、近年のコンピュータの高性能化と情報技術の目覚ましい発展に伴い、HTML 文書に代表されるような多くの情報が巷に溢れている。これらの情報を効率よく取り扱うためには、クラスタリングなどのデータマイニング技術を用いて、あらかじめ情報を分類しておくことが考えられる。これはリスク解析においても同様のことが言える。つまり、存在する様々なリスク情報をあらかじめ分類しておくことによって、取り扱うリスク情報へのアクセスを容易にすることができるのである。

ここでは、存在する様々なリスク情報（リスクに関する文書情報）に対してクラスタリングを行うことにより、グループ化された情報の中からどのような特徴が抽出できるのか、また、実データに対する提案手法の有効性について、考察を行っていく。

データの説明

ここでは、リスクに関する情報を取り扱うことを目的としているので、2002年にヨーロッパで行われたリスクに関する国際会議で発表された文献 127 件 [102] (付録: A.2 参照) を用いて文書情報クラスタリングを行う。

選択した文献、用語、用語-帰属度の設定ルールについて以下に示す。

1. 文献: 「Society for Risk Analysis Annual Meeting 2002」で発表された論文 127 件

2. 用語:

t_1 : Risk

t_2 : Perception

t_3 : Management

t_4 : Communication

t_5 : Assessment

t_6 : Uncertainty

3. 用語-帰属度を定めるルール

(a) 用語が文献のタイトルにあれば、帰属度 1.0 とし、なければ帰属度 0 とする。

(b) 用語が文献のキーワードにあれば、帰属度 0.6 とし、なければ帰属度 0 とする。

(c) 用語が文献のアブストラクトにあれば、帰属度 0.2 とし、なければ帰属度 0 とする。

なお、選択した用語については、発表された論文に使われていたキーワードについて、使用頻度の高い順に 6 つ選択した。

分類結果

クラスター数の決定

まず始めに、いくつかのクラスターに分類することが適当なのかを分析する。用語として選択したキーワードは全部で 6 つあるので、最大で 6 つのクラスターに分類することとし、クリस्प c -平均法により、 $c = 2$ から $c = 6$ で分類を行った。そのクラスタリング結果を表 5.1 に示す。

ここには、クラスタリングによりそれぞれのクラスに分類された文献数が示してある。すなわち、 $c = 2$ で分類したときには、Class 1 には 47 件の文献が、Class 2 には 80 件の文献が分類されたことを示している。

次に、クラスター数をいくつに設定するのが適当なのかを分析する。クラスター数を $c = 2$ から $c = 6$ までのそれぞれの分類結果において、各クラスに含まれている用語の種類を見て

表 5.1: クラスタリング結果

	$c = 2$	$c = 3$	$c = 4$	$c = 5$	$c = 6$
Class 1	47	46	38	37	35
Class 2	80	38	33	28	26
Class 3	—	43	47	21	19
Class 4	—	—	9	9	9
Class 5	—	—	—	32	28
Class 6	—	—	—	—	10
合計	127	127	127	127	127

みる .

表 5.2 から表 5.6 は、それぞれのクラスタリング結果において、選択した用語がどのくらいの割合で各クラスに含まれているのかを示したものである。表中において R, P, M, C, A, U はそれぞれ選択した用語の頭文字を表している。文献数というのは、用語として選択した 6 つの言葉が、分類された文献中のタイトルもしくはキーワードに含まれている場合の文献の数を表しており、割合というのは、選択した用語が含まれている文献の数がどれくらいの割合で分類された文献の中に含まれているかを示したものである。図 5.8 から図 5.12 は表中の割合を各クラスごとに表したものである。また、これらの結果から、まわりの用語に比べて高い割合で存在している用語をそのクラスの特徴を表す用語として抽出した。

表 5.2: クラスタリング結果 ($c = 2$)

		R	P	M	C	A	U	特徴を表す用語
Class 1 (47 件)	文献数	37	0	37	1	1	3	Risk, Management
	割合	79 %	0 %	79 %	2 %	2 %	6 %	
Class 2 (80 件)	文献数	70	30	2	14	17	4	Risk, Communication, Perception, Assessment
	割合	88 %	38 %	3 %	18 %	21 %	5 %	

R:Risk, P:Perception, M:Management, C:Communication, A:Assessment, U:Uncertainty

表 5.3: クラスタリング結果 ($c = 3$)

		R	P	M	C	A	U	特徴を表す用語
Class 1 (46 件)	文献数	38	0	35	1	0	3	Risk, Management
	割合	83 %	0 %	74 %	2 %	0 %	6 %	
Class 2 (38 件)	文献数	32	28	1	2	0	1	Risk, Perception
	割合	84 %	35 %	1 %	3 %	0 %	1 %	
Class 3 (43 件)	文献数	35	2	3	12	18	2	Risk, Communication, Assessment
	割合	81 %	5 %	7 %	28 %	42 %	5 %	

R:Risk, P:Perception, M:Management, C:Communication, A:Assessment, U:Uncertainty

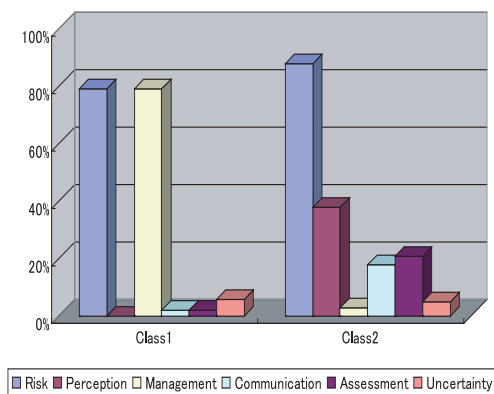


図 5.8: クラスタリング結果 ($c = 2$)

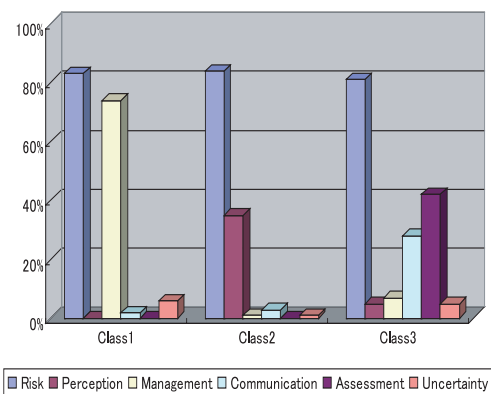


図 5.9: クラスタリング結果 ($c = 3$)

表 5.4: クラスタリング結果 ($c = 4$)

		R	P	M	C	A	U	特徴を表す用語
Class 1 (38 件)	文献数	32	0	35	1	0	3	Risk, Management
	割合	89 %	0 %	97 %	3 %	0 %	8 %	
Class 2 (33 件)	文献数	28	28	1	2	0	1	Risk, Perception
	割合	85 %	85 %	3 %	6 %	0 %	3 %	
Class 3 (47 件)	文献数	38	2	1	4	18	3	Risk, Assessment
	割合	81 %	4 %	2 %	9 %	38 %	6 %	
Class 4 (9 件)	文献数	8	0	2	8	0	0	Risk, Communication
	割合	89 %	0 %	22 %	89 %	0 %	0 %	

R:Risk, P:Perception, M:Management, C:Communication, A:Assessment, U:Uncertainty

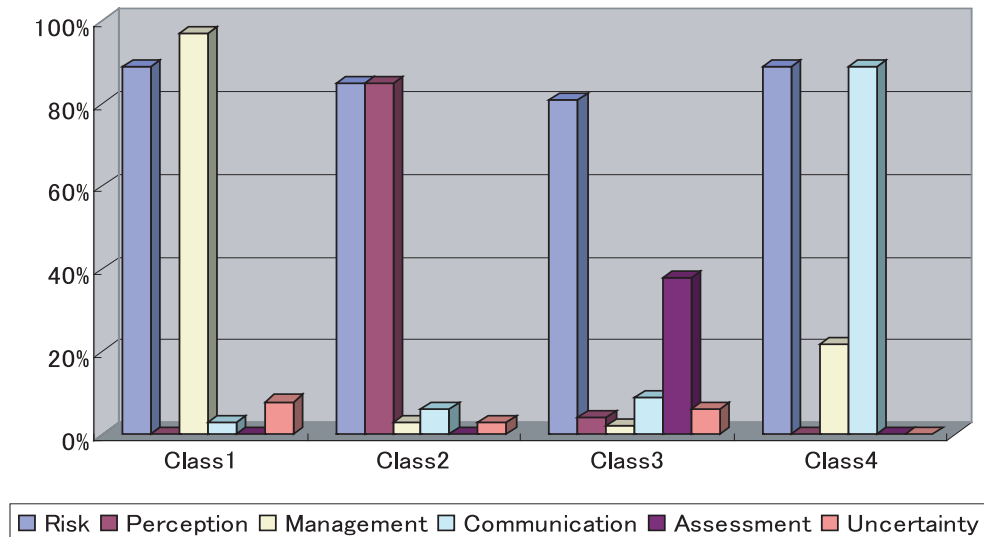


図 5.10: クラスタリング結果 ($c = 4$)

表 5.5: クラスタリング結果 ($c = 5$)

		R	P	M	C	A	U	特徴を表す用語
Class 1 (37 件)	文献数	29	0	32	1	0	3	Risk, Management
	割合	78 %	0 %	86 %	3 %	0 %	8 %	
Class 2 (28 件)	文献数	25	26	1	0	0	0	Risk, Perception
	割合	89 %	93 %	4 %	0 %	0 %	0 %	
Class 3 (21 件)	文献数	16	2	1	2	18	2	Risk, Assessment
	割合	76 %	10 %	5 %	10 %	86 %	10 %	
Class 4 (9 件)	文献数	8	0	2	8	0	0	Risk, Communication
	割合	89 %	0 %	22 %	89 %	0 %	0 %	
Class 5 (32 件)	文献数	28	2	3	4	0	1	Risk
	割合	88 %	6 %	9 %	13 %	0 %	3 %	

R:Risk, P:Perception, M:Management, C:Communication, A:Assessment, U:Uncertainty

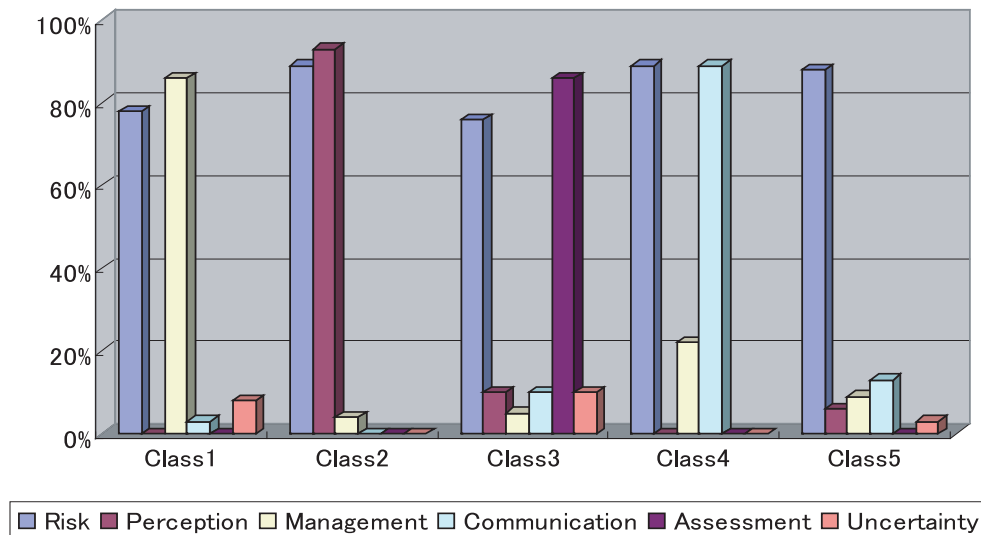


図 5.11: クラスタリング結果 ($c = 5$)

表 5.6: クラスタリング結果 ($c = 6$)

		R	P	M	C	A	U	特徴を表す用語
Class 1 (35 件)	文献数	28	0	31	1	0	3	Risk, Management
	割合	80 %	0 %	89 %	3 %	0 %	9 %	
Class 2 (26 件)	文献数	25	26	1	0	0	0	Risk, Perception
	割合	96 %	100 %	4 %	0 %	0 %	0 %	
Class 3 (19 件)	文献数	15	2	0	2	17	2	Risk, Assessment
	割合	79 %	11 %	0 %	11 %	89 %	11 %	
Class 4 (9 件)	文献数	8	0	2	8	0	0	Risk, Communication
	割合	89 %	0 %	22 %	89 %	0 %	0 %	
Class 5 (28 件)	文献数	26	0	5	0	1	1	Risk
	割合	93 %	0 %	18 %	0 %	4 %	4 %	
Class 6 (10 件)	文献数	4	2	0	4	0	1	Risk, Communication
	割合	40 %	20 %	0 %	40 %	0 %	10 %	

R:Risk, P:Perception, M:Management, C:Communication, A:Assessment, U:Uncertainty

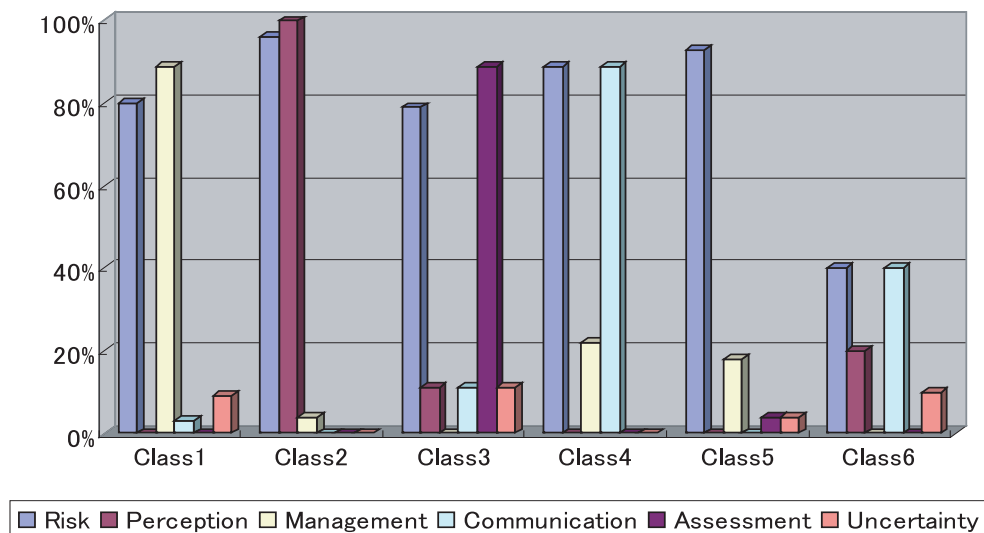


図 5.12: クラスタリング結果 ($c = 6$)

すなわち、表 5.2 及び図 5.8 について見てみると、「Class 1」に分類された文献は全部で 47 件あり、そのうち用語「Risk」が含まれている文献は 37 件で全体の 79%、用語「Perception」が含まれている文献は無し、用語「Management」が含まれている文献は 37 件で全体の 79%、用語「Communication」が含まれている文献は 1 件で全体の 2%、用語「Assessment」が含まれている文献は 1 件で全体の 2%、用語「Uncertainty」が含まれている文献は 3 件で全体の 6%であることを示している。また、これらの結果から「Class 1」の特徴を表す用語として、「Risk」、「Management」の 2 つを抽出したことが分かる。

これらの結果から適切なクラスター数の決定について考察してみる。まず用語「Risk」について見てみると、クラスター数を変えてもどのクラスにも高い割合で存在していることが分かり、クラスター数の決定には依存していないとみなすことができる。また用語「Uncertainty」についても同様のことが考えられ、クラスター数を変えてもどのクラスにも低い割合で存在していることが分かり、この用語についてもクラスター数の決定には依存していないとみなすことができる。つまり、選択した 6 つの用語のうち「Risk」と「Uncertainty」を除く 4 つの用語のクラスに分類することが妥当ではないか、ということが考えられる。

実際に 4 つのクラスに分類した結果 (表 5.4 及び図 5.10) について見てみると、各クラスの特徴を表す用語がそれぞれ、

- Class 1 — Risk, Management
- Class 2 — Risk, Perception
- Class 3 — Risk, Assessment
- Class 4 — Risk, Communication

であることが分かる。すなわち、今回分類に用いた文献については上記の 4 つのクラスに分類することが妥当であるということが分かる。

各手法による分類結果の考察

ここでは、本研究で提案した各手法の有効性についての考察を行う。考察を行う手順としては以下のとおりである。

1. クリスプ c -平均法と競合学習クラスタリングにより分類を行い、どちらのモデルが適当であるかを検証する。
2. 1の結果を用いて、ファジィの導入について考察を行う。
3. 1, 2の結果を用いて、カーネル関数の利用について考察を行う。

これまでの分析でクラスター数を4つに分類することが適当であることが分かったので、実際にそれぞれの文献がどのクラスに属するのかを調べた。その結果を表 5.7 及び図 5.13 に示す。

表 5.7: 各クラスに属する文献の数

クラスター	文献数
Class1 : Risk Management	50
Class2 : Risk Perception	36
Class3 : Risk Assessment	25
Class4 : Risk Communication	16
合計	127

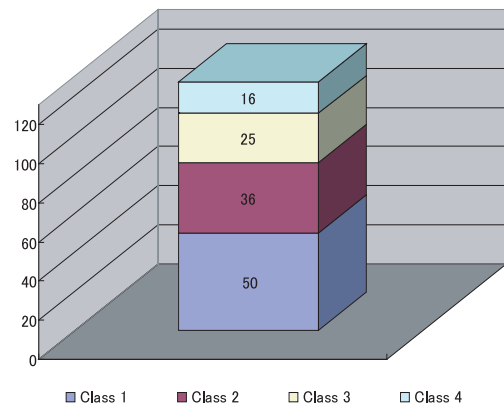


図 5.13: 各クラスに属する文献の数

すなわち、分類に用いた 127 件の文献について、Risk Management に関する文献については 50 件、Risk Perception に関する文献については 36 件、Risk Assessment に関する文献については 25 件、Risk Communication に関する文献については 16 件であったことを示している。

まず始めに、アルゴリズム CCM とアルゴリズム CCL の分類性能の比較を行い、どちらのモデルが今回のデータに対して有効であるかを検証する。ここでは、分類性能を比較するために誤分類数での比較を行う。アルゴリズム CCM とアルゴリズム CCL のクラスタリング結果を表 5.8 に示す。ここには、それぞれの手法によって分類した結果の誤分類された文献の数を示してある。誤分類された文献とは、それぞれのクラスに分類されるべき文献が他のクラスに分類されてしまった文献のことを指す。すなわち、アルゴリズム CCM についてみる

と，Risk Management に関する文献 50 件のうち 12 件が他のクラスに分類されてしまい誤分類となっており，同様に，Risk Perception に関する文献 36 件中 5 件，Risk Communication に関する文献 16 件中 7 件，合計すると分類に用いた文献 127 件中 24 件が誤分類とみなされたことを示している．この 2 つの手法による誤分類数を比較すると，アルゴリズム CCL の分類結果の方が 20 件と少ないことが分かる．この結果から，今回の文献データに関しては競合学習モデルの方が適していることが分かる．

次に，ファジィ化の必要性について考察する．先程の分析から競合学習モデルを用いることにし，ここではアルゴリズム CCL とファジィを導入したアルゴリズム FCCL との比較を行う．アルゴリズム CCL とアルゴリズム FCCL のクラスタリング結果を表 5.9 に示す．この 2 つの手法による誤分類数を比較すると，アルゴリズム CCL の分類結果が 20 件だったのに対してアルゴリズム FCCL の分類結果は 16 件であった．この結果から，アルゴリズムをファジィ化したほうが分類性能が良いことが分かる．

最後に，カーネル関数の利用について考察する．これまでの分析からアルゴリズム FCCL を用いることにし，ここではアルゴリズム FCCL とカーネル・ファジィ競合学習クラスタリング（アルゴリズム K-FCCL）との比較を行う．アルゴリズム FCCL とアルゴリズム K-FCCL のクラスタリング結果を表 5.10 に示す．この 2 つの手法による誤分類数を比較すると，アルゴリズム FCCL の分類結果が 16 件だったのに対してアルゴリズム K-FCCL の分類結果では 12 件と最も少ない結果となった．これらの結果から，本研究で提案したカーネル・ファジィ競合学習クラスタリングが最も性能が良いことが分かる．

表 5.8: アルゴリズム CCM と CCL の比較

	CCM	CCL
Risk Management	12	9
Risk Perception	5	2
Risk Assessment	0	2
Risk Communication	7	7
誤分類数の合計	24	20

表 5.9: アルゴリズム CCL と FCCL の比較

	CCL	FCCL
Risk Management	9	12
Risk Perception	2	2
Risk Assessment	2	0
Risk Communication	7	2
誤分類数の合計	20	16

表 5.10: アルゴリズム FCCL と K-FCCL の比較

	FCCL	K-FCCL
Risk Management	12	0
Risk Perception	2	5
Risk Assessment	0	0
Risk Communication	2	7
誤分類数の合計	16	12

また、それぞれの手法におけるクラスタリング結果の詳細について、表 5.11 から表 5.14 及び図 5.14 から図 5.17 に示す。ここには、それぞれの手法により、どのクラスに分類されたのかが示してある。すなわち、アルゴリズム CCM についてみると、Risk Management に関する文献のうち 38 件については正しく Class 1 の Risk Management のクラスに分類されているが、誤分類とされた 12 件の文献については Class 2 の Risk Perception のクラスに 1 件、Class 3 の Risk Assessment のクラスに 11 件の文献が分類されているということを示している。同様に、Risk Perception に関する文献については、誤分類とされた 5 件の文献が Class 3 の Risk Assessment のクラスに分類されており、また、Risk Communication に関する文献については、誤分類とされた 7 件の文献が Class 3 の Risk Assessment のクラスに分類されているということを示している。

表 5.11: アルゴリズム CCM による分類結果

	Class 1	Class 2	Class 3	Class 4	合計
Risk Management	38	1	11	0	50
Risk Perception	0	31	5	0	36
Risk Assessment	0	0	25	0	25
Risk Communication	0	0	7	9	16
合計	38	32	48	9	127

表 5.12: アルゴリズム CCL による分類結果

	Class 1	Class 2	Class 3	Class 4	合計
Risk Management	41	0	9	0	50
Risk Perception	1	34	1	0	36
Risk Assessment	2	0	23	0	25
Risk Communication	4	0	3	9	16
合計	48	34	36	9	127

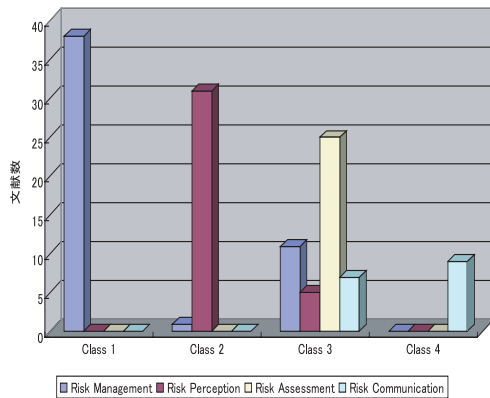


図 5.14: アルゴリズム CCM による分類結果

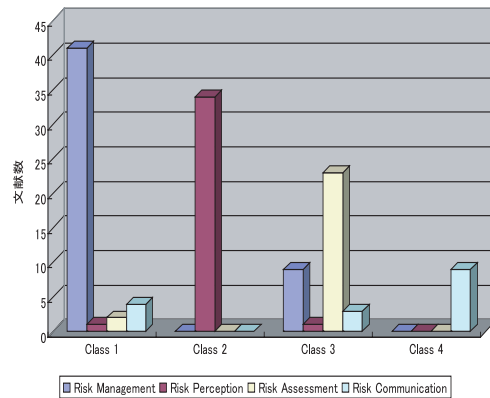


図 5.15: アルゴリズム CCL による分類結果

表 5.13: アルゴリズム FCCL による分類結果 ($\lambda = 10$)

	Class 1	Class 2	Class 3	Class 4	合計
Risk Management	38	1	11	0	50
Risk Perception	0	34	2	0	36
Risk Assessment	0	0	25	0	25
Risk Communication	0	0	2	14	16
合計	38	35	40	14	127

表 5.14: アルゴリズム K-FCCL による分類結果 ($\lambda = 10, cnst = 0.01$)

	Class 1	Class 2	Class 3	Class 4	合計
Risk Management	50	0	0	0	50
Risk Perception	5	31	0	0	36
Risk Assessment	0	0	25	0	25
Risk Communication	7	0	0	9	16
合計	62	31	25	9	127

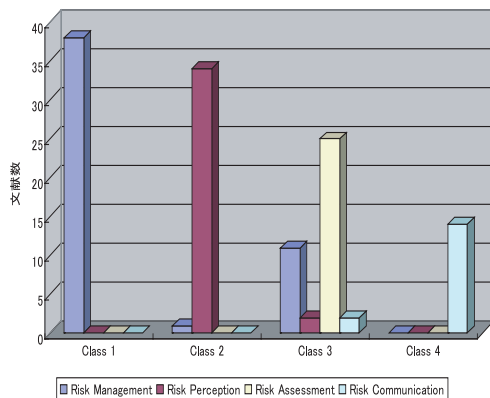


図 5.16: アルゴリズム FCCL による分類結果

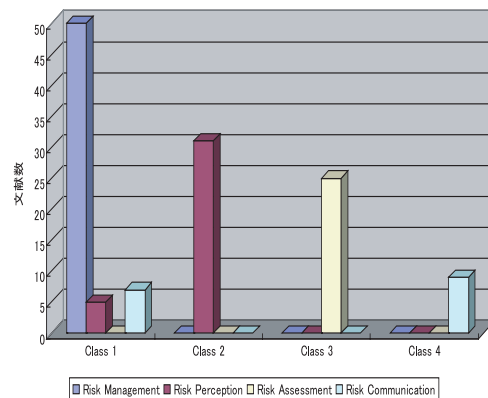


図 5.17: アルゴリズム K-FCCL による分類結果

結果の考察

本実験では、リスクに関連する文献情報を分類し、得られた結果からどのようにして各グループの特徴を抽出していくのかについて考察を行った。その結果、今回の文献情報に関しては、Risk Management, Risk Perception, Risk Assessment, Risk Communication の4つのクラスに分類することが適当であるということが分かった。

また、本研究で提案した手法が有効であるかどうかを検証するために、実際にこれらの文献がどのクラスに所属しているのかを調べ、考察を行った。その結果、今回の文献情報に関しては、非線形化された競合学習モデルによるファジィクラスタリングアルゴリズムが最も優れた分類性能を発揮しており、本研究で提案した手法の有効性を確かめることができた。

5.6 まとめ

本章では、ファジィマルチ集合をモデルとした文書情報クラスタリングを行うために、ファジィマルチ集合間の類似度として一般化したコサイン相関を定義し、ファジィマルチ集合をクラスタリングする方法について述べた。

また、提案した手法の有効性を検証するために、実際の文献情報を用いてそこから得られる用語と文献の関連性からファジィマルチ集合を用いてモデル化し、上記のクラスタリングアルゴリズムにより分類を行った。その結果、数値例 1 において、分類性能を向上させることを目的としたカーネル関数を利用することの有効性を確認することができた。また、数値例 2 においては、リスクに関連する情報を分析することを目的として提案した手法を用いてリスクに関する文献情報のクラスタリングを行い、クラスター数の決定方法、及びそれらのクラスターがどのような特徴をもっているのかを抽出することができた。

さらに分類性能を向上させるためには、用語と文献の関連性から得られるファジィマルチ集合をどのように作成していくか、すなわち、選択する用語数を増やしたり、より適切な帰属度の付与の仕方について検討していくことが考えられる。また、文書情報だけでなく Web 上の情報に対しての本理論の適用性を把握するために、本研究で提案した理論（モデル）を用いた情報検索システムを構築し、実在する Web 情報に対してどのくらいの性能を発揮できるのかということについても検証していく必要がある。

第6章 クラスタリングによる日本周辺の軍事情勢の分析

6.1 はじめに

21世紀に入り、国際社会の流れがさらにスピードアップしグローバル化しつつある中、世界は国際テロや大量破壊兵器あるいは弾道ミサイルの拡散など、予測困難でかつグローバルな脅威にさらされるようになってきた [101]。

2001年9月のアメリカ同時多発テロ以降、国際テロとの闘いにおける各国の努力は着実な効果を挙げ一方、今なおテロの脅威は世界に拡散し、各地でテロ事件が発生している。特に、国際テロ組織が従来の組織の枠にとらわれない活動を実施するようになったため、いつどこでテロが起きるのかを予測することが難しく、各国は対処困難な状況が続いている。

また、大量破壊兵器や弾道ミサイルの移転・拡散は冷戦後の大きな脅威の一つとして認識されている。その背景には、このような大量破壊兵器などを求める国家の多くが自国の国土や国民を危険にさらすことに対する抵抗が少なく、その国土において国際テロ組織の活動が指摘されるなど政府の統治能力が低いこと、大量破壊兵器などが使用される可能性が高いと考えられるからである。さらに、このような国家では、関連技術や物質の管理体制にも不安があり、結果として、化学物質や核物質などが移転・流出する可能性も高くなっているからである。具体的な問題としては、イランや北朝鮮¹の核問題が国際社会にとっての差し迫った課題となっている。

このような諸課題に対して、一国のみで対処することは困難であることから、現在、国際社会においては、多国間の枠組みを活用した各種の協力体制も構築されている。中でも、各国間の協力を推進するための主要な場として機能することが期待されている国連²、特に安保理³の意義・役割は重要であり、実効性と信頼性を高める形で改革することが急務の課題となっている。我が国としても、イラク復興支援などに代表されるような各種の取り組みにより、積極的にこれに取り組んでいくこととしている [11]。

国際社会は、より効果的に諸課題へ対応する方策を模索し、テロ情報を含む各種の情報交換体制の強化、テロリストを厳正に処罰するための国際的な法的枠組み強化、大量破壊兵器

¹朝鮮民主主義人民共和国

²国際連合

³安全保障理事会

不拡散への取組強化，途上国などに対するテロ対処能力向上支援などのための取組を行っていく必要がある．

一方，アジア太平洋地域に目を向けてみると，その地域の特性上，民族・宗教など多様性に富み，各国の安全保障観も様々であることから，依然として領土問題や統一問題といった従来からの問題が残されている．さらに，最近では東南アジア地域におけるテロや海賊行為などの問題が地域の安全保障に深刻な影響を及ぼすようになっている．具体的には，朝鮮半島における軍事的対峙，中国・台湾問題，インドネシアやタイ国内での分離・独立運動，中国・東南アジア各国間での南沙群島の領有権をめぐる対立の問題などが挙げられる．我が国においては，日本固有の領土である北方領土や竹島の領土問題が依然として未解決のまま存在している．また，この地域の多くの国においては，国防費の増額や新装備の導入など軍事力の拡充・近代化を行ってきており，特に，政治的・経済的に地域の大国として重要な影響力をもつ中国については，軍事面においても，各国がその動向に注目する存在となっている [11, 101]．

このような情勢の中，我が国を取り巻く戦略環境と地域情勢を正しく理解することは非常に重要なことであり，そのために常に変転する全世界の軍事情勢の流れを把握することは必要不可欠であると考えられる．

本章では，我が国を取り巻く軍事情勢の動向の変化をクラスタリングにより分析を行う．すなわち，クラスタリングにより世界の国々をグループ化し，そこから得られる情報を利用して我が国に対する軍事的リスクの分析を行っていく．

6.2 クラスタリングによる分析

世界各国の軍事データを用いたクラスタリングによる分析を，次の手順で行う．

1. 冷戦終焉期（1985 年）における世界各国の軍事データを用いてクラスタリングを行い，各グループ（各クラスター）の特徴について分析を行う．また，各地域ごとの特徴⁴，特に東アジア地域の特徴について分析する．
2. 現在の世界の軍事データ（2003 年）を用いてクラスタリングを行い，各グループ（各クラスター）の特徴について分析を行う．ここでも，各地域ごとの特徴⁵，特に東アジア地域の特徴について分析する．
3. 1985 年と 2003 年のクラスタリング結果および各国の軍事データを比較することで，日本周辺における軍事的リスクについての分析を行う．併せて，各国の国防政策などに触れることでクラスタリング結果の妥当性評価も行う．

⁴1985 年においては世界を 8 つの地域に分割している．

⁵2003 年においては世界を 7 つの地域に分割している．

6.2.1 クラスタリングに用いるデータ

クラスタリングに用いるデータとして、各国の国防支出ならびに兵力、各種兵器の数などの 17 次元のデータを用いて分析を行う（付録：B.1 参照）なお、本研究においては各変数に対する重みは考慮しておらず、データをそのまま使用してクラスタリングを行った。これらのデータの出所は英国国際戦略研究所から発行されているミリタリー・バランスである [49, 50]。以下に、今回用いた各変数の説明を記す。

- 国防費
 - － 国防支出費
 - － 国民一人当たりの国防支出費
 - － GDP に占める国防支出の割合
- 兵力
 - － 現役兵力
フルタイム勤務の全ての男女将兵
 - － 予備役
平時には完全充足あるいは即応体制にはないが、緊急時には予備役将兵を再呼集することによって動員し得る編成・部隊
 - － 準軍隊
正規軍部隊に対する支援戦力として、あるいはそれに代わり得る戦力として活用できると思われる組織
- 陸上戦力
 - － 戦車
無積載量が少なくとも 16.5 トン以上で、少なくとも口径 75 ミリ以上の 360 度旋回可能な砲を装備している戦闘用の装甲装軌車両
 - － 軽戦車
無積載量が少なくとも 6 トン以上で、少なくとも口径 75 ミリ以上の直接射撃式の砲を組み込むか積載している装甲戦闘車両で、上記の戦車に含まれないもの。
 - － 装甲歩兵戦闘車
少なくとも口径 20 ミリ以上の砲を組み込むか積載し、歩兵 1 個分隊を輸送するよう設計された装甲戦闘車両
 - － 攻撃ヘリコプター
攻撃を目的とした武装ヘリコプター

- 海上戦力

- 潜水艦
満載排水量 1,000 トン以上で、全長 60 メートル以上の水中で活動することのできる船
- 主要水上戦闘艦艇
満載排水量 1,000 トン以上で、自艦防衛用以外の兵装を搭載した全ての水上艦で、空母、巡洋艦（8,000 トン以上）、駆逐艦（8,000 トン以下）、フリゲート艦（8,000 トン以下）が含まれる。
- 哨戒艦艇・沿岸戦闘艦艇
領海侵入対処として沿岸警備を主任務とする全ての艦艇

- 航空戦力

- 作戦機（空軍）… 空軍が持つ、攻撃能力を保有する航空機
- 作戦機（海軍）… 海軍が持つ、攻撃能力を保有する航空機
- 武装ヘリコプター（空軍）… 空軍が持つ、攻撃能力を保有するヘリコプター
- 武装ヘリコプター（海軍）… 海軍が持つ、攻撃能力を保有するヘリコプター

6.2.2 クラスタリングアルゴリズム

クラスタリングを行うための手法はいくつかあるが、本分析では世界各国の軍事データをクラスタリングするための手法として、本研究で提案したカーネル・競合学習クラスタリング（アルゴリズム K-CCL）を用いる。その理由について以下に述べる。

まず、ユークリッド距離の 2 乗を用いた場合のアルゴリズムについて考察する。本分析で用いるデータは様々な性質を持つ変数が混在しているため、各変数間の単位および値の大きさが大きく異なるという特徴を持っている。特に、国防支出費の値について見てみると、先進国においては他の変数に比べて圧倒的に大きく、クラスタリングを行う際に、値の大きさに左右されるというユークリッド距離の 2 乗を用いた場合の欠点が、クラスタリング結果に現れてしまう可能性がある。つまり、他の変数を用いている意味がなくなってしまうのである。

実際に、全 17 次元のデータを用いた場合の結果と国防支出費のみのデータを用いてクラスタリングした結果は、ほぼ同じ結果になっており、クラスタリング結果が国防支出費の値にのみ依存してしまう。表 6.1 は、2003 年の軍事データをユークリッド距離の 2 乗を用いてクラスタリングした結果の一部を示したものであり、表 6.2 は、2003 年における国防支出費の上位 10ヶ国を示したものである。これらの表を比較すると、クラスタリング結果における Cluster 1 から Cluster 3 に含まれている国については、国防支出費の高い国が上から順番に

分類されている様子が分かる．つまり，国防支出費の値のみに依存してクラスタリングされていることがこれらの結果から明らかである．このことから，今回のような特徴を持つデータを取り扱うときには，内積に基づく類似度を利用した方が全ての変数を考慮した結果を得られることが期待できる．

表 6.1: 2003 年の軍事データをユークリッド距離の 2 乗を用いてクラスタリングした結果の一部（国名については順不同）

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
アメリカ	中国	イタリア	カナダ	チェコ共和国
	ロシア	サウジアラビア	ベルギー	デンマーク
	フランス	インド	ギリシャ	ハンガリー
	日本	韓国	オランダ	ルクセンブルク
	イギリス		ノルウェー	ポルトガル
	ドイツ		ポーランド	アルバニア
			スペイン	アルメニア
			⋮	⋮
			(合計 21ヶ国)	(合計 125ヶ国)

表 6.2: 2003 年における国防支出費の上位 10 カ国

	国名	国防支出費 (単位 100 万ドル)
1	アメリカ	329616
2	中国	48380
3	ロシア	48040
4	フランス	38005
5	日本	37070
6	イギリス	35249
7	ドイツ	31465
8	イタリア	24210
9	サウジアラビア	20981
10	インド	13073

次に、クラスタリングアルゴリズムの非線形化の必要性について考察する。カーネル化されていない従来のアルゴリズムでは初期値依存性が高く、適切な初期値が選択されないとうまくクラスタリングされない場合が生起してしまう。うまくクラスタリングされない場合は、ほとんどの国が一つのクラスターに固まってしまうたり、同じ国が複数のクラスターにまたがって分類されるという結果である。これらの結果については、目的関数の値も正確に求められず、収束回数も平均 2 回程度であることから、明らかに適切な結果ではないことが分かる。ちなみに、クラスタリングの成功回数はアルゴリズム CCM において、100 回の試行中 32 回である。

ここで、成功したクラスタリング結果において、目的関数値の値が最も大きかった分類結果について見てみることにする。表 6.3 は、2003 年の軍事データをアルゴリズム CCM で分類した結果である。この結果から各クラスターの特徴を分析してみると以下のことが分かる。

表 6.3: 2003 年の軍事データをコサイン相関を用いてクラスタリングした結果の一部（国名については順不同）

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
アメリカ	チェコ共和国	エジプト	キプロス	グルジア
カナダ	ギリシャ	イラク	エストニア	北朝鮮
フランス	ハンガリー	ヨルダン	マルタ	ベトナム
ドイツ	ポーランド	シリア	ベリーズ	パラグアイ
中国	トルコ	アフガニスタン	ガイアナ	タンザニア
日本	ロシア	パキスタン	スリナム	
⋮	⋮	⋮	⋮	
(合計 71ヶ国)	(合計 35ヶ国)	(合計 27ヶ国)	(合計 19ヶ国)	

● Cluster 1

国防支出費の割合が他の変数に比べて極端に大きくなっている国が集まっているクラスター。最も多くの国（157カ国中 71カ国）がこのクラスターに分類されている。

● Cluster 2

Cluster 1 に分類された国に次いで国防支出費の割合が他の変数に比べて高いが、併せて、戦車や装甲歩兵戦闘車などの陸上兵器の割合も比較的高い国が集まっているクラスターである。

● Cluster 3

他の変数に比べて、戦車の割合が比較的大きい国が集まっているクラスターである。特に、中東諸国の国が多く分類されている。

● Cluster 4

他の変数に比べて、一人当たりの国防支出費の割合が最も大きい国が集まっているクラスターである。しかし、国防支出費は全般的に低く、また各種兵器についても装備していない国が多く含まれている。

● Cluster 5

国防支出の割合が半分を超えているものの、予備役の割合も比較的高い国が集まっているクラスターである。すなわち、予備役の数もクラスタリング結果に大きな影響を与えている。

ここで、各変数の割合というのは、ある国において、ある変数が他の変数と比べてどの程度の割合をもっているかということである。すなわち、ここでは全 17 次元のデータを正規化して得られた値を見て、割合としている。例えば、表 6.4 は日本における 2003 年の軍事データの一部とそれらの正規化した値が示されているが、国防支出の値が 0.9995 と他の変数に比べて非常に高くなっていることが分かる。このように、クラスター 1 に分類された国については、国防支出の割合が他の変数に比べて非常に大きくなっている様子が分かる。

表 6.4: 日本における 2003 年の軍事データの一部とそれらの正規化した値

	国防支出	現役兵力	戦車	潜水艦	空軍作戦機
元のデータ	37070 (100 万ドル)	239.9 (1,000 人)	1020 (両)	16 (隻)	270 (機)
正規化した値	0.9995	0.0064	0.0276	0.0004	0.0073

以上の考察から、アルゴリズム CCM によるクラスタリングではユークリッド距離の 2 乗を用いた場合とは異なる結果を得てはいるものの、各変数間の比重をもとにクラスタリングがされており、各国の軍事的能力を比較するための情報を得ることが難しいことが分かった。ちなみに、アルゴリズム CCL によりクラスタリングした結果についても、ほぼ同様の結果を得ている。

一方、非線形化された（カーネル化された）クラスタリングアルゴリズムは初期値依存性も少なく、またガウシアンカーネルのパラメータを小さくすることで全ての変数を考慮した

クラスタリング結果を得ることが可能であるので、軍事力の大きさを見る今回の分析においては非常に効果的な手法であると言える。また、計算時間の観点からも、第 4 章において競合学習モデルの方が c -平均クラスタリングよりも望ましい結果を得ているので、本研究においてはカーネル・競合学習クラスタリングを用いて分析を行うことにした。参考までに、以下にアルゴリズム K-CCL を再度示しておく⁶。また、本分析で使用した各種パラメータを表 6.5 に示す。

Algorithm of Kernel-based Competitive Learning Clustering

```

begin initialize  $\eta, s_{ik}(i = 1, \dots, c, k = 1, \dots, n), T$ 

  do select pattern  $x_k$ 
     $i \leftarrow \arg \max_j s_{jk}$  (allocate  $x_k$  to cluster  $G_i$ )

     $s_{ik} \leftarrow \frac{s_{ik} + \eta K_{jk}}{\sqrt{1 + 2\eta s_{ij} + \eta^2 K_{jj}}}$  (update similarity measure)
  until no significant change  $s_{ik}$  in  $T$  attempts

return  $s_{ik}$ 

end
    
```

表 6.5: クラスタリングに使用した各種パラメータ

	1985 年	2003 年
クラスター数 (c)	5	5
国の数 (n)	138	157
学習率係数 初期値 (η)	0.5	0.5
学習率係数 減少幅	5.0×10^{-6}	5.0×10^{-6}
カーネル関数パラメータ ($cnst$)	5.0×10^{-7}	1.0×10^{-6}
最大学習回数 (T)	100,000	100,000

本研究ではクラスター数を $c = 5$ とし、世界の国々をその国の軍事力に応じて 5 つのグループに分類し分析を行う。参考までに、クラスター数を $c = 2$ から $c = 5$ まで変化させて分類を行ったが、1985 年のデータを用いた場合と 2003 年のデータを用いた場合の両方において、特

⁶アルゴリズム K-CCL におけるパターン x_k というのが、各国（日本やアメリカなどの国）を指すことになる。

異なる傾向の変化は見られなかった。実際に、世界の軍事 3 大国と呼ばれているアメリカ、中国、ロシア（1985 年においてはソ連）は、クラスター数を $c = 2$ から $c = 5$ に変化させても同一のクラスターに分類される結果となり、このことは世界の情勢から考えても妥当性があるものと考えられる。また、クラスター数を $c = 5$ としたときに、アメリカやロシアと同一クラスターに分類された国、すなわち軍事力能力が高いと判断される国については、クラスター数を $c = 2$ から $c = 4$ まで変化させた場合においても、同一クラスターに分類されている（付録：図 B.1 ~ 図 B.12 を参照）

6.3 冷戦終焉期における世界とアジア地域の軍事情勢

冷戦期（1945 年から 1989 年）とは、ソビエト連邦を盟主とする共産主義（社会主義）陣営とアメリカ合衆国を盟主とする資本主義陣営の対立構造が続いた時期のことである。この間、世界は大きく 3 種類の国に分かれていた。すなわち、共産主義を支持する国、資本主義を支持する国、どちらにも属さない国である。

1985 年、ソ連共産党書記長に就任したミハイル・ゴルバチョフは、改革（ペレストロイカ）および新思考外交を掲げて大胆な軍縮提案を行い、西側との関係改善に乗り出した。また、東欧諸国に対しても改革を促し、1989 年に入るとポーランドやハンガリーを皮切りに共産党体制が相次いで倒れ、11 月には冷戦の象徴ともいべきベルリンの壁が崩壊すると、チェコスロバキアやルーマニアでも共産党政権が崩壊した（東欧革命）。その結果、12 月には地中海のマルタ島でゴルバチョフとジョージ・H・W・ブッシュが会談し、冷戦の終結を宣言することとなったのである。つまり、1985 年という年は世界が軍縮の方向に向かい始めるきっかけとなった年であるとみなすことができる。

したがって、ここでは第 2 次世界大戦後、最も軍事力が充実していたと思われる 1985 年における世界各国の軍事データ（138ヶ国）を用いてクラスタリングを行い、冷戦終焉期、つまり世界が軍縮に向かい始めた時の分析を行う。

6.3.1 クラスタリング結果

クラスタリング結果を図 6.1、図 6.2 および表 6.6 に示す。図 6.1 は各クラスターごとに国名を記したものであり、図 6.2 はクラスタリング結果を国別に各クラスターごと色分けをして、世界地図にプロットしたものである。また、表 6.6 は各クラスターに含まれている各地域ごとの国数を表したものである。なお、ここでは世界をその特性上から 8 つの地域に分けている。

Cluster 1 (9)	Cluster 2 (15)	Cluster 3 (8)	Cluster 4 (28)	Cluster 5 (78)		
アメリカ	フランス	ギリシャ	ベルギー	ルクセンブルク	コスタリカ	セーシェル
カナダ	イタリア	トルコ	デンマーク	ポルトガル	エルサルバドル	タンザニア
(東・西)ドイツ	オランダ	ブルガリア	ハンガリー	アルバニア	グアテマラ	ウガンダ
イギリス	ポーランド	チェコスロバキア	ノルウェー	キプロス	ホンジュラス	ベナン
ソ連	スペイン	イラン	オーストリア	アイルランド	ニカラグア	ブルキナ・ファソ
イラク	エジプト	リビア	フィンランド	マルタ	パナマ	コートジボアール
インド	イスラエル	パキスタン	ルーマニア	バーレーン	ボリビア	ガンビア
中国	サウジアラビア	ベトナム	スウェーデン	レバノン	コロンビア	ガーナ
ブラジル	シリア		スイス	カタール	エクアドル	ギニア
	オーストラリア		ユーゴスラビア	チュニジア	ガイアナ	ギニア・ビサウ
	日本		アルジェリア	(南・北)イエメン	パラグアイ	リベリア
	北朝鮮		ヨルダン	アフガニスタン	スリナム	マリ
	韓国		クウェート	バングラディシュ	ウルグアイ	ニジェール
	台湾		モロッコ	ネパール	ジブチ	ナイジェリア
	アルゼンチン		オマーン	スリランカ	エチオピア	セネガル
			アラブ首長国連邦	ブルネイ	スーダン	シエラレオネ
			インドネシア	フィジー	ブルンジ	トーゴ
			マレーシア	ラオス	カメルーン	ボツワナ
			ミャンマー	モンゴル	カーボ・ヴェルデ	マラウイ
			シンガポール	ニュージーランド	中央アフリカ共和国	モザンビーク
			タイ	パプアニューギニア	チャド	ザンビア
			キューバ	フィリピン	コンゴ	ジンバブエ
			メキシコ	パハマ	コンゴ民主共和国	
			チリ	ドミニカ	赤道ギニア	
			ペルー	ハイチ	ガボン	
			ベネズエラ	ジャマイカ	ルワンダ	
			アンゴラ	トリニダード・トバコ	ケニア	
			南アフリカ	ベリーズ	マダガスカル	

図 6.1: 1985 年における各国の軍事データを用いたクラスタリングの結果を、各クラスターごとと色分けをし(赤, 橙, 黄, 緑, 青), そのクラスターに分類された国名を記している. 括弧内の数字は、各クラスターに含まれている国の数である.

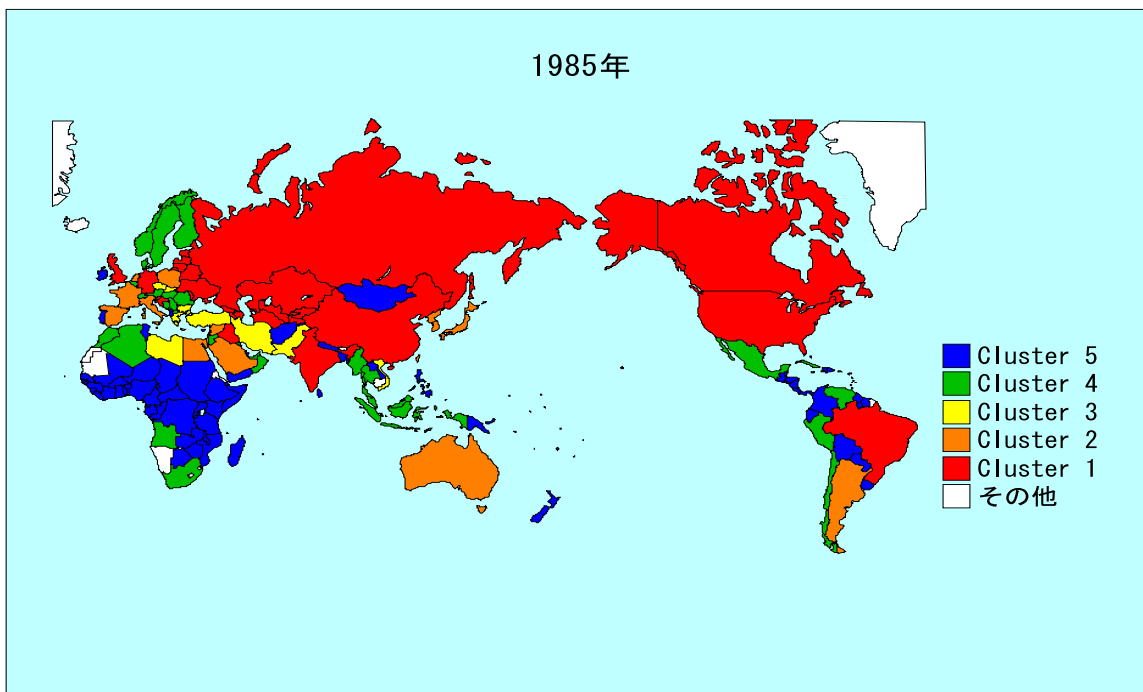


図 6.2: 1985 年における各国の軍事データを用いたクラスタリングの結果を世界地図にプロットした図。各クラスターの色は図 6.1 と同じである。

表 6.6: 各クラスターに含まれている各地域ごとの国の数 (1985 年)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
NATO 諸国	4	4	2	3	2
ワルシャワ条約機構	1	1	2	2	0
その他の欧州諸国	0	0	0	5	4
中東・北アフリカ諸国	1	3	2	6	5
中央・南アジア諸国	1	0	1	0	4
東アジア・大洋州諸国	1	3	1	5	4
中央・ラテンアメリカ諸国	1	1	0	4	19
サハラ以南のアフリカ諸国	0	0	0	2	37
合計	9	15	8	28	78

各クラスターの特性

図 6.3 から図 6.7 は、それぞれ国防費、現役兵力、戦車、主要水上戦闘艦艇、空軍作戦機における各クラスターごとに平均をとったものの割合を表したものである。あわせて、各クラスターごとの各変数の平均が示されている。

すなわち、図 6.3 における国防費について見てみると、Cluster 1 に分類された 9ヶ国の国防費の平均額が 985 億 6,400 万ドルであるということを示しており、同様に Cluster 2 については 125 億 6,900 万ドル（15ヶ国平均）、Cluster 3 については 30 億 3,300 万ドル（8ヶ国平均）、Cluster 4 については 20 億 8,300 万ドル（28ヶ国平均）、Cluster 5 については 1 億 7,000 万ドル（78ヶ国平均）であるということを示している。

以下に、これらのグラフから読み取ることができる各クラスターごとの特徴を示す。

- Cluster 1
どのカテゴリーにおいても 3 分の 2 以上の割合を占めており、軍事的能力が高い国のクラスターである。
- Cluster 2
現役兵力および戦車の数以外のカテゴリーにおいては Cluster 1 に次ぐ割合を占めており、軍事的能力が比較的高い国のクラスターである。
- Cluster 3
国防費と主要水上戦闘艦艇を除いては、Cluster 2 とほぼ同じ割合で存在しており、Cluster 1, 2 と Cluster 4, 5 には含まれた中間国のクラスターである。
- Cluster 4
各カテゴリーにおいて Cluster 2, 3 よりも割合が少なく、また Cluster 5 よりも割合が多いため軍事的能力が比較的低い国のクラスターである。
- Cluster 5
各カテゴリーにおいて割合が最も少なく、他のクラスターと比較して軍事的能力が低い国のクラスターである。

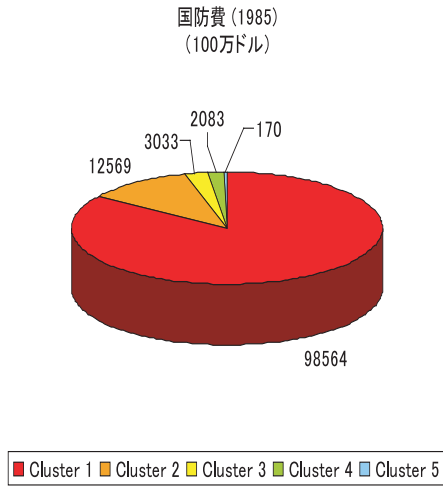


図 6.3: 1985 年における各クラスターごとの国防費の比較 (単位: 100 万ドル)

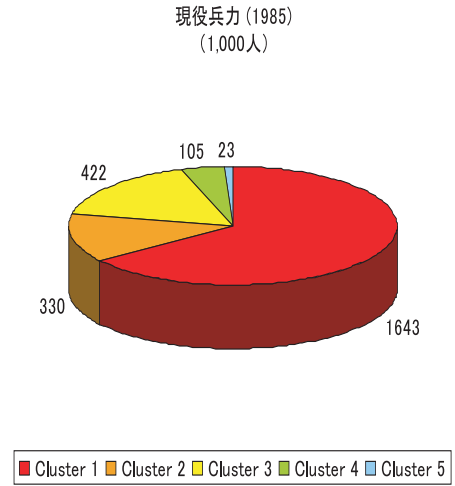


図 6.4: 1985 年における各クラスターごとの現役兵力の比較 (単位: 1,000 人)

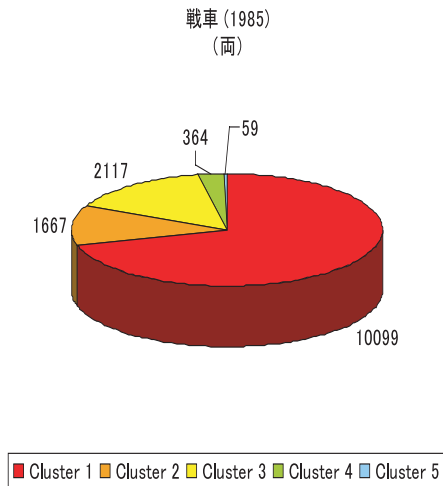


図 6.5: 1985 年における各クラスターごとの戦車の数の比較 (単位: 両)

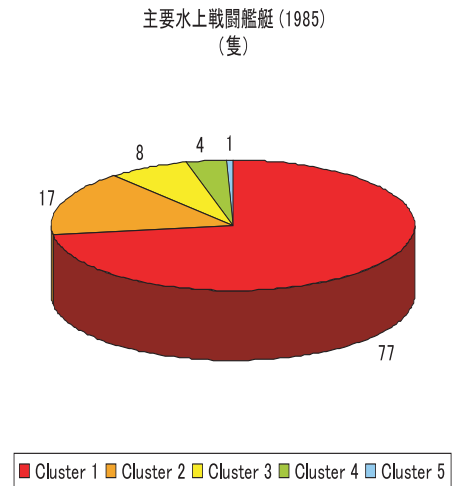


図 6.6: 1985 年における各クラスターごとの主要水上戦闘艦艇の数の比較 (単位: 隻)

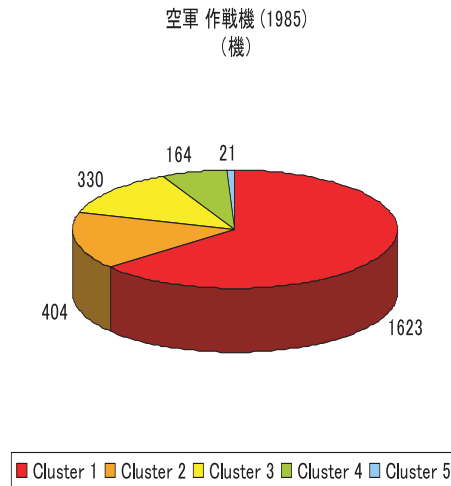


図 6.7: 1985 年における各クラスターごとの空軍作戦機の数比較 (単位: 機)

6.3.2 世界の軍事情勢の分析

NATO (北大西洋条約機構) 諸国

NATO⁷ とは、アメリカ合衆国を中心としたアメリカ・ヨーロッパ諸国によって結成された軍事同盟のことである。Cluster 1 に分類された国の約半数が、また Cluster 2 に分類された国の約 4 分の 1 がこの NATO 諸国の国であることから、軍事的能力の高い国が集まっている組織であることが分かる。

ワルシャワ条約機構

ワルシャワ条約機構 (Warsaw Treaty Organization/Warsaw Pact : WP) とは、1955 年にワルシャワ条約に基づきソビエト社会主義共和国連邦を盟主とした東ヨーロッパ諸国が結成した軍事同盟のことである。クラスタリング結果から判断すると、NATO 諸国に比べると軍事的能力に劣ると見ることができが、実際にはソ連一国で、アメリカ・カナダを除く NATO 諸国の合計を上回った軍事力を保有している点に注目すべきである (図 6.8 参照)

⁷NATO : North Atlantic Treaty Organization

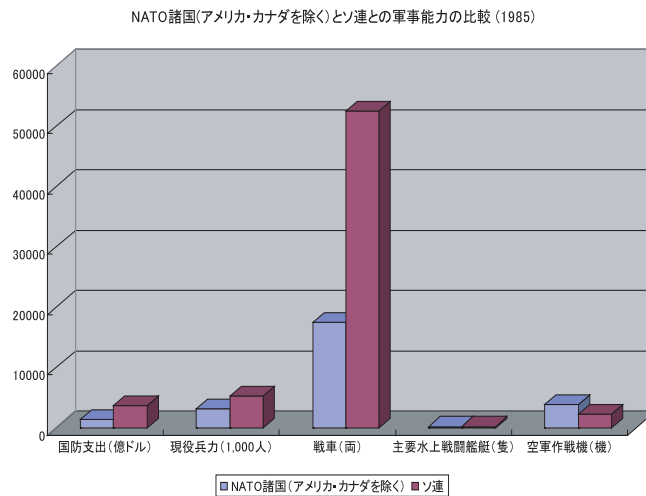


図 6.8: NATO 諸国 (アメリカ・カナダを除く) の合計とソ連の保有する軍事能力の比較を示したグラフ。特に、ソ連一国の保有する戦車の数は NATO 諸国の 3 倍にもものぼる。

その他の欧州諸国

Cluster 1 から Cluster 3 に含まれている国が無く、NATO 諸国と WP の国を除いた欧州諸国は比較的軍事的能力が低い国であるとみなすことができる。特に、地域の特性上、潜水艦や水上戦闘艦艇を保有している国が少ないため、このような結果になったものと考えられる。

中東・北アフリカ諸国

世界の火薬庫と呼ばれている中東諸国においては、イラクが Cluster 1 に分類されており、次いで、エジプト、イスラエル、サウジアラビア、シリアが Cluster 2 に分類されている。注目すべきは、これらの国の GDP に対する国防支出の割合の平均が 22.48% と非常に高い割合を示していることである。特に、イラクにおいては対 GDP の割合が 42.2% と、世界の中で最も高い割合を示している。

中央・南アジア諸国

この地域で最も軍事力の高いインドが Cluster 1 に分類されているほかは、パキスタン (Cluster 3) を除いて、軍事的能力が低いとされる Cluster 5 に分類されており、軍事的能力が比較的低い国が集まっている地域であると見なすことができる。しかしながら、インドの軍事的能力は非常に高く、この地域はインドの一国集中型の地域ともいえる。

中央・ラテンアメリカ諸国

この地域で最も高い軍事力を保有するブラジルが Cluster 1 に分類されており、次いでアルゼンチンが Cluster 2 に分類されているほかは、Cluster 5 に分類された国の約 4 分の 1 がこの中央・ラテンアメリカ地域の国であることから、この地域の軍事的能力は比較的低いと見なすことができる。なお、アルゼンチンが Cluster 2 に分類されてはいるが実際の軍事力を比較してみると各変数においてブラジルと約 2 倍ほどの差があり、中央・アジア諸国と同様の、ブラジルの一国集中型の地域といえる。

アフリカ諸国

ほとんどの国が軍事的能力の低いとされる Cluster 5 に分類されており、軍事的能力の低い国が集まっている地域であるといえる。この背景には、アフリカ諸国の中には戦車を保有していない国もあり、ほとんどの国が潜水艦や主要水上戦闘艦艇を保有しておらず、海上作戦能力の低さがこの結果になっているものと考えられる。

6.3.3 東アジア地域の軍事情勢の分析

東アジア地域においては、強大な軍事力を誇る中国が Cluster 1 に分類されており、次いで、日本、北朝鮮、韓国、台湾が Cluster 2 に分類される結果となった。実際に東アジア各国のデータを比較してみても、この様子を見て取ることができる。

図 6.9 は 1985 年における東アジア地域の国防費の比較を表したもので、Cluster 1, 2 に分類された中国、日本、北朝鮮、韓国、台湾の国防費が他の東アジア諸国に比べて多く支出されていることが分かる。

また、図 6.10 から図 6.13 は 1985 年における東アジア諸国の兵力、および陸・海・空の各戦力の比較を表したもので、特に、中国の圧倒的な軍事力を見て取ることができる。

1985 年における東アジア地域の情勢をまとめると、この時期は中国の軍事力が圧倒的であり、陸上、海上、航空戦力とも群を抜いていたことが分かった。また、日本と同じクラスターに分類された国については、韓国、台湾の各戦力は日本と同等、もしくはそれ以下であったが、北朝鮮については兵力および各戦力とも日本より高い割合で保有しており、我が国は各種装備の保有数において北朝鮮より劣っていたことが分かった。

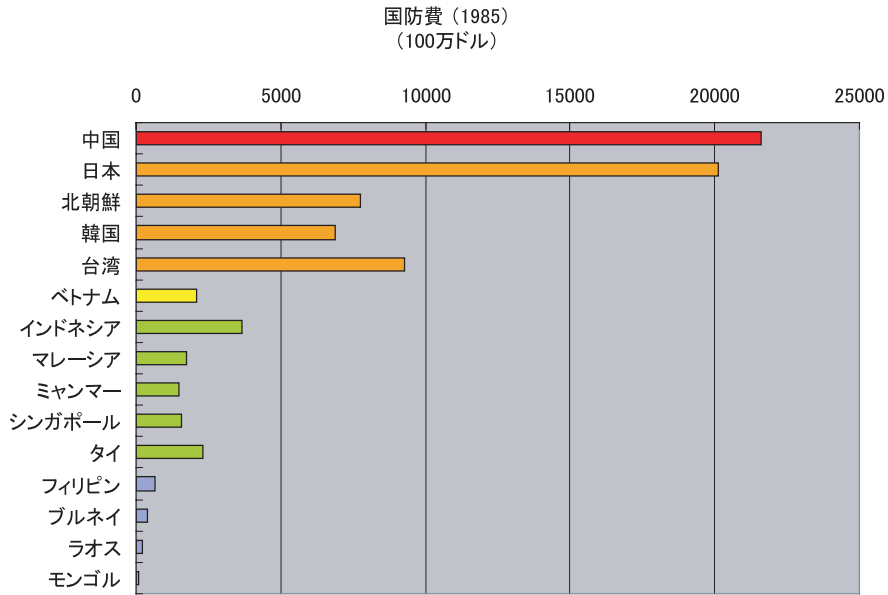


図 6.9: 1985 年における東アジア地域の国防費を比較したグラフ (単位: 100 万ドル)。棒グラフの各色はそれぞれのクラスターの色を示している。

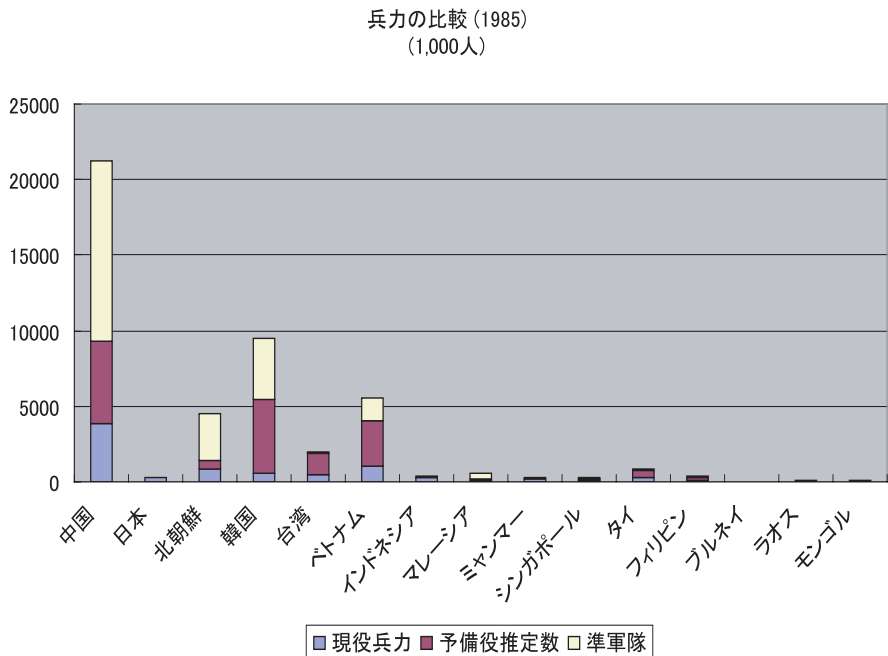


図 6.10: 1985 年における東アジア地域の兵力を比較したグラフ (単位: 1,000 人)

陸上戦力の比較 (1985)

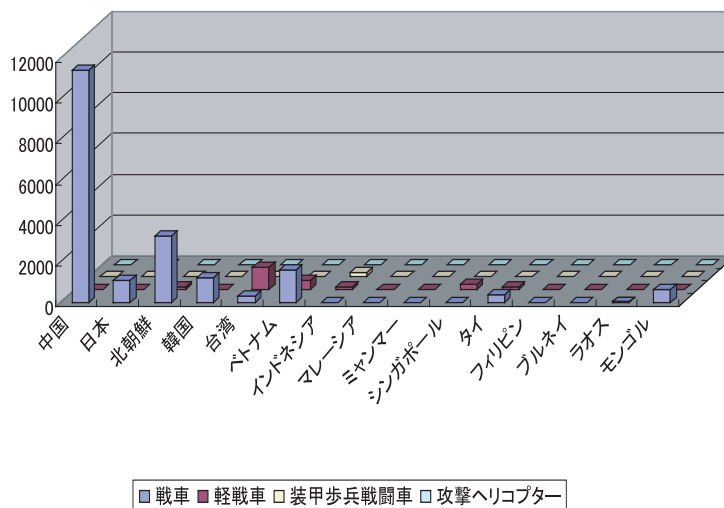


図 6.11: 1985 年における東アジア地域の陸上戦力を比較したグラフ。各国の保有する戦車，軽戦車，装甲歩兵戦闘車，攻撃ヘリコプターの数が比較されている。

海上戦力の比較 (1985)

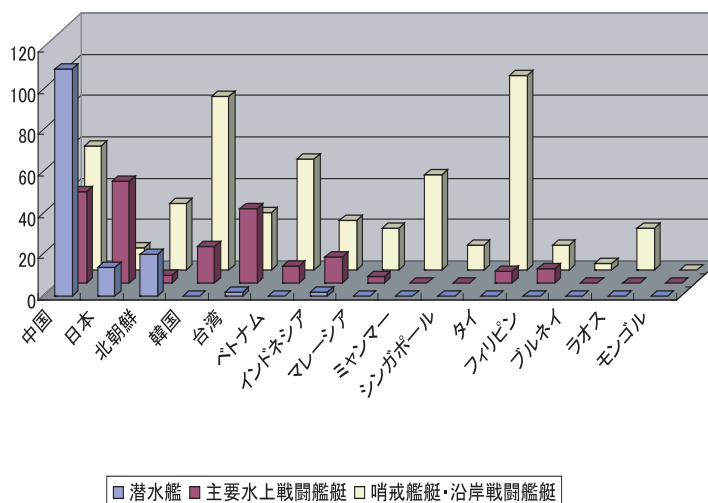


図 6.12: 1985 年における東アジア地域の海上戦力を比較したグラフ。各国の保有する潜水艦，主要水上戦闘艦艇，哨戒艦艇・沿岸戦闘艦艇の数が比較されている。

航空戦力 (1985)

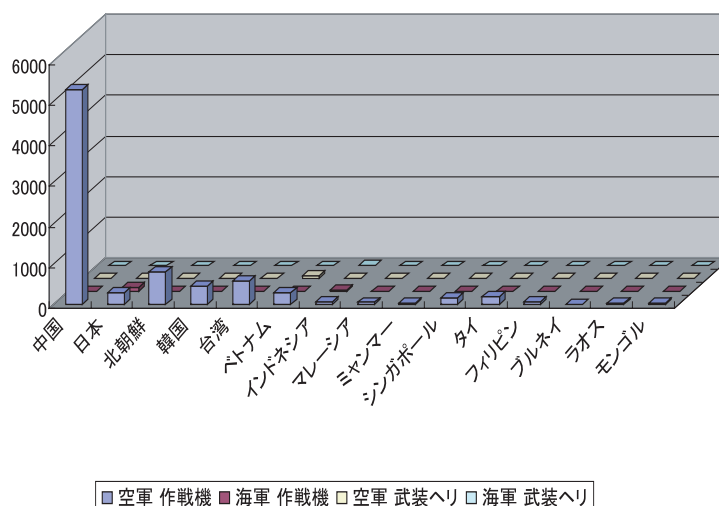


図 6.13: 1985 年における東アジア地域の航空戦力を比較したグラフ。各国の保有する空軍作戦機，海軍作戦機，空軍武装ヘリコプター，海軍武装ヘリコプターの数と比較されている。

6.4 現在の世界とアジア地域の軍事情勢

冷戦終結後，安全保障環境を改善するため，世界各国がその保有する軍事力や国防政策の透明性を高め，防衛当局者間の対話・交流，各種共同訓練などを通じて相互の信頼関係を深めることで，無用な軍備増強や不測の事態の発生とその拡大を抑えることが重要との認識が共有されることとなった。さらに，近年，国際テロリストや大量破壊兵器などのグローバルな脅威に対処するため，国家間の相互依存が拡大・深化したことに伴い，二国間及び多国間の連携・協力関係の充実・強化が図られるようになってきた。

こうした中，アメリカは本土防衛を国防の最優先事項とするとともに，海外においても米国の安全と行動の自由の確保，国益を重視する現実主義的な姿勢を示しており，必要な軍の変革を推進するとともに，世界的に米軍の展開態勢を見直す作業を進めている。また，ロシアについては，兵員の削減と軍種の統合，装備面での軍の近代化，即応態勢の立て直しなどが進められてきており，今後は，軍の効率化・近代化や即応態勢の向上を含めた軍改革の残された課題達成に取り組んでいくものと考えられている [11]。

一方，アジア太平洋地域においては，北朝鮮の核問題，特に中国における軍事費の大幅な増加，核・ミサイル戦力や海・空軍を中心とした全軍の近代化や台湾問題など，日本の安全を脅かす新たな要因が発生している。

このような情勢の下，ここでは 2003 年における世界各国の軍事データ (157ヶ国) を用い

てクラスタリングを行い，現在の世界各国の軍事情勢の分析を行う．

6.4.1 クラスタリング結果

クラスタリング結果を図 6.14，図 6.15 および表 6.7 に示す．図 6.14 は各クラスターごとに国名を記したものであり，図 6.15 はクラスタリング結果を国別に各クラスターごと色分けをして，世界地図にプロットしたものである．また，表 6.7 は各クラスターに含まれている地域ごとの国数を表したものである．なお，ここでは世界をその特性上から 7 つの地域に分けている⁸．

表 6.7: 各クラスターに含まれている各地域ごとの国の数 (2003 年)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
NATO 諸国	5	5	5	2	1
ロシア・欧州諸国	3	0	2	6	15
中東・北アフリカ諸国	3	4	4	4	3
中央・南アジア諸国	2	0	0	3	6
東アジア・大洋州諸国	4	3	4	2	6
中央・ラテンアメリカ諸国	2	0	2	4	18
サハラ以南のアフリカ諸国	0	0	0	3	36
合計	19	12	17	24	85

各クラスターの特性

各クラスターの特性を見るために，クラスタリングに使用した各変数，すなわち国防費，兵力，主要陸上戦力，主要海上戦力，主要航空戦力の 5 つのカテゴリーを用いて分析を行う．図 6.16 から図 6.20 は，それぞれ国防費，現役兵力，戦車，主要水上戦闘艦艇，空軍作戦機における各クラスターごとに平均をとったものの割合を表したものである．あわせて，各クラスターごとの各変数の平均が示されている．

⁸1985 年においては世界を 8 つの地域に分けていたが，ワルシャワ条約機構が，1989 年の冷戦終結に伴い 1991 年 7 月 1 日に解散したため，ロシアを欧州諸国に入れて 7 地域とした．

第 6 章 クラスタリングによる日本周辺の軍事情勢の分析

Cluster 1 (19)	Cluster 2 (12)	Cluster 3 (17)	Cluster 4 (24)	Cluster 5 (85)		
アメリカ	カナダ	ベルギー	チェコ共和国	ルクセンブルク	ニューージーランド	赤道ギニア
フランス	ギリシャ	デンマーク	ハンガリー	アルバニア	パプアニューギニア	ガボン
ドイツ	オランダ	ノルウェー	オーストリア	アルメニア	バハマ	ルワンダ
イタリア	スペイン	ポーランド	アゼルバイジャン	ボスニア・ヘルツェゴビナ	ドミニカ	ケニア
イギリス	トルコ	ポルトガル	ブルガリア	クロアチア	ハイチ	マダガスカル
ロシア	エジプト	スウェーデン	フィンランド	キプロス	ジャマイカ	セーシェル
ペラルーシ	イラン	スイス	ルーマニア	エストニア	トリニダード・トバゴ	タンザニア
ウクライナ	イスラエル	アルジェリア	ユーゴスラビア	マケドニア	ベリーズ	ウガンダ
イラク	サウジアラビア	クウェート	ヨルダン	グルジア	コスタリカ	ベナン
カタール	オーストラリア	オマーン	リビア	アイルランド	エルサルバドル	ブルキナファソ
シリア	台湾	アラブ首長国連邦	モロッコ	ラトビア	グアテマラ	コートジボアール
インド	ベトナム	インドネシア	イエメン	リトアニア	ホンジュラス	ガンビア
パキスタン		マレーシア	カザフスタン	マルタ	ニカラグア	ガーナ
中国		ミャンマー	トルクメニスタン	モルドバ	パナマ	ギニア
日本		シンガポール	ウズベキスタン	スロバキア	ボリビア	ギニアビサウ
北朝鮮		チリ	フィリピン	スロベニア	エクアドル	リベリア
韓国		コロンビア	タイ	バーレーン	ガイアナ	マリ
メキシコ			キューバ	レバノン	パラグアイ	ニジェール
ブラジル			アルゼンチン	チュニジア	スリナム	ナイジェリア
			ペルー	アフガニスタン	ウルグアイ	セネガル
			ベネズエラ	バングラディシュ	ジブチ	シエラレオネ
			コンゴ民主共和国	キルギスタン	エチオピア	トーゴ
			アンゴラ	ネパール	スーダン	ボツワナ
			南アフリカ	スリランカ	ブルンジ	マラウイ
				タジキスタン	カメルーン	モザンビーク
				ブルネイ	カーボヴェルデ	ザンビア
				フィジー	中央アフリカ共和国	ジンバブエ
				ラオス	チャド	
				モンゴル	コンゴ	

図 6.14: 2003 年における各国の軍事データを用いたクラスタリングの結果を、各クラスターごと色分けをし(赤, 橙, 黄, 緑, 青), そのクラスターに分類された国名を記している。括弧内の数字は、各クラスターに含まれている国の数である。

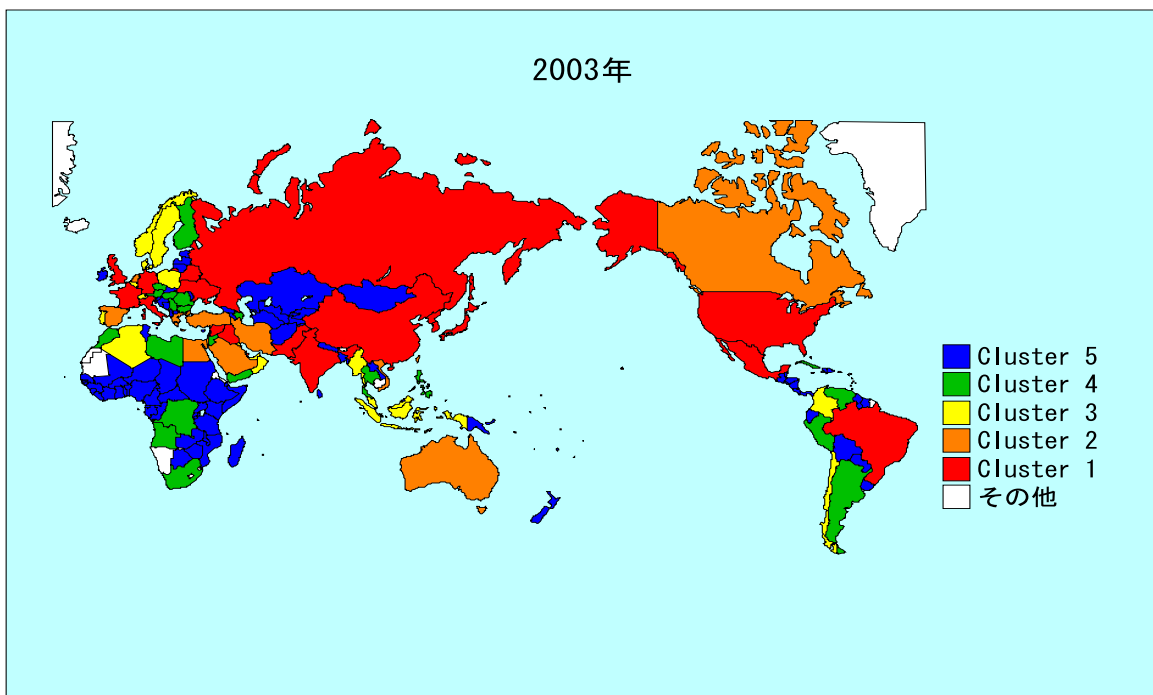


図 6.15: 2003 年における各国の軍事データを用いたクラスタリングの結果を世界地図にプロットした図 . 各クラスターの色は図 6.14 と同じである .

以下に、これらのグラフから読み取ることができる各クラスターごとの特徴を示す。

- Cluster 1
すべてのカテゴリーにおいて 2 分の 1 以上の割合を占めており、軍事的能力が高い国のクラスターである。
- Cluster 2
すべてのカテゴリーにおいて Cluster 1 に次ぐ割合を占めており、軍事的能力が比較的高い国のクラスターである。
- Cluster 3
各種戦力においては Cluster 4 とほぼ同等の割合であるが、国防費および兵力において勝っており、中間国のクラスターとして見なすことができる。
- Cluster 4
各カテゴリーにおいて Cluster 2, 3 よりも割合が少なく、また Cluster 5 よりも割合が多いため軍事的能力が比較的低い国のクラスターである。
- Cluster 5
各カテゴリーにおいて割合が最も少なく、他のクラスターと比較して軍事的能力が低い国のクラスターである。

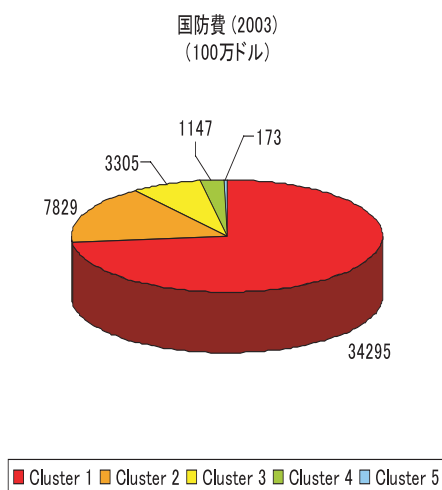


図 6.16: 2003 年における各クラスターごとの国防費の比較 (単位: 100 万ドル)

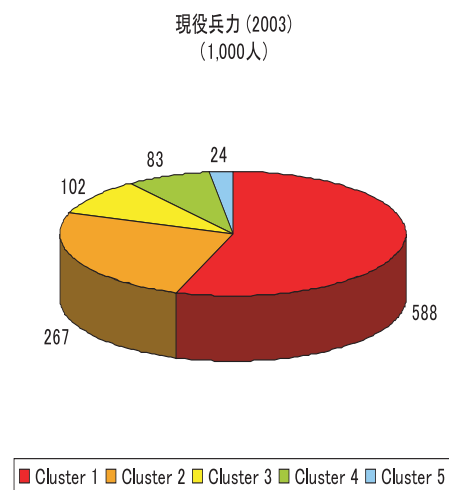


図 6.17: 2003 年における各クラスターごとの現役兵力の比較 (単位: 1,000 人)

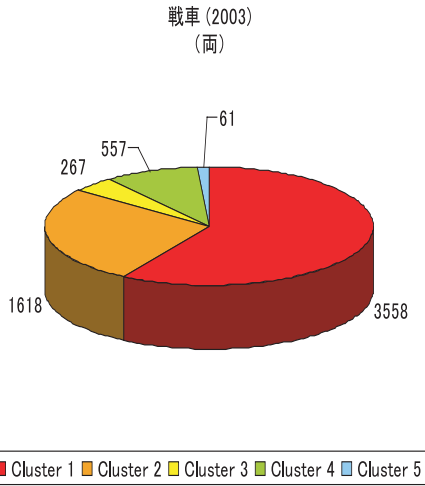


図 6.18: 2003 年における各クラスターごとの戦車の数の比較 (単位: 両)

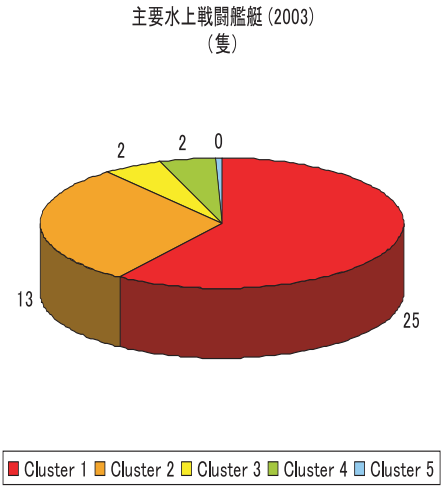


図 6.19: 2003 年における各クラスターごとの主要水上戦闘艦艇の数の比較 (単位: 隻)

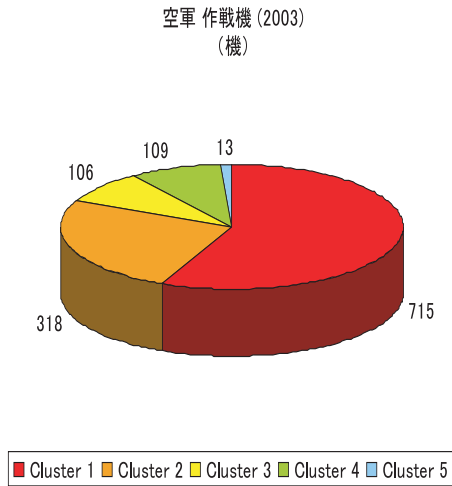


図 6.20: 2003 年における各クラスターごとの空軍作戦機の数の比較 (単位: 機)

6.4.2 世界の軍事情勢の分析

NATO（北大西洋条約機構）諸国

Cluster 1 と Cluster 2 に分類されている国のうち約 3 分の 1（31ヶ国中 10ヶ国）が NATO 諸国の国であることから、現在においても軍事的能力の高い国が集まっている組織であることが分かる。

ロシア・欧州諸国

Cluster 1 に含まれている国は、ロシア、ベラルーシ、ウクライナといずれも旧ソ連を形成していた国であり、いまだに高い軍事能力を保有していることが分かる。それ以外の欧州諸国は半数以上が Cluster 5 に分類されており、1985 年の結果とあまり変化がみられず、比較的軍事的能力が低い地域であるとみなすことができる。

中東・北アフリカ諸国

この地域の特性として、GDP に占める国防支出の割合が他の地域に比べて非常に高い点あげられる。特に、Cluster 1 に分類されたイラク、カタール、シリアの 3 カ国の平均が 10.2% と 2 桁台に達しており、軍事力の拡充に力を注いでいることが分かる。また、カタールにいたっては国民一人当たりの国防費が 2,857 ドル（世界平均：173 ドル）と世界でもトップである。

中央・南アジア諸国

この地域で最も軍事力の高いインドと、インドとの対立が続くパキスタンが Cluster 1 に分類されているほかは、Cluster 4, 5 に分類されており、この地域は比較的軍事能力の低い国が集まっている地域である。

中央・ラテンアメリカ諸国

この地域で最も高い軍事力を保有するブラジル、および海上戦力が最も高いメキシコが Cluster 1 に分類されているほかは、大部分が Cluster 4, 5 に分類されており、ブラジルなどの一部を除いて軍事的能力の低い国が集まっている地域である。

アフリカ諸国

ほとんどの国が Cluster 5 に分類されており、軍事的能力の低い国が集まっている。その背景には、アフリカ諸国の中には戦車を保有していない国もあり、ほとんどの国が潜水艦や主要水上戦闘艦艇を保有しておらず、海上作戦能力の低さがこの結果になっているものと考えられる。

6.4.3 東アジア地域の軍事情勢の分析

東アジア地域においては、世界軍事 3 大国の一つである中国が Cluster 1 に分類されているほか、日本、北朝鮮、韓国も Cluster 1 に分類される結果となった。これらについて、台湾、ベトナムが Cluster 2 に分類されている。

図 6.21 は 2003 年における東アジア地域の国防費の比較を表したもので、Cluster 1 に分類された国の中でも特に中国、日本の国防費が他の東アジア諸国に比べて多く支出されている様子が分かる。

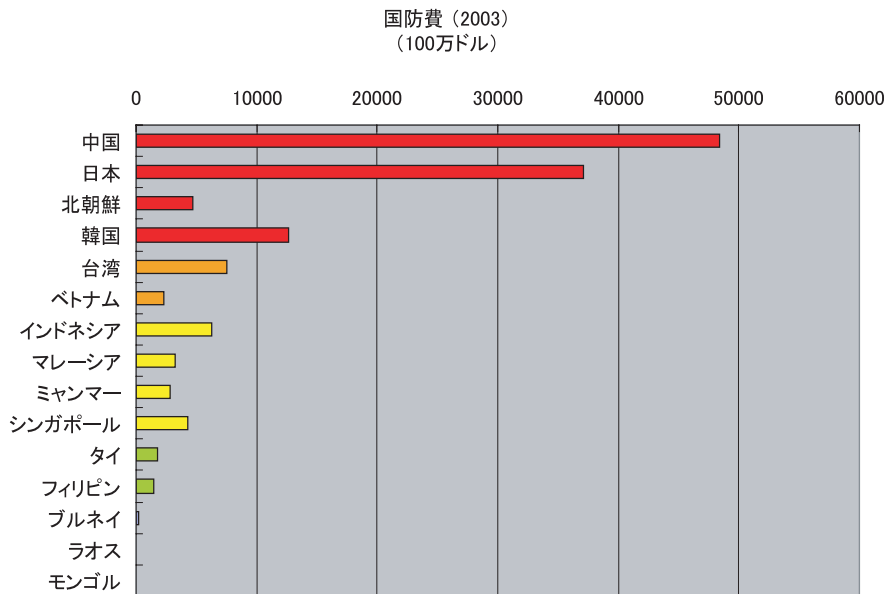


図 6.21: 2003 年における東アジア地域の国防費を比較したグラフ (単位: 100 万ドル)。棒グラフの各色はそれぞれのクラスターの色を示している。

また、図 6.22 から図 6.25 は 2003 年における東アジア諸国の兵力、および陸・海・空の各戦力の比較を表したもので、中国をトップに、北朝鮮、韓国、日本と続いていることが読み取れる。注目すべきは、韓国、北朝鮮の兵力で、現役兵力では中国の半分以下なのに対して、有事の際には兵力として加わることでできる予備役や準軍隊を含めると、中国の 1.5 倍から 2 倍近くの兵力に膨れ上がることがわかる。

2003 年における東アジア地域の情勢をまとめると、中国の軍事力の優勢は変わらないが、1985 年においては圧倒的であった中国軍の兵力において、予備役や準軍隊を含めたときの兵力では韓国や北朝鮮が中国をも上回るような傾向が見られるようになってきた。これは、中国が軍縮を進める一方でその周りの国、特に朝鮮半島の国々において軍拡を進める動きがあるためである。また、全体的に軍事的能力が低い国のクラスターに属する国が少なくなっており、この地域は世界的にみても軍事的能力が高い国が集まっている地域であるとみなすことができる。

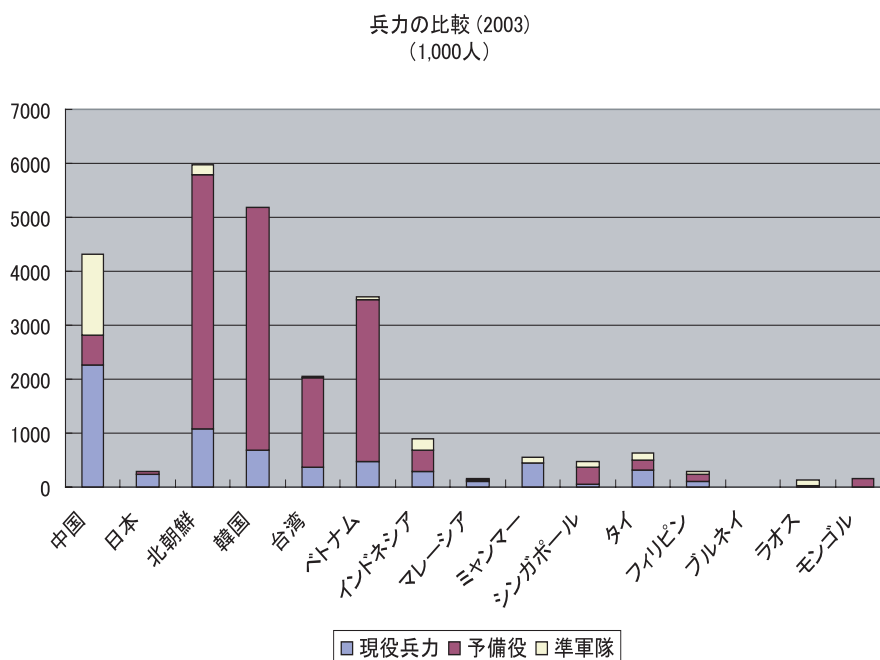


図 6.22: 2003 年における東アジア地域の兵力を比較したグラフ (単位: 1,000 人)

陸上戦力の比較 (2003)

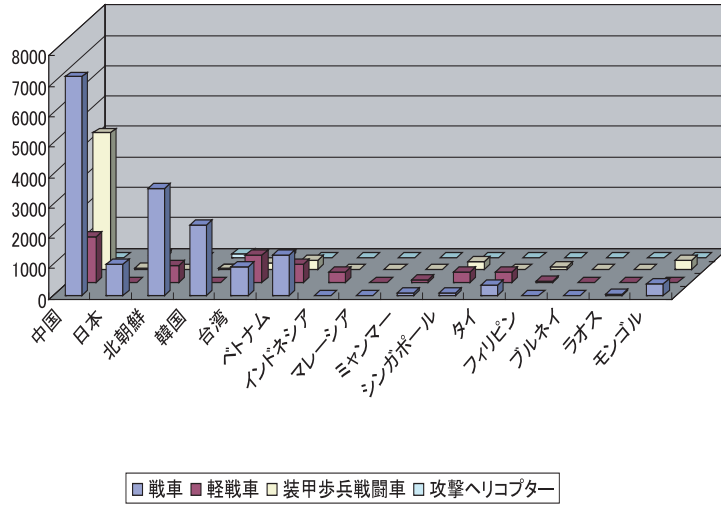


図 6.23: 2003 年における東アジア地域の陸上戦力を比較したグラフ。各国の保有する戦車，軽戦車，装甲歩兵戦闘車，攻撃ヘリコプターの数が比較されている。

海上戦力の比較 (2003)

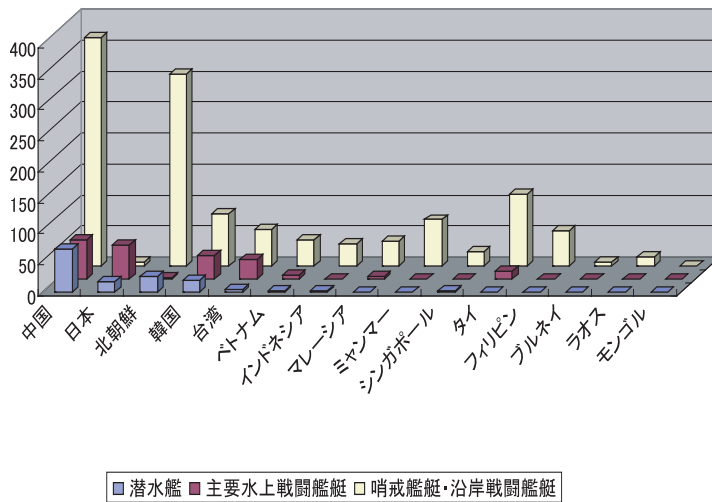


図 6.24: 2003 年における東アジア地域の海上戦力を比較したグラフ。各国の保有する潜水艦，主要水上戦闘艦艇，哨戒艦艇・沿岸戦闘艦艇の数が比較されている。

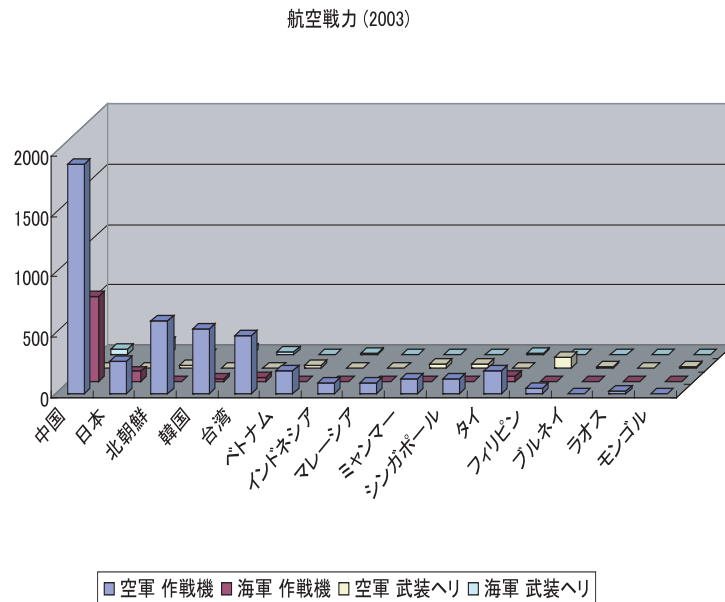


図 6.25: 2003 年における東アジア地域の航空戦力を比較したグラフ。各国の保有する空軍作戦機，海軍作戦機，空軍武装ヘリコプター，海軍武装ヘリコプターの数と比較されている。

6.5 日本をとりまく軍事的リスク

本節では日本を取り巻く軍事的リスク分析を行う。そのため、日本周辺地域、すなわち東アジア地域に注目をし、クラスター分析により軍事的能力の高い国と判断された国、言い換えれば軍事的脅威の高い国として分類された中国などの軍事情勢、あるいは国防政策について調べ、日本に対する脅威（リスク）について分析を行う。

6.5.1 東アジア地域における主要国の軍事情勢の変化

ここでは 2003 年のクラスター分析により Cluster 1 および Cluster 2 に分類された国、すなわち中国、日本、北朝鮮、韓国、台湾、ベトナムの 6ヶ国のデータを用いて 1985 年から 2003 年にかけてどのように軍事力が推移したのかについて見ていく。

図 6.26 から図 6.28 には、1985 年と 2003 年における東アジア主要 6ヶ国の国防費関係の比較が示されている。図 6.26、図 6.27 から、中国、日本、韓国には国防費の増加が見られるが、北朝鮮、台湾については減少していることがわかる。特に中国については、国防支出額において約 2.2 倍もの増加が見られる。

一方、図 6.28 は、東アジア主要 6ヶ国 GDP に占める国防支出費の割合の比較を示したものであるが、北朝鮮の国防費の GDP に占める割合をしてみると、他国が減少もしくは変化が無いのに対して 5%もの増加が見られる。

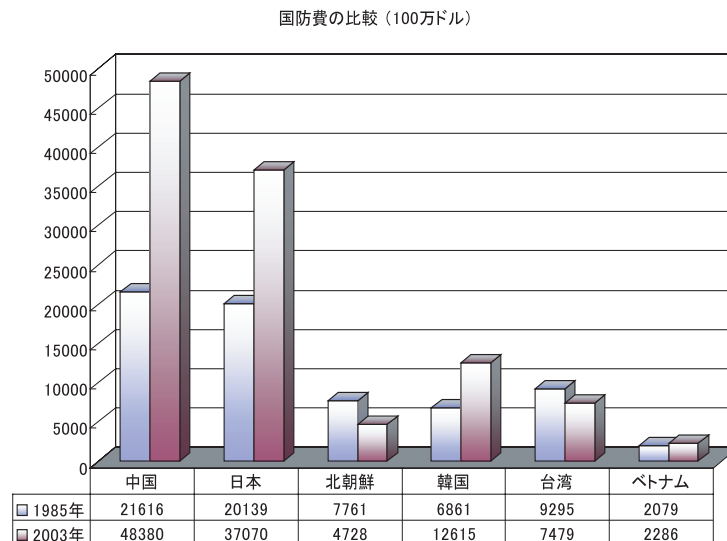


図 6.26: 1985 年と 2003 年における東アジア主要 6ヶ国における国防費の比較を表したグラフ (単位：100 万ドル)

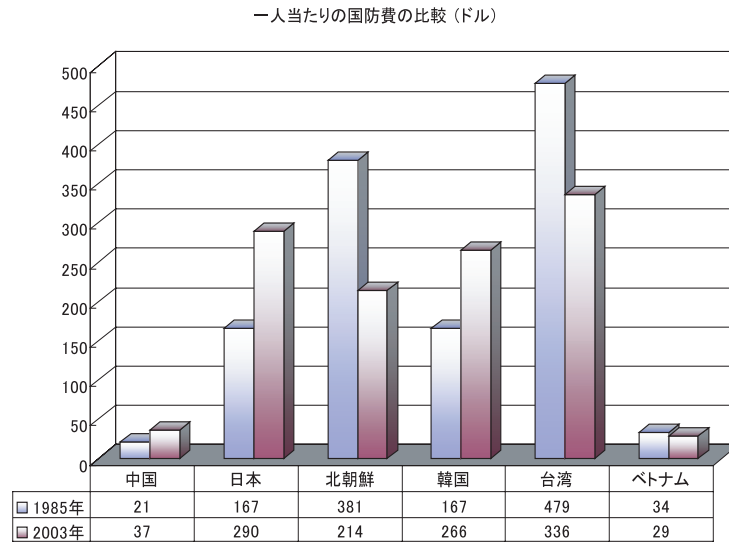


図 6.27: 1985 年と 2003 年における東アジア主要 6ヶ国における国民一人当たりの国防費の比較を表したグラフ (単位: ドル)

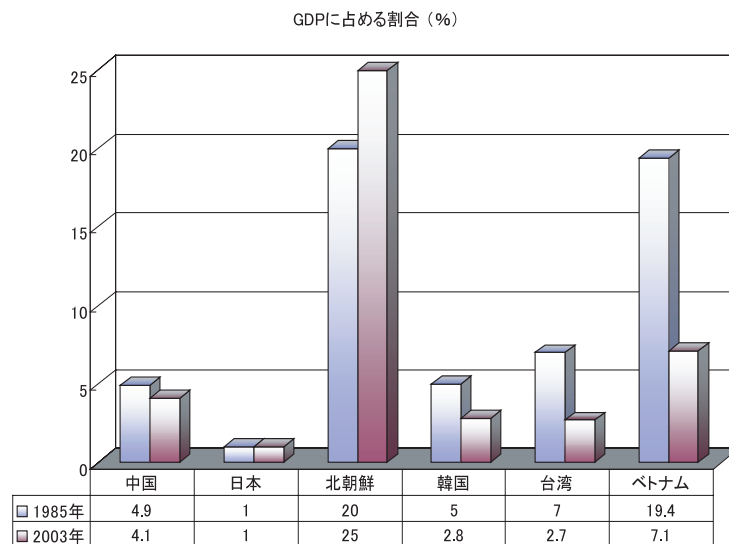


図 6.28: 1985 年と 2003 年における東アジア主要 6ヶ国における GDP に占める国防費の割合を比較したグラフ (単位: %)

ここで、北朝鮮の国防費関係に注目してみる。北朝鮮の国防支出が減少している理由としては経済的な問題が考えられる。実際に、図 6.29 に示されているように北朝鮮の GDP は年々減少し、2003 年は 1985 年に比べて 199 億ドルの減少がある。その結果、国防支出についても約 30 億ドルの減少があったと考えられる。このような状況の中で、GDP に占める国防費の割合が増加しているということは、北朝鮮がより軍事力に力を注いでいることに他ならず、今後の動向に注目すべき国であることがわかる。

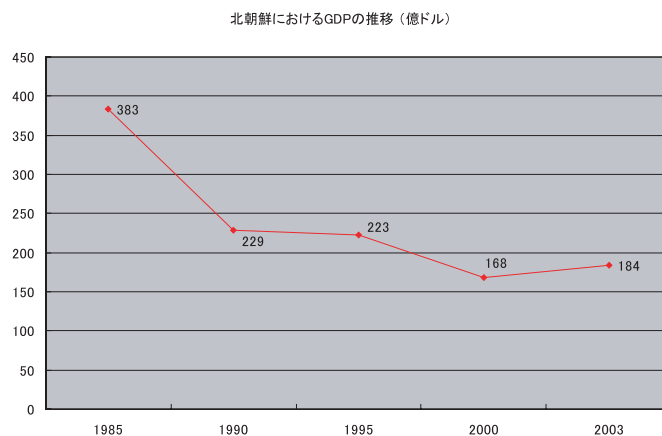


図 6.29: 1985 年から 2003 年にかけての北朝鮮における GDP の推移（単位：億ドル）[55]

また、図 6.30 は各国の現役兵力、予備役、準軍隊を足し合わせた総兵力の比較を示したものであり、日本、台湾においてはほとんど変化が無く、また、中国、韓国、ベトナムにおいては多くの兵力が削減されている。中でも中国は、1985 年の約 5 分の 1 まで総兵力を削減している。しかしながら、北朝鮮においては総兵力が増加傾向にあり、特に図 6.31 に示されているように総人口に占める総兵力の割合が 27% と群を抜いている。これは、日本において自衛官、予備自衛官、海上保安庁などに勤務している人間が 500 人に 1 人という割合に対して、北朝鮮では 4 人に 1 人の割合であるということを示している。また、韓国にいたっては 1985 年に比べて大幅に減ってはいるものの、10 人に 1 人という割合であり、台湾についても同様の割合である。この背景には、これらの国は徴兵制度をとっているためであると考えられる。

図 6.32 から図 6.34 は、陸・海・空の各戦力の増減を示したものである。これらのグラフから分かることは、中国は陸上戦力の主力である戦車や海上戦力の主力でもある潜水艦、あるいは航空戦力の主力である空軍作戦機など、数の上では大幅な削減を行っているということである。一方、韓国に目を向けてみると、陸上・海上・航空の全ての戦力において装備の増加が見られる。特に陸上戦力の主力である戦車の数が大幅に増加している。また、北朝鮮について見てみると、陸上戦力の増加が見られ、また、哨戒艦艇・沿岸戦闘艦艇の大幅な増加も見られるが、その他の装備に関してはあまり変化が見られない傾向にある。

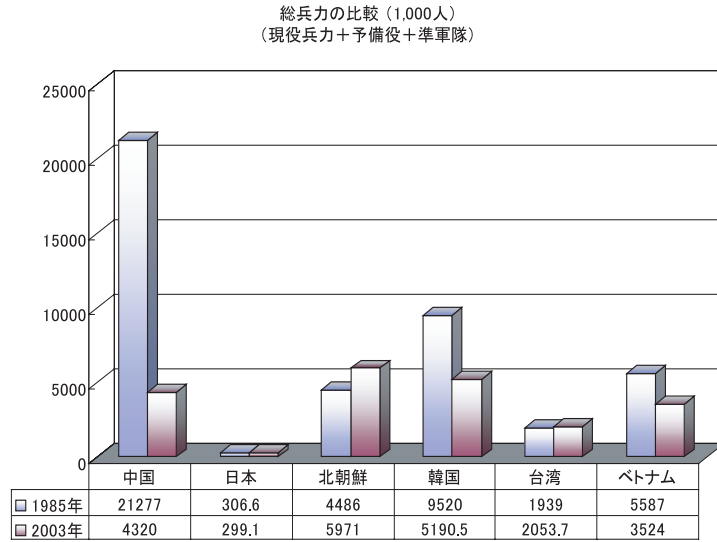


図 6.30: 1985 年と 2003 年における東アジア主要 6ヶ国の総兵力の比較 (単位: 1,000 人)

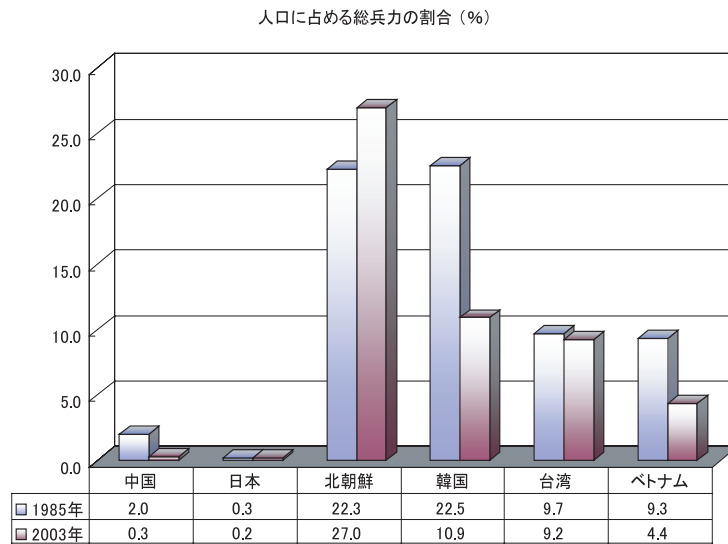


図 6.31: 1985 年と 2003 年における東アジア主要 6ヶ国の総人口に占める総兵力の割合 (単位: %)

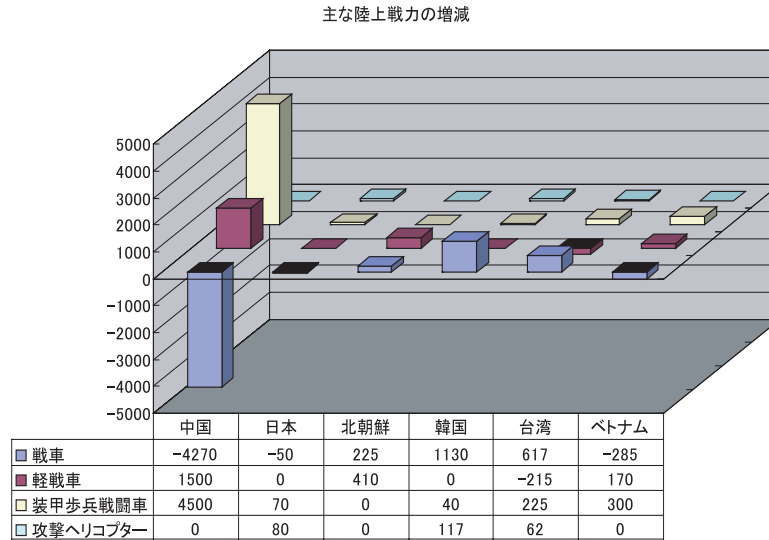


図 6.32: 1985 年と 2003 年における東アジア主要 6ヶ国の陸上戦力の増減 . 1985 年よりも減少したものはマイナスで表現されている .

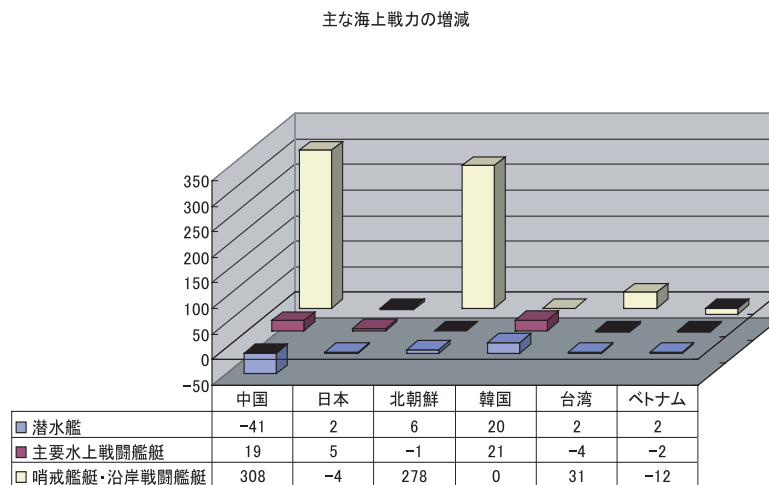


図 6.33: 1985 年と 2003 年における東アジア主要 6ヶ国の海上戦力の増減 . 1985 年よりも減少したものはマイナスで表現されている .

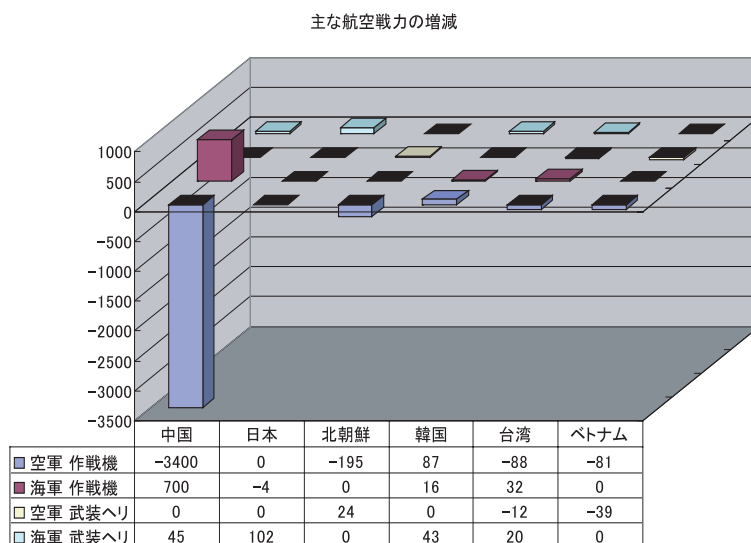


図 6.34: 1985 年と 2003 年における東アジア主要 6ヶ国の航空戦力の増減．1985 年よりも減少したものはマイナスで表現されている．

6.5.2 日本に対する軍事的脅威

これまでの分析から，日本に対する軍事的な意味での脅威としては，東アジア諸国においても特に，中国，北朝鮮，韓国の 3ヶ国に注目する必要があると判断できる．これは，クラスタリング結果においても裏付けられており，2003 年におけるクラスタリング結果では，全て軍事的能力の高い国のクラスターである Cluster 1 に分類されている．上記 3ヶ国の軍事情勢の変化とそれぞれの国における国防政策を踏まえた上で，日本に対する軍事的脅威についてまとめると，以下のようなになる．

中国

1985 年における中国は，東アジア地域において他国を寄せ付けない圧倒的な軍事力を保有していた．しかしながら，冷戦後の世界の軍縮の動きに合わせて，この約 15 年間で大幅な削減を実施している．具体的には，陸上戦力の主力である戦車については 4,270 両の削減，海上戦力の主力でもある潜水艦については 41 隻の削減，また空軍作戦機にいたっては 3,400 機もの作戦機を削減している（図 6.32 ～ 図 6.34 参照）．

この背景には，世界の軍事発展の趨勢に対応するために「中国の特色ある軍事変革」を積極的に推し進めるとい中国の国防政策が関係している．これにより，中国は継続する高い国防費の伸びを背景に，陸軍を中心とした兵員の削減と核・ミサイル戦力や海・空軍を中心

とした全軍の更なる近代化を進めるとともに、高い能力を持つ人材の育成に努めている。

中国は政治・経済的にアジア地域の大国として着実に成長し続けているが、一方で、最近では日本周辺での動きも活発化してきている。特に、昨年 11 月の中国原子力潜水艦による我が国領海内潜没航行事案をはじめとする海軍艦艇の活動や、海洋調査船による活動などが顕著な例である。これは、中国海軍が近海において防御作戦空間を拡大し総合的作戦能力を増強することを目指しているためであると指摘されており、特に領海が隣接している我が国においては常にその動向に注目し、軍事衝突が生起するリスクを回避するためにも、国としての具体的な方策を打ち立てていくことが急務の課題であるといえる。

北朝鮮

北朝鮮は「先軍政治」という政治方式の下、現在も軍事面に対して資源を重点的に配分し戦力・即応態勢の維持・強化に努めている。このことは図 6.28 における GDP に占める国防支出の割合が 25%であるということからも見てわかる。しかも、本年 4 月の最高人民会議における北朝鮮の公式発表によれば、北朝鮮の本年の国家予算に占める国防費の割合は 15.9%となっており、国家の運営において軍事を重要視し、かつ軍事に依存する状況は今後も継続すると考えられる。

また、北朝鮮は大規模な特殊部隊を保持するなど、いわゆる非対称的な軍事能力を維持・強化していると考えられている。特に、総人口の約 5%が現役の軍人であり、予備役、準軍隊を含めると約 25%もの兵力になる（図 6.31）。

特に最近では、北朝鮮の核問題にも注目が集められており、このことは我が国の安全保障に重大な影響を及ぼす問題であるのみでならず、国際社会全体にとっても重要な問題であることから、6ヶ国協議を通じた具体的解決、日・米・韓の緊密な連携、中国及びロシアさらには他の諸国や国際機関などによりその対応がとられている。

また、北朝鮮の弾道ミサイルについてはなお不明な点が多いが、政治外交的観点や外貨獲得の観点などからも弾道ミサイルに高い優先度を与えており、長射程化を着実に進めてきているということが考えられる。北朝鮮のこうした軍事的な動きは朝鮮半島の緊張を高めており、我が国を含む東アジア全域の安全保障にとって重大な不安定要因となっている。特に我が国においては、北朝鮮のミサイルの射程圏内にあり直接被害が及ぶ可能性も高い。この国は、依然として閉鎖的な体制をとっているため、その動向を明確に把握することは困難であると言われているが、その真の意図がなんであるのかを見極めることが重要である。そのためには、中国と同様に常にその動向に注目し、最新の注意を払っていく必要がある。未だに解決されていない拉致問題も含めて、日本の国益が北朝鮮によって損なわれるリスクは高いものであると判断できる。

韓国

韓国における 1985 年の国防支出額は、68 億 6,100 万ドルと東アジア地域において 5 番目であったのが、2003 年には 126 億 1,500 万ドルと中国、日本に次いで 3 番目に多い国となっている（図 6.26）。特に、近年における韓国の目覚ましい経済発展を背景に、海・空軍の近代化が推し進められている。また、韓国の「2004 国防白書」[56]によると、国防政策の重点事項として次の 4 点が設定されている。

1. 確固たる国防態勢の確立：
どのような軍事的脅威と侵略に対しても即座に対応できるよう備え、敵の戦争挑発を抑制し、挑発時には勝利を保障できる態勢を維持する。
2. 協力的自主国防推進：
自主国防と米韓同盟が安全保障の重要な 2 本の軸であるとし、安全保障の自主的な力量を備えていくとともに、米韓同盟をより堅固かつ未来志向的に発展させていく。
3. 一貫した国防改革推進：
絶え間なく変化する安全保障環境と新しい挑戦に備え、持続的な自己評価と改革を通して軍事的対応態勢と能力を強化していく。
4. 信頼される国軍像確立：
以上のためには何よりも軍自らが内部改革を通して望ましい軍隊文化を発展・定着させていかねばならない。

このような政策から、韓国は北朝鮮に対する警戒を最重要事項とし、日本と同様にアメリカとの関係を維持していこうとしていることがわかる。実際に、在韓米軍は朝鮮半島の軍事バランスを維持し、朝鮮半島における大規模な武力紛争の発生を抑止する上で大きな役割を果たしている。アメリカは、米韓相互防衛条約に基づき第 2 歩兵師団と第 7 空軍などを中心とする約 36,000 人の部隊を韓国に配備し、韓国軍とともに米韓連合軍司令部を設置している。

以上のことから判断すると、1985 年以降徐々に軍事力を高めてきてはいるものの、日本に対する軍事的リスクは低いものと判断できる。

6.6 クラスタリング結果の妥当性評価

本節では、本分析で行ったクラスタリング結果の妥当性を、主成分分析により得られた結果と、核兵器および弾道ミサイルの拡散・保有状況の変化から評価する。

6.6.1 主成分分析による妥当性評価

ここでは、2003 年の軍事データを用いて主成分分析を行い、本分析で行ったクラスタリング結果の妥当性を主成分分析の結果を用いることで評価していく。なお、本分析においては、数理システム社の S-PLUS を用いて主成分分析を行った。

図 6.35 は 2003 年の軍事データを用いて主成分分析した結果得られた、第 10 主成分までの固有値と累積寄与率を表しており、図 6.36 は第 5 主成分までの各変数の主成分負荷量を表している。一般的に、主成分分析においては以下の 3 項目のどれかに該当する主成分を解析対象とする。

1. 固有値が 1 以上の成分
2. 寄与率に大幅な変化が見られたところ
3. 累積寄与率が 80%を超えたところ

これらの観点から見ると、図 6.35 から、第 4 主成分までを解析対象とすればよいことが分かる。

しかしながら、図 6.36 から各主成分の持つ特徴を読み取る際に、第 1 主成分および第 2 主成分については解釈が容易であるが、第 3 主成分および第 4 主成分については解釈が困難である。また、寄与率の変化については第 1 主成分と第 2 主成分との間が最も大きいということから、本分析においては第 1 主成分と第 2 主成分を解析対象とすることにする。

さて、ここで各主成分の特徴を見ていくと、図 6.36 に示されている第 1 主成分および第 2 主成分の主成分負荷量から以下のように読み取ることができる。

第 1 主成分

航空戦力の要である空軍作戦機や海軍作戦機、および海上戦力の要である潜水艦や主要水上戦闘艦艇などの変数が大きく正の方向に示されている。これらの兵器はいずれも強力な戦闘力を発揮できる兵器であるため、軍事的能力の総合指標を表す成分であると解釈できる。すなわち、値の大きな国ほど軍事的能力が高い国であることを示している。

第 2 主成分

哨戒艦艇・沿岸警備艦艇や軽戦車、および準軍隊の変数が正の方向に示されており、海軍の武装ヘリコプターや攻撃ヘリコプター、国防支出費などの変数が負の方向に示されている。哨戒艦艇や沿岸警備艦艇、あるいは準軍隊などは正規の戦闘力ではなく、2 次的な軍事力として解釈することができる。また、国防支出費が多いということは高価な装備品を購入することが可能であり、近代的な最新鋭の兵器を装備することが可能である。したがって、第 2 主成分は値の大きな国ほど 2 次的な軍事能力が高く、値が低い国ほど、近代的な戦闘力を保持した国であると解釈できる。

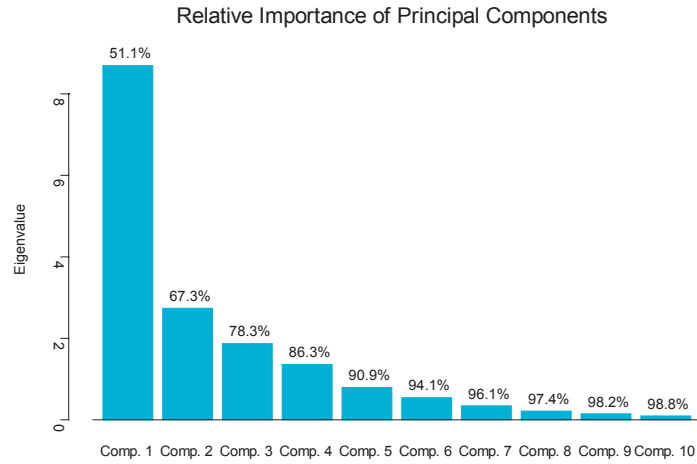


図 6.35: 2003 年の軍事データを用いて主成分分析した結果得られた、第 10 主成分までの固有値と累積寄与率を表したグラフ。各主成分の棒グラフが固有値の値を示しており、棒グラフの上部に記されている数字が累積寄与率を表している。

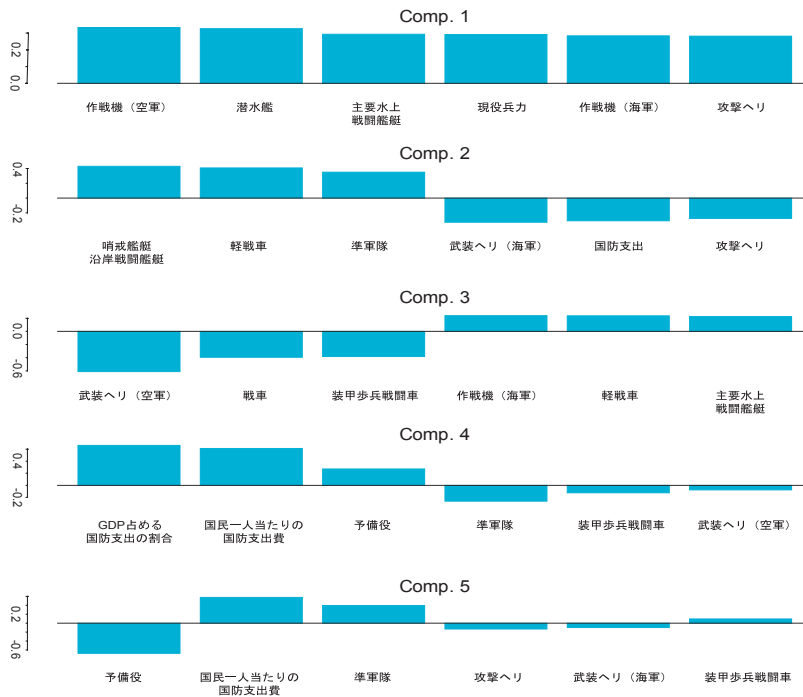


図 6.36: 2003 年の軍事データを用いて主成分分析した結果得られた、第 5 主成分までの主成分負荷量を表したグラフ。各主成分の主成分負荷量については値（絶対値）の大きいものから 6 つの変数が示されている。

図 6.37 は第 1 主成分と第 2 主成分の主成分得点の分布を表しており、各国の国名が記されている。先程の各主成分の分析から判断すると、アメリカ (U.S.A) やロシア (Russia)、中国 (China) は第 1 主成分の値が他の国に比べて大きい位置に分布しており、軍事的能力が高い国であると解釈できる。これは、現在アメリカ・ロシア・中国が世界の軍事 3 大国と呼ばれていることから、妥当性のある解釈であることが分かる。

また、第 2 主成分に注目してみると、アメリカは値が小さい位置に分布しており、近代的な装備品を保有した国であることが分かる。反対に、中国や北朝鮮 (North Korea) は値が大きい位置に分布しており、2 次的な軍事能力、すなわち正規の軍事力ではないが、いざという時には軍事力となり得る力を保有している国であることが分かる。

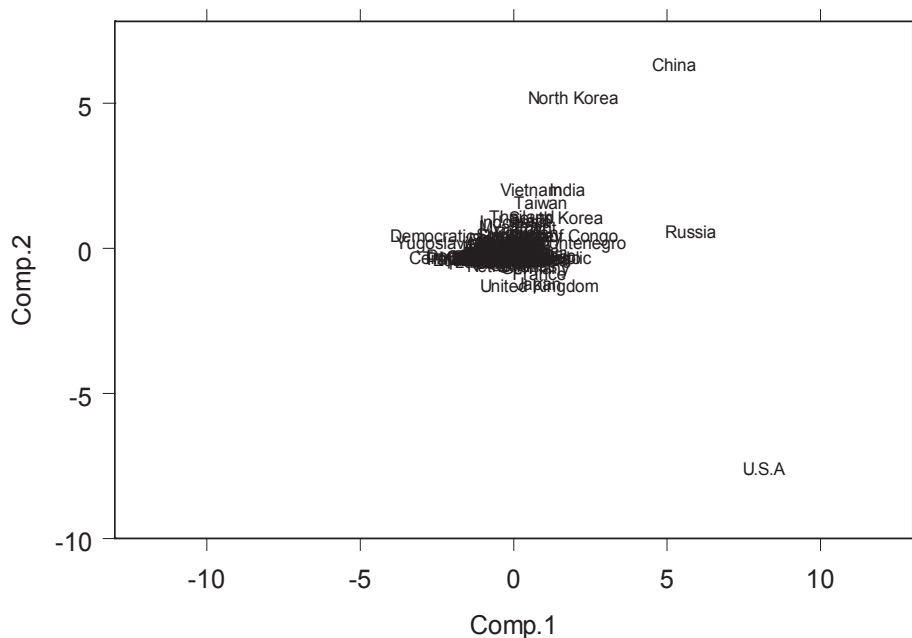


図 6.37: 2003 年の軍事データを用いて主成分分析した結果得られた、第 1 主成分と第 2 主成分の主成分得点の分布を表した図

ここで、主成分分析により得られた結果とクラスタリングにより得られた結果を融合してみる。図 6.38 および図 6.39 は、主成分分析の結果とクラスタリング結果を一つにまとめたものである。すなわち、国名がクラスタリング結果により得られたクラスターごとに色分けされており、赤に近づくほど軍事的能力が高い国のクラスターに分類された国、逆に青に近づくほど軍事的能力が低い国のクラスターに分類された国であることを示している。

この結果から判断すると、赤に近づくほど軍事的能力が高い国、すなわち暖色系の色で示された国は軍事的能力が高い国であり、これらの国は第 1 主成分の値が大きいほど多く分布

している。反対に、青に近づくほど軍事的能力が低い国、すなわち寒色系の色で示された国は軍事的能力が低い国であり、これらの国は第 1 主成分の値が小さいほど多く分布している。

これらの結果はいずれも主成分分析で得られた結果、すなわち第 1 主成分の値の大きさは各国の軍事的能力の大きさを表したものであるとの解釈と一致しており、本分析で行ったクラスタリング結果は妥当性があるものと評価できる。

6.6.2 核兵器および弾道ミサイルの拡散・保有状況の変化による妥当性評価

ここでは、現在も世界で問題となっている核兵器、核開発問題に注目して、今回行ったクラスタリング結果の妥当性を評価していく。1985 年においては、国連安全保障理事会の常任理事国であり、NPT⁹ 上の核保有国でもあるアメリカ、ロシア、イギリス、フランス、中国、1974 年に核実験を行い核保有を宣言したインド、公式には認められていないが保有の可能性が極めて高いとされているイスラエルの合計 7ヶ国が核兵器を保有していた。これらの国々は 1985 年のクラスタリング結果において、軍事的能力が高い国のクラスターである Cluster 1 および Cluster 2 に分類されている。

一方、1998 年 5 月に核実験を行い事実上核保有を宣言したパキスタン、および現在核開発の疑惑が指摘されているイランや北朝鮮に注目してみると、1985 年におけるクラスタリング結果では、北朝鮮が Cluster 2、イラン、パキスタンが Cluster 3 に分類されていたのに対し、2003 年におけるクラスタリング結果では、Cluster 1 にパキスタン、北朝鮮、Cluster 2 にイランが分類される結果となり、いずれの国も軍事的能力の高い方のクラスターに移動していることが確認できる（表 6.8 参照）

また、東アジア地域についてみると、ここでは核兵器と同等の問題になっている弾道ミサイルの拡散状況に注目してみる。すると、1985 年には中国のみしか保有していなかったのが、2003 年には中国を含む、韓国、北朝鮮、台湾、ベトナムが保有するようになった [11]。これらの国々についても、1985 年のクラスタリング結果よりも 2003 年の結果の方が軍事的能力が高い方のクラスターに分類されていることが確認できる。

以上の分析から判断すると、今回のクラスタリング結果、および結果の解釈については妥当性があるものと評価できる。

⁹核拡散防止条約 (NPT : Nuclear Non-Proliferation Treaty) は、1967 年 1 月 1 日前に核兵器その他の核爆発装置を製造し、かつ爆発させた国を核保有国として定義している。

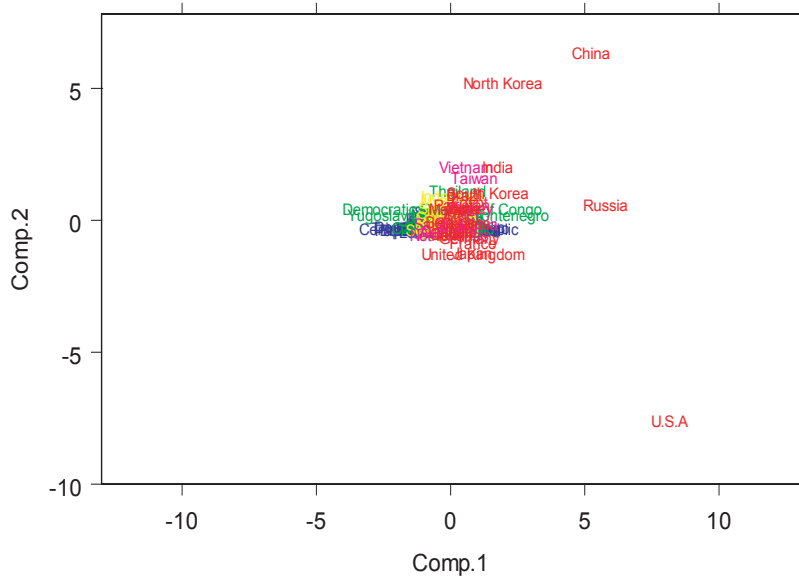


図 6.38: 2003 年の軍事データを用いて主成分分析した結果とクラスタリングによる結果を一つにまとめた図．国名をクラスタリング結果により得られたクラスターごとに色分けをして表している．すなわち，赤に近づくほど軍事的能力が高い国のクラスターに分類された国であり，逆に青に近づくほど軍事的能力が低い国のクラスターに分類された国である．

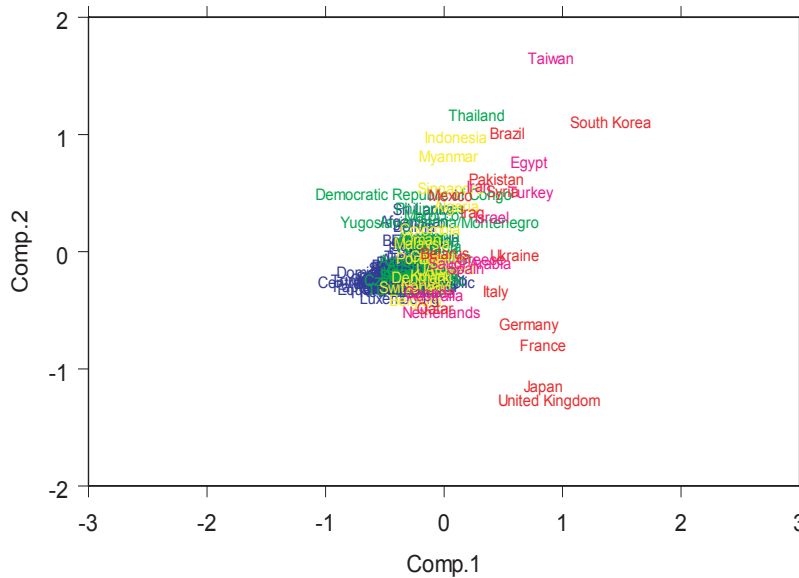


図 6.39: 図 6.38 の拡大図．第 1 主成分の主成分得点の幅を-3 から 3，第 2 主成分の主成分得点の幅を-2 から 2 として表した図

表 6.8: 現在の世界の核兵器保有状況とクラスタリング結果の比較．太字の国は 1985 年には核兵器を保有していなかった国であることを表している．

	国 名	クラスター番号	
		1985 年	2003 年
NPT 上の核保有国	アメリカ	1	1
	ロシア	1	1
	イギリス	1	1
	フランス	2	1
	中 国	1	1
核保有宣言国	インド	1	1
	パキスタン	3	1
保有の可能性が極めて高い国	イスラエル	2	2
核開発疑惑国	イラン	3	2
	北朝鮮	2	1

6.7 まとめ

我が国を取り巻く戦略環境と地域情勢を正しく理解するために、常に変転する全世界の軍事情勢の流れを把握することは必要不可欠であると考えられる．そのため、本章では日本を取り巻く軍事情勢の動向を把握するためにクラスタリングを用いた分析を行った．

その結果、各地域の軍事情勢の動向や各地域ごとの特性などをクラスタリング結果から読みとることができ、特に、東アジア地域においては、最終的には日本に対する脅威のある国として3ヶ国まで絞ることができ、クラスタリングによる解析で必要な情報を効率よく取り出すことができるということが確認できた．また、今回のように変数の値の大きさに偏りがあるようなデータを取り扱う場合には、内積に基づく類似度を利用することで全ての変数を考慮した分類結果が得られるということが分かった．また、この類似度を用いた場合、各変数間の比重をもとにデータがクラスタリングされるので、データの質的な部分の特徴を抽出するのに利用できると思われる．さらに、ガウシアンカーネルの分散を調整するパラメータの値を小さくしていくと、データの量的な部分に関してのクラスタリングが可能となることも分かった．

日本を取り巻く軍事的リスクに関しては、ある国の軍事能力を見るとき、よく国防費などを用いて評価されるが、国防費だけでその国の軍事力を判断することは北朝鮮のような危険な国を見逃すことになり、軍事的リスクを高める要因となり得ることが分かった．また、我が

国に脅威を及ぼす国はクラスタリングにより中国、韓国、北朝鮮の 3ヶ国に絞ることができ、この中でも特に中国、北朝鮮に対する警戒を強化しなければならないことが明らかになった。韓国については、この国の国防政策から軍事力の増加は北朝鮮を意識したものであり、日本を対象にしたものではないため、韓国に対する警戒はさほど必要ないと考えられる。

一方、我が国に目を向けてみると、2003 年のクラスタリング結果においては Cluster 1 に分類されており、世界的に見ても軍事的能力の高い国として見ることができる。データで見ると、現役兵力や戦車の数などは他国に比べて少ないのだが、国防支出費が非常に多く、また他の国において比較的装備されていない潜水艦や主要水上戦闘艦艇の数においても世界水準を大幅に上回る数を保有しているため、このような結果になったものと考えられる。特に、2003 年における国防支出費には、約 370 億ドルと世界で 5 番目に多い額が使われている。この背景には、日本がアメリカなどの先進国と同様に質の高い装備品を保有している、あるいは開発を行っているということがあり、日本が Cluster 1 に分類された要因の一つであると考えられる。

また、本分析で行ったクラスタリング結果の妥当性を、主成分分析と核兵器および弾道ミサイルの拡散・保有状況の変化から評価し、いずれの場合においても本研究で提案したクラスタリング結果は妥当性があるものと評価できた。

主成分分析やクラスタリングは、複数の変数に関するデータをもとにこれらの変数間の相互関連を分析する多変量解析と呼ばれる手法の一つである。各手法の持つ特徴はそれぞれ異なり、通常はデータ解析の目的・用途に応じて使い分けられる。すなわち、主成分分析は、多くの量的説明変数がある場合に、これを少数の総合指標で表わす目的で用いられる。一方、クラスタリングは、解析対象となるデータの中から互いに似たものを集めてクラスターを作り、データを分類するという目的で用いられる。したがって、主成分分析はデータの持つ特徴を容易に把握することが可能であるが、データを分類するには適していない。また、クラスタリングはデータを分類する目的で用いられるため、この手法のみではそのデータの持つ特徴を把握することは困難である。このため、それぞれの手法の持つ弱点を補うために、各手法を組み合わせることで分析を行うということが考えられる。

本章の最後では、クラスタリング結果の妥当性を評価する目的で主成分分析を行った。すなわち、本研究で提案したクラスタリングアルゴリズムにより得られた各クラスターがどのような特徴を持つのかを、主成分分析により確かめた。その結果、今回分析に用いたデータの持つ特性とクラスタリングにより得られた結果の特徴が一致していることが確認できた。以上のことから、クラスタリングは主成分分析あるいは他の多変量解析の方法などと組み合わせることによって、データマイニングの目的である「発見的な知識獲得」に大きく寄与できるものであると考えられる。

ここで、我が国における国防政策について簡単に触れておくと、2004年12月に策定された新防衛大綱¹⁰において、前述した脅威に対処するための方針が掲げられている。新防衛大綱によると我が国の安全保障の基本方針としては、

1. 我が国に直接脅威が及ぶことを防止し、脅威が及んだ場合にはこれを排除するとともに、その被害を最小化する
2. 国際的な安全保障環境を改善し、我が国に脅威が及ばないようにする

という2つの目標が掲げられており、また、これらの目標を達成するための防衛力の役割として

1. 新たな脅威や多様な事態への実効的な対応
2. 本格的な侵略事態への対処への備え
3. 国際的な安全保障環境の改善のための主体的・積極的な取組

の3つが掲げられている。この中でも特に「新たな脅威や多様な事態への実効的な対応」を重要視しており、防衛力の第一の役割として位置づけている。ここで示されている新たな脅威には、

- 弾道ミサイル攻撃への対応
- ゲリラや特殊部隊による攻撃等への対応
- 島嶼部に対する侵略への対応
- 周辺海空域の警戒監視及び領空侵犯対処や武装工作船等への対応
- 大規模・特殊災害等への対応

といった5項目が挙げられており、これらはいずれも中国や北朝鮮を意識したものであると判断することができる。このことは、今回クラスタリングにより東アジア地域を分析した結果、すなわち我が国にとって脅威を及ぼすであろう国は中国、北朝鮮であるという結果と一致しており、ここでも本研究の妥当性が確認できる。

今後はより詳細な分析を行うために、各国の軍事データだけでなく、その国の経済状況をも考慮に入れた分析をしていく必要があると考えられる。また、軍事データにおいても、ただ単に兵器の数だけを用いて比較するのではなく、近年の科学技術の発展に伴う装備のハイテク化により、各種兵器の性能も考慮に入れた分析、あるいは、今回はクラスタリングの妥当性をみるために用いた核戦力や弾道ミサイルなどの要素も考慮に入れたより広範囲かつ詳細な分析を行っていく必要があると考えられる。

¹⁰平成17年度以降に係る防衛計画の大綱について [12]

第7章 結論

本研究では、内積に基づく類似度を利用したクラスタリングアルゴリズムにおいて、近年注目が集められているカーネル法と呼ばれるテクニックを利用することで、非線形分離可能なクラスタリングアルゴリズムへの拡張を行った。これにより、高次元特徴空間への写像を利用したクラスタリングが可能となり、内積に基づく類似度を利用した場合においても従来の手法では分類不可能なデータを分類することが可能となった。類似度として、 c -平均法を中心としたバッチ型のクラスタリングアルゴリズムにおいてはコサイン相関を利用し、競合学習モデルに基づくオンライン型のクラスタリングアルゴリズムにおいては内積を用いた。

また、内積に基づく類似度を利用したクラスタリングアルゴリズムの応用研究として、ファジィマルチ集合を利用した文書情報クラスタリングの方法について示し、非線形化されたクラスタリングアルゴリズムの有効性について検証を行った。さらに、日本周辺の軍事情勢の動向の変化をクラスタリングを用いることにより分析を行い、日本に対する軍事的リスクをクラスタリング結果と各国の国防政策との両面から検証を行うことで、非線形化されたクラスタリングアルゴリズムの妥当性について評価を行った。

7.1 まとめ

第1章では、本研究の背景として、クラスター分析の持つ役割について触れ、内積に基づく類似度を利用したクラスタリング手法の位置づけについて示した。また、従来の研究における問題点について指摘し、本研究の意義、目的について簡潔に述べた。

第2章では、まず、非階層的クラスタリングの代表的な手法である c -平均クラスタリングを行うための類似度としてコサイン相関を定義し、その性質について示した。次に、コサイン相関を類似度とした場合のクリスプ c -平均法およびクリスプ c -平均法をファジィ化したファジィ c -平均法について示した。ファジィ c -平均法においては、Bezdek らによって提案された標準的ファジィ c -平均法と、宮本らによって提案されたエントロピー正則化ファジィ c -平均法について紹介し、コサイン相関を用いた場合の目的関数やアルゴリズムの変化について示した。このとき、ファジィクラスタリングの特徴でもあるファジィ分類関数について述べ、両手法の持つ性質の違いについて示した。さらに、エントロピー正則化ファジィ c -平均法の拡張手法として知られている、クラスターサイズ調整変数を導入したアルゴリズムと、この手法

をクリスプ化する方法について示し、クラスター領域が異なる場合においても適切な分類境界が得られることを人工データを使用した数値例を用いて示した。

第 3 章では、クラスタリングアルゴリズムをカーネル法を利用することで非線形化する方法について述べた。従来の c -平均法やファジィ c -平均法が持つ問題点として、非線形な分類境界を持つようなデータに対しては最適な分類を行うことができず、特に、コサイン相関を類似度とした場合、同一直線上の個体はすべて同じであるとみなされてしまうため、線形な分類境界を持つようなデータでさえ最適な分類を行うことができないということが挙げられる。なぜならば、分類境界が各クラスター中心を結ぶ垂直 2 等分線を形成するように求められ、また、内積を類似度とした場合には、分類境界が各クラスター中心間の角度を 2 等分するように形成されるという特徴を持っているからである。

そこで本章では、カーネル法と呼ばれる線形モデルで表される手法を容易に非線形に拡張できる手法を、エントロピー正則化ファジィ c -平均法に基づくクラスタリングアルゴリズムに利用することでアルゴリズムを非線形化し、上記の問題を解決する方法について述べ、提案手法の効果、有効性を人工データによる数値実験により視覚的に示した。また、アルゴリズムをファジィ化することによりカーネル法を利用した際のファジィ分類関数の式を導出し、従来の手法とは全く異なった分類規則が存在することを数値例により示した。

しかしながら、カーネル関数を計算する際のパラメータについては、本研究では実験的・経験的にしか得ることができておらず、厳密な意味でクラスタリングを行っているとは言えない。よって、今後はカーネル関数の最適なパラメータの決定をも考慮したアルゴリズムの開発が必要となる。

第 4 章では、 c -平均法とは異なるアプローチである競合学習という概念からクラスタリングを考察する方法について示した。競合学習モデルに基づくクラスタリングアルゴリズムはオンライン型のクラスタリング手法とも呼ばれ、大量なデータを取り扱う際に有効な手法であることが知られている。この手法は、 c -平均法のような目的関数の最適化を行うのではなく、入力データとクラスター中心との位置関係だけでクラスタリングが行われるのが特徴である。この考えを利用することにより、カーネル法を利用することによるクラスタリングアルゴリズムの非線形化の際の計算量の増加を、内積を利用したクラスタリングアルゴリズムにおいても防げることを示した。

また、第 3 章で用いたものよりもより複雑な形状をした人工データを用いた数値実験、および実データとして統計学の世界でよく用いられている Iris データを用いた数値実験により、競合学習モデルの有効性を示した。さらに、井口らによって提案された自己組織化マップのアルゴリズムをカーネル化することで、高次元特徴空間上に写像されたデータの位置関係の視覚化を、内積を利用した場合でも可能であることを示した。

しかしながら、競合学習モデルを用いた場合においても、カーネル法を利用する場合にはカーネル関数のパラメータの決定方法が重要であると言える。よって、ここでもカーネル関

数のパラメータをも同時に学習させていけるようなアルゴリズムの開発が必要である。

第 5 章では、情報検索におけるクラスター型検索モデルへの応用として、ファジィマルチ集合という新たな数理モデルを用いる方法について述べた。ファジィマルチ集合の特徴は要素の数と帰属度を同時に表現する構造を持つため、同一の用語が異なる度合いを持って複数回出現しているような Web 上の情報あるいは文献情報などを適切に表現することが可能であると考えられている。

本章では、このような特徴を持つファジィマルチ集合をモデルとした文書情報クラスタリングの方法について述べ、実際の文書情報を用いた実験により、従来のクラスタリング手法との分類性能を比較することで非線形化されたクラスタリングアルゴリズムの有効性を示した。また、リスク解析における情報クラスタリングの位置づけについても述べ、リスクに関する文献情報をクラスタリングすることで、提案手法の妥当性の評価も行った。

第 6 章では、リスク解析におけるクラスタリングの利用方法の一例として、クラスタリングによる日本周辺の軍事情勢の分析を行う方法について述べた。各国の軍事データを使用してクラスタリングを行い、世界の国々をクラスタリングによりグループ化し、そこから得られる情報を利用して我が国を取り巻く軍事的リスク（脅威）の分析を行った。

その結果、クラスタリングにより日本に脅威のある国を 3ヶ国まで絞ることができ、クラスタリングによる解析で必要な情報を効率よく取り出すことができるということが確認できた。また、主成分分析および世界の核保有状況の変化あるいは大量破壊兵器の拡散状況から、本研究で提案したクラスタリングアルゴリズムの結果は妥当性があることを示した。さらに、クラスタリングは主成分分析あるいは他の多変量解析の方法などと組み合わせることによって、データマイニングの目的である「発見的な知識獲得」に大きく寄与できるものであることを示した。

以上のことから、本研究の主な成果をまとめると以下のようなになる。

- I. 内積に基づく類似度を利用したクラスタリング手法は、距離を用いる場合とは異なる性質を持つことについて明らかにし、データクラスタリングにおける現状の問題点について指摘した。
- II. 内積に基づく類似度を用いた c -平均法およびその拡張手法、あるいは競合学習モデルに基づく従来のクラスタリング手法にカーネル法を適用することで、非線形分離可能なアルゴリズムに拡張可能である。
- III. 内積を利用した類似度とカーネル法（ガウシアンカーネル）を組み合わせることによって、ユークリッド距離では非線形分離できなかった複雑なデータについても分類可能なアルゴリズムを開発した。

- IV. 競合学習モデルを用いたクラスタリングは、カーネル法を用いても計算量が大幅に増加せず非線形分離が行えるため、従来の手法に比べて優れた性能を持つアルゴリズムである。
- V. ファジィマルチ集合は、情報検索における新たな数理モデルとして適用可能であり、文書情報あるいは Web 情報を適切に表現することが可能である。
- VI. ベクトル間の角度を類似度としたクラスタリングアルゴリズムは、その性質から情報検索における文書分類等に用いることができ、非線形化されたアルゴリズムの分類性能は従来の手法よりも優れている。
- VII. クラスタリングはリスク解析において情報を集約するためのデータマイニングのツールとして利用可能であり、また、他の多変量解析の方法と組み合わせることにより、そのデータが持つ特徴を容易に、かつ詳細に把握することが可能である。

7.2 今後の課題と展望

本研究では、データクラスタリングの研究においてはあまり検討されていない内積に基づく類似度に焦点を当ててきた。カーネル法を利用することにより、ユークリッド距離を用いる場合よりも分類性能において優れている場合があることが確認されたが、残された課題は多い。以下では本研究で得られた主な諸課題についてまとめる。

- 実データに基づく検証と評価

本研究では、提案手法と従来の手法との比較を容易にするため、数値例においては人工データを多く用いて実験を行った。また、実データとしてはパターン認識の世界においてよく用いられている Iris データを用いて実験を行った。しかしながら、実世界において本提案手法がどれくらいの性能をもっているのかを検証するためには、実データを用いたより多くの数値実験により評価していく必要がある。

- パラメータの決定方法

本研究で用いたカーネル関数のパラメータの決定方法については未だに決定的な方法が確立しておらず、経験的あるいは実験的に決定するのみである。カーネル関数のパラメータは非常にセンシティブであることが知られており、特にデータの形が分からないような実データに応用する際にはその決定方法が重要となるため、最適なパラメータをも同時に決められるようなアルゴリズムを開発していく必要がある。

- 情報検索システムの構築

本研究では、新たな情報検索の理論として、ファジィマルチ集合を用いた文書分類の方

法を提案したが、文書情報だけでなく Web 上の情報に対しての本理論の適用性を把握するために、本研究で提案した理論（モデル）を用いた情報検索システムを構築し、実在する Web 情報に対してどのくらいの性能を発揮できるのかを検証していくことも重要であると考えられる。

- クラスタリングによる軍事情勢分析

本研究では、軍事力の大きさの変化、つまり量的な変化によって分析を行ったが、より詳細な分析を行うためには、その国の経済状況をも考慮に入れた分析をしていく必要があると考えられる。また、軍事データにおいてもただ単に兵器の数だけを用いて比較するのではなく、近年の科学技術の発展に伴う装備のハイテク化により各種兵器の性能も考慮に入れた分析、つまり質的な変化についても見ていく必要がある。

内積に基づく類似度を使用したクラスタリングアルゴリズムはあまり研究がされていない領域であり、多くの研究の余地があると言えよう。特に、非線形分離可能となったクラスタリングアルゴリズムは、実世界におけるデータ解析においても分類性能を向上させるものであり、理論的にも実用的にも発展の余地が大きく、今後のさらなる研究が望まれる。

謝辞

本論文は、著者が筑波大学大学院システム情報工学研究科在学中に行った研究成果をまとめたものです。この間、ご指導・ご助言ならびにご支援を頂いた多くの方々に心から感謝の意を表します。

筑波大学大学院 博士課程 システム情報工学研究科リスク工学専攻に在籍してから4年間という長きにわたり、ご指導を賜りました筑波大学大学院 システム情報工学研究科 宮本定明 教授に深く感謝致します。クラスタリングの知識など皆無であった私を、温かく、時には厳しくご指導して頂きました。また、研究の方向性から細部に至る多くの的確なアドバイスは、研究者としての論理的なものの考え方を学ぶのに非常に有益なものとなりました。そして、常にリスク工学的な視点に立ち、様々な観点からリスクを解析するための方法論を確立しようとする先生のお姿を見て、非常に多くのことを学ばせて頂きました。

また、システム情報工学研究科の稲垣敏之 教授、鬼沢武久 教授からの論文審査での厳しいご講評やご指摘は、本論文を作成するにあたり、非常に有益なものとなりました。さらに、システム情報工学研究科 イリチュ(佐藤)美佳 助教授、遠藤靖典 助教授には、ソフトウェア基礎グループでの研究発表会、国際会議への参加から論文審査に至るまで、あらゆる機会を通してご指導して頂きました。ここに感謝の意を表します。

ソフトウェア基礎グループの仲間たちには、楽しく快適な研究環境を提供して頂きました。特に、宮本研究室の井口亮 氏にはゼミなどを通じてたくさんの貴重なご意見を頂きましたし、Arnold C. Alanzado 氏には国際会議への参加の際に何度も英語の添削をして頂きました。また、研修期間中、共に学んだ工藤功介 氏には、同期として多大なる協力をして頂き、また緩みがちな私の気持ちに活を入れてくれる存在でもありました。ここに感謝の意を表します。

さらに、本研修を陰で支えて頂きました、防衛庁陸上幕僚監部装備部開発課の皆様、霞ヶ浦駐屯地業務隊の皆様に深く感謝の意を表します。また、筑波大学でこのようなすばらしい研修を行う機会を与えて下さいました、防衛大学校 渡邊啓二 教授、筑波大学大学院 システム情報工学研究科 水谷孝一 教授に深く感謝致します。

このように、本論文は名前を書き上げられないくらい多くの方々のご指導・ご協力により完成したものです。ここに深く感謝の意を表します。

最後になりましたが、4年間の研究生活を陰ながら支え続け、いつも温かく見守ってくれた家族に心から感謝致します。

参考文献

- [1] P. Adriaans, D. Zantinge, *Data Mining*, Addison-Wesley Longman Publishing, 1996.
- [2] M.R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, 1973.
- [3] 麻生 英樹, 津田 宏治, 村田 昇, *パターン認識と学習の統計学*, 岩波書店, 2003.
- [4] T. Bedford, R. Cooke, *Probabilistic Risk Analysis: Foundations and Methods*, Cambridge University Press, 2001.
- [5] A. Ben-Hur, D. Horn, H.T. Siegelmann, V.N. Vapnik, Support vector clustering, *Journal of Machine Learning Research*, Vol.2, pp. 125–137, 2001.
- [6] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York, 1981.
- [7] J.C. Bezdek, J. Keller, R. Krisnapuram, N.R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Kluwer Academic Publishers, Boston, 1999.
- [8] J.P. Bigus, *Data Mining with Neural Networks*, McGraw-Hill, 1996.
- [9] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, T. Vetter, Comparison of view-based object recognition algorithms using realistic 3D models, C. von der Malsburg *et al.*, (eds.), *Artificial Neural Networks, ICANN 1996*, LNCS 1112, Springer Verlag, pp. 251–256, 1996.
- [10] W. Blizard, Real-valued multisets and fuzzy sets, *Fuzzy Sets and Systems*, Vol.33, pp. 77–97, 1989.
- [11] 防衛庁, *日本の防衛 — 防衛白書 (平成 17 年版)*, 2005 .
- [12] 防衛庁ホームページ, “平成 17 年度以降に係る防衛計画の大綱について”, URL address: <http://www.jda.go.jp/>

-
- [13] C. Burges, B. Schölkopf, Improving the accuracy and speed of support vector learning machines, M. Mozer *et al.*, (eds.), *Advances in Neural Information Processing Systems 9*, MIT Press, pp. 375–381, 1997.
- [14] C. Burges, A Tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, Vol.2, pp. 121–167, 1998.
- [15] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, A. Zanasi, *Discovering Data Mining*, Prentice Hall PTR, New Jersey, 1998.
- [16] Clusty, URL address: <http://clusty.com/>
- [17] R.N. Dave, R. Krishnapuram, Robust clustering methods: a unified view, *IEEE Trans. Fuzzy. Syst.*, Vol.5, No.2, pp. 270–293, 1997.
- [18] D. Decoste, B. Schölkopf, Training invariant support vector machines, *Machine Learning*, Vol.46, No.4, pp. 161–190, 2002.
- [19] H. Drucker, D. Wu, V. Vapnik, Support vector machines for span categorization, *IEEE Trans. on Neural Networks*, Vol.10, No.5, pp. 1048–1054, 1999.
- [20] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification: Second Edition*, Wiley Interscience, 2001.
- [21] S. Dumais, J. Platt, D. Heckerman, M. Sahami, Inductive learning algorithms and representations for text categorization, *Proc. of the 7th International Conference on Information and Knowledge Management*, 1998.
- [22] D. Dumitrescu, B. Lazzerini, L.C. Jain, *Fuzzy Sets and Their Application to Clustering and Training*, CRC Press, 2000.
- [23] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics*, Vol.3, pp. 32–57, 1974.
- [24] J.C. Dunn, Well-separated clusters and optimal fuzzy partitions, *Journal of Cybernetics*, Vol.4, pp. 95–104, 1974.
- [25] B.S. Everitt, A.J. Gourlay, R.E. Kendell, An attempt at validation of traditional psychiatric syndromes by cluster analysis, *The British Journal of Psychiatry*, Vol.119, pp. 399–412, 1971.

-
- [26] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, Knowledge discovery and data mining: towards a unifying framework, *Proc. of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, Portland, Oregon, Aug. 2–4, 1996.
- [27] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, *Advanced in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [28] E.W. Forgy, Cluster analysis of multivariate data: efficiency vs. interpretability of classifications, *Biometrics*, Vol. 21, pp. 768-769 (Abstract), 1965.
- [29] R.E. Frank, P.E. Green, Numerical taxonomy in marketing analysis: a review article, *Journal of Marketing Research*, Vol.5, pp. 83-98, 1968.
- [30] 福田 剛志 , 森本 康彦 , 徳山 豪 , データマイニング , 共立出版 , 2001.
- [31] I. Gath, A.B. Geva, Unsupervised optimal fuzzy clustering, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.11, No.7, pp. 773–781, 1989.
- [32] M. Girolami, Mercer kernel-based clustering in feature space, *IEEE Trans. on Neural Networks*, Vol.13, No.2, 2002.
- [33] P.E. Green, R.E. Frank, P.J. Robinson, Cluster analysis in test market selection, *Management Science*, Vol.13, No.8, pp. B387-B400, 1967.
- [34] J.A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, New York, 1975.
- [35] 春山 秀行 , 遠藤 靖典 , 大久保 貴義 , カーネル関数を用いた階層的クラスタリング , 日本知能情報ファジィ学会誌 , Vol.17, No.4, pp. 459–467, 2005.
- [36] R.J. Hathaway, J.C. Bezdek, W. Pedrycz, A parametric model for fusing heterogeneous fuzzy data, *IEEE Transactions on Fuzzy Systems*, Vol.4, No.3, pp. 270-281, 1996.
- [37] D. Haussler, Convolution kernels on discrete structures, Technical Report UCSC-CRL-99-10, UC Santa, Cruz, 1999.
- [38] F.G.J. Hayhoe, D. Quaglino, R. Doll, *The Cytology and Cytochemistry of Acute Leukaemias: a study of 140 cases*, Medical Research Council Special Report Series No.304, Her Majesty's Stationery Office, London, 1964.
- [39] F.R. Hodson, Searching for structure within multivariate archaeological data, *World Archaeology* , Vol.1, pp. 90-105, 1969.

-
- [40] F.R. Hodson, Cluster analysis and archaeology: some new developments and applications, *World Archaeology*, Vol.1, pp. 299-320, 1969.
- [41] F.R. Hodson, D.G. Kendall, P. Tautu (eds.), *Mathematics in the Archaeological and Historical Sciences*, Edinburgh University Press, Edinburgh, 1971.
- [42] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis*, Wiley, Chichester, 1999.
- [43] H. Ichihashi, K. Honda, N. Tani, Gaussian mixture PDF approximation and fuzzy c -means clustering with entropy regularization, *Proc. of the 4th Asian Fuzzy System Symposium*, May 31-Jun. 3, 2000, Tsukuba, Japan, pp. 212-221, 2000.
- [44] H. Ichihashi, K. Miyagishi, K. Honda, Fuzzy c -Means clustering with regularization by K-L information, *Proc. of 10th IEEE International Conference on Fuzzy Systems*, pp. 924-927, 2001.
- [45] H. Ichihashi, K. Honda, Noise fuzzy clustering in kernel feature spaces for robust PCA, *Proc. of SCIS & ISIS 2004*, Sep. 21-24, 2004, Yokohama, Japan, 2004.
- [46] R. Inokuchi, S. Miyamoto, LVQ clustering and SOM using a kernel function, *Proc. of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004)*, July 25-29, 2004, Budapest, Hungary, pp. 182-185, 2004.
- [47] 井口 亮, 宮本 定明, カーネル関数を利用した LVQ クラスタリングと SOM, 日本知能情報ファジィ学会誌, Vol.17, No.1, pp. 88-94, 2005.
- [48] R. Inokuchi, S. Miyamoto, On-line learning approach to kernel fuzzy LVQ, *Proc. of the 6th International Symposium on Advanced Intelligent Systems (ISIS 2005)*, Sep. 25-Oct. 1, 2005, Yeosu, Korea, pp. 484-487, 2005.
- [49] The International Institute for Strategic Studies, *The MILITARY BALANCE 1985 - 1986*, The International Institute for Strategic Studies, 1985.
- [50] The International Institute for Strategic Studies, *The MILITARY BALANCE 2003 - 2004*, Oxford University Press, 2003.
- [51] T. Jaakkola, D. Haussler, Exploiting generative models in discriminative classifiers, *Advances in Neural Information Processing Systems*, Vol.11, pp. 487-493, 1999.
- [52] T. Jaakkola, M. Diekhans, D. Haussler, A discriminative framework for detecting remote protein homologies, *Journal of Computational Biology*, Vol.7, pp. 95-114, 2000.

-
- [53] T. Joachims, Text categorization with support vector machines: learning with many relevant features, *Proc. of the European Conference on Machine Learning*, pp. 137–142, 1998.
- [54] D.M. Kammen, D.M. Hassenzahl, *Should We Risk It ?*, Princeton University Press, 1999.
- [55] 韓国銀行, 北朝鮮の GDP 推定結果, 各年度 .
- [56] 韓国国防省, 2004 年度版国防白書, 2005 .
- [57] N.B. Karayiannis, J.C. Bezdek, An integrated approach to fuzzy learning vector quantization and fuzzy *c*-means clustering, *IEEE Trans. on Fuzzy Systems*, Vol.5, No.4, pp. 622-628, 1997.
- [58] N.B. Karayiannis, Soft learning vector quantization and clustering algorithms based on mean-type aggregation operators, *International Journal of Fuzzy Systems*, Vol.4, No.3, pp. 739–751, 2002.
- [59] 岸田 和明, 情報検索の理論と技術, 勁草書房, 1998.
- [60] 北 研二, 津田 和彦, 獅々堀 正幹, 情報検索アルゴリズム, 共立出版, 2002.
- [61] D.E. Knuth, *The art of Computer Programming, Vol.2 / Seminumerical Algorithms*, Addison Wesley, Reading, Massachusetts, 1969.
- [62] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Heidelberg, 1995.
- [63] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, *IEEE Trans. Fuzzy Systems.*, Vol.1, No.1, pp. 98–110, 1993.
- [64] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Müller, E. Säckinger, P. Simard, V. Vapnik, Learning algorithms for classification: a comparison on handwritten digit recognition, J.H. Oh *et al.*, (eds.), *Neural Networks: The Statistical Mechanics Perspective*, World Scientific Publishing, pp. 261–276, 1998.
- [65] Z.-Q. Liu, S. Miyamoto (eds.), *Soft Computing and Human-Centered Machines*, Springer, Tokyo, 2000.
- [66] J.B. MacQueen, Some methods of classification and analysis of multivariate observations, *Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.*, pp. 281-297, 1967.

- [67] 前田 英作, 村瀬 洋, カーネル非線形部分空間方によるパターン認識, 信学論, J82-D-II(4), pp. 600–612, 1999.
- [68] Z. Manna, R. Waldinger, *The Logical Basis for Computer Programming, Vol. 1: Deductive Reasoning*, Addison-Wesley, Reading, Massachusetts, 1985.
- [69] D. Mattera, S. Haykin, Support vector meshines for dynamic reconstruction of a chaotic system, B. Schölkopf *et al.*, (eds.), *Advanced in Kernel Methods — Support Vector Learning*, MIT Press, pp. 211–242, 1999.
- [70] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.-R. Müller, Fisher discriminant analysis with kernels, Y.-H. Hu *et al.*, (eds.), *Neural Network for Signal Processing IX*, IEEE, pp. 41–48, 1999.
- [71] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, Dordrecht, 1990.
- [72] 宮本 定明, ファジィマルチ集合の基本演算について, 日本ファジィ学会誌, Vol.8, No.4, pp. 639–645, 1996.
- [73] S. Miyamoto, Fuzzy multisets with infinite collections of memberships, *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA '97)*, June 25–29, 1997, Prague, Czech Republic, Vol.1, pp. 61–66, 1997.
- [74] S. Miyamoto, M. Mukaidono, Fuzzy c -means as a regulation and maximum entropy approach, *Proc. of 7th International Fuzzy Systems Association World Congress (IFSA '97)*, Vol.2, pp. 86–92, 1997.
- [75] 宮本 定明, 馬屋原 一孝, 向殿 政男, ファジィ c -平均法とエントロピ - 正則化法におけるファジィ分類関数, 日本ファジィ学会誌, Vol.10, No.3, pp. 548–557, 1998.
- [76] 宮本 定明, クラスタ分析入門, 森北出版, 1999.
- [77] 宮本 定明, 馬屋原 一孝, クラスタサイズの調整機能をもつファジィ c -平均法, 第 16 回ファジィシステムシンポジウム講演論文集, pp. 537–538, 2000.
- [78] S. Miyamoto, Fuzzy multiset and their generalizations, in C.S. Calude *et al.*, eds., *Multiset Processing*, Lecture Notes in Computer Science, LNCS 2235, Springer, Berlin, pp. 225–235, 2001.
- [79] S. Miyamoto, Fuzzy multisets and fuzzy clustering of documents, *Proc. of FUZZ-IEEE 2001*, Dec. 2–5, 2001, Melbourne, Australia, pp. 1539–1542, 2001.

-
- [80] S. Miyamoto, Information clustering based on fuzzy multisets, *Information Processing and Management*, Vol.39, pp. 195–213, 2003.
- [81] S. Miyamoto, Y. Nakayama, Algorithms of hard c -means clustering using kernel functions in support vector machines, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.7, No.1, pp. 19–24, 2003.
- [82] S. Miyamoto, D. Suizu, Fuzzy c -means clustering using kernel functions in support vector machines, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.7, No.1, pp. 25–30, 2003.
- [83] S. Miyamoto, D. Suizu, O. Takata, Methods of fuzzy c -means and possibilistic clustering using a quadratic term, *International Journal of Approximate Reasoning*, Vol.34, pp. 181–199, 2003.
- [84] S. Miyamoto, Multisets and fuzzy multisets as a framework of information systems, *Proc. of Modeling Decisions for Artificial Intelligence (MDAI 2004)*, Aug. 2–4, 2004, Barcelona, Spain, LNAI 3131, pp. 27–40, 2004.
- [85] S. Mukherjee, E. Osuna, F. Girosi, Nonlinear prediction of chaotic time series using a support vector machine, J. Principe *et al.*, (eds.), *Neural Networks for Signal Processing VII – Proc. of the 1997 IEEE Workshop*, IEEE, pp. 511–520, 1997.
- [86] K.-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machines, W. Gerstner *et al.*, (eds.), *Artificial Neural Networks – ICANN 1997*, LNCS 1327, Springer Verlag, pp. 999–1004, 1997.
- [87] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms, *IEEE Transactions on Neural Networks*, Vol.12, No.4, pp. 181–201, 2001.
- [88] H. Murata, T. Onoda, Applying kernel based subspace classification to a non-intrusive monitoring of household electric appliances, *Artificial Neural Networks – ICANN 2001*, G. Dorffner *et al.*, (eds.), LNCS 2130, Springer Verlag, pp. 692–698, 2001.
- [89] 日本リスク研究学会, リスク学事典, TBS ブリタニカ, 2002.
- [90] 西田 春彦, 吉田 光雄, 平松 闊, 田中 邦夫, クラスタ分析, マイクロソフトウェア, 1983.
- [91] 西田 英郎, 佐藤 嗣二, 实例クラスタ分析, 内田老鶴圃, 1992.

-
- [92] 西田 英郎, クラスタ分析とその応用, 内田老鶴圃, 1998 .
- [93] N.R. Pal, J.C. Bezdek, R.J. Hathaway, Sequential competitive learning and the fuzzy c -means clustering algorithm, *Neural Networks*, Vol.9, No.5, pp. 787–796, 1996.
- [94] C.J.van Rijsbergen, *Information Retrieval: Second edition*, Butter Worths, London, 1979.
- [95] D. Roobaert, M.V. Hulle, View-based 3D object recognition with support vector machines, *Proc. of the IEEE International Workshop on 1999*, pp. 77-84, 1999.
- [96] G. Salton, *Automatic Information Organization and Retrieval*, McGraw-Hill, 1968.
- [97] G. Salton, M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.
- [98] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, Vol.10, pp. 1299–1319, 1998.
- [99] B. Schölkopf, A. Smola, *Learning with Kernels*, MIT Press, 2002.
- [100] 渋谷 和宏, 宮本 定明, 高田 治, 馬屋原 一孝, ファジィ c -平均クラスタリングにおける制約条件と正則化および可能性クラスタリング, 日本ファジィ学会誌, Vol.13, No.6, pp. 707–715, 2001.
- [101] (財)史料調査会, 2002年版世界軍事情勢, 原書房, 2002 .
- [102] “Society for risk analysis annual meeting 2002 preliminary programme”, URL address: http://www.sraeurope.org/pdfdocs/final_programme.pdf
- [103] K. Tsuda, Subspace classifier in the Hilbert space, *Pattern Recognition Letters*, Vol.20, pp. 513–519, 1999.
- [104] K. Tsuda, M. Kawanabe, G. Rätsch, S. Sonnenburg, K.-R. Müller, A new discriminative kernel from probabilistic models, *Advances in Neural Information Processing Systems*, Vol.14, pp. 977–984, 2002.
- [105] “UCI repository of machine learning databases and domain theories”, FTP address: <ftp://ftp.ics.uci.edu/pub/machine-learning-databases>
- [106] V.N. Vapnik, *The Nature of Statistics Learning Theory*, Springer, New York, 1995.
- [107] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

-
- [108] M.P. Windham, Cluster validity for fuzzy c -means clustering algorithm, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, No.4, pp. 357–363, 1982.
- [109] P. Winkel, N. Tygstrup, The Copenhagen Study Group for Liver Diseases, Classification of cirrhosis: the resolution of data modes and their recovery in an independent material, *Computers and Biomedical Research*, Vol.4, pp. 417–426, 1971.
- [110] R.R. Yager, On the theory of bags, *Int. J. General Systems*, Vol.13, pp. 23–37, 1986.
- [111] 山本 英子 , 梅村 恭司 , データマイニング , 共立出版 , 1998.
- [112] T. Yasukouchi, S. Miyamoto, A family of hard clustering algorithms derived from the entropy regularization methods, *Proc. of International Workshop on Fuzzy Systems & Innovative Computing 2004*, June 2–3, 2004, Kitakyushu, Japan, pp. 374–377, 2004.
- [113] L.A. Zadeh, Fuzzy sets, *Information and Control*, Vol.8, pp.338–353, 1965.
- [114] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, K.-R. Müller, Engineering support vector machine kernels that recognize translation initiation sites, *Bioinformatics*, Vol.16, No.9, pp. 799–807, 2000.

研究業績

査読付き学術雑誌論文

1. Kiyotaka Mizutani, Ryo Inokuchi and Sadaaki Miyamoto,
Algorithms of Nonlinear Document Clustering based on Fuzzy Multiset Model,
International Journal of Intelligent Systems. (to appear)
2. 水谷 清隆, 張 西偉, 宮本 定明,
情報検索のためのファジィマルチ集合モデルとカーネル関数を利用したクラスタリング,
日本知能情報ファジィ学会誌, Vol.15, No.5, pp. 525–533, 2003.

査読付き国際会議論文

1. Kiyotaka Mizutani and Sadaaki Miyamoto,
Possibilistic Approach to Kernel-based Fuzzy c -Means Clustering with Entropy Regularization,
Proc. of the Modeling Decisions for Artificial Intelligence (MDAI 2005), Jul. 25–27, 2005, Tsukuba, Ibaraki, Japan, V. Torra *et al.*, (eds.), *Lecture Note in Artificial Intelligence*, LNAI 3558, pp. 144–155, 2005.
2. Kiyotaka Mizutani and Sadaaki Miyamoto,
Kernel-based Fuzzy Competitive Learning Clustering,
Proc. of the 2005 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2005), May. 22–25, 2005, Reno, Nevada, U.S.A, pp. 636–639, 2005.
3. Kiyotaka Mizutani, Ryo Inokuchi and Sadaaki Miyamoto,
LVQ Clustering and SOM Based on Inner Product by using a Kernel Function,
Proc. of the Joint 2nd International Conference on Soft Computing and Intelligent Systems and 5th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2004), Sep. 21–24, 2004, Yokohama, Japan, pp. 1–4 (CD-ROM), 2004.
★ SCIS & ISIS 2004 Session Best Presentation Award

4. Sadaaki Miyamoto and Kiyotaka Mizutani,
Fuzzy Multiset Model and Methods of Nonlinear Document Clustering for Information Retrieval,
Proc. of the Modeling Decisions for Artificial Intelligence (MDAI 2004), Aug. 2–4, 2004, Barcelona, Catalonia, Spain, V. Torra *et al.*, (eds.), *Lecture Note in Artificial Intelligence*, LNAI 3131, pp. 273–283, 2004.
5. Kiyotaka Mizutani and Sadaaki Miyamoto,
Fuzzy Multiset Model for Information Retrieval and Clustering using a Kernel Function,
Proc. of the 14th International Symposium on Methodologies for Intelligent Systems (ISMIS 2003), Oct. 28–31, 2003, Maebashi, Japan, N. Zhong *et al.*, (eds.), *Lecture Note in Artificial Intelligence*, LNAI 2871, pp. 417–421, 2003.
6. Sadaaki Miyamoto and Kiyotaka Mizutani,
Fuzzy Multiset Space and c -Means Clustering Using Kernels with Application to Information Retrieval,
Proc. of the 10th International Fuzzy Systems Association World Congress (IFSA 2003), Jun. 29 – Jul. 2, 2003, Istanbul, Turkey, T. Bilgic *et al.*, (eds.), *Lecture Note in Artificial Intelligence*, LNAI 2715, pp. 387–395, 2003.

口頭発表

1. 水谷 清隆, 井口 亮, 宮本 定明,
カーネル法を利用した競合学習モデルに基づくオンライン型クラスタリング,
第 15 回インテリジェント・システム・シンポジウム講演論文集, pp. 449–452, 2005.
★ 第 15 回インテリジェント・システム・シンポジウム ベストプレゼンテーション賞
2. 水谷 清隆, 宮本 定明,
競合学習モデルによる非線形クラスタリングと文書情報への応用,
第 21 回ファジィシステムシンポジウム講演論文集, pp. 39–42, 2005.
3. 水谷 清隆, 宮本 定明,
ファジィマルチ集合とカーネル関数を利用したクラスタリング,
第 19 回ファジィシステムシンポジウム講演論文集, pp. 141–144, 2003.
★ 2004 年度 日本知能情報ファジィ学会 奨励賞

4. 水谷 清隆, 張 西偉, 宮本 定明,
ファジィマルチ集合に基づく情報検索モデルとクラスタリング,
ラフ集合研究部会第1回ワークショップ, 2002.

その他の査読付き論文

1. Kiyotaka Mizutani, Xi Wei Zhang and Sadaaki Miyamoto,
Information Retrieval and Clustering Based on Fuzzy Multiset Model,
Proc. of the First Workshop on Rough Sets and Kansei Engineering in Japan, Dec.
14–15, 2002, Tokyo, Japan, S. Tsumoto *et al.*, (eds.), *Bulletin of International Rough
Set Society*, Vol.7, No.1/2, pp. 45–50, 2003.

付録A 文書情報クラスタリングに用いた文献のタイトル(第5章)

A.1 数値例1に用いた文献のタイトル

ここでは、第5章の文書情報クラスタリングの数値例1で用いた文献のタイトルについて示す。なお、クラスタリング以外のアルゴリズムをテーマとする論文(9件)については、タイトルの最後に♣をつけて表示している。

1. L_1 空間に基づくファジィ c -平均クラスタリング
2. 固有値を用いた次元係数の正則化による線形ファジィクラスタリング
3. ファジィ集合とLVQを用いた関数近似手法 ♣
4. 正則化ファジィ c -平均法におけるいくつかの目的関数と2つの制約
5. データの部分構造を考慮した数量化分析1類 ♣
6. ファジィ c -平均クラスタリングにおける制約条件と正則化
7. Viterbi復号アルゴリズムにおけるファジィ推論を用いたチャンネルの推定 ♣
8. 関数近似問題におけるファジィ遺伝的機械学習アルゴリズム ♣
9. ファジィパターン識別のための教師あり、教師なし離散化
10. 局所的なマイナー成分を用いた最小絶対誤差に基づく線形ファジィクラスタリング
11. 類似度が与えられた場合のファジィクラスタリング
12. ファジィルールにより交叉演算子を選択する遺伝的アルゴリズムについて ♣
13. 可能性クラスタリングにおける分類関数の考察と計算手法の検討
14. 最小自乗誤差規範と最小絶対誤差規範にもとづく c -回帰法
15. 不確定性を含むデータのファジィクラスタリングにおける中心の計算
16. ファジィデータに対する L_1 距離に基づくファジィ c -平均法
17. ファジィマルチ集合とファジィクラスタリング
18. クラスタリングにおける c -回帰法の分類関数とプロトタイプ関数
19. ファジィクラスタリングの評価基準について
20. 正則化ファジィクラスタリングと分類関数 - エントロピー正則化と2次正則化 -
21. 次元係数の正則化による線形ファジィクラスタリング
22. ファジィ符号化を用いたクラスタリングアルゴリズム

23. ファジィ c -平均クラスタリングアルゴリズムによるファジィ規則生成法の改善
24. ファジィクラスタリングを用いた都市道路網の色彩評価に基づくイメージ分析
25. ファジィデータに対する L_1 距離を用いたファジィクラスタリング
26. ニューロファジィモデルにおける種々の自己増殖型アルゴリズム ♣
27. クラスタワイズ回帰型主観的評価モデル ♣
28. 相転移を利用した階層化ファジィクラスタリング
29. SOM を用いた階層型分類手法 ♣
30. ファジィ情報検索について ♣

A.2 数値例 2 に用いた文献のタイトル

ここでは, 第 5 章の文書情報クラスタリングの数値例 2 で用いた 127 件の文献のタイトルについて示す.

1. Understanding public and other stakeholder of risk during the siting process for a waste transfer station in an urban neighbourhood.
2. Safety reports : Integrated russian experience.
3. Normative regulation of emergencies prevention on the design stage of potentially dangerous objects on the territory of Russian Federation.
4. Risk assessment for steel buildings and occupants under fire situation.
5. Decision making for accidental risk management : Uncertainty level quantification.
6. Time will tell : Changes in risk perception and the processing of risk information about the inevitable of the millennium bug.
7. Social science in a nuclear research centre : Dream or Reality?
8. Safety culture : an integrative concept for managing risks?
9. Urban air pollution in the regional press : a dramatic representation.
10. Did people believe in the Twin Tower attack on September the 11th?
11. Risk evaluation : The noise exposure.
12. Researching risk.
13. Beliefs related to natural disasters and flooding. An example from Assam, India.
14. Technical limitations of cost-effectiveness analysis.
15. Neglected regulation : the institutional attention of risk.
16. Space, power and public constructions of risk : a case of study the foot and mouth disease outbreak in the UK.
17. Obstacles to Holistic Risk Management in the Energy Sector.
18. Competitive health risk model based on the mortality age-dependent model and population heterogeneity assumption.
19. A formal to health risk assessment for cases of acute and chronicle exposure.
20. Risk analysis within integrated coastal zone management.
21. Scientific risk construct : economic aspects.
22. Trust and confidence in the applied field of EMF.
23. The use of mental models in chemical risk protection - the need for multi-method evaluation.
24. Climate change, political-administrative regulation and coastal protection "PAS".
25. IT-Operation criteria selection for risk analysis support.

26. The psychometric qualities of the employee safety, health and environment questionnaire.
27. The effects on traffic safety attitudes and behaviour of a safety campaign among adolescents.
28. Social trust in UK risk regulators : The role of practical reasoning strategies in the formation of public attitudes.
29. Scenario-based risk assessment comparing occupational exposure to perchloroethylene and trichloroethylene in mental-degreasing over the last 30 years.
30. Trust as a prerequisite of low risk perception in relation to biotechnology : case Finland.
31. Climate change, risk constructs and coastal protection : Aim and approach of the interdisciplinary project KRIM.
32. Social representations of uranium mining residues.
33. e-Risk management.
34. The interrelation between the internal organisational processes and external risk communication.
35. The growth and development of risk communication through an analysis of the literature in the field.
36. Temporal discounting in environmental risk evaluation.
37. Media resonance and the collective action of experts in the conflict of nuclear waste management.
38. Questioning social amplification : learning about risk communication.
39. Short term and long term management of a nuclear accident : a CD-ROM for the construction of a common risk culture between stakeholders and stake-seekers.
40. Zha methodology - A powerful tool.
41. The KRIM decision support system - policy support for integrated coastal protection and risk management.
42. New Approach to early recognition of risks and rare events : The application of the theory of catastrophes and the mathematical modeling.
43. Elementary my dear Watson : On condition and case in catastrophe risk.
44. Holes in holistic risk management : Financial institution's approaches to operational risk.
45. Assessment of risk in tunnels Bayesian Networks.
46. Transparency and public hearings. An evaluation of dialogs in the process of building a Swedish nuclear waste repository.

47. Exchange of nuclear safety information across expert group. A study of trouble shooting.
48. Inherent flight safety vs operating safety aspect.
49. The Japanese consensus conference on GM crops : an experiment towards deliberative policy-making.
50. The 5S technique for small to medium-sized businesses : A risk management perspective.
51. Consumer exposure to chemicals widely used in personal care products : A scenario-based exposure assessment of polycyclic musk compounds.
52. Prospective risk recognition through Petri Net simulation and reachable-states analysis.
53. Managing occupational risk in the health care facilities.
54. An integrated approach in managing risk in industrial areas.
55. Mobile telecommunication technology in the UK-multi-way engagement or self-regulation gone wrong?
56. Living with radiation : Perception of radiation risk amongst adolescents born after chernobyl in Belarus and Ukraine.
57. The neglect of risk : Result from a study of experts risk assessments.
58. Financial provision of emergency risk management.
59. The problem of municipal solid waste in the Moscow region.
60. Weighing ecological risk and environmental impact : Developing a holistic risk management strategy for wetland sediment remediation.
61. Basic hydromechanical information for risk assessment of coastal defences in the Jade-Weser area (Germany).
62. Application of risk analysis technologies for optimization of administrative decisions in safety when for storage of irradiated nuclear fuel.
63. Communicating European food safety : British and Dutch newspapers covering the BSE-crisis.
64. Scientific risk construct : Environmental aspects.
65. Evaluation of operating experience for early recognition of risks.
66. Stakeholders involvement in the vicinity of industrial sites.
67. Are rules safe? A study of rule management in railways.
68. Decision making and bids in construction industry.
69. The requirement for holistic risk management.

70. The knowledge potential of open, integrated social decision making : the case of waste management facilities in the United Kingdom.
71. Ratifying Kyoto : applying the precautionary principal to the climate change issue controversy in Canada.
72. Image of local environmental uncertainty and an indeterminable future : A Swedish case study.
73. Risk communication on organic foods : effects of collaboration between sources.
74. Land use planning and industrial risk : tools for prevention and disaster mitigation.
75. External emergency plans : a challenging risk management tool for local communities.
76. Coping capacity of people at risk : Predictors of households coping with the risk of flooding.
77. Lessons learned from scientific dialogue on EMF risk assessment.
78. Plugging leaks sinking ships.
79. Cultural dimensions of risk perception.
80. Risk communication for chemical risk management. An OECD guidance document.
81. Risk management as internalization. The example of risk management by insurance.
82. Stakeholder views in transboundary water contamination problems : the case of the January 2000 Tisza River cyanide spill.
83. Integrated flood risk management for Hungary's Tisza River.
84. SERF3 - Socio-economic research on fusion. The narrow path for ITER.
85. Risk and sustainability. The claim of sustainability.
86. Pagination or animation? Examples of emf risk information tools for the public.
87. Attitudes towards genetically modified food in six european countries.
88. We didn't see it coming : the subjective nature of signal detection.
89. Public opinion as constraint for coastal risk management : The role of the media.
90. The perception of the Galtur avalanche disaster.
91. ARAMIS project : development of a accidental risk assessment methodology in the context of the Seveso 2 directive.
92. Predicting exposure concentrations of cadmium in vegetables and fruits by probabilistic modelling : Improved selection of input variables and plausibility check.
93. Exposure assessment for children : Is there a great risk than in adults?
94. The principle of two-dimensional probabilistic modelling - talking extrapolation factors as an example.
95. Uncertainty analysis in population based environmental risk assessment with respect to traffic noise related cardiovascular mortality.

96. Fast and dynamic security risk assessment and presentation for telecommunication : security risk radar.
97. Information regulation of major industrial accidents : The case of the U.S. LEPCs.
98. Post mining management in the Lorraine region : Multi-criteria decision-aid method for the hierarchization of mining subsidence risk.
99. Terrorism as a generalized threat.
100. Social aspects of risk management : Integrated perspectives after the first ten years.
101. Scientific and practical aspect of risk analyses and management on oil and gas production enterprises.
102. Estimating quality-quantity risk trade-off value : Decision-making in hospital management.
103. MIRIAM : an integrated approach to organise major risk control in hazardous chemical establishments.
104. Predictive analysis for project portfolio risk management.
105. Consciousness difference between doctors-nurses on the medical information that affects the risk assessment of the word.
106. Sensitive factors associated with cancer risk due to inhalation exposures to volatile organic compounds in drinking water during showering.
107. Paradigmatic study of risk perception from three student groups specialising in radiation and environment studies leading to new views of optimistic bias.
108. Assessing websites for risk information and disaster preparedness.
109. The baycol/lipobay issue - a case study - the politics of product recall.
110. Reduction of pollution using technologies and methods for identification, estimation and limitation of industrial risk.
111. The impact of information frames on the laypersons risk appraisal.
112. An advanced strategic risk management solution : The inclusion of precautionary principle in the strategic decision process to enhance global risk management and foster sustainable development : Concepts, tools, methodology and results.
113. Consequence-based approach for land-use planning : a precautionary principle in accidental risk control?
114. Trust in science and regulation : The case of biotechnological risks.
115. Risk as perceived by the public : Disparities of qualitative and quantitative findings.
116. Consumers as Guinea pigs? The relative importance of the risk dimension in public discourses and individual evaluations of biotechnology.
117. Adored and feared? The public's perception of mobile telephony.

118. Blood work : Where is the risk on the hospital waste? risk perception hospital waste environmental behaviour.
119. Radioactive contamination of wild mushrooms : A cross-cultural risk perception study using World Wide Web.
120. Biotechnology : the perennial challenge.
121. Classifying corporate occupational safety cultures : lessons learned from six case studies in Germany and the UK.
122. Communication - as necessary as it is useful - a short guideline.
123. Beyond stigma? Risk, encroachment and sense of place.
124. Innovative packaging methodology construction.
125. Affected people's flood risk perception and local experience of disaster management.
126. Early risk in environmental health – Is it feasible? A German perspective.
127. Risk of maritime traffic accidents at the Turkish Straits.

付録B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

B.1 クラスタリングに用いたデータ

B.1.1 1985年

1985年における世界各国の軍事データ（国防費関係および兵力関係）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP比 ²	現役兵力	予備役	準軍隊
		100万ドル	ドル	%	1,000人	1,000人	1,000人
アメリカ	NATO	380899	1592	6.1	2156.6	1212.3	178.7
カナダ		11164	440	2.2	83	24.7	7.9
ベルギー		3538	363	2.9	91.6	178.5	15.9
デンマーク		1858	363	2.1	29.6	162.2	0
フランス		30659	556	3.9	464.3	393	89.5
(東・西)ドイツ		29393	387	3.2	478	1420	97.5
ギリシャ		3439	346	7	201.5	404	25
イタリア		13729	240	2.2	385.1	799	6.6
ルクセンブルク		74	203	1	0.7	0	0.5
オランダ		5731	396	2.9	105.5	176.3	8.7
ノルウェー		2651	638	2.8	37	291	0
ポルトガル		964	94	3.2	73	173	37.3
スペイン		5856	152	2.4	320	1085	110.5
トルコ		3489	69	3.1	630	936	125
イギリス	35003	619	5.2	334	294.5	0	
ソ連	ワルシャワ 条約機構	368327	1321	16.1	5300	5400	600
ハンガリー		2060	193	6.8	106	143	75
ポーランド		21644	582	20.5	319	501	218
ブルガリア		1424	159	2.9	148.5	795	172.5
チェコスロバキア		2038	131	8.2	203.3	575	131
ルーマニア		1204	53	1.7	189.5	565	37
アルバニア	ヨーロッパ	164	55	4.1	40.4	155	12.5
オーストリア		1124	149	1.2	54.7	170	0

¹国民一人当たりの国防支出費

²GDPに占める国防支出費の割合

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP 比 ²	現役兵力	予備役	準軍隊
		100 万ドル	ドル	%	1,000 人	1,000 人	1,000 人
キプロス		75	113	2.2	10	60	3
フィンランド		1308	267	1.2	36.5	700	3.5
アイルランド		278	78	1	13.7	16.4	0
マルタ		16	44	1.1	0.8	0	0.9
スウェーデン		2776	332	1.9	65.7	735.5	0.6
スイス		1681	260	1.2	20	625	480
ユーゴスラビア/ セルビア/モンテネグロ		2904	125	3.7	241	500	15
アルジェリア	中東・ 北アフリカ	1056	48	1.2	170	150	30.6
バーレーン		223	534	4.1	2.8	0	2.2
エジプト		4617	95	13	445	380	139
イラン		7952	178	7.7	610	350	2570
イラク		18664	1174	42.2	1000	75	654.8
イスラエル		11468	2709	21.2	142	370	4.5
ヨルダン		915	261	15.9	70.3	35	11
クウェート		2215	1295	9.1	12	0	0
レバノン		173	65	9	17.4	0	0
リビア		1170	311	6.2	73	40.4	7
モロッコ		1475	67	5.4	149	0	31
オマーン		2858	1786	20.8	29.2	0	3.6
カタール		260	854	6	6	0	0
サウジアラビア		31485	2728	19.6	62.5	0	43.5
シリア		8014	763	16.4	402.5	272.5	14.3
チュニジア		472	66	5	35.1	0	9.5
アラブ首長国連邦 (南・北) イエメン	2805	2003	7.6	43	0	0	
	478	47	9.9	64.1	85	70	
アフガニスタン	中央・ 南アジア	425	24	8.7	47	0	30
バングラディッシュ		215	2	1	91.3	30	55
インド		10948	14	3.5	1260	250	260
ネパール		31	2	0.8	25	25	22
パキスタン		2669	28	6.1	482.8	513	164
スリランカ		257	16	2.9	21.6	16.1	19.5
オーストラリア	東アジア・ 太平洋州	6613	420	2.7	70.4	31.5	0
ブルネイ		390	1739	8	4.1	0	2.7
中国		21616	21	4.9	3900	5377	12000
フィジー		19	28	1.2	2.7	0.3	0
インドネシア		3674	23	2.8	278.1	0	82
日本		20139	167	1	243	43.6	20
北朝鮮		7761	381	20	838	540	3108
韓国		6861	167	5	598	4822	4100
ラオス		196	54	7.1	53.7	0	0
マレーシア		1748	112	3.8	110	46.4	372.3
モンゴル		71	37	8	33	0	15
ミャンマー		1475	40	5	186	0	73

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP比 ²	現役兵力	予備役	準軍隊
		100万ドル	ドル	%	1,000人	1,000人	1,000人
ニュージーランド		664	204	2	12.4	9.6	0
バブアニューギニア		31	9	1	3.2	0	4.6
フィリピン		633	12	1.4	114.8	123	112
シンガポール		1567	613	6	55	162	37.5
台湾		9295	479	7	444	1470	25
タイ		2296	44	4	235.3	500	86
ベトナム		2079	34	19.4	1027	3000	1560
バハマ	カリブ	9	40	0.3	0.5	0	0.5
キューバ		1384	137	6.3	161.5	165	1418.5
ドミニカ		89	14	1.2	22.2	0	1
ハイチ		27	4	0.9	6.9	0	15
ジャマイカ		24	11	0.8	2.1	1.3	1.5
トリニダード・トバゴ		63	54	0.6	2.1	0	0.4
ベリーズ	中央アメリカ	3	18	1	0.6	0.3	0
コスタリカ		25	10	0.4	0	0	8
エルサルバドル		478	100	5.7	41.7	0	18
グアテマラ		261	33	1.6	31.7	10	26.6
ホンジュラス		190	43	3.5	16.6	50	5
メキシコ		1890	24	0.7	129.1	250	0
ニカラグア		307	94	17.3	62.9	57	45
パナマ		159	73	2	12	0	0
アルゼンチン	ラテンアメリカ	4670	153	3.8	108	377	21
ボリビア		292	46	3	27.6	0	0
ブラジル		11066	82	3	276	1340	220
チリ		2125	176	10	101	100	25
コロンビア		823	29	1.6	66.2	116.6	50
エクアドル		657	70	2.8	42.5	0	0.2
ガイアナ		50	63	7.7	6.6	0	5
パラグアイ		72	20	1.1	14.4	36.3	6
ペルー		3262	175	6.7	128	175	51.6
スリナム		7	19	0.9	2	0	0.7
ウルグアイ		207	69	2.8	31.9	0	2.7
ベネズエラ		1855	107	2.1	49	0	22
ジブチ	サハラ以南 のアフリカ	46	106	7.9	3	0	1.5
エチオピア		99	2	6.6	217	0	169
スーダン		27	1	2.9	56.6	0	0
ブルンジ		52	11	3	5.2	0	1.5
カメルーン		137	13	1.1	7.3	0	4
カーボ・ヴェルデ		3	9	1.4	7.7	0	0
中央アフリカ共和国		15	6	1.4	2.3	0	10
チャド		32	6	2.5	12.2	0	11.4
コンゴ		49	26	1.5	8.7	0	6.1
コンゴ民主共和国		71	2	0.9	48	0	22
赤道ギニア		3	8	2.4	2.2	0	0

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP 比 ²	現役兵力	予備役	準軍隊
		100 万ドル	ドル	%	1,000 人	1,000 人	1,000 人
ガボン		68	68	1.3	2.4	0	4.8
ルワンダ		28	5	1.1	5.2	0	1.2
ケニア		221	11	2.5	13.7	0	0
マダガスカル		47	5	1.1	21.1	0	8
セーシェル		7	113	3	1.2	0	0.9
タンザニア		359	16	3.8	40.4	0	50.5
ウガンダ		112	8	2	20	0	23
ベナン		18	4	1.1	4.5	0	3.5
ブルキナ・ファソ		30	4	1.4	4	40	2.1
コートジボアール		122	12	1.2	13.2	12	7.8
ガンビア		1	2	0.4	0.5	0	0
ガーナ		93	7	1	15.1	0	0
ギニア		46	7	1.1	9.9	0	9
ギニア・ビサウ		9	10	3	8.6	0	2
リベリア		43	19	2.6	6.8	50	2
マリ		52	7	2.8	4.9	0	12.3
ニジェール		10	2	0.5	2.2	0	2.5
ナイジェリア		651	7	1.7	94	0	12
セネガル		55	8	1.4	10.1	0	6.8
シェラレオネ		9	2	0.7	3.1	0	0.8
トーゴ		28	9	2.5	3.6	0	1.6
アンゴラ		1328	152	15	49.5	50	64
ボツワナ		32	30	2.2	4	0	1
マラウイ		32	5	2	5.3	0.5	1
モザンビーク		330	24	9.9	15.8	0	9.5
南アフリカ		3252	97	3.8	106.4	317	130
ザンビア		34	5	0.9	16.2	0	1.2
ジンバブエ		475	57	5.7	41	0	26

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

1985 年における世界各国の軍事データ（陸上戦力および海上戦力）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
アメリカ	NATO	14139	0	2150	1891	95	200	89
カナダ		114	0	0	0	3	23	0
ベルギー		334	116	10	0	0	4	0
デンマーク		208	48	0	0	4	20	27
フランス		1260	342	780	68	18	47	3
(東・西)ドイツ		6162	0	3544	0	26	34	6
ギリシャ		1747	270	240	8	10	21	9
イタリア		1770	0	0	0	10	23	0
ルクセンブルク		0	0	0	0	0	0	0
オランダ		1146	0	0	0	5	18	3
ノルウェー		100	70	135	0	14	5	30
ポルトガル		66	0	0	0	3	17	19
スペイン		779	180	250	0	7	23	87
トルコ		2922	0	0	0	16	18	29
イギリス		1030	271	0	314	28	60	0
ソ連	ワルシャワ 条約機構	52600	0	27000	1250	371	289	0
ハンガリー		1230	100	200	0	0	0	0
ポーランド		3450	110	0	0	3	4	8
ブルガリア		1860	0	60	0	2	0	13
チェコスロバキア		3500	0	1100	0	0	0	0
ルーマニア		1230	0	0	0	0	6	49
アルバニア	ヨーロッパ	100	0	0	0	3	0	2
オーストリア		170	0	0	0	0	0	0
キプロス		8	0	0	0	0	0	2
フィンランド		1	1	1	0	0	0	5
アイルランド		0	12	0	0	0	0	1
マルタ		0	0	0	0	0	0	0
スウェーデン		670	200	0	0	13	1	4
スイス		846	0	0	0	0	0	0
ユーゴスラビア/ セルビア/モンテネグロ		920	0	200	0	9	2	18
アルジェリア		中東・ 北アフリカ	700	0	650	0	2	3
バーレーン	0		0	0	0	0	0	0
エジプト	2159		30	417	0	14	10	30
イラン	1000		50	180	0	0	7	7
イラク	2900		250	500	65	0	1	3
イスラエル	3600		0	0	0	3	0	45
ヨルダン	795		0	0	0	0	0	9
クウェート	240		0	0	0	0	0	48
レバノン	50		60	0	0	0	0	4
リビア	2800		0	700	0	6	1	5

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
モロッコ		120	70	0	0	0	1	1
オマーン		33	36	6	0	0	0	4
カタール		24	0	30	0	0	0	6
サウジアラビア		450	0	350	0	0	4	1
シリア		4200	0	600	0	0	2	7
チュニジア		59	50	0	0	0	1	17
アラブ首長国連邦		136	60	1	0	0	0	9
(南・北)イエメン		1114	0	100	0	0	0	10
アフガニスタン	中央・ 南アジア	450	60	40	0	0	0	0
バングラディシュ		50	6	0	0	0	3	6
インド		2500	150	350	0	8	28	8
ネパール		0	16	0	0	0	0	0
パキスタン		1506	0	0	2	11	8	24
スリランカ		0	0	0	0	0	0	2
オーストラリア	東アジア・ 太平洋州	103	0	63	0	6	13	23
ブルネイ		0	16	0	0	0	0	3
中国		11450	0	0	0	110	44	60
フィジー		0	0	0	0	0	0	0
インドネシア		0	111	200	0	2	13	24
日本		1070	0	0	10	14	49	11
北朝鮮		3275	150	0	0	20	4	32
韓国		1200	0	0	0	0	18	84
ラオス		30	25	0	0	0	0	20
マレーシア		0	26	0	0	0	3	20
モンゴル		650	0	0	0	0	0	0
ミャンマー		24	0	0	0	0	0	46
ニュージーランド		0	26	0	0	0	4	8
バブアニューギニア		0	0	0	0	0	0	5
フィリピン		0	28	45	0	0	7	12
シンガポール		0	270	0	0	0	0	12
台湾		309	1120	0	0	2	36	28
タイ		390	144	0	0	0	6	94
ベトナム		1600	450	0	0	0	8	54
バハマ	カリブ	0	0	0	0	0	0	0
キューバ		835	55	50	0	4	2	17
ドミニカ		0	14	0	0	0	1	20
ハイチ		0	6	0	0	0	0	14
ジャマイカ		0	0	0	0	0	0	0
トリニダード・トバゴ		0	0	0	0	0	0	6
ベリーズ	中央アメリカ	0	0	0	0	0	0	0
コスタリカ		0	0	0	0	0	0	0
エルサルバドル		0	12	0	0	0	0	0
グアテマラ		0	23	0	0	0	0	13

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
ホンジュラス	ラテンアメリカ	0	12	0	0	0	0	9
メキシコ		0	40	0	0	0	8	66
ニカラグア		122	30	0	0	0	0	2
パナマ		0	0	0	0	0	0	2
アルゼンチン		230	50	450	0	4	7	6
ボリビア		0	0	0	0	0	0	37
ブラジル		0	575	0	0	7	17	6
チリ		171	65	0	0	2	8	10
コロンビア		12	0	0	0	2	4	5
エクアドル		0	145	0	0	2	1	10
ガイアナ		0	0	0	0	0	0	11
パラグアイ		3	10	0	0	0	0	11
ペルー		250	110	0	0	12	16	4
スリナム		0	0	0	0	0	0	9
ウルグアイ		0	67	0	0	0	3	7
ベネズエラ		80	422	0	0	2	8	0
ジブチ	サハラ以南 のアフリカ	0	0	0	0	0	0	0
エチオピア		935	20	40	0	0	2	14
スーダン		170	183	0	0	0	0	13
ブルンジ		0	0	0	0	0	0	0
カメルーン		0	0	12	0	0	0	3
カーボ・ヴェルデ		0	0	0	0	0	0	1
中央アフリカ共和国		4	0	0	0	0	0	0
チャド		0	0	0	0	0	0	0
コンゴ		50	17	0	0	0	0	9
コンゴ民主共和国		50	0	0	0	0	0	46
赤道ギニア		0	0	0	0	0	0	1
ガボン		0	0	0	0	0	0	2
ルワンダ		0	0	0	0	0	0	0
ケニア		76	0	0	0	0	0	3
マダガスカル		12	0	0	0	0	0	1
セーシェル		0	0	0	0	0	0	1
タンザニア		30	66	0	0	0	0	12
ウガンダ		13	0	0	0	0	0	0
ベナン		0	10	0	0	0	0	0
ブルキナ・ファソ		0	0	0	0	0	0	0
コートジボアール		0	5	0	0	0	0	8
ガンビア		0	0	0	0	0	0	0
ガーナ		0	0	0	0	0	0	6
ギニア		53	20	0	0	0	0	16
ギニア・ビサウ		10	20	0	0	0	0	13
リベリア		0	0	0	0	0	0	5
マリ		21	12	0	0	0	0	3
ニジェール	0	0	0	0	0	0	0	

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
ナイジェリア		76	50	0	0	0	2	9
セネガル		0	0	0	0	0	0	8
シエラレオネ		0	0	0	0	0	0	0
トーゴ		9	0	0	0	0	0	2
アンゴラ		465	50	0	0	0	0	12
ボツワナ		0	0	0	0	0	0	0
マラウイ		0	0	0	0	0	0	0
モザンビーク		285	0	0	0	0	0	26
南アフリカ		250	0	1500	0	3	1	4
ザンビア		30	50	0	0	0	0	0
ジンバブエ		31	0	0	0	0	0	0

1985年における世界各国の軍事データ（航空戦力）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
アメリカ	NATO	1350	160	3700	0
カナダ		0	0	181	32
ベルギー		0	0	181	0
デンマーク		0	0	102	0
フランス		104	23	475	0
(東・西)ドイツ		122	12	966	70
ギリシャ		0	0	314	0
イタリア		0	83	315	0
ルクセンブルク		0	0	0	0
オランダ		15	17	218	0
ノルウェー		0	0	92	0
ポルトガル		0	0	116	0
スペイン		10	40	177	0
トルコ		20	7	368	0
イギリス		32	121	599	0
ソ連	ワルシャワ 条約機構	875	310	2350	2650
ハンガリー		0	0	155	12
ポーランド		0	0	675	12
ブルガリア		0	3	226	20
チェコスロバキア ルーマニア		0	0	474	24
		0	0	378	0
アルバニア	ヨーロッパ	0	0	100	0
オーストリア		0	0	32	0
キプロス		0	0	0	0
フィンランド		0	0	64	0
アイルランド		0	0	15	0
マルタ		0	0	0	0
スウェーデン		0	0	524	0
スイス		0	0	299	0
ユーゴスラビア/ セルビア/モンテネグロ		0	0	420	0
アルジェリア	中東・ 北アフリカ	0	0	332	35
バーレーン		0	0	0	0
エジプト		0	0	427	48
イラン		2	12	80	0
イラク		0	0	500	100
イスラエル		0	0	684	60
ヨルダン		0	0	121	0
クウェート		0	0	76	23
レバノン		0	0	7	4
リビア		0	0	535	42

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
モロッコ		0	0	105	0
オマーン		0	0	52	0
カタール		0	0	17	2
サウジアラビア		0	0	205	0
シリア		0	0	500	100
チュニジア		0	0	20	0
アラブ首長国連邦 (南・北)イエメン		0	0	42	7
		0	0	179	15
アフガニスタン	中央・ 南アジア	0	0	150	30
バングラディシュ		0	0	23	0
インド		36	26	846	60
ネパール		0	0	0	0
パキスタン		3	6	375	0
スリランカ		0	0	0	2
オーストラリア	東アジア・ 太平洋州	0	0	138	0
ブルネイ		0	0	6	0
中国		0	0	5300	0
フィジー		0	0	0	0
インドネシア		19	14	68	0
日本		84	0	270	0
北朝鮮		0	0	800	0
韓国		0	0	451	0
ラオス		0	0	20	0
マレーシア		0	0	42	0
モンゴル		0	0	12	0
ミャンマー		0	0	22	0
ニュージーランド		0	0	44	0
パプアニューギニア		0	0	0	0
フィリピン		0	0	64	0
シンガポール		0	0	164	0
台湾		0	0	567	12
タイ	0	0	183	0	
ベトナム	0	0	270	65	
バハマ	カリブ 中央アメリカ	0	0	0	0
キューバ		0	0	250	38
ドミニカ		0	0	36	0
ハイチ		0	0	7	0
ジャマイカ		0	0	0	0
トリニダード・トバコ		0	0	0	0
ベリーズ		0	0	0	0
コスタリカ		0	0	0	0
エルサルバドル		0	0	32	4
グアテマラ		0	0	16	4

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
ホンジュラス	ラテンアメリカ	0	0	25	0
メキシコ		0	0	85	0
ニカラグア		0	0	17	0
パナマ		0	0	0	0
アルゼンチン		54	24	157	6
ボリビア		0	0	14	9
ブラジル		0	16	166	0
チリ		6	0	101	0
コロンビア		0	0	49	17
エクアドル		0	0	72	0
ガイアナ		0	0	0	0
パラグアイ		0	0	5	0
ペルー		13	8	108	42
スリナム		0	0	0	0
ウルグアイ		7	0	41	0
ベネズエラ		4	6	91	0
ジブチ		サハラ以南 のアフリカ	0	0	0
エチオピア	0		0	150	30
スーダン	0		0	45	0
ブルンジ	0		0	3	0
カメルーン	0		0	15	2
カーボ・ヴェルデ	0		0	0	0
中央アフリカ共和国	0		0	2	0
チャド	0		0	2	0
コンゴ	0		0	21	0
コンゴ民主共和国	0		0	40	0
赤道ギニア	0		0	2	0
ガボン	0		0	11	0
ルワンダ	0		0	4	0
ケニア	0		0	0	0
マダガスカル	0		0	12	0
セーシェル	0		0	0	0
タンザニア	0		0	29	0
ウガンダ	0		0	0	0
ベナン	0		0	0	0
ブルキナ・ファソ	0		0	0	0
コートジボアール	0		0	6	0
ガンビア	0		0	0	0
ガーナ	0		0	10	0
ギニア	0		0	6	0
ギニア・ビサウ	0		0	0	0
リベリア	0		0	0	0
マリ	0		0	5	0
ニジェール	0		0	0	0

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
ナイジェリア		0	0	49	0
セネガル		0	0	2	0
シエラレオネ		0	0	0	0
トーゴ		0	0	11	0
アンゴラ		0	0	141	12
ボツワナ		0	0	5	0
マラウイ		0	0	0	0
モザンビーク		0	0	18	0
南アフリカ		0	0	356	16
ザンビア		0	0	44	0
ジンバブエ		0	0	53	0

B.1.2 2003 年

2003 年における世界各国の軍事データ（国防費関係および兵力関係）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP 比 ²	現役兵力	予備役	準軍隊
		100 万ドル	ドル	%	1,000 人	1,000 人	1,000 人
アメリカ	NATO	329616	1138	3.3	1414	1259.3	53
カナダ		7771	247	1.1	52.3	35.4	9.3
ベルギー		3435	332	1.3	39.2	100.5	0
チェコ共和国		1401	136	2.1	49.4	0	5.6
デンマーク		2564	483	1.6	22.7	64.9	
フランス		38005	636	2.5	260.4	100	101.4
ドイツ		31465	383	1.5	296	390.3	0
ギリシャ		6154	579	4.4	177.6	291	4
ハンガリー		1083	110	1.8	33.4	90.3	14
イタリア		24210	421	1.9	216.8	65.2	254.3
ルクセンブルク		193	438	0.9	0.9	0	0.6
オランダ		7330	459	1.6	49.6	32.2	3.3
ノルウェー		3434	759	1.9	26.6	219	0
ポーランド		3400	88	1.9	163	234	21.4
ポルトガル		2945	294	2.3	43.6	210.9	46.4
スペイン		8253	206	1.2	177.9	328.5	73.3
トルコ		8727	127	5.1	514.8	378.7	150
イギリス	35249	590	2.4	210.4	256.7	0	
ロシア	ヨーロッパ (NATO 以外)	48040	333	4.8	988.1	2400	409.1
アルバニア		107	35	2.5	27	0	0
アルメニア		615	162	6.4	44.6	210	1
オーストリア		1699	209	0.8	34.6	72	0
アゼルバイジャン		851	104	3.3	72.1	300	15
ベラルーシ		1891	188	2.5	79.8	289.5	110
ボスニア・ヘルツェゴビナ		178	44	3.8	19.8	0	0
ブルガリア		378	48	2.5	68.4	303	34
クロアチア		517	110	2.4	51	140	10
キプロス		227	245	2.4	10	60	0.1
エストニア		93	68	1.6	5.5	24	2.6
フィンランド		1970	378	1.4	31.8	485	3.1
マケドニア		96	49	2.7	12.3	60	7.6
グルジア		236	46	1.7	17.5	250	11.7
アイルランド		718	187	0.6	10.5	14.8	0
ラトビア		141	60	1.8	5.5	14	3.2
リトアニア		233	63	1.8	13.5	25	13.8
マルタ	25	62	0.7	2.1	0	0	
モルドバ	151	35	1.7	7.2	66	3.4	
ルーマニア	999	45	2.3	99.2	130	79.9	
スロバキア	439	81	2	26.2	20	4.7	
スロベニア	311	156	1.5	9	20	4.5	

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP比 ²	現役兵力	予備役	準軍隊
		100万ドル	ドル	%	1,000人	1,000人	1,000人
スウェーデン		3947	448	1.7	33.9	262	35.6
スイス		2852	395	1.1	27.6	351.2	0
ウクライナ		4728	97	2.2	302.3	1000	112.5
ユーゴスラビア/ セルビア/モンテネグロ		706	68	5.3	74.5	400	50
アルジェリア	中東・ 北アフリカ	2970	95	5.9	136.7	150	181.2
バーレーン		314	435	4	10.7	0	10.2
エジプト		3121	44	3.9	443	254	330
イラン		4865	67	4.6	520	350	40
イラク		1470	64	9.7	389	650	43
イスラエル		9437	1499	9.7	161.5	425	8
ヨルダン		844	162	9.3	100.2	35	10
クウェート		3295	1582	10.7	15.5	23.7	6.6
レバノン		509	144	3.2	71.8	0	13
リビア		531	98	3.8	76	40	0.5
モロッコ		1307	42	3.6	196.3	150	50
オマーン		2568	944	13.4	41.7	0	4.4
カタール		1741	2857	10.6	12.4	0	0
サウジアラビア		20981	981	12	199.5	0	15.5
シリア		1819	107	10.3	319	354	108
チュニジア		383	39	1.8	35	0	12
アラブ首長国連邦	2706	949	4	41.5	0	0	
イエメン	486	24	5.7	66.5	40	70	
アフガニスタン	中央・ 南アジア	245	10	13	50	0	70
バングラディシュ		638	4	1.4	137	0	63.2
インド		13073	13	2.7	1298	535	1089.7
カザフスタン		1986	124	2	60	237	34.5
キルギスタン		265	52	1.9	10.9	57	5
ネパール		99	4	1.9	51	0	40
パキスタン		2541	17	3.9	620	513	289
スリランカ		504	26	3.2	157.9	5.5	88.6
タジキスタン		123	20	1.7	6	0	1.2
トルクメニスタン		284	58	0.8	17.5	0	0
ウズベキスタン	1797	70	2.9	52	0	20	
オーストラリア	東アジア・ 太平洋州	7549	388	2	50.9	20.3	0
ブルネイ		253	741	5.2	7	0.7	3.7
中国		48380	37	4.1	2270	550	1500
フィジー		26	32	1.6	3.5	6	0
インドネシア		6245	29	3.7	297	400	195
日本		37070	290	1	239.9	47	12.2
北朝鮮		4728	214	25	1082	4700	189
韓国		12615	266	2.8	686	4500	4.5
ラオス		14	3	0.8	29.1	0	100
マレーシア	3260	145	3.6	100	41.6	20.1	

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP 比 ²	現役兵力	予備役	準軍隊
		100 万ドル	ドル	%	1,000 人	1,000 人	1,000 人
モンゴル		23	9	2.2	9.1	137	7.2
ミャンマー		2837	58	5	444	0	100.3
ニュージーランド		630	164	1.2	8.7	5.9	0
バプアニューギニア		13	3	0.5	3.1	0	0
フィリピン		1511	19	2.1	106	131	44
シンガポール		4334	1010	5.2	60.5	312.5	96.3
台湾		7479	336	2.7	370	1657	26.7
タイ		1730	27	1.5	306	200	113
ベトナム		2286	29	7.1	484	3000	40
バハマ	カリブ	25	9	0.5	0.9	0	0
キューバ		1040	93	3.9	46	39	26.5
ドミニカ		153	18	0.7	24.5	0	15
ハイチ		31	4	1	0	0	5.3
ジャマイカ		35	13	0.5	2.8	1	0.2
トリニダード・トバゴ		64	49	0.7	2.7	0	0
ベリーズ	中央アメリカ	17	70	2.3	1.1	0.7	0
コスタリカ		89	21	0.6	0	0	8.4
エルサルバドル		155	24	1.2	16.8	0	12
グアテマラ		181	15	0.8	31.4	35	19
ホンジュラス		108	16	1.7	8.3	60	6
メキシコ		5319	52	0.9	192.8	300	11
ニカラグア		31	6	1.3	14	0	0
パナマ		131	44	1.3	0	0	11.8
アルゼンチン	ラテンアメリカ	1386	36	1.4	69.9	0	31.2
ボリビア		119	14	1.6	31.5	0	37.1
ブラジル		9651	55	2.3	287.6	1115	385.6
チリ		2557	160	4.1	80.5	50	36.8
コロンビア		2840	64	3.7	158	60.7	104.6
エクアドル		685	52	3	59.5	100	0.3
ガイアナ		5	6	0.7	1.6	1.5	1.5
パラグアイ		54	9	1	18.6	164.5	14.8
ペルー		865	33	1.6	110	188	77
スリナム		8	19	5	2	0	0
ウルグアイ		212	62	1.8	23.9	0	0.9
ベネズエラ		1081	43	1.3	82.3	8	23
ジブチ	サハラ以南 のアフリカ	21	34	3.8	9.8	0	2.5
エチオピア		442	7	8	252.5	0	0
スーダン		629	19	4.9	117	0	7
ブルンジ		38	6	5.9	45.5	0	5.5
カメルーン		122	8	1.2	23.1	0	9
カーボヴェルデ		8	19	3.2	1.2	0	0.1
中央アフリカ共和国		20	5	2.1	2.6	0	1
チャド		14	2	0.8	30.4	0	4.5
コンゴ		88	28	3.1	10	0	2

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	国防費			兵力		
		国防支出	一人当たり ¹	GDP 比 ²	現役兵力	予備役	準軍隊
		100 万ドル	ドル	%	1,000 人	1,000 人	1,000 人
コンゴ民主共和国		946	18	21.7	81.4	0	0
赤道ギニア		4	10	0.2	2.3	0	0
ガボン		75	58	1.7	4.7	0	2
ルワンダ		68	9	4.1	70	0	10
ケニア		348	11	3.2	24.4	0	5
マダガスカル		45	3	1	13.5	0	8.1
セーシェル		11	140	1.7	0.5	0	0.3
タンザニア		127	3	1.5	27	80	1.4
ウガンダ		158	6	2.7	55	0	1.8
ベナン		46	7	1.8	4.6	0	2.5
ブルキナファソ		41	3	1.5	10.2	0	4.2
コートジボアール		136	8	1.4	17	10	8.9
ガンビア		3	2	1	0.8	0	0
ガーナ		30	1	0.5	7	0	0
ギニア		57	7	1.8	9.7	0	9.6
ギニアビサウ		3	3	1.6	9.3	0	2
リベリア		24	7	4.5	15	0	0
マリ		68	6	2.3	7.4	0	7.8
ニジェール		33	3	1.6	5.3	0	5.4
ナイジェリア		546	5	1.2	78.5	0	82
セネガル		65	7	1.3	9.4	0	5.8
シエラレオネ		17	4	2.2	13	0	0
トーゴ		23	5	1.8	9.5	0	0.8
アンゴラ		946	68	9.8	100	0	10
ボツワナ		254	157	4.5	9	0	1.5
馬拉ウイ		12	1	0.7	5.3	0	1.5
モザンビーク		76	4	2	11	0	0
南アフリカ		1697	37	1.7	60	73.4	0
ザンビア		25	2	0.7	21.6	0	1.4
ジンバブエ		637	56	3.4	36	0	21.8

2003 年における世界各国の軍事データ（陸上戦力および海上戦力）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
アメリカ	NATO	8023	0	7116	1349	72	118	21
カナダ		114	0	0	0	2	16	14
ベルギー		143	0	236	46	0	3	0
チェコ共和国		541	0	880	0	0	0	0
デンマーク		220	0	0	12	4	3	27
フランス		614	0	384	292	10	34	35
ドイツ		2398	0	2122	199	12	13	20
ギリシャ		1723	0	501	20	8	14	40
ハンガリー		743	0	490	0	0	0	0
イタリア		1183	0	27	45	6	18	19
ルクセンブルク		0	0	0	0	0	0	0
オランダ		283	0	345	0	4	15	0
ノルウェー		170	0	53	0	6	3	15
ポーランド		974	0	1248	43	4	4	23
ポルトガル		187	0	0	0	2	6	28
スペイン		552	0	122	28	6	16	37
トルコ		4205	0	650	37	13	19	49
イギリス		543	0	575	126	15	34	23
ロシア		ヨーロッパ (NATO 以外)	21870	150	14700	1000	53	32
アルバニア	373		0	0	0	0	0	20
アルメニア	110		0	104	0	0	0	0
オーストリア	114		137	0	0	0	0	0
アゼルバイジャン	320		0	226	0	0	0	6
ベラルーシ	1586		0	1588	0	0	0	0
ボスニア・ヘルツェゴビナ	192		1	35	0	0	0	0
ブルガリア	1474		0	214	0	1	1	23
クロアチア	287		0	104	0	1	0	7
キプロス	104		0	43	12	0	0	0
エストニア	0		0	0	0	0	0	3
フィンランド	235		0	276	0	0	0	9
マケドニア	115		0	11	12	0	0	0
グルジア	86		0	78	0	0	0	11
アイルランド	14		14	0	0	0	0	8
ラトビア	3		0	0	0	0	4	0
リトアニア	0		0	0	0	0	0	5
マルタ	0		0	0	0	0	0	0
モルドバ	0		0	45	0	0	0	0
ルーマニア	1258		0	177	0	0	7	38
スロバキア	271		0	404	0	0	0	0
スロベニア	70		0	26	8	0	0	0
スウェーデン	280		0	1623	0	7	0	36
スイス	556		0	435	0	0	0	0

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
ウクライナ		3784	0	3043	205	1	3	8
ユーゴスラビア/ セルビア/モンテネグロ		931	0	525	0	4	3	31
アルジェリア	中東・ 北アフリカ	1000	0	989	0	2	3	25
バーレーン		140	0	25	0	0	1	10
エジプト		3655	0	1455	0	4	11	47
イラン		1565	80	750	20	3	3	56
イラク		2600	0	1200	100	0	0	6
イスラエル		3950	0	0	0	3	0	53
ヨルダン		1018	19	26	0	0	0	3
クウェート		368	0	450	0	0	0	10
レバノン		310	40	0	0	0	0	7
リビア		800	0	1000	0	1	1	9
モロッコ		744	100	115	0	0	2	27
オマーン		117	37	0	0	0	0	13
カタール		30	0	40	0	0	0	7
サウジアラビア		1055	0	970	12	0	8	26
シリア		4500	0	2200	0	0	2	18
チュニジア		84	54	0	0	0	0	19
アラブ首長国連邦		81	76	430	0	0	2	16
イエメン	790	0	200	0	0	0	11	
アフガニスタン	中央・ 南アジア	500	0	120	0	0	0	0
バングラディッシュ		180	40	0	0	0	5	33
インド		3898	90	1600	0	19	29	45
カザフスタン		930	0	573	0	0	0	0
キルギスタン		215	0	387	0	0	0	0
ネパール		0	0	0	0	0	0	0
パキスタン		2368	0	0	19	10	8	9
スリランカ		62	0	52	0	0	0	61
タジキスタン		40	0	34	0	0	0	0
トルクメニスタン		702	0	930	0	0	0	0
ウズベキスタン	340	0	405	0	0	0	0	
オーストラリア	東アジア・ 太平洋州	71	0	0	0	6	11	15
ブルネイ		0	20	0	0	0	0	6
中国		7180	1500	4500	0	69	63	368
フィジー		0	0	0	0	0	0	9
インドネシア		0	365	11	0	2	0	36
日本		1020	0	70	90	16	54	7
北朝鮮		3500	560	0	0	26	3	310
韓国		2330	0	40	117	20	39	84
ラオス		25	10	0	0	0	0	16
マレーシア		0	26	0	0	0	4	41
モンゴル	370	0	310	0	0	0	0	

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
ミャンマー		100	105	0	0	0	0	75
ニュージーランド		0	0	0	0	0	3	4
パプアニューギニア		0	0	0	0	0	0	4
フィリピン		0	65	85	0	0	1	58
シンガポール		80	350	272	0	2	0	24
台湾		926	905	225	62	4	32	59
タイ		333	354	0	3	0	13	115
ベトナム		1315	620	300	0	2	6	42
バハマ	カリブ	0	0	0	0	0	0	7
キューバ		900	0	0	0	0	0	5
ドミニカ		0	12	0	0	0	0	15
ハイチ		0	0	0	0	0	0	0
ジャマイカ		0	0	0	0	0	0	5
トリニダード・トバコ		0	0	0	0	0	0	12
ベリーズ	中央アメリカ	0	0	0	0	0	0	0
コスタリカ		0	0	0	0	0	0	0
エルサルバドル		0	0	0	0	0	0	5
グアテマラ		0	0	0	0	0	0	9
ホンジュラス		0	12	0	0	0	0	16
メキシコ		0	0	0	0	0	11	109
ニカラグア		127	10	0	0	0	0	5
パナマ		0	0	0	0	0	0	14
アルゼンチン	ラテンアメリカ	200	150	105	0	3	13	14
ボリビア		0	36	0	0	0	0	60
ブラジル		178	286	0	0	4	20	47
チリ		260	0	20	0	2	6	27
コロンビア		0	12	0	0	4	4	27
エクアドル		30	108	0	0	2	2	11
ガイアナ		0	0	0	0	0	0	1
パラグアイ		12	0	0	0	0	0	10
ペルー		275	110	0	0	6	5	10
スリナム		0	0	0	0	0	0	3
ウルグアイ		15	68	15	0	0	3	8
ベネズエラ		81	191	0	7	2	6	6
ジブチ	サハラ以南 のアフリカ	0	0	0	0	0	0	7
エチオピア		270	0	0	0	0	0	0
スーダン		200	70	75	0	0	0	6
ブルンジ		0	0	0	0	0	0	0
カメルーン		0	0	14	0	0	0	11
カーボヴェルデ		0	0	0	0	0	0	0
中央アフリカ共和国		3	0	0	0	0	0	0
チャド		60	0	0	0	0	0	0
コンゴ		40	10	0	0	0	0	3

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	陸上戦力				海上戦力		
		戦車	軽戦車	装甲歩兵 戦闘車	攻撃ヘリ	潜水艦	主要水上 戦闘艦艇	哨戒艦艇 沿岸戦闘艦艇
		両	両	両	機	隻	隻	隻
コンゴ民主共和国		60	0	0	0	0	0	2
赤道ギニア		0	0	0	0	0	0	2
ガボン		0	0	12	0	0	0	2
ルワンダ		30	0	0	0	0	0	0
ケニア		78	0	0	0	0	0	4
マダガスカル		0	12	0	0	0	0	0
セーシェル		0	0	0	0	0	0	0
タンザニア		45	55	0	0	0	0	6
ウガンダ		190	20	0	0	0	0	0
ベナン		0	18	0	0	0	0	1
ブルキナファソ		0	0	0	0	0	0	0
コートジボアール		10	5	10	0	0	0	2
ガンビア		0	0	0	0	0	0	0
ガーナ		0	0	50	0	0	0	6
ギニア		38	15	0	0	0	0	2
ギニアビサウ		10	15	0	0	0	0	3
リベリア		0	0	0	0	0	0	0
マリ		33	18	0	0	0	0	0
ニジェール		0	0	0	0	0	0	0
ナイジェリア		250	100	0	0	0	1	10
セネガル		0	0	0	0	0	0	10
シエラレオネ		0	0	0	0	0	0	5
トーゴ		2	9	20	0	0	0	2
アンゴラ		400	0	400	0	0	0	7
ボツワナ		0	60	0	0	0	0	0
マラウイ		0	0	0	0	0	0	0
モザンビーク		80	0	40	0	0	0	0
南アフリカ		168	0	1200	0	2	0	8
ザンビア		30	30	0	0	0	0	0
ジンバブエ		40	0	0	0	0	0	0

2003 年における世界各国の軍事データ（航空戦力）

国 名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
アメリカ	NATO	1705	693	3513	0
カナダ		0	0	140	0
ベルギー		0	0	90	0
チェコ共和国		0	0	34	34
デンマーク		0	0	68	0
フランス		58	30	478	0
ドイツ		65	22	376	0
ギリシャ		0	18	418	0
ハンガリー		0	0	37	49
イタリア		18	80	263	6
ルクセンブルク		0	0	0	0
オランダ		10	21	137	30
ノルウェー		0	0	61	0
ポーランド		26	12	224	0
ポルトガル		0	0	50	0
スペイン		17	37	186	0
トルコ		0	16	483	0
イギリス		34	178	415	0
ロシア		ヨーロッパ (NATO 以外)	217	102	2151
アルバニア	0		0	26	0
アルメニア	0		0	8	10
オーストリア	0		0	52	11
アゼルバイジャン	0		0	47	15
ベラルーシ	0		0	210	50
ボスニア・ヘルツェゴビナ	0		0	0	0
ブルガリア	0		10	177	25
クロアチア	0		0	24	10
キプロス	0		0	0	0
エストニア	0		0	0	0
フィンランド	0		0	63	0
マケドニア	0		0	0	0
ゲルジア	0		0	7	3
アイルランド	0		0	0	0
ラトビア	0		0	0	0
リトアニア	0		0	0	0
マルタ	0		0	0	0
モルドバ	0		0	0	0
ルーマニア	0		0	106	0
スロバキア	0		0	71	19
スロベニア	0		0	0	0
スウェーデン	0		0	207	0
スイス	0		0	123	0

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
ウクライナ		0	13	499	0
ユーゴスラビア/ セルビア/モンテネグロ		0	0	103	44
アルジェリア	中東・ 北アフリカ	0	0	175	93
バーレーン		0	0	34	40
エジプト		0	24	579	121
イラン		5	19	306	0
イラク		0	0	316	0
イスラエル		0	0	438	100
ヨルダン		0	0	101	22
クウェート		0	0	81	20
レバノン		0	0	6	5
リビア		0	7	400	41
モロッコ		0	0	95	24
オマーン		0	0	40	0
カタール		0	0	18	19
サウジアラビア		0	21	294	0
シリア		0	16	548	71
チュニジア		0	0	29	15
アラブ首長国連邦		0	0	106	59
イエメン	0	0	76	8	
アフガニスタン	中央・ 南アジア	0	0	5	5
バングラディシュ		0	0	83	0
インド		35	50	744	40
カザフスタン		0	0	164	2
キルギスタン		0	0	52	9
ネパール		0	0	0	0
パキスタン		6	9	374	0
スリランカ		0	0	22	24
タジキスタン		0	0	0	4
トルクメニスタン		0	0	89	0
ウズベキスタン	0	0	135	42	
オーストラリア	東アジア・ 太平洋州	0	16	161	0
ブルネイ		0	0	0	5
中国		700	45	1900	0
フィジー		0	0	0	0
インドネシア		0	17	90	2
日本		80	102	270	0
北朝鮮		0	0	605	24
韓国		16	43	538	0
ラオス		0	0	24	0
マレーシア		0	6	95	0
モンゴル		0	0	0	13

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第6章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
ミャンマー		0	0	123	29
ニュージーランド		0	5	6	0
パプアニューギニア		0	0	0	0
フィリピン		0	0	44	87
シンガポール		0	0	126	28
台湾		32	20	479	0
タイ		44	8	194	0
ベトナム		0	0	189	26
バハマ	カリブ	0	0	0	0
キューバ		0	0	130	45
ドミニカ		0	0	16	0
ハイチ		0	0	0	0
ジャマイカ		0	0	0	0
トリニダード・トバゴ		0	0	0	0
ベリーズ	中央アメリカ	0	0	0	0
コスタリカ		0	0	0	0
エルサルバドル		0	0	21	5
グアテマラ		0	0	10	12
ホンジュラス		0	0	18	0
メキシコ		8	0	107	71
ニカラグア		0	0	0	15
パナマ		0	0	0	0
アルゼンチン	ラテンアメリカ	25	21	128	28
ボリビア		0	0	37	16
ブラジル		24	54	254	0
チリ		10	11	79	0
コロンビア		0	0	57	23
エクアドル		0	0	92	0
ガイアナ		0	0	0	0
パラグアイ		0	0	13	0
ペルー		0	9	118	25
スリナム		0	0	7	0
ウルグアイ		1	0	31	0
ベネズエラ		3	9	125	31
ジブチ	サハラ以南 の 아프리카	0	0	0	0
エチオピア		0	0	50	25
スーダン		0	0	27	10
ブルンジ		0	0	0	0
カメルーン		0	0	15	7
カーボヴェルデ		0	0	0	0
中央アフリカ共和国		0	0	0	0
チャド		0	0	2	2
コンゴ		0	0	12	0

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

国名	地域・組織	航空戦力			
		作戦機 (海軍)	武装ヘリ (海軍)	作戦機 (空軍)	武装ヘリ (空軍)
		機	機	機	機
コンゴ民主共和国		0	0	8	0
赤道ギニア		0	0	0	0
ガボン		0	0	10	5
ルワンダ		0	0	0	5
ケニア		0	0	29	34
マダガスカル		0	0	12	0
セーシェル		0	0	0	0
タンザニア		0	0	19	0
ウガンダ		0	0	15	6
ベナン		0	0	0	0
ブルキナファソ		0	0	5	0
コートジボアール		0	0	9	8
ガンビア		0	0	0	0
ガーナ		0	0	19	0
ギニア		0	0	8	0
ギニアビサウ		0	0	3	0
リベリア		0	0	0	0
マリ		0	0	16	0
ニジェール		0	0	0	0
ナイジェリア		0	0	84	15
セネガル		0	0	8	0
シエラレオネ		0	0	0	0
トーゴ		0	0	16	0
アンゴラ		0	0	105	16
ボツワナ		0	0	30	0
マラウイ		0	0	0	0
モザンビーク		0	0	0	4
南アフリカ		0	0	84	12
ザンビア		0	0	63	0
ジンバブエ		0	0	50	12

B.2 クラスタリング結果 ($c = 2 \sim c = 4$): 1985 年

2つのクラスターに分類した結果

Cluster 1 (58)			Cluster 2 (80)		
アメリカ	クウェート	アンゴラ	ルクセンブルク	トリニダード・トバゴ	ケニア
カナダ	リビア	南アフリカ	ポルトガル	ベリーズ	マダガスカル
ベルギー	モロッコ		アルバニア	コスタリカ	セーシェル
デンマーク	オマーン		オーストリア	エルサルバドル	タンザニア
フランス	サウジアラビア		キプロス	グアテマラ	ウガンダ
(東・西)ドイツ	シリア		アイルランド	ホンジュラス	ベナン
ギリシャ	アラブ首長国連邦		マルタ	ニカラグア	ブルキナ・ファソ
ハンガリー	インド		バーレーン	パナマ	コートジボアール
イタリア	パキスタン		ヨルダン	ボリビア	ガンビア
オランダ	オーストラリア		レバノン	コロンビア	ガーナ
ノルウェー	中国		カタール	エクアドル	ギニア
ポーランド	インドネシア		チュニジア	ガイアナ	ギニア・ビサウ
スペイン	日本		(南・北)イエメン	パラグアイ	リベリア
トルコ	北朝鮮		アフガニスタン	スリナム	マリ
イギリス	韓国		バングラディシュ	ウルグアイ	ニジェール
ソ連	マレーシア		ネパール	ジブチ	ナイジェリア
ブルガリア	ミャンマー		スリランカ	エチオピア	セネガル
チェコスロバキア	シンガポール		ブルネイ	スーダン	シエラレオネ
フィンランド	台湾		フィジー	ブルンジ	トーゴ
ルーマニア	タイ		ラオス	カメルーン	ボツワナ
スウェーデン	ベトナム		モンゴル	カーボ・ヴェルデ	マラウイ
スイス	キューバ		ニュージーランド	中央アフリカ共和国	モザンビーク
ユーゴスラビア	メキシコ		バプアニューギニア	チャド	ザンビア
アルジェリア	アルゼンチン		フィリピン	コンゴ	ジンバブエ
エジプト	ブラジル		バハマ	コンゴ民主共和国	
イラン	チリ		ドミニカ	赤道ギニア	
イラク	ペルー		ハイチ	ガボン	
イスラエル	ベネズエラ		ジャマイカ	ルワンダ	

図 B.1: 1985 年における各国の軍事データを $c = 2$ でクラスタリングした結果

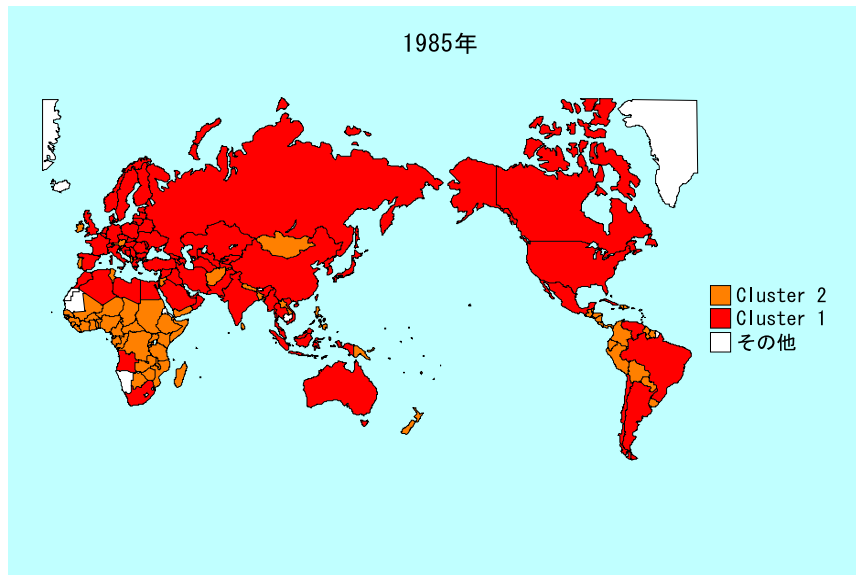


図 B.2: 1985 年における各国の軍事データを $c = 2$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.1 と同じである。

3つのクラスターに分類した結果

Cluster 1 (25)	Cluster 2 (35)		Cluster 3 (78)		
アメリカ	ベルギー	キューバ	ルクセンブルク	コスタリカ	セーシェル
カナダ	デンマーク	メキシコ	ポルトガル	エルサルバドル	タンザニア
フランス	ギリシャ	チリ	アルバニア	グアテマラ	ウガンダ
(東・西)ドイツ	ハンガリー	ペルー	キプロス	ホンジュラス	ベナン
イタリア	ノルウェー	ベネズエラ	アイルランド	ニカラグア	ブルキナ・ファソ
オランダ	トルコ	アンゴラ	マルタ	パナマ	コートジボアール
ポーランド	オーストリア	南アフリカ	バーレーン	ボリビア	ガンビア
スペイン	ブルガリア		レバノン	コロンビア	ガーナ
イギリス	チェコスロバキア		カタール	エクアドル	ギニア
ソ連	フィンランド		チュニジア	ガイアナ	ギニア・ビサウ
エジプト	ルーマニア		(南・北)イエメン	パラグアイ	リベリア
イラン	スウェーデン		アフガニスタン	スリナム	マリ
イラク	スイス		バングラディシュ	ウルグアイ	ニジェール
イスラエル	ユーゴスラビア		ネパール	ジブチ	ナイジェリア
サウジアラビア	アルジェリア		スリランカ	エチオピア	セネガル
シリア	ヨルダン		ブルネイ	スーダン	シエラレオネ
インド	クウェート		フィジー	ブルンジ	トーゴ
オーストラリア	リビア		ラオス	カメルーン	ボツワナ
中国	モロッコ		モンゴル	カーボ・ヴェルデ	マラウイ
日本	オマーン		ニュージーランド	中央アフリカ共和国	モザンビーク
北朝鮮	アラブ首長国連邦		バプアニューギニア	チャド	ザンビア
台湾	パキスタン		フィリピン	コンゴ	ジンバブエ
ベトナム	インドネシア		バハマ	コンゴ民主共和国	
アルゼンチン	韓国		ドミニカ	赤道ギニア	
ブラジル	マレーシア		ハイチ	ガボン	
	ミャンマー		ジャマイカ	ルワンダ	
	シンガポール		トリニダード・トバゴ	ケニア	
	タイ		ベリーズ	マダガスカル	

図 B.3: 1985 年における各国の軍事データを $c = 3$ でクラスタリングした結果

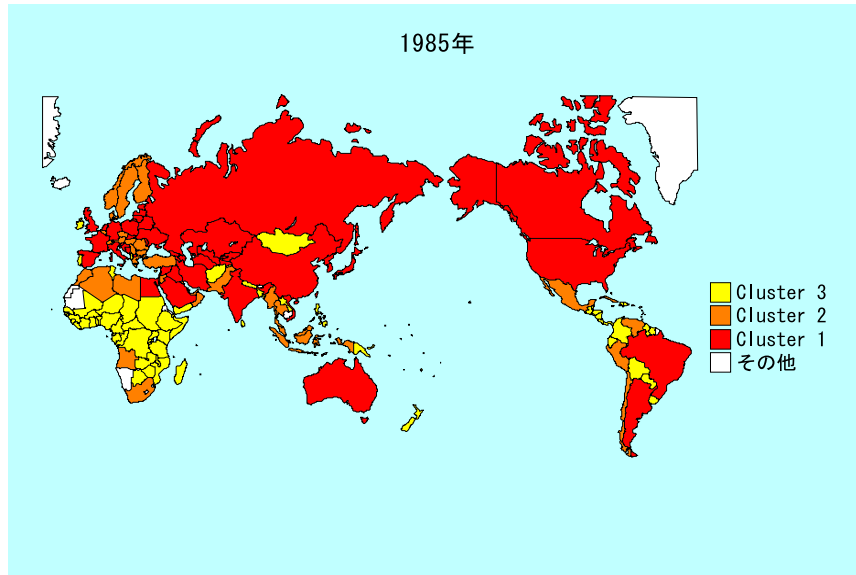


図 B.4: 1985 年における各国の軍事データを $c = 3$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.3 と同じである。

4 つのクラスターに分類した結果

Cluster 1 (13)	Cluster 2 (13)	Cluster 3 (34)		Cluster 4 (78)		
アメリカ	フランス	ベルギー	メキシコ	ルクセンブルク	コスタリカ	セーシェル
カナダ	オランダ	デンマーク	チリ	ポルトガル	エルサルバドル	タンザニア
(東・西)ドイツ	ポーランド	ギリシャ	ペルー	アルバニア	グアテマラ	ウガンダ
イタリア	スペイン	ハンガリー	ベネズエラ	キプロス	ホンジュラス	ベナン
イギリス	エジプト	ノルウェー	アンゴラ	アイルランド	ニカラグア	ブルキナ・ファソ
ソ連	イラン	トルコ	南アフリカ	マルタ	パナマ	コートジボアール
イラク	イスラエル	オーストリア		バーレーン	ボリビア	ガンビア
サウジアラビア	オーストラリア	ブルガリア		レバノン	コロンビア	ガーナ
シリア	日本	チェコスロバキア		カタール	エクアドル	ギニア
インド	韓国	フィンランド		チュニジア	ガイアナ	ギニア・ビサウ
中国	台湾	ルーマニア		(南・北)イエメン	パラグアイ	リベリア
北朝鮮	ベトナム	スウェーデン		アフガニスタン	スリナム	マリ
ブラジル	アルゼンチン	スイス		バングラディシュ	ウルグアイ	ニジェール
		ユーゴスラビア		ネパール	ジブチ	ナイジェリア
		アルジェリア		スリランカ	エチオピア	セネガル
		ヨルダン		ブルネイ	スーダン	シエラレオネ
		クウェート		フィジー	ブルンジ	トーゴ
		リビア		ラオス	カメルーン	ボツワナ
		モロッコ		モンゴル	カーボ・ヴェルデ	マラウイ
		オマーン		ニュージーランド	中央アフリカ共和国	モザンビーク
		アラブ首長国連邦		バプアニューギニア	チャド	ザンビア
		パキスタン		フィリピン	コンゴ	ジンバブエ
		インドネシア		バハマ	コンゴ民主共和国	
		マレーシア		ドミニカ	赤道ギニア	
		ミャンマー		ハイチ	ガボン	
		シンガポール		ジャマイカ	ルワンダ	
		タイ		トリニダード・トバゴ	ケニア	
		キューバ		ベリーズ	マダガスカル	

図 B.5: 1985 年における各国の軍事データを $c = 4$ でクラスタリングした結果

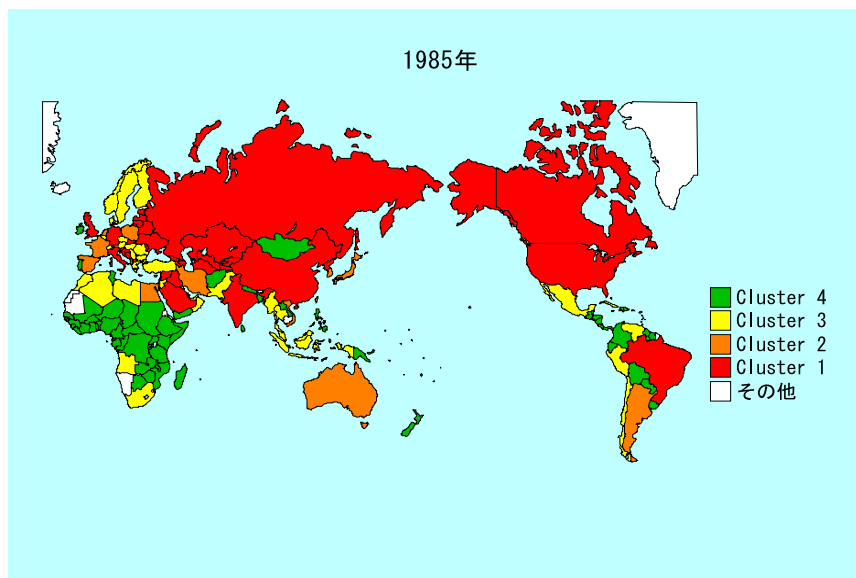


図 B.6: 1985 年における各国の軍事データを $c = 4$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.5 と同じである。

B.3 クラスタリング結果 ($c = 2 \sim c = 4$): 2003 年

2 つのクラスターに分類した結果

Cluster 1 (66)			Cluster 2 (91)			
アメリカ	イラン	キューバ	ルクセンブルク	フィジー	カメルーン	マラウイ
カナダ	イラク	メキシコ	アルバニア	ラオス	カーボベルデ	モザンビーク
ベルギー	イスラエル	アルゼンチン	アルメニア	モンゴル	中央アフリカ共和国	ザンビア
チェコ共和国	ヨルダン	ブラジル	アゼルバイジャン	ニュージーランド	チャド	ジンバブエ
デンマーク	クウェート	チリ	ボスニア・ヘルツェゴビナ	ババアニューギニア	コンゴ	
フランス	リビア	コロンビア	クロアチア	バハマ	コンゴ民主共和国	
ドイツ	モロッコ	アンゴラ	キプロス	ドミニカ	赤道ギニア	
ギリシャ	オマーン	南アフリカ	エストニア	ハイチ	ガボン	
ハンガリー	カタール		マケドニア	ジャマイカ	ルワンダ	
イタリア	サウジアラビア		グルジア	トリニダード・トバゴ	ケニア	
オランダ	シリア		アイルランド	ベリーズ	マダガスカル	
ノルウェー	アラブ首長国連邦		ラトビア	コスタリカ	セーシェル	
ポーランド	インド		リトアニア	エルサルバドル	タンザニア	
ポルトガル	カザフスタン		マルタ	グアテマラ	ウガンダ	
スペイン	パキスタン		モルドバ	ホンジュラス	ベナン	
トルコ	ウズベキスタン		スロバキア	ニカラガ	ブルキナファソ	
イギリス	オーストラリア		スロベニア	パナマ	コートジボアール	
ロシア	中国		バーレーン	ボリビア	ガンビア	
オーストリア	インドネシア		レバノン	エクアドル	ガーナ	
ベラルーシ	日本		チュニジア	ガイアナ	ギニア	
ブルガリア	北朝鮮		イエメン	パラグアイ	ギニアビサウ	
フィンランド	韓国		アフガニスタン	ベルー	リベリア	
ルーマニア	マレーシア		バングラディシュ	スリナム	マリ	
スウェーデン	ミャンマー		キルギスタン	ウルグアイ	ニジェール	
スイス	フィリピン		ネパール	ベネズエラ	ナイジェリア	
ウクライナ	シンガポール		スリランカ	ジブチ	セネガル	
ユーゴスラビア	台湾		タジキスタン	エチオピア	シエラレオネ	
アルジェリア	タイ		トルクメニスタン	スーダン	トーゴ	
エジプト	ベトナム		ブルネイ	ブルンジ	ボツワナ	

図 B.7: 2003 年における各国の軍事データを $c = 2$ でクラスタリングした結果

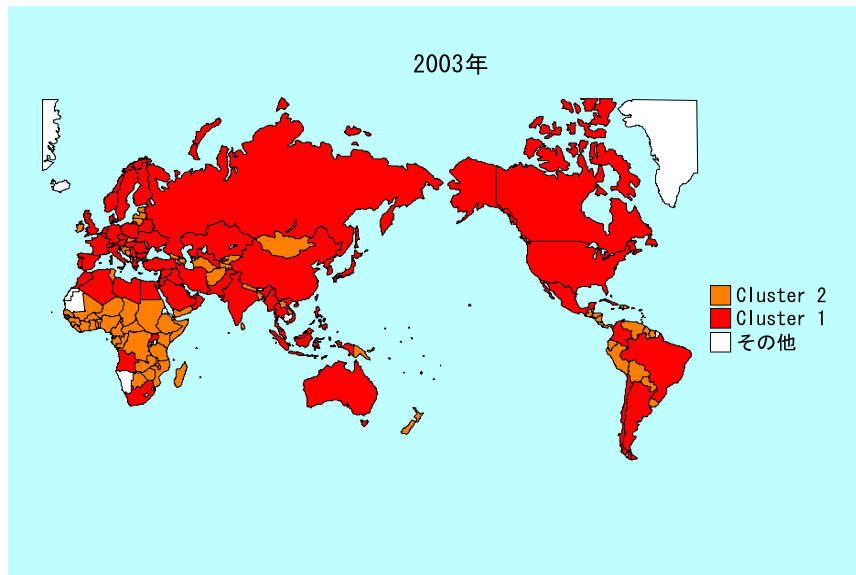


図 B.8: 2003 年における各国の軍事データを $c = 2$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.7 と同じである。

3つのクラスターに分類した結果

Cluster 1 (48)		Cluster 2 (28)		Cluster 3 (85)	
アメリカ	シリア	チェコ共和国	ルクセンブルク	ニュージーランド	赤道ギニア
カナダ	アラブ首長国連邦	ハンガリー	アルバニア	バブアニューギニア	ガボン
ベルギー	インド	オーストリア	アルメニア	バハマ	ルワンダ
デンマーク	パキスタン	アゼルバイジャン	ボスニア・ヘルツェゴビナ	ドミニカ	ケニア
フランス	オーストラリア	ブルガリア	クロアチア	ハイチ	マダガスカル
ドイツ	中国	フィンランド	キプロス	ジャマイカ	セーシェル
ギリシャ	インドネシア	ルーマニア	エストニア	トリニダード・トバゴ	タンザニア
イタリア	日本	ユーゴスラビア	マケドニア	ベリーズ	ウガンダ
オランダ	北朝鮮	ヨルダン	グルジア	コスタリカ	ベナン
ノルウェー	韓国	リビア	アイルランド	エルサルバドル	ブルキナファソ
ポーランド	マレーシア	モロッコ	ラトビア	グアテマラ	コートジボアール
ポルトガル	ミャンマー	イエメン	リトアニア	ホンジュラス	ガンビア
スペイン	シンガポール	カザフスタン	マルタ	ニカラグア	ガーナ
トルコ	台湾	トルクメニスタン	モルドバ	パナマ	ギニア
イギリス	ベトナム	ウズベキスタン	スロバキア	ボリビア	ギニアビサウ
ロシア	メキシコ	フィリピン	スロベニア	エクアドル	リベリア
ベラルーシ	ブラジル	タイ	バレーン	ガイアナ	マリ
スウェーデン	チリ	キューバ	レバノン	パラグアイ	ニジェール
スイス	コロンビア	アルゼンチン	チュニジア	スリナム	ナイジェリア
ウクライナ		ペルー	アフガニスタン	ウルグアイ	セネガル
アルジェリア		ベネズエラ	バングラディシュ	ジブチ	シエラレオネ
エジプト		コンゴ民主共和国	キルギスタン	エチオピア	トーゴ
イラン		アンゴラ	ネパール	スーダン	ボツワナ
イラク		南アフリカ	スリランカ	ブルンジ	マラウイ
イスラエル			タジキスタン	カメルーン	モザンビーク
クウェート			ブルネイ	カーボベルデ	ザンビア
オマーン			フィジー	中央アフリカ共和国	ジンバブエ
カタール			ラオス	チャド	
サウジアラビア			モンゴル	コンゴ	

図 B.9: 2003 年における各国の軍事データを $c = 3$ でクラスタリングした結果

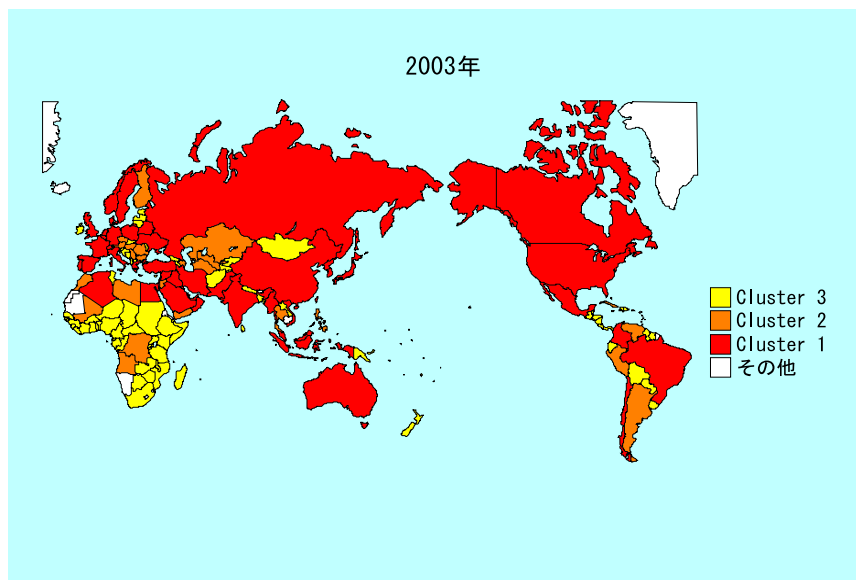


図 B.10: 2003 年における各国の軍事データを $c = 3$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.9 と同じである。

4 つのクラスターに分類した結果

Cluster 1 (28)	Cluster 2 (20)	Cluster 3 (24)	Cluster 4 (85)
アメリカ	ベルギー	チェコ共和国	ルクセンブルク
カナダ	デンマーク	ハンガリー	アルバニア
フランス	ノルウェー	オーストリア	アルメニア
ドイツ	ポーランド	アゼルバイジャン	ボスニア・ヘルツェゴビナ
ギリシャ	ポルトガル	ブルガリア	クロアチア
イタリア	トルコ	フィンランド	キプロス
オランダ	スウェーデン	ルーマニア	エストニア
スペイン	スイス	ユーゴスラビア	マケドニア
イギリス	アルジェリア	ヨルダン	グルジア
ロシア	イラン	リビア	アイルランド
ペルー	イスラエル	モロッコ	ラトビア
ウクライナ	クウェート	イエメン	リトアニア
エジプト	オマーン	カザフスタン	マルタ
イラク	アラブ首長国連邦	トルクメニスタン	モルドバ
カタール	インドネシア	ウズベキスタン	スロバキア
サウジアラビア	マレーシア	フィリピン	スロベニア
シリア	ミャンマー	タイ	バーレーン
インド	シンガポール	キューバ	レバノン
パキスタン	チリ	アルゼンチン	チュニジア
オーストラリア	コロンビア	ペルー	アフガニスタン
中国		ベネズエラ	バングラディシュ
日本		コンゴ民主共和国	キルギスタン
北朝鮮		アンゴラ	ネパール
韓国		南アフリカ	スリランカ
台湾			スリランカ
ベトナム			タジキスタン
メキシコ			ブルネイ
ブラジル			フィジー
			ラオス
			モンゴル
			ニュージーランド
			バプアニューギニア
			バハマ
			ハイチ
			ジャマイカ
			トリニダード・トバゴ
			ベリーズ
			コスタリカ
			エルサルバドル
			グアテマラ
			ホンジュラス
			ニカラグア
			パナマ
			ボリビア
			スロバキア
			エクアドル
			ガイアナ
			パラグアイ
			スリナム
			ウルグアイ
			ジブチ
			エチオピア
			スーダン
			ブルンジ
			カメルーン
			カーボベルデ
			中央アフリカ共和国
			チャド
			コンゴ
			赤道ギニア
			ガボン
			ルワンダ
			ケニア
			マダガスカル
			セーシェル
			タンザニア
			ウガンダ
			ベナン
			ブルキナファソ
			コートジボアール
			ガンビア
			ガーナ
			ギニア
			ギニアビサウ
			リベリア
			マリ
			ニジェール
			ナイジェリア
			セネガル
			シエラレオネ
			トーゴ
			ボツワナ
			マラウイ
			モザンビーク
			ザンビア
			ジンバブエ

図 B.11: 2003 年における各国の軍事データを $c = 4$ でクラスタリングした結果

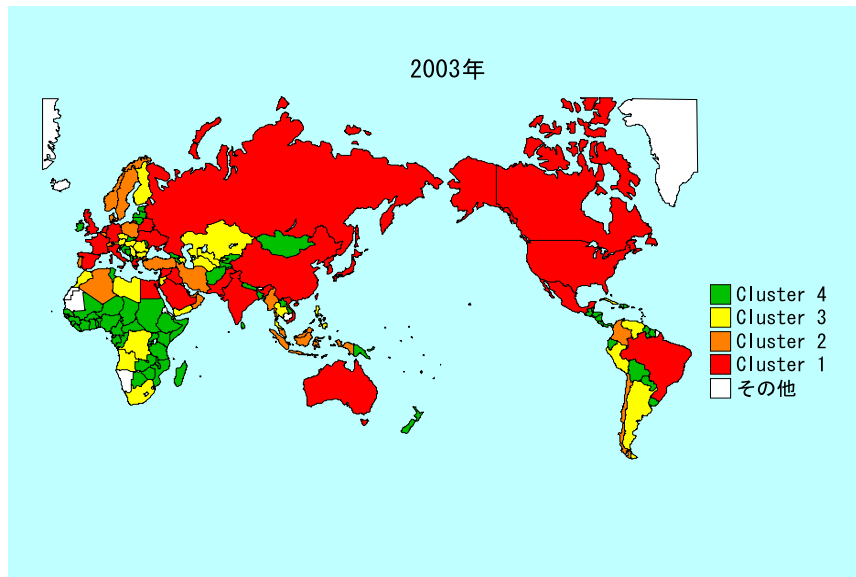


図 B.12: 2003 年における各国の軍事データを $c = 4$ でクラスタリングした結果を世界地図にプロットした図。各クラスターの色は図 B.11 と同じである。

B.4 各変数における上位 10ヶ国

B.4.1 1985 年

国防支出費 (100 万ドル)			国民一人当たり (ドル)			GDP に占める割合 (%)		
1	アメリカ	380899	1	サウジアラビア	2728	1	イラク	42.2
2	ソ連	368327	2	イスラエル	2709	2	イスラエル	21.2
3	イギリス	35003	3	アラブ首長国連邦	2003	3	オマーン	20.8
4	サウジアラビア	31485	4	オマーン	1786	4	ポーランド	20.5
5	フランス	30659	5	ブルネイ	1739	5	北朝鮮	20.0
6	(東・西)ドイツ	29393	6	アメリカ	1592	6	サウジアラビア	19.6
7	ポーランド	21644	7	ソ連	1321	7	ベトナム	19.4
8	中国	21616	8	クウェート	1295	8	ニカラグア	17.3
9	日本	20139	9	イラク	1174	9	シリア	16.4
10	イラク	18664	10	カタール	854	10	ソ連	16.1

現役兵力 (1,000 人)			予備役 (1,000 人)			準軍隊 (1,000 人)		
1	ソ連	5300	1	ソ連	5400	1	中国	12000
2	中国	3900	2	中国	5377	2	韓国	4100
3	アメリカ	2152	3	韓国	4822	3	北朝鮮	3108
4	インド	1260	4	ベトナム	3000	4	イラン	2570
5	ベトナム	1027	5	台湾	1470	5	ベトナム	1560
6	イラク	1000	6	(東・西)ドイツ	1420	6	キューバ	1419
7	北朝鮮	838	7	ブラジル	1340	7	イラク	655
8	トルコ	630	8	アメリカ	1212	8	ソ連	600
9	イラン	610	9	スペイン	1085	9	スイス	480
10	韓国	598	10	トルコ	936	10	マレーシア	372

戦車 (両)			軽戦車 (両)			装甲歩兵戦闘車 (両)		
1	ソ連	52600	1	台湾	1120	1	ソ連	27000
2	アメリカ	14139	2	ブラジル	575	2	(東・西)ドイツ	3544
3	中国	11450	3	ベトナム	450	3	アメリカ	2150
4	(東・西)ドイツ	6162	4	ベネズエラ	422	4	南アフリカ	1500
5	シリア	4200	5	フランス	342	5	チェコスロバキア	1100
6	イスラエル	3600	6	イギリス	271	6	フランス	780
7	チェコスロバキア	3500	7	ギリシャ	270	7	リビア	700
8	ポーランド	3450		シンガポール	270	8	アルジェリア	650
9	北朝鮮	3275	9	イラク	250	9	シリア	600
10	トルコ	2922	10	スウェーデン	200	10	イラク	500

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

攻撃ヘリコプター		(機)
1	アメリカ	1891
2	ソ連	1250
3	イギリス	314
4	フランス	68
5	イラク	65
6	日本	10
7	ギリシャ	8
8	パキスタン	2

潜水艦		(隻)
1	ソ連	371
2	中国	110
3	アメリカ	95
4	イギリス	28
5	(東・西)ドイツ	26
6	北朝鮮	20
7	フランス	18
8	トルコ	16
9	ノルウェー	14
	エジプト	14
	日本	14

主要水上戦闘艦艇		(隻)
1	ソ連	289
2	アメリカ	200
3	イギリス	60
4	日本	49
5	フランス	47
6	中国	44
7	台湾	36
8	(東・西)ドイツ	34
9	インド	28
10	カナダ	23
	イタリア	23
	スペイン	23

哨戒艦艇・沿岸戦闘艦艇		(隻)
1	タイ	94
2	アメリカ	89
3	スペイン	87
4	韓国	84
5	メキシコ	66
6	中国	60
7	ベトナム	54
8	ルーマニア	49
9	クウェート	48
10	ミャンマー	46
	コンゴ民主共和国	46

作戦機（海軍）		(機)
1	アメリカ	1350
2	ソ連	875
3	(東・西)ドイツ	122
4	フランス	104
5	日本	84
6	アルゼンチン	54
7	インド	36
8	イギリス	32
9	トルコ	20
10	インドネシア	19

武装ヘリコプター（海軍）		(機)
1	ソ連	310
2	アメリカ	160
3	イギリス	121
4	イタリア	83
5	スペイン	40
6	インド	26
7	アルゼンチン	24
8	フランス	23
9	オランダ	17
10	ブラジル	16

作戦機（空軍）		(機)
1	中国	5300
2	アメリカ	3700
3	ソ連	2350
4	(東・西)ドイツ	966
5	インド	846
6	北朝鮮	800
7	イスラエル	684
8	ポーランド	675
9	イギリス	599
10	台湾	567

武装ヘリコプター（空軍）		(機)
1	ソ連	2650
2	イラク	100
	シリア	100
4	(東・西)ドイツ	70
5	ベトナム	65
6	イスラエル	60
	インド	60
8	エジプト	48
9	リビア	42
	ペルー	42

B.4.2 2003 年

国防支出費 (100 万ドル)			国民一人当たり (ドル)			GDP に占める割合 (%)		
1	アメリカ	329616	1	カタール	2857	1	北朝鮮	25.0
2	中国	48380	2	クウェート	1582	2	コンゴ民主共和国	21.7
3	ロシア	48040	3	イスラエル	1499	3	オマーン	13.4
4	フランス	38005	4	アメリカ	1138	4	アフガニスタン	13.0
5	日本	37070	5	シンガポール	1010	5	サウジアラビア	12.0
6	イギリス	35249	6	サウジアラビア	981	6	クウェート	10.7
7	ドイツ	31465	7	アラブ首長国連邦	949	7	カタール	10.6
8	イタリア	24210	8	オマーン	944	8	シリア	10.3
9	サウジアラビア	20981	9	ノルウェー	759	9	アンゴラ	9.8
10	インド	13073	10	ブルネイ	741	10	イラク	9.7

現役兵力 (1,000 人)			予備役 (1,000 人)			準軍隊 (1,000 人)		
1	中国	2270	1	北朝鮮	4700	1	中国	1500
2	アメリカ	1414	2	韓国	4500	2	インド	1090
3	インド	1298	3	ベトナム	3000	3	ロシア	409
4	北朝鮮	1028	4	ロシア	2400	4	ブラジル	386
5	ロシア	988	5	台湾	1657	5	エジプト	330
6	韓国	686	6	アメリカ	1259	6	パキスタン	289
7	パキスタン	620	7	ブラジル	1115	7	イタリア	254
8	イラン	520	8	ウクライナ	1000	8	インドネシア	195
9	トルコ	515	9	イラク	650	9	北朝鮮	189
10	ベトナム	484	10	中国	550	10	アルジェリア	181

戦車 (両)			軽戦車 (両)			装甲歩兵戦闘車 (両)		
1	ロシア	21870	1	中国	1500	1	ロシア	14700
2	アメリカ	8023	2	台湾	905	2	アメリカ	7116
3	中国	7180	3	ベトナム	620	3	中国	4500
4	シリア	4500	4	北朝鮮	560	4	ウクライナ	3043
5	トルコ	4205	5	インドネシア	365	5	シリア	2200
6	イスラエル	3950	6	タイ	354	6	ドイツ	2122
7	インド	3898	7	シンガポール	350	7	スウェーデン	1623
8	ウクライナ	3784	8	ブラジル	286	8	インド	1600
9	エジプト	3655	9	ベネズエラ	191	9	ベラルーシ	1588
10	北朝鮮	3500	10	ロシア	150	10	エジプト	1455

付録 B クラスタリングによる日本周辺の軍事情勢の分析：参考資料（第 6 章）

攻撃ヘリコプター		(機)
1	アメリカ	1349
2	ロシア	1000
3	フランス	292
4	ウクライナ	205
5	ドイツ	199
6	イギリス	126
7	韓国	117
8	イラク	100
9	日本	90
10	台湾	62

潜水艦		(隻)
1	アメリカ	72
2	中国	69
3	ロシア	53
4	北朝鮮	26
5	韓国	20
6	インド	19
7	日本	16
8	イギリス	15
9	トルコ	13
10	ドイツ	12

主要水上戦闘艦艇		(隻)
1	アメリカ	118
2	中国	63
3	日本	54
4	韓国	39
5	フランス	34
	イギリス	34
7	ロシア	32
	台湾	32
9	インド	29
10	ブラジル	20

哨戒艦艇・沿岸戦闘艦艇		(隻)
1	中国	368
2	北朝鮮	310
3	タイ	115
4	メキシコ	109
5	ロシア	88
6	韓国	84
7	ミャンマー	75
8	スリランカ	61
9	ボリビア	60
10	台湾	59

作戦機（海軍）		(機)
1	アメリカ	1705
2	中国	700
3	ロシア	217
4	日本	80
5	ドイツ	65
6	フランス	58
7	タイ	44
8	インド	35
9	イギリス	34
10	台湾	32

武装ヘリコプター（海軍）		(機)
1	アメリカ	693
2	イギリス	178
3	ロシア	102
	日本	102
5	イタリア	80
6	ブラジル	54
7	インド	50
8	中国	45
9	韓国	43
10	スペイン	37

作戦機（空軍）		(機)
1	アメリカ	3513
2	ロシア	2151
3	中国	1900
4	インド	744
5	北朝鮮	605
6	エジプト	579
7	シリア	548
8	韓国	538
9	ウクライナ	499
10	トルコ	483

武装ヘリコプター（空軍）		(機)
1	ロシア	700
2	エジプト	121
3	イスラエル	100
4	アルジェリア	93
5	フィリピン	87
6	シリア	71
	メキシコ	71
8	アラブ首長国連邦	59
9	ベラルーシ	50
10	ハンガリー	49