

# 交友関係データと脳波データからの関連性自動抽出

谷 研究室 小林 彩  
Aya Kobayashi

## 概要

本研究では、小学生を対象に行った脳波簡易測定及び、友人関係やコンピュータゲームのプレイ時間等に関するアンケートにより収集されたデータに対して、データマイニング手法を用いて、脳波データパターンとプレイ時間の関係性の抽出を試みる。

## 1 はじめに

データマイニングとは、多量のデータから有用な知識を発掘するデータ処理技術の総称である。顧客管理情報、売り上げデータ等の管理にも用いられる。代表的な解析手法として頻出パターン抽出、クラス分類、回帰分析、クラスタリングが知られている [2, 3]。

近年、コンピュータ、携帯電話、スマートフォン、コンピュータゲーム機の普及に伴い、コンピュータゲームやコンピュータ操作等が人間の脳に与える影響について議論されており、影響がないという主張と、影響があるという主張の双方が存在する。又、影響があるという主張の中にも、好影響があるという主張 [1]、悪影響があるという主張 [4] が存在する。

本研究では、日本大学文理学部のプロジェクトチームが 2007 年 9 月に埼玉県内の小学生 5,6 年生 103 人を対象に行った脳波簡易測定および友人関係やコンピュータゲームのプレイ時間等のデータに対して、データマイニング手法を用いて、脳波データパターンとプレイ時間の関係性の抽出を試みる。

## 2 クラスタリング

本研究では、最短距離法、最遠点距離法、 $k$ -medoids 法 [2, 3] の 3 つのクラスタリング手法を採用した。最短距離法、最遠点距離法はクラスタリング手法の中でも階層的な手法の凝集型に分類される。また、 $k$ -medoids 法は分割最適化手法に分類される。各クラスタはデータの集合であり、データ  $d, d'$  の距離を  $\text{dis}(d, d')$  とする。クラスタ  $C, C'$  の最短距離を  $\min_{d \in C, d' \in C'} \text{dis}(d, d')$  とする。クラスタ間の距離として、最短距離法では最短距離を用いる。以下で、最短距離法、最遠点距離法、 $k$ -medoids 法それぞれのアルゴリズム概要を述べる。

### ◇ 最短距離法 .

1. 入力の  $n$  個のデータ集合  $D = \{d_1, \dots, d_n\}$  に対して、 $n$  個のクラスタ  $C_1 = \{d_1\}, \dots, C_n = \{d_n\}$  を作成し、 $C = \{C_1, \dots, C_n\}$  とする。
2. クラスタの各ペアの距離を求め最も距離の小さいクラスタのペア  $C_i, C_j$  を選択。
3.  $C$  から  $C_i, C_j$  を削除し、 $C$  に  $C_i \cup C_j$  を追加。
4.  $|C|$  が入力されたクラスタ数になるまで 2. から 4. を繰り返す。

### ◇ 最遠点法 .

1. 入力の  $n$  個のデータ集合  $D = \{d_1, \dots, d_n\}$  から任意の 1 点  $d_i$  をとり、 $C = \{d_i\}$  とする。
2.  $\max_{d \in D - C, d' \in C} \text{dis}(d, d')$  となる  $d$  を  $C$  に加える。
3.  $C$  のサイズが  $k$  未満なら 2. に戻る。
4.  $C$  の点に対応した  $k$  個のクラスタを作る。このとき、 $D$  の各点は、最も近い  $C$  の点に対応したクラスタに分類される。

### ◇ $k$ -medoids 法 .

1. 入力の  $n$  個のデータ集合  $D = \{d_1, \dots, d_n\}$  の各要素  $d$  に対してクラスタ  $C_1, \dots, C_k$  をランダムに選び  $d$  をそのクラスタの要素とする。
2. 各クラスタ  $C_i$  に対して  $\min_{d \in C_i} \sum_{d' \in C_i - d} \text{dis}(d, d')$  となる要素を 1 つ選び medoid  $m_i$  とする。
3.  $D = \{d_1, \dots, d_n\}$  の各要素  $d$  に対して  $\min_{i \in \{1, \dots, k\}} \text{dis}(d, m_i)$  となる要素を 1 つ選び  $d$  をクラスタ  $C_i$  の要素とする。
4. 属するクラスタが変わった要素が  $D = \{d_1, \dots, d_n\}$  に存在すれば 2. に戻る。

### 3 実験概要

小学生のゲームのプレイ時間と脳波データパターンをクラス数 3, 5, 10 で最短距離法,  $k$ -medoids 法, 最遠点法を用いてクラスタリングを行う。

#### 3.1 マイニング対象データ

日本大学文理学部のプロジェクトチームが 2007 年 9 月に埼玉県内の小学生 5,6 年生 103 人を対象に, 脳波簡易測定, および, コンピュータゲームのプレイ時間等のアンケート調査を実施した。コンピュータゲームのプレイ時間に関する質問項目は, 1 日あたりの携帯型ゲーム機の平均プレイ時間, 据え置き型ゲーム機の平均プレイ時間, 1 日の最長プレイ時間である。また, 日本大学文理学部のプロジェクトチームは, 簡易脳波測定で得られたデータを 3 つの脳波パターン  $P_1, P_2, P_3$  に分類した。ここで,  $i < j$  ならば, 脳波パターン  $P_i$  は脳波データ  $P_j$  より悪影響が少ない。

### 4 実験結果

各表において, 1 列目をクラス, 2 列目をクラスに属するデータの数, 3 列目から 5 列目をクラスに属する各脳波パターン  $P_1, P_2, P_3$  のデータの数 (上段) と割合の百分率 (下段) とする。ただし,  $i < j$  ならば, クラスタ  $C_i$  の平均プレイ時間はクラス  $C_i$  の平均プレイ時間以上とする。

	数	$P_1$	$P_2$	$P_3$
$C_1$	2	0	100	0
		0	2	0
$C_2$	3	0	33	67
		0	1	2
$C_3$	80	39	48	13
		31	38	11

表 1: 最遠点法  $k = 3$

	数	$P_1$	$P_2$	$P_3$
$C_1$	2	0	100	0
		0	2	0
$C_2$	3	0	33	67
		0	1	2
$C_3$	17	29	47	24
		5	8	4
$C_4$	5	60	40	0
		3	2	0
$C_5$	58	40	48	12
		23	28	12

表 2: 最遠点法  $k = 5$

	数	$P_1$	$P_2$	$P_3$
$C_1$	2	0	100	0
		0	2	0
$C_2$	2	0	50	50
		0	1	1
$C_3$	14	21	50	29
		3	7	4
$C_4$	5	60	40	0
		3	2	0
$C_5$	3	67	33	0
		2	1	0
$C_6$	4	75	25	0
		3	1	0
$C_7$	7	57	29	14
		4	2	1
$C_8$	1	0	0	100
		0	0	1
$C_9$	29	31	55	14
		9	16	4
$C_{10}$	18	39	50	11
		7	9	2

表 3: 最遠点法  $k = 10$

	数	$P_1$	$P_2$	$P_3$
$C_1$	1	0	100	0
		0	1	0
$C_2$	7	29	42	29
		2	3	2
$C_3$	77	14	48	38
		11	37	29

表 4: 最短距離法  $k = 3$

	数	$P_1$	$P_2$	$P_3$
$C_1$	1	0	100	0
		0	1	0
$C_2$	1	0	100	0
		0	1	0
$C_3$	5	40	20	40
		2	1	2
$C_4$	1	0	100	0
		0	1	0
$C_5$	77	37	48	14
		29	37	11

表 5: 最短距離法  $k = 5$

	数	$P_1$	$P_2$	$P_3$
$C_1$	1	100	0	0
		1	0	0
$C_2$	1	0	100	0
		0	1	0
$C_3$	1	100	0	0
		1	0	0
$C_4$	3	0	67	33
		0	2	1
$C_5$	1	0	100	0
		0	1	0
$C_6$	1	0	100	0
		0	1	0
$C_7$	2	50	50	0
		1	1	0
$C_8$	25	44	44	12
		11	11	3
$C_9$	3	33	0	67
		1	0	2
$C_{10}$	47	34	51	14
		16	24	7

表 6: 最短距離法  $k = 10$

	数	$P_1$	$P_2$	$P_3$
$C_1$	21	33	48	19
		7	10	4
$C_2$	30	37	43	20
		11	13	6
$C_3$	34	38	53	9
		13	18	3

表 7:  $k$ -medoids 法  $k = 3$

	数	$P_1$	$P_2$	$P_3$
$C_1$	19	26	53	21
		5	10	4
$C_2$	18	33	50	17
		6	9	3
$C_3$	11	36	36	28
		4	4	3
$C_4$	5	60	40	0
		3	2	0
$C_5$	32	41	50	9
		13	16	3

表 8:  $k$ -medoids 法  $k = 5$

	数	$P_1$	$P_2$	$P_3$
$C_1$	5	40	60	0
		2	3	0
$C_2$	11	9	64	27
		1	7	3
$C_3$	11	27	55	18
		3	6	2
$C_4$	10	40	30	30
		4	3	3
$C_5$	6	33	33	33
		2	2	2
$C_6$	2	50	50	0
		1	1	0
$C_7$	8	75	25	0
		6	2	0
$C_8$	5	40	60	0
		2	3	0
$C_9$	13	23	62	15
		3	8	2
$C_{10}$	14	50	43	7
		7	6	1

表 9:  $k$ -medoids 法  $k = 10$

## 5 考察

最遠点法においては,  $k = 3$  と  $k = 5$  の場合はクラスタごとのサイズに偏りが見られたが,  $k = 10$  の場合のクラスタごとのサイズに偏りがあまり見られなかった. またクラスタと脳波パターンの関連性は見られなかった (表 1, 2, 3 参照).

最短距離法においては,  $k = 3$ ,  $k = 5$ ,  $k = 10$  の場合にクラスタごとのサイズに偏りが見られた. またクラスタと脳波パターンの関連性は見られなかった (表 4, 5, 6 参照).

$k$ -medoids 法においては,  $k = 3$  と  $k = 5$  の  $k = 10$  場合はクラスタごとのサイズに偏りがあまり見られなかった. またクラスタと脳波パターンの関連性は見られなかった (表 7, 8, 9 参照).

よって, 3つの手法を用いて行った本研究における, プレイ時間の長さによるクラスタリング結果からは, プレイ時間と脳波パターンの関連性は認められなかった. 今後の他手法を用いたさらなる検証が待たれる.

## 参考文献

- [1] Rui Nouchi, Yasuyuki Taki, Hikaru Takeuchi, Hiroshi Hashizume, Yuko Akitsuki, Yayoi Shigemune, Atsushi Sekiguchi, Yuka Kotozaki, Takashi Tsukiura, Yukihito Yomogida, Ryuta Kawashima, Brain Training Game Improves Executive Functions and Processing Speed in the Elderly: A Randomized Controlled Trial.
- [2] 福田 剛志, 徳山 豪, 森本 康彦, データマイニング, 共立出版, 2001 年.
- [3] 元田 浩, 山口 高平, 津本 周作, 沼尾 正行, データマイニングの基礎, オーム社, 2006 年.

[4] 森 昭雄, ゲーム脳の恐怖, NHK 出版, 2007 年 .

[5] Wolfram Research, Partitioning Data into Clusters,

<http://reference.wolfram.com/mathematica/tutorial/PartitioningDataIntoClusters.en.html>