大規模データのクラスタリング

クラスタ中心を再計算しない非階層的クラスタリング：ドラッグストアの会員分類を例として

出口慎二[1)]　○中山厚穂[2)]　高崎祐哉[3)]

1　データエクスプローリング
2　首都大学東京大学院社会科学研究科経営学専攻
3　カスタマー・コミュニケーションズ株式会社

要旨：手続き

1. ID付POSデータから，「会員（レコード）」×「商品カテゴリ（フィールド）」のデータを作る．値は，ここでは「購買金額（値）」とする．
2. 「会員」×「商品カテゴリ」のデータから，ちいさな（具体的には，N=2,000 の）データを複数（ここでは，5つ）作る．これは，単純に，等間隔抽出で行う．
3. 複数の（5つの）サンプリングデータの階層的クラスタリングを行い，その結果から，全データのクラスタ構造を仮定する．
4. 仮定したクラスタ構造をもとに，クラスタ中心行列を作る．具体的には以下の通り．
    1. 複数のサンプリングデータ（5つ）それぞれを同じ数のクラスタ（ここでは6つ）に分ける．
    2. いずれも性質が同等と思われる複数のクラスタについて，算術平均によりクラスタ中心を計算する．
5. 作成したクラスタ中心行列を使って，あらためて，全データ（N=4,946,955）のクラスタリング（クラスタ中心の再計算を行わない）を行う．

要旨：環境

■ 使用したマシンスペック

◆ パーソナルコンピュータ（個人で持てる，普通のPC（パーソナルコンピュータ）．）

- CPU　core-i7（インテル(R) Core(TM) i7-4790 プロセッサー）
- メモリ　16GB（PC3-10600）
- HDD　1TB

■ 使用したソフトウェア

◆ MySQL　データの加工に使用（オープンソース） http://www-jp.mysql.com/

◆ SQLite　データの加工に使用（パブリックドメイン）http://www.sqlite.org/

◆ PostgreSQL　データの加工に使用（フリー）http://www.postgresql.jp/

◆ Excel VBA　主に全データ（N=4,946,955）のクラスタリングに使用（商用ソフトウェア）

◆ SPSS　サンプルデータ（N=2,000）のクラスタリング，および全データ（N=4,946,955）のクラスタリング後のクラスタ平均の計算に使用（商用ソフトウェア）

◆ 秀丸　データの目視に使用（テキストエディタ（シェアウェア））http://hide.maruo.co.jp/

※Access は大きな（2GBを超える）データには向かない．
※FileMakerは動くが時間がかかる．

要旨：結果

■ 全データのクラスタリング結果（表1）

- ◆ クラスタ番号「0」は無視（日用雑貨の購買金額が年間で1千万円を超えるなど，異常値．店舗ごと使う社内使い用の会員番号と思われる）
- ◆ ドラッグストアということもあり，「生鮮食品」はほぼ無い．
- ◆ いずれの購買金額とも少ないクラスタ（クラスタ番号「1」）に大多数が割り振られた（全体の81%）．それでも，金額でみて，「化粧品」と「日用雑貨」は多少購買が見られる．
- ◆ やや購買があるがいずれともあまり買わないクラスタ（クラスタ番号「2」）がそれに続く（全体の17%）．「化粧品」「日用雑貨」には他よりお金を使う傾向は同様．
- ◆ 次いで全体の2%ほどのクラスタ（クラスタ番号「3」）が，「化粧品」に年間73,000円ほど．
- ◆ クラスタ番号「4」以降は全体の1%未満．

表1．クラスタ中心（全データ）

クラスタ番号	度数	年齢	性別	加工食品	生鮮食品	菓子類	飲料酒類	その他食品	日用雑貨	ＯＴＣ医薬品類	化粧品	家庭用品	ＤＩＹ用品	ペット用品	その他日用品
0	3	38	1.33	1,220	0	1,755	54,030	4,288,613	12,400,989	15,794	97,232	60	0	0	0
1	3,993,876	52	1.82	315	0	352	489	417	1,844	1,341	2,077	157	21	71	4
2	843,341	54	1.89	2,301	0	2,150	2,860	3,379	16,149	9,914	17,752	1,359	72	581	32
3	81,427	53	1.95	2,605	0	2,887	3,927	5,458	24,327	13,148	73,690	2,066	68	805	53
4	20,959	63	1.60	23,014	1	13,176	88,774	11,743	20,393	11,787	11,370	2,719	162	1,705	99
5	4,106	56	1.93	3,086	0	3,917	6,968	15,243	28,040	30,493	264,224	1,935	58	803	65
6	3,243	66	1.64	4,677	0	4,565	8,689	22,597	36,249	204,351	24,346	2,663	195	1,459	99
全体	4,946,955	53	1.83	792	0	754	1,378	1,063	4,668	3,129	6,023	398	31	175	10

手続き1:データの作成

■ 実はこれが一番手間.

- ◆ 元のデータは1年分約2億3千万件少々の(会員ID付) POSデータ(15GB).
- ◆ 会員ID(行:500万弱※1) × 商品ジャンル(列:12), 値は購買金額, のデータを作る.
  - この集計(クラスタリング用データの作成)に関連するデータの加工に手間と時間を要した.
    - 最終的にはSQLの使用で, かなりの時間を短縮できた.
    - データが横に長くなると, 通常はSQLではデータを確認しにくい(SQLで, 1レコードごと, 縦に表示させることも可能). 簡単な目視には秀丸(シエアウェアのテキストエディタ)を使用した(データの一部分だけを読み込むことができる).

  ※1 2013年1年間に購買記録があった会員は500万人弱.
- ◆ 分析用データはテキストファイル.
  - 以降の分析手続きはSQLのを使わずに, テキストファイルに対して, Excel VBA から行った(全データ分割後のクラスタ平均計算にはSPSS使用).

表2. 商品ジャンル

1	加工食品
2	生鮮食品
3	菓子類
4	飲料酒類
5	その他食品
6	日用雑貨
7	ＯＴＣ医薬品類
8	化粧品
9	家庭用品
10	ＤＩＹ用品
11	ペット用品
12	その他日用品

手続き2・3：サンプリングデータの階層的クラスタリング

■ サンプリングデータを作成

◆ 全データ（N=4,946,955）から，n=2,000のサンプリングデータを5つ作成．

- N=2,000とした根拠
  - 標本調査を参考にした．（値が単純に二項分布する変数であり，知りたい真の値が50%と仮定すると，無限母集団からのサンプリング誤差は，95%の確からしさで，n=2,000のとき，およそ±2%．）
  - ここで使う変数（商品ジャンルごとの年間購買金額）は，おそらく正規分布しない（左に偏った分布）．
  - サンプリングデータは複数（ここでは5つ）作り，それぞれをクラスタリングして判断した．

◆ サンプリングデータの階層的クラスタリング．

- 手法は，Ward法．結果概観（5つの樹形図）は次々ページ（図1）参照．

◆ クラスタ数の決定

- 4～7クラスタに分けた場合をそれぞれ見比べて判断．
  - すべてのクラスタについて，クラスタ平均を計算し，特徴を書き出す．
  - 複数のクラスタに共通する特徴のみ拾う．→その特徴の数（重複は除く）がクラスタ数．

  詳しくは次ページ「クラスタ数の検討」参照．
- ここでの判断の正しさの事後的な確認
  - サンプリングデータの検討から作ったクラスタ平均と，全データのクラスタリング後に計算したクラスタ平均を比較することで，ここでの判断（サンプリングデータからのクラスタ平均の決定）が全データに沿っていたかどうかを判断できる．著しく異なる場合は，クラスタ平均を決める部分（サンプリングデータを取るところ）からやり直せば良い．

◆ クラスタ数の検討

- 5つのサンプリングデータすべてを4～7クラスタに切り，そのすべてのクラスタ（計110クラスタ:(4クラスタ×5つのデータ)＋(5クラスタ×5つのデータ)＋(6クラスタ×5つのデータ)＋(7クラスタ×5つのデータ)，計110クラスタ）についてクラスタ平均を計算.
- それぞれ（計110クラスタ）解釈して，まずクラスタ数を決める．そのうえで，複数（5つ）のサンプリングデータに重複して出現したクラスタを拾う．複数のサンプリングデータで見られたクラスタは5つと判断（下記）.
  - 飲料・酒類（10万円/年くらい？）
  - 日用雑貨（5万円/年くらい？）
  - OTC医薬品（9万円/年くらい？）
  - 化粧品（12万円/年くらい？）
  - 全部ほぼ0円/年（正確には数百円～2千円くらい/年）
- （念のため1つ増やして（何かありそうなので））6クラスタ解で考えることとした.

◆ （定期的なチェックの必要）

- 定期的に（定期分の，たとえば12か月分の．期間が変わると，金額の比較がむつかしい（時期の違いによっても，たとえば期間内のセールの回数が違うと，どう比較してよいかわからなくなる））データ全体のクラスタ平均の計算，チェックを行う.

←クラスタ中心の再計算を行わないので，新規会員が増えるにつれ，過去に計算したクラスタ中心が，現データの実情と乖離する可能性がある.

サンプリングデータ(5つ)の階層的クラスタリング結果(1)

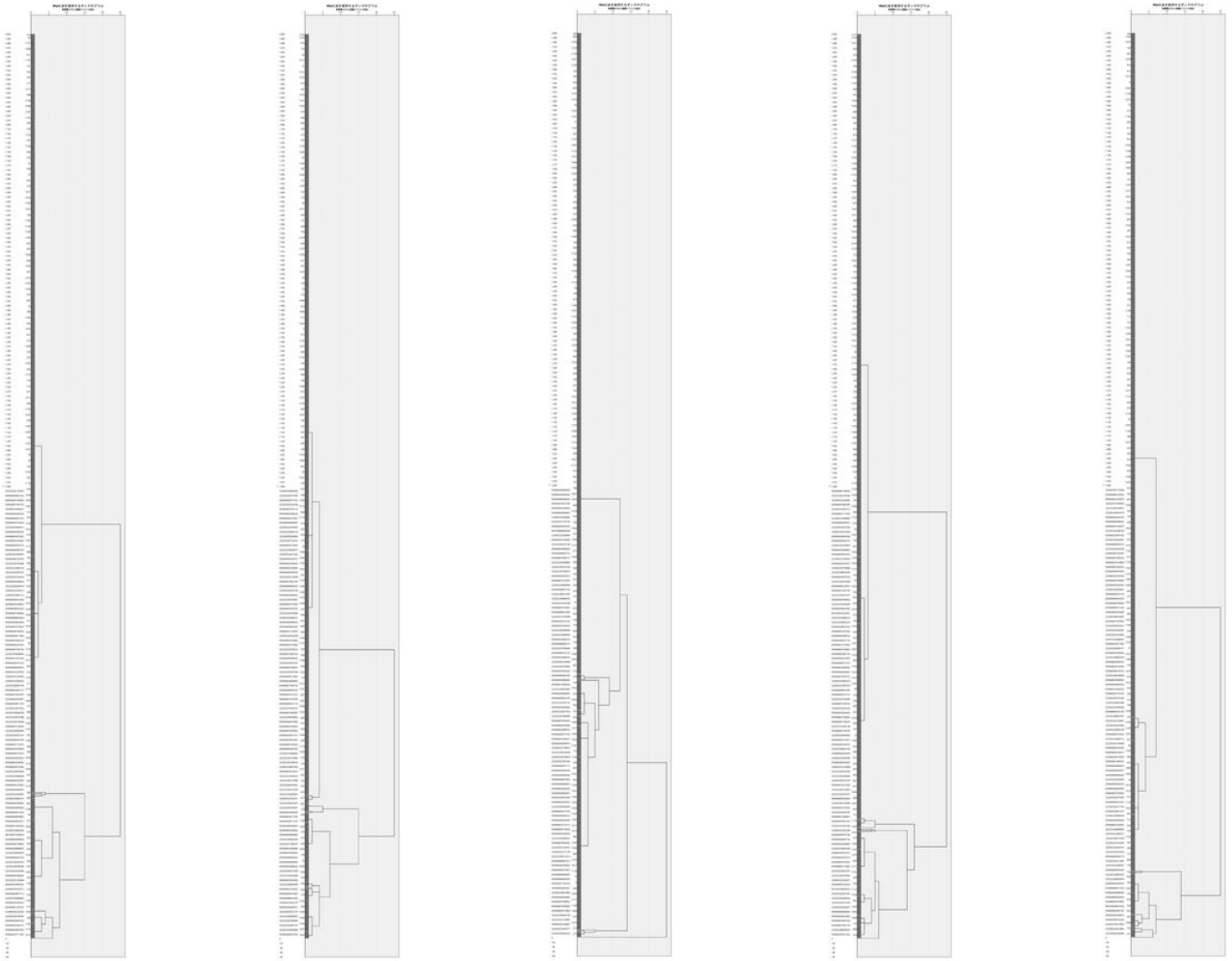

図1．サンプリングデータの階層的クラスタリング(ウォード法)の結果(樹形図):1

サンプリングデータ(5つ)の6クラスタ解

■ サンプリングデータのクラスタリング結果

表3. 5つのサンプリングデータそれぞれのクラスタリング結果(6クラスタ解. クラスタ平均)

		11 加工食品	12 生鮮食品	13 菓子類	14 飲料・酒類	19 その他食品	21 日用雑貨	22 ＯＴＣ医薬品類	23 化粧品	24 家庭用品	25 ＤＩＹ用品	26 ペット用品	29 その他日用品	
サンプリングデータ#1	c1(n=1,677)	332	0	398	615	498	2,075	1,718	2,200	176	19	103	6	少ない
	c2(n=200)	1,850	0	1,944	2,527	1,781	14,031	3,824	16,565	1,128	27	224	12	化粧品(低)/日用雑貨
	c3(n=62)	2,899	0	2,811	3,636	5,468	36,319	27,685	20,326	2,667	328	541	56	日用雑貨
	c4(n=46)	614	0	1,972	1,853	3,761	13,540	7,538	50,477	1,125	84	165	104	化粧品(中)
	c5(n=14)	16,756	0	14,338	94,378	21,267	15,839	21,785	18,621	1,395	38	103	32	飲料・酒類
	c6(n=1)	641	0	5,779	781	7,952	5,679	237,911	5,155	1,330	0	0	0	OTC医薬品
サンプリングデータ#2	c1(n=1,683)	321	0	393	485	448	2,333	1,619	2,316	198	13	73	5	少ない
	c2(n=152)	579	0	739	655	906	6,107	2,550	23,235	482	12	136	4	化粧品(低)/日用雑貨
	c3(n=126)	3,491	0	3,623	5,333	9,815	27,272	19,465	18,862	1,920	59	563	100	日用雑貨
	c4(n=19)	19,660	0	8,218	36,421	5,735	8,552	3,616	4,109	1,382	133	1,475	24	飲料・酒類
	c5(n=17)	1,624	0	1,299	1,013	2,486	18,036	7,826	88,278	1,037	0	350	0	化粧品(中)
	c6(n=3)	672	0	1,904	635	34,518	25,225	37,763	200,220	2,250	0	0	73	化粧品(高)/OTC医薬品
サンプリングデータ#3	c1(n=1,418)	205	0	253	260	220	1,546	1,137	1,450	125	23	50	5	少ない
	c2(n=550)	1,861	0	1,762	2,735	2,327	11,716	6,958	15,705	1,060	37	220	41	化粧品(低)/日用雑貨
	c3(n=16)	24,586	0	13,600	109,013	8,877	26,518	10,985	24,676	5,416	71	479	462	飲料・酒類
	c4(n=12)	375	0	955	1,561	4,730	15,286	14,767	118,244	800	0	0	0	化粧品(中)
	c5(n=3)	726	0	1,592	1,761	12,326	33,368	175,312	7,240	2,554	0	0	0	OTC医薬品
	c6(n=1)	0	0	0	0	0	621,241	0	0	0	0	0	0	日用雑貨
サンプリングデータ#4	c1(n=1,735)	478	0	475	611	514	2,665	1,463	2,650	235	14	41	4	少ない
	c2(n=235)	2,199	0	2,394	3,985	3,415	16,779	14,829	23,673	1,348	17	1,039	17	化粧品(低)/日用雑貨
	c3(n=20)	1,182	0	1,025	1,588	4,913	21,421	10,552	110,453	1,585	38	368	20	化粧品(中)
	c4(n=8)	25,957	0	35,114	142,492	3,554	9,071	5,516	7,388	1,972	31	2,194	0	飲料・酒類
	c5(n=1)	0	0	17,562	6,063	45,843	19,489	130,703	319,173	475	0	0	0	化粧品(高)/OTC医薬品
	c6(n=1)	849	0	1,421	737	7,470	36,810	278,087	14,560	1,778	0	0	0	OTC医薬品
サンプリングデータ#5	c1(n=1,514)	185	0	298	290	251	1,395	1,148	1,810	119	21	41	6	少ない
	c2(n=339)	1,976	0	1,973	1,948	2,555	9,526	8,421	11,277	821	28	321	6	化粧品(低)/日用雑貨
	c3(n=117)	2,030	0	2,249	2,673	3,395	26,086	11,096	34,307	2,987	183	2,496	38	化粧品(中)
	c4(n=22)	12,787	0	7,800	52,216	17,320	18,239	11,978	8,650	2,028	207	104	0	飲料・酒類
	c5(n=7)	4,955	0	6,691	4,908	14,648	25,628	42,561	173,469	3,636	304	740	204	化粧品(高)/OTC医薬品
	c6(n=1)	1,125	0	2,408	1,615	5,624	22,275	259,753	2,988	770	0	0	0	OTC医薬品
													単位:円	

手続き4：クラスタ中心の作成

■ 30クラスタ→6クラスタ

◆ 5つのサンプリングデータそれぞれについて6クラスタ解を計算，そのそれぞれ（計30）についてクラスタ中心を求める（表3）．

- 5つのデータセット，全30のクラスタすべてを解釈．
- そのなかから，複数（ここでは3つ以上）のデータセットで共通して現れた特徴的なクラスタを6つ選び出す．
- それぞれ同内容を持つクラスタ，3つ～5つ，のクラスタ中心の算術平均値を求める．
  - （どれも）少ない　5つのデータに在り
  - 化粧品（低）/日用雑貨　5つのデータに在り
  - 化粧品（中）　5つのデータに在り
  - 飲料・酒類　5つのデータに在り
  - 化粧品（高）/OTC医薬品　3つのデータに在り
  - OTC医薬品　4つのデータに在り

表4．作成したクラスタ平均

	11 加工食品	12 生鮮食品	13 菓子類	14 飲料・酒類	19 その他食品	21 日用雑貨	22 ＯＴＣ医薬品類	23 化粧品	24 家庭用品	25 ＤＩＹ用品	26 ペット用品	29 その他日用品
（どれも）少ない	304	0	363	452	386	2,003	1,417	2,085	171	18	62	5
化粧品（低）/日用雑貨	1,693	0	1,762	2,370	2,197	11,632	7,316	18,091	968	24	388	16
化粧品（中）	1,165	0	1,500	1,738	3,857	18,874	10,356	80,352	1,507	61	676	33
飲料・酒類	19,949	0	15,814	86,904	11,351	15,644	10,776	12,689	2,438	96	871	104
化粧品（高）/OTC医薬品	1,876	0	8,719	3,869	31,670	23,447	70,342	230,954	2,120	101	247	92
OTC医薬品	835	0	2,800	1,223	8,343	24,533	237,766	7,486	1,608	0	0	0

単位：円（小数点以下はまるめて表示しています）

手続き5：全データのクラスタリング （＊）

■ 中心の再計算を行わない非階層的クラスタリング

◆ クラスタ中心をすでに作ってあるので，クラスタ中心の再計算は行わない．
＝データ件数（i）が増えても影響は無い（計算時間が増えるだけ．変数の数が著しく大きくなければ，メモリは要らない）

◆ あらかじめ決めたクラスタ中心（μ：前掲）にしたがって，全データを，もっとも「距離（δ_k）」が近いクラスタ中心を持つk番目のクラスタに割り振る．

- k番目のクラスタ平均との距離（大小関係だけ分かれは良いので pで除する必要はない）

$$\delta_k = \sum_1^p \left(x_p - \mu_{pk}\right)^2$$

※ i（個人）は無い∵再計算しないから

※$\left(x_p - \mu_{pk}\right)$には負もあり得るので二乗は必須

※ kはクラスタの番号／ pは変数の番号

→変数（p）回の繰り返しはあるが，個人（i）回の繰り返しは無い．

→これをクラスタの数（k回）計算し，クラスタとの距離が最少の（もっとも近い）クラスタを選ぶ．

◆ ここでは，「距離」をシンプルに定義したが，距離の定義は必要に応じ変更可能．

◆ また変数ごとに重みを変えた計算も可能．

全データのクラスタリングの結果(1)

■ クラスタ中心の比較

- ◆ サンプリングデータから作ったクラスタ平均は表4(10ページ)参照.
- ◆ 全データのクラスタリングの結果については表1(4ページ)参照(下記に再掲).
- ◆ いずれとも, 金額が少ないセルは青, 金額は高いセルは赤, になるよう色を付けている.
- ◆ 表1には, クラスタ番号「0」および「全体」行があるので注意(無視).
- ◆ おおむね, 事前にサンプリングデータから作成したクラスタ平均と, 全データのクラスタリング後に計算した, 全データに基づくクラスタ平均は, 同等の性格を示した.

【再掲】表1. クラスタ中心(全データ)

クラスタ番号	度数	年齢	性別	加工食品	生鮮食品	菓子類	飲料酒類	その他食品	日用雑貨	ＯＴＣ医薬品類	化粧品	家庭用品	ＤＩＹ用品	ペット用品	その他日用品
0	3	38	1.33	1,220	0	1,755	54,030	4,288,613	12,400,989	15,794	97,232	60	0	0	0
1	3,993,876	52	1.82	315	0	352	489	417	1,844	1,341	2,077	157	21	71	4
2	843,341	54	1.89	2,301	0	2,150	2,860	3,379	16,149	9,914	17,752	1,359	72	581	32
3	81,427	53	1.95	2,605	0	2,887	3,927	5,458	24,327	13,148	73,690	2,066	68	805	53
4	20,959	63	1.60	23,014	1	13,176	88,774	11,743	20,393	11,787	11,370	2,719	162	1,705	99
5	4,106	56	1.93	3,086	0	3,917	6,968	15,243	28,040	30,493	264,224	1,935	58	803	65
6	3,243	66	1.64	4,677	0	4,565	8,689	22,597	36,249	204,351	24,346	2,663	195	1,459	99
全体	4,946,955	53	1.83	792	0	754	1,378	1,063	4,668	3,129	6,023	398	31	175	10

全データのクラスタリングの結果(2)

■ 結果の解釈

＜クラスタの特徴＞	＜人数(%)＞	＜平均年齢＞	＜男性割合＞
◆ (どれも)少ない	n=3,993,876 (81%)	52歳	18%
◆ 化粧品(低)/日用雑貨	n= 843,341 (17%)	54歳	11%
◆ 化粧品(中)	n= 81,427 (2%)	53歳	5%
◆ 飲料・酒類	n= 20,959 (0%)	63歳	40%
◆ 化粧品(高)/OTC医薬品	n= 4,106 (0%)	56歳	7%
◆ OTC医薬品	n= 3,243 (0%)	66歳	36%

■ すべての個人についてクラスタがわかっている

◆ 全会員(N=4,946,955)について所属クラスタがわかっているので，当該クラスタに属する層の属性を(推測ではなく)正確に把握できる.

- あるクラスターにアクセスしたいとき，そのクラスターにはどんな人たちが含まれるのか把握できていると，アクションを起こし易い.

補遺(1)

- 使用したデータ
  - 「大規模データ」であることを重視して，複数の企業，複数の店舗におけるPOSデータを使用した．
  - 店舗・企業を超えて存在するルールを見つけるためには，今回のように店舗間をマージしたデータで良いのだろう．
  - 個別の会社でマーケティング施策に使うためには，1店舗ないし1社分のデータだけを使い，商品は，カテゴリにまとめずに商品そのままの品目を使うべきかもしれない．
    - 複数の会社間で，商品そのものを変数にとると，会社ごと取扱商品に違いがある場合，結果として，取扱品目でクラスタが分かれる(会社で分かれる)ということが起こりうる．
- "「クラスタ中心を再計算しない」非階層的クラスタリング"の意義
  - データをサンプリングしないですべて見る，という発想がある．
  - 必ずしも特別なマシンスペックなどを要さずに，大量のデータを処理できる方法には意義があるだろう．
    - サンプリングデータのクラスタリングが示す通り，クラスターを見るだけなら必ずしも全てのデータを見る必要はない．ただし，全データをクラスタリングすると，全会員にクラスタ番号が割り振られる(クラスタクロスを見られる)．

補遺(2)

- データの「加工」と「分析」はそれぞれに専門家が必要？
  - POSデータは基本的に横(列方向)にはあまり大きくない.
    - たとえば, SQLでは, 列の上限が1,000～2,000程度だったりする. (一般にビッグデータは縦に大きくても, 横にはさほど大きくない)
  - しかし分析用のデータは横方向におおきいこともしばしば.
    - POSデータでは, 購入した商品のコードは, 1列にならぶ. その1列にいくつものコードが(JICFSコードでもコードの数は3000弱)値として入っている. →1列の設計
    - 分析では, コードを横に並べる必要もしばしばある(JICFSコードだと3000列近く必要になる)→どうしても変数の数の列(3,000)が必要
  - データが大きくなってくると, 分析用データを作るのもスキルが必要(分析者が片手間でできるような域を超えてくる)
    - ただし, SQLを使う人も, 通常は, 列に大きいデータというのは扱わない. 普段データベースを扱っていても, 分析用データを作れるわけではない.

※ここでの「分析」は多変量解析を想定(集計ではない).

単独企業の顧客クラスタリング

■ 企業A

会社が違えば取扱商品が異なる.
→ジャンルではなく個々の商品ごとの購買を見ても，企業をまたいで客層が分かれるのではなく，しばしば単純に，企業単位のクラスタができてしまう.
→企業単位で分けて見てみる必要も

- ◆ 会員番号の数　　:94,674
- ◆ 商品コードの数　　:745（JICFSコードの数．商品そのものの種類の数ではない）

商品コードの例）
醤油
砂糖
低カロリー甘味料
味噌
食塩

- ◆ サンプリングデータ（5つ）のクラスタリング結果（階層的）から，5～7クラスタと判断.

- 7クラスタ解を5つのサンプリングデータについてみたところ，下記7クラスタが複数のデータで確認された．ただしこの他にも，複数のクラスタには現れないが，ある程度人数のいるクラスタが散見された．（→階層的な検討の必要性 cf. AID）
  - 「衣料用合成洗剤」
  - 「シャンプー」
  - 「化粧水」
  - 「健康食品（高）」
  - 「ファンデーション」
  - 「化粧水/ファンデーション」
  - 「大人用オムツ」
- 複数のクラスタ平均を単純に算術平均して，全データのクラスタリング用にクラスタ中心を作成（変数が多い（745）ので，表は割愛）
- 作成したクラスタ平均を使って全データ（N=94,674）をクラスタリング．
  - 結果は右記．
  - 全変数（p=745）のうち，6変数だけ抜き出して，サンプリングデータのクラスタリング結果と，全データのクラスタリング結果を見ると次ページのとおり．

表5．全データのクラスタリング結果

クラスタ	件数	%
全体	94,674	100%
1．衣料用合成洗剤	46,701	49%
2．化粧水	2,035	2%
3．健康食品（高）	2,549	3%
4．シャンプー	35,253	37%
5．化粧水/ファンデーション	2,244	2%
6．大人用オムツ	420	0%
7．ファンデーション	5,472	6%

■ 作成したクラスタ平均と全データのクラスタリング結果から計算したクラスタ平均

表6. サンプリングデータから作成したクラスタ平均

クラスタ番号	健康食品	大人用オムツ	衣料用合成洗剤	化粧水	ファンデーション	シャンプー
1 衣料用合成洗剤	44	19	112	41	23	81
2 化粧水	599	301	646	4,708	2,170	764
3 健康食品(高)	8,086	1,662	1,245	946	738	1,052
4 シャンプー	219	57	314	223	184	362
5 化粧水/ファンデーション	354	19	736	6,908	4,205	850
6 大人用オムツ	1,236	31,223	764	761	1,665	300
7 ファンデーション	840	383	1,035	2,162	3,224	1,151

表7. 全データのクラスタリング後に計算したクラスタ平均

クラスタ番号	健康食品	大人用オムツ	衣料用合成洗剤	化粧水	ファンデーション	シャンプー
1 衣料用合成洗剤	41	46	73	35	36	52
2 化粧水	438	152	575	4,534	727	891
3 健康食品(高)	13,308	365	785	692	533	910
4 シャンプー	339	168	505	336	291	588
5 化粧水/ファンデーション	764	180	689	12,414	5,625	1,116
6 大人用オムツ	1,638	41,170	1,084	988	694	601
7 ファンデーション	513	200	1,184	1,141	3,938	1,541