

2012.11.06 S-PLUS学生研究奨励賞

数量化理論第Ⅱ類とアソシエーションルール解析による自動車バナーデザインの分析

大阪大学大学院 システム創成専攻

乾口研究室 大木基至

和歌山大学 システム工学部

デザインシステム計画研究室 堀切川和也

00. 目次

01. 背景と目的

02. アンケート調査

03. 数量理論第II類とクラスター分析

04. アソシエーションルール解析

05. まとめ

01. 背景：ネット広告市場の肥大化

- ・ バナー広告を中心としたインターネット広告(図1)市場の伸び
- ・ 2017年には1兆円規模の成長が見込まれる(図2)



図1. ネット広告の例

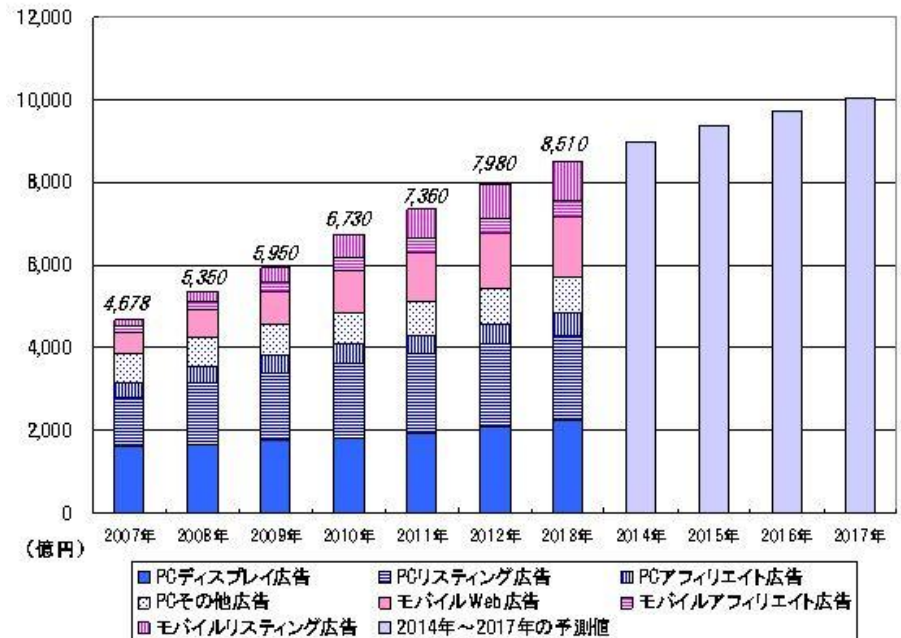
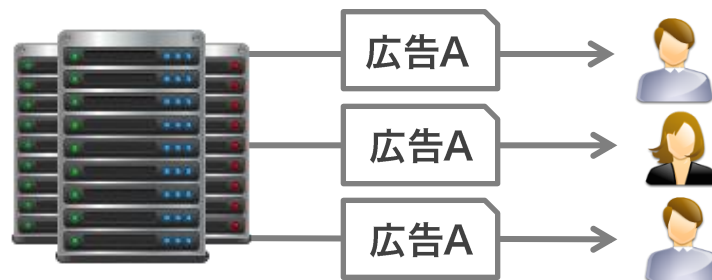


図2. インターネット広告費の推移予測[1]

01. 背景：広告配信の仕組み

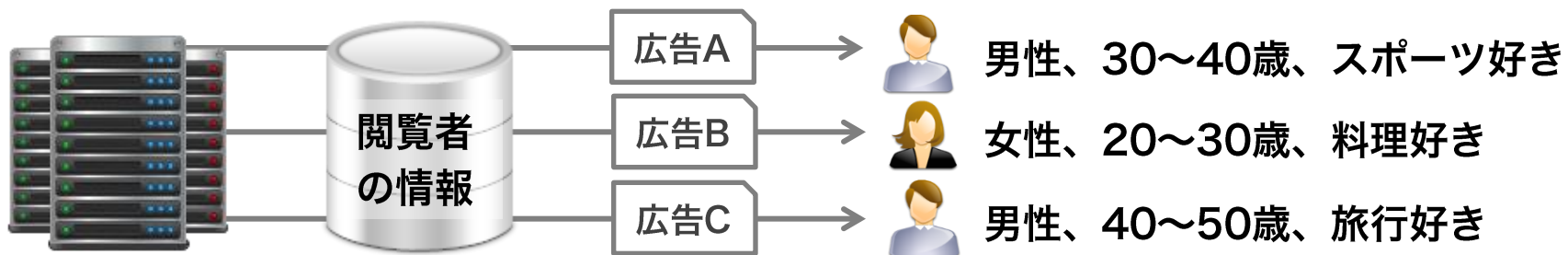
従来の方法



配信方法：すべてのWeb閲覧者に同じ広告を配信する

問題点：クリック率の低い閲覧者にも同じ価格で配信してしまう

現在の方法：DSP(Demand-Side Platform)



配信方法：各Web閲覧者ごとにクリック率の高い広告のみ配信する

課題点：クリック率の高い広告の推定方法

01. 本研究の目的

広告配信の課題

- **どの閲覧者にどの広告を配信すべきか**
 - 各広告に対して、クリック率が高い閲覧者の特徴を把握する必要がある[2]
- **生成する広告デザイン要素の選定が困難**
 - ロゴ、文字のサイズ、位置、配色など多数のデザインパターンが考えられる[3]
 - デザイン要素間が独立であるとは限らないため、線形的手法(多変量解析)が正しいとは限らない[4]



本研究の取り組み

- **好まれるデザイン要素の推定と閲覧者のクラスタリング**
- **非線形的手法であるアソシエーションルールによる分析**

02. アンケート調査

調査概要

- ・ 被験者：大学生75名（平均年齢21.9歳、標準偏差1.85）
- ・ 分析対象：39種の自動車のバナー広告(図3)
 - 自動車の種類は多く、バナーの種類も様々であるため



図3. 自動車のバナー広告の例

- ・ 評価方法：「魅力的」についての3段階SD法
 - 「魅力的である」と感じるバナーはクリック率が高いと予測できるため
- ・ アンケート収集方法：パソコン上で回答

02. アンケート調査

属性(アイテム)と属性値(カテゴリ)

ロゴ：有り or 無し

車の向き：正面 or
側面 or 斜め正面

車面積：
1/4 以上 or 1/8以上
or 1/8未満

シーン：
走行中 or 停止中

フォントの種類：
1種類 or 2種類以上

文字面積：
1/4 以上 or 1/8以
上 or 1/8未満

フォント：
明朝体 or ゴシック体

見出し語の位置：
上 or 中 or 下



サブバナー：
有り or 無し

見出し語の大きさ：
大 or 中 or 小

人・キャラクター：
有り or 無し

背景：自然 or 都会的
or 仮想空間 or 無地

背景色：ブルー系 or レッド系
or グリーン系 or グレー
or ブラック or ホワイト

03. 本研究の分析の流れ

好まれるデザイン要素の推定と閲覧者のクラスタリング

数量化理論第II類による分析

カテゴリースコアによるクラスター分析

各クラスタが好むデザイン要素を推定する

非線形的手法であるアソシエーションルール解析

if-thenルールを抽出し、クラスタ毎にまとめる

各クラスタが好むデザイン要素の組み合わせを推定する

03. 数量化理論第II類

数量化理論第II類とは

- ・ 統計数理研究所所長の林知己夫によって1940年代に開発された数量化理論の1つ
- ・ 説明変数と目的変数が共に質的変数で与えられているデータに対して、説明変数から目的変数を予測する方法
- ・ S-PLUSによって第II類を行い、説明変数と目的変数の関連度(偏相関係数とカテゴリースコア)を算出する[a]

表1. 数量化理論第II類で用いるデータの例

	ロゴ	書体の種類	書体	...	評価
トヨタ 86	ロゴ無し	フォント2種類以上	明朝体	...	魅力的である
マツダ BIANTE	ロゴ有り	フォント2種類以上	ゴシック体	...	どちらでもない
日産 BLUEBIRD	ロゴ無し	フォント1種類	ゴシック体	...	魅力的でない
トヨタ CAMRY	ロゴ無し	フォント1種類	ゴシック体	...	魅力的である
...

03. クラスター分析：準備

- 数量化理論第II類により算出した偏相関係数とカテゴリースコアをもとに、凝集階層型クラスター分析を行う。
- S-PLUSで用意されている5つの手法を **Agglomerative Coefficient(凝集係数)[5]**の点で比較する。

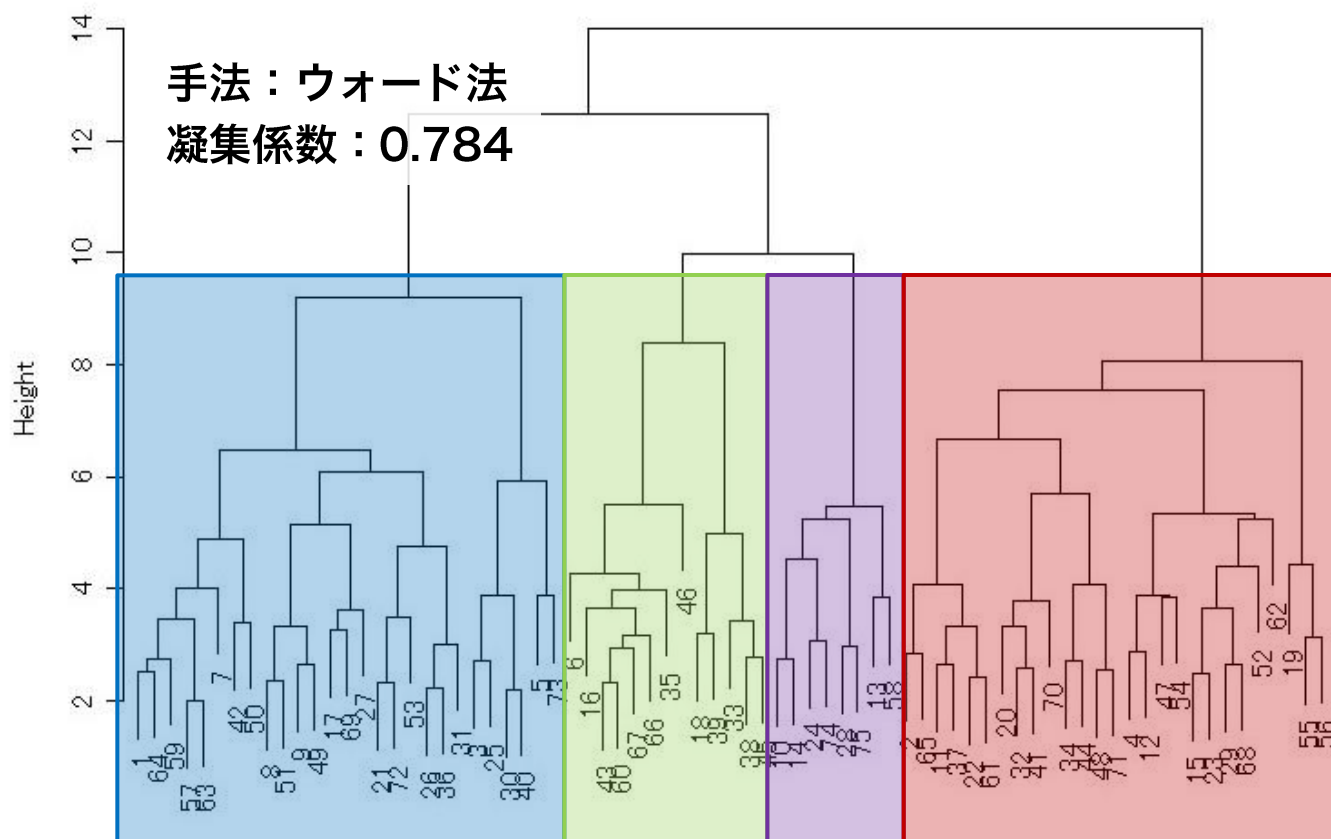
表2. カテゴリースコアを用いた結果

手法	凝集係数
郡平均法	0.494
完全連結法	0.676
単連結法	0.289
ウォード法	0.784
重心法	0.577

- ウォード法の凝集係数が高いため、本研究では**ウォード法**を使用し、クラスタリング結果の解釈を行う。

03. クラスタ分析：樹形図

カテゴリースコアによるクラスタ分析の階層図



4つのクラスターに分類

03. クラスタ分析：考察

各カテゴリースコアの影響度の考察

表3. 「ロゴ」のカテゴリースコア

	ロゴ	
	無し	有り
クラスタ 1	0.044	-0.057
クラスタ 2	0.139	-0.179
クラスタ 3	-0.040	0.051
クラスタ 4	-0.108	0.140
平均値	0.009	-0.011

- 各クラスタごとに影響度は正と負に分類される。
- 各クラスタでレンジが小さいため影響度は小さい。

表4. 「フォントの種類」のカテゴリースコア

	フォントの種類	
	1種類	2種類以上
クラスタ 1	-0.101	0.256
クラスタ 2	-0.014	0.035
クラスタ 3	0.015	-0.037
クラスタ 4	0.101	-0.257
平均値	0.000	-0.001

- 各クラスタごとに影響度は正と負に分類される。
- クラスタ1, 4ではレンジが大きい。

03. クラスタ分析：考察

各カテゴリースコアの影響度の考察

表5. 「フォント」のカテゴリースコア

	フォント	
	ゴシック体	明朝体
クラスタ 1	-0.073	0.244
クラスタ 2	0.138	-0.459
クラスタ 3	-0.001	0.005
クラスタ 4	-0.229	0.762
平均値	-0.041	0.138

- 各クラスタごとに影響度は正と負に分類される。
- クラスタ2, 4ではレンジが大きい。

表6. 「見出し語の位置」のカテゴリースコア

	見出しの位置		
	下	上	中
クラスタ 1	0.871	-0.016	-0.357
クラスタ 2	-0.314	-0.033	0.324
クラスタ 3	-1.678	0.084	0.420
クラスタ 4	0.339	0.100	-0.667
平均値	-0.196	0.034	-0.070

- 各クラスタでカテゴリースコアの絶対値が大きく、レンジも大きいため、バナーの評価に大きく影響を与えるカテゴリーである。

03. クラスター分析：考察

各カテゴリースコアの影響度の考察

表7. 「背景」のカテゴリースコア

	背景			
	仮想空間	自然	都会的	無地
クラスタ 1	-0.653	-0.123	0.000	0.314
クラスタ 2	-0.524	0.129	0.466	-0.316
クラスタ 3	1.083	-0.167	-0.207	-0.020
クラスタ 4	0.601	0.121	-0.003	-0.294
平均値	0.127	-0.010	0.064	-0.079

- 各クラスタで影響するカテゴリが異なり、絶対値も大きい。
- レンジの値も大きいいため、バナーの評価に大きく影響を与えるカテゴリである。

表8. 「背景色」のカテゴリースコア

	背景色					
	グリーン系	グレー	ブラック	ブルー系	ホワイト	レッド系
クラスタ 1	0.027	0.169	0.100	0.044	-0.253	-0.226
クラスタ 2	-0.097	-0.369	0.107	-0.251	0.372	0.553
クラスタ 3	0.158	0.059	-0.042	-0.078	0.080	-0.195
クラスタ 4	-0.211	0.340	-0.370	0.375	-0.437	0.093
平均値	-0.030	0.050	-0.051	0.023	-0.059	0.056

- 「背景」のカテゴリ同様に、バナーの評価に大きく影響を与えるカテゴリである。

03. クラスタ分析：考察

各カテゴリースコアの影響度の考察

表9. 「車の向き」のカテゴリースコア

	車の向き		
	斜め正面	正面	側面
クラスタ 1	-0.025	0.042	0.039
クラスタ 2	-0.074	0.372	-0.173
クラスタ 3	0.032	-0.348	0.288
クラスタ 4	0.079	-0.410	0.199
平均値	0.003	-0.086	0.088

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ2, 3, 4ではレンジが大きい。

表10. 「シーン」のカテゴリースコア

	シーン	
	走行中	停止中
クラスタ 1	0.323	-0.083
クラスタ 2	-0.112	0.029
クラスタ 3	-0.058	0.015
クラスタ 4	-0.764	0.197
平均値	-0.153	0.039

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ1, 4ではレンジが大きい。

03. クラスタ分析：考察

各カテゴリースコアの影響度の考察

表11. 「サブバナー」のカテゴリースコア

	サブバナー	
	無し	有り
クラスタ 1	0.103	-0.165
クラスタ 2	-0.045	0.072
クラスタ 3	-0.062	0.099
クラスタ 4	0.099	-0.158
平均値	0.024	-0.038

- 各クラスタごとで影響度は正と負に分類される。
- 各クラスタでレンジが小さいため影響度は小さい。

表12. 「見出し語の大きさ」のカテゴリースコア

	見出し語の大きさ		
	小	大	中
クラスタ 1	0.443	-0.426	-0.012
クラスタ 2	-0.251	0.592	-0.221
クラスタ 3	0.405	-0.285	-0.078
クラスタ 4	-0.003	-0.072	0.049
平均値	0.149	-0.048	-0.065

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ1, 2, 3ではレンジが大きい
ため、バナーの評価に大きく影響を与えるカテゴリーである。

03. クラスタ分析：考察

各カテゴリースコアの影響度の考察

表13. 「文字面積」のカテゴリースコア

	文字面積		
	1/4以上	1/8以上	1/8未満
クラスタ 1	-0.100	0.036	0.052
クラスタ 2	-0.175	0.064	0.090
クラスタ 3	0.172	0.105	-0.245
クラスタ 4	0.574	0.036	-0.526
平均値	0.118	0.061	-0.157

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ3, 4ではレンジが大きい。

表14. 「車面積」のカテゴリースコア

	車面積		
	1/4以上	1/8以上	1/8未満
クラスタ 1	0.170	-0.073	-0.280
クラスタ 2	-0.417	0.304	0.238
クラスタ 3	0.265	-0.044	-0.688
クラスタ 4	-0.017	-0.088	0.372
平均値	0.000	0.025	-0.090

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ2, 3, 4ではレンジが大きい。

03. クラスター分析：考察

各カテゴリースコアの影響度の考察

表15. 「人・キャラクター」のカテゴリースコア

	人・キャラクター	
	無し	有り
クラスタ 1	0.041	-0.066
クラスタ 2	-0.006	0.009
クラスタ 3	-0.145	0.233
クラスタ 4	0.277	-0.444
平均値	0.042	-0.067

- 各クラスタごとで影響度は正と負に分類される。
- クラスタ3, 4ではレンジが大きい。

各カテゴリースコアの影響度の考察のまとめ

- 全クラスタに影響度が大きいカテゴリは「見出し語の位置」「背景」「背景色」「見出し語の大きさ」
- 各クラスタごとでレンジが大きいカテゴリは異なるため、カテゴリごとに各クラスタの解釈を行う必要がある。

03. クラスター分析：考察

クラスター1の評価が高いバナー

クラスター1の最大 カテゴリースコア	ロゴ	フォントの種類	フォント	見出し語の位置	背景	背景色
	無し	2種類以上	明朝体	下	無地	ブラック
車の向き	シーン	サブバナー	見出し語の大きさ	文字面積	車面積	人・キャラ
側面	走行中	無し	小	1/8未満	1/4以上	無し



見出し語小さく、車面積1/4以上
⇒ 高級感のある・高品質な
バナーを好むクラスターである

クラスター2の評価が高いバナー

クラスター2の最大 カテゴリースコア	ロゴ	フォントの種類	フォント	見出し語の位置	背景	背景色
	無し	2種類以上	ゴシック体	中	都会的	レッド系
車の向き	シーン	サブバナー	見出し語の大きさ	文字面積	車面積	人・キャラ
正面	停止中	有り	大	1/8以上	1/8以上	有り



レッド系、都会的、見出し語大
⇒ 力強い、インパクトのある
バナーを好むクラスターである

03. クラスター分析：考察

クラスタ3の評価が高いバナー

クラスタ3の最大 カテゴリースコア	ロゴ	フォントの種類	フォント	見出し語の位置	背景	背景色
	有り	1種類	明朝体	中	仮想空間	グリーン系
車の向き	シーン	サブバナー	見出し語の大きさ	文字面積	車面積	人・キャラ
側面	停止中	有り	小	1/4以上	1/4以上	有り



仮想空間、見出し語小
⇒ クールな、シンプルな
バナーを好むクラスタである

クラスタ4の評価が高いバナー

クラスタ4の最大 カテゴリースコア	ロゴ	フォントの種類	フォント	見出し語の位置	背景	背景色
	有り	1種類	明朝体	下	仮想空間	ブルー系
車の向き	シーン	サブバナー	見出し語の大きさ	文字面積	車面積	人・キャラ
側面	停止中	無し	中	1/4以上	1/8未満	無し



ブルー系、停止中
⇒ 清々しい・爽やかな
バナーを好むクラスタである

04. アソシエーションルール解析：概要

アソシエーションルール解析とは

- If – then ルールの形式でデータ内にある変数間の関係を抽出する解析方法

(背景色=グレー) かつ (フォント=明朝体) なら (魅力的である)
条件(前提部)
結論(帰結部)

- 本研究では**アプリアリ**のアルゴリズムでif-thenルールを抽出する。**アプリアリ**とは、確信度(confidence)と支持度(support)に基づきルール抽出を行う方法である。

$$\text{support} = \frac{\text{条件と結論を満たす対象数}}{\text{全対象数}} \quad \text{confidence} = \frac{\text{条件と結論を満たす対象数}}{\text{条件を満たす対象数}}$$

- S-PLUSでは“bigdata”ライブラリからアプリアリのアルゴリズムを実行できる[b]。

04. アソシエーションルール解析：準備

- 4つのクラスタごとに、結論部「魅力的である」に関するif-thenルールを抽出する。
- 抽出方法
 - support = 0.1, confidence = 0.9 に設定し、アプリアリのアルゴリズムでif-thenルールを抽出
 - 条件部の長さが短い上位3つのif-thenルールを被験者ごとに抽出する。
 - 条件部の長さが同じルールが3つ以上ある場合、その中でLift値が高いルールを取得する。
 - Lift値も同じ場合、すべてのルールを取得する。
 - 抽出後、1人の被験者内で、あるif-thenルールに包含されているif-thenルールは削除する。

(例)ルール1：「ロゴ無し」→「魅力的である」
ルール2：「ロゴ無し」かつ「背景色ブルー」→「魅力的である」
この場合、ルール2を削除する。

04. アソシエーションルール解析：考察

クラスタ1の考察

- クラスタ1内の被験者数：26人、抽出したルール総数：111

条件部での出現頻度が多いカテゴリ5種

出現頻度の多いif-thenルールの条件部

上位カテゴリ5種
人・キャラ無し
ロゴ無し
背景色がブラック
フォントが明朝体
背景が都会的

背景色がブラック ∧ 人・キャラ無し(5人)
 ロゴ無し ∧ フォントが明朝体(5人)
 背景が都会的 ∧ 車面積1/8未満(3人)



- 「背景色がブラック」「背景が都会的」などII類と同じカテゴリを持つルールが多い。
- したがって、II類と同じ解釈ができる。クラスタ1は好きなカテゴリがあるバナーを好むグループであると考えられる。

04. アソシエーションルール解析：考察

クラスタ2の考察

- クラスタ2内の被験者数：13人、抽出したルール総数：51

条件部での出現頻度が多いカテゴリ5種

出現頻度の多いif-thenルールの条件部

上位カテゴリ5種
ロゴ無し
人・キャラ無し
背景色がレッド系
フォントが明朝体
背景が自然

ロゴ無し ∧ 背景色がブラック(4人)
 背景色がレッド系 ∧ 停止中(2人)
 背景が自然 ∧ 車面積1/8以上(2人)



- 背景色がレッド系に関するルールが他クラスタと比べ多い。
- II類で明らかでなかった「背景が自然」に関するルールが多い。
- レッド系の配色と背景が自然のバナーを好む可能性が高い。

04. アソシエーションルール解析：考察

クラスタ3の考察

- クラスタ3内の被験者数：8人、抽出したルール総数：29

条件部での出現頻度が多いカテゴリ5種

出現頻度の多いif-thenルールの条件部

上位カテゴリ5種
人・キャラ無し
フォントが明朝体
背景色がグレー
背景が無地
車面積1/8未満

背景色がグレー (2人)
 フォントが明朝体 ∧ 車向き斜め正面(2人)
 背景が無地 ∧ 人・キャラ無し(1人)



- カテゴリースコアの小さかった「背景色がグレー」「背景が無地」「キャラ無し」などのカテゴリが多い。
- しかし、これらのカテゴリはクールでシンプルなバナーとなるデザイン要素であるため妥当な結果だと考えられる。

04. アソシエーションルール解析：考察

クラスタ4の考察

- クラスタ4内の被験者数：27人、抽出したルール総数：117

条件部での出現頻度が多いカテゴリ5種

出現頻度の多いif-thenルールの条件部

上位カテゴリ5種
ロゴ無し
サブバナー無し
背景色がブルー系
フォントが明朝体
車の向きが正面

ロゴ無し ∧ フォントが明朝体(5人)
 サブバナー無し ∧ 車の向き 正面(4人)
 ロゴ無し ∧ 背景色がブルー系(3人)



- 背景色がブルー系に関するルールが他クラスタと比べ多い。
- カテゴリースコアの小さかった「車の向き 正面」に関するルールが多いため、車の顔が見やすいものを好む傾向にあると考えられる。

04. アソシエーションルール解析：考察

アソシエーションルール解析全体の考察

- ・ 全クラスタを通じて、「ロゴ無し」「フォントが明朝体」「見出し語の位置が上」「車の向き 斜め正面」「停止中」「サブバナー無し」「人・キャラ無し」を条件に持つルールが多い。
- ・ 全クラスタを通じて、バナー内の情報量が少ないものを好む傾向にあった。
- ・ 被験者のバナーを好むカテゴリとして、全体の配色傾向・車の色など、カラーイメージの影響が高いと考えられる。
- ・ II類でカテゴリースコアの低かったカテゴリがif-thenルールの条件部に多く抽出されたこともあった。これは、カテゴリの組み合わせによる相乗効果による結果だと考えられるため、II類では発見が難しいカテゴリの組み合わせ効果を発見できると考えられる。

05. まとめ

本研究の成果

- 数量化理論第II類、クラスター分析により各クラスタ毎の被験者が好むデザイン要素を考察した。
- アソシエーションルール解析により、被験者が好むデザイン要素の組み合わせを考察した。

今後の課題

- 各クラスタ毎に推定した好まれるバナーの再評価
 - 本研究の考察結果の妥当性を確認する必要がある
- 各ユーザのクリック率を目的変数とした分析行うこと
 - 広告配信サービスを行う企業との協力が必要である
- If-then ルール解析と他の多変量解析手法との比較
 - 非線形的な手法であるif-thenルール解析の有用性を線形的手法と比較し、確認する必要がある

00. S-PLUSによる解析コード

数量化理論第II類の解析コード[a]

```
options(contrasts=c("contr.treatment","contr.poly")) # 1と0のダミー変数に変換するためのコマンド
matD <- model.matrix(~Logo + FontNum + FontType + TitleLocation + Background + BackgroundColor + CarDirection +
CarScene + Banner + CharaSize + CharaArea + CarArea + Person, data=Data) # 行列形式のデータとして読み込む
matD <- matD[,-1] # 定数項の列を削除
evaluation <- Data$Evaluation # 評価属性のインデックスを用意
vecN <- nrow(Data) # サンプル数を求める
vecn <- table(evaluation) # 各評価属性ごとのサンプル数を求める
vecb <- apply(matD,2,sum) # 各列ごとに1の数を数える
vec1 <- apply(matD[evaluation==1,],2,sum) # 評価属性が1のもので各列ごとに1の数を数える
vec2 <- apply(matD[evaluation==2,],2,sum) # 評価属性が2のもので各列ごとに1の数を数える
vec3 <- apply(matD[evaluation==3,],2,sum) # 評価属性が3のもので各列ごとに1の数を数える
matT <- t(matD)%*%matD - vecb%*%t(vecb)/vecN # 以下、カテゴリースコアの算出式
matB <- vec1%*%t(vec1)/vecn[1] + vec2%*%t(vec2)/vecn[2] + vec3%*%t(vec3)/vecn[3] - vecb%*%t(vecb)/vecN
i <- eigen(solve(matT)%*%matB)
```

[参考文献：どうすればS-PLUSで数量化理論2類ができるか。 , <http://www.msi.co.jp/splus/support/qa/a-13.html>]

アソシエーション分析の解析コード[b]

```
library(bigdata) #bigdataライブラリを読み込む
data.cv <- importData("data.txt", colNameRow=1, stringsAsFactors=F) #データを読み込む
bd.assoc.rules(data.cv, input.format="column.value", min.support=0.1, min.confidence=1.0) #実行
```

[参考文献：Big Data User's Guide for TIBCO Sportfire S+ 8.2, PP.59-77]

00. 参考文献

- [1]インターネット広告市場は2013年に8,500億円規模へ：
<http://japan.internet.com/wmnews/20090127/5.html>
- [2]廣瀬 寛：最適化アルゴリズムとクラウドソーシングプラットフォームによる広告クリエイティブの効果改善、電子情報通信学会技報、pp.13-18、2010
- [3]和田修平、椎塚久雄：バナー広告における色彩のイメージ伝達性の検証-コンセプトと情報の信頼性に注目して-、工学院大学研究報告、Vol.109、pp.129-134、2010
- [4]山川義介、管由紀子：広告クリエイティブの最適化の実際、第40回日本行動計量学会、2012
- [5]Agglomerative Coefficient : <http://stat.ethz.ch/R-manual/R-devel/library/cluster/html/coef.hclust.html>
- 対象間をはじめに併合したときの非類似度と最後に併合したときの非類似度により、クラスター全体の適合度を測る指標