

# 対応分析による関心領域マッピングと その評価軸の社会的価値に関する考察

A Mapping of Scope of People's Interest with Correspondence Analysis and  
a Discussion about The Evaluation Metric and it's Societal Value

森下 壮一郎 \*1      竹中 毅 \*1      武田 英明 \*1\*2  
MORISHITA, Soichiro      TAKENAKA, Takeshi      TAKEDA, Hideaki

\*1 東京大学人工物工学研究センター  
RACE, The Univ. of Tokyo.

\*2 国立情報学研究所  
National Institute of Informatics

In this paper, diversification of values among the people is discussed. We focus attention on lifestyle of the people, and surveyed scope of people's interest with a questionnaire about lifestyle. Additionally, we did a mapping of scope of people's interest with correspondence analysis. As a result, we found that the distribution of scope of people's interest was decided with according to gender difference and it's variance.

## 1. はじめに

人間の社会的活動は価値観の共有により営まれている。人々はそれぞれが別々の価値観を持つ一方で、貨幣や商品、サービス、労働などについての価値観を共有している。すなわち一般に人々は財やサービスについての価値観を共有していることが前提とされており、さもなくば交換が成り立たず、社会的活動の、特に経済的活動に支障を来す。しかしながら本来は人々の価値観は多様であり、また等価でなくとも価値がより高いものを手に入れることができると互いが判断すれば交換は行われるのであるから、厳密な意味での等価交換が行われることは極めて稀である。例えば、二者間で財やサービスの交換が行われたことを第三者が観察したとき、二者間で共有されている価値観に応じて等価と見なされて交換が行われたと解釈するのが自然であるが、実際はその逆である。すなわち、交換された財やサービスが等価となるような価値観をもって、それが二者間で共有されているものと見なしているに過ぎない。財やサービスの社会的価値は、以上のような手続きにより決まる価値観により規定されると考えられるが、その価値観は往々にして交換を行った二者のいずれの価値観とも合致しないであろう。一般に社会的活動の分析が困難である理由はこの点にあり、いずれにしろ個人の価値観を知ることが極めて重要である。しかしながら価値観は個人の内面に關わるものであり、しかもそれは本質的には観測不能である。価値観を直接知ることが不可能であるので、外部から観測可能な情報により個人をモデル化し、それに基づいて推定せざるを得ない。

以上のことから本稿では価値観推定のための個人のモデル化を目指し、個人の内面に關わる事項として趣味や娯楽についての人々の関心領域に着目して、ライフスタイル調査の結果に基づく対応分析 [Benzécri 92] によりそのマッピングを行う。そして観測可能な情報としての性別、観測不可能な情報としての関心領域との対応付けについて考察する。

## 2. ライフスタイル調査

ライフスタイル調査の一環として携帯電話を用いた登録制調査サービスを利用し、様々な日常行動や性格、情報機器への接触態度、情報への関心、経済的価値観などについてのアンケート

調査を実施した [Takenaka 08]。アンケートの回答者には割引クーポンなどに交換できるポイントが付与される。調査は2007年6月に主に首都圏の10代から60代までを対象として行った。回答数は  $N = 8,177$  であった。

### 2.1 世代および性別毎の比率

表1に回答者の世代および性別毎の度数分布表を示す。なお16歳以上65歳以下の各クラスについては、日本国内における世代および性別毎の分布についての統計資料に基づき期待値を計算して括弧内の数値として示した。

表1: 世代および性別毎の集計表

年齢	男性		女性		計
-15	49	—	88	—	137
16-20	303	(321)	669	(304)	972 (626)
21-25	282	(358)	805	(343)	1087 (701)
26-30	453	(400)	829	(389)	1282 (789)
31-35	718	(470)	815	(459)	1533 (930)
36-40	723	(420)	558	(413)	1281 (833)
41-45	524	(387)	285	(383)	809 (770)
46-50	287	(369)	157	(368)	444 (736)
51-55	209	(418)	160	(421)	369 (838)
56-60	111	(484)	47	(493)	158 (977)
61-65	66	(396)	13	(418)	79 (814)
66-	19	—	7	—	26
計	3744	—	4433	—	8177

実際の日本国内の人口比に対して高齢層の度数が低い傾向にある。これは、携帯電話を用いたアンケートであるために、携帯電話のユーザ比率が反映していると考えられる。男女比を見ると、女性は16歳から30歳までのクラスにおいて男性を大きく上回る。一方36歳を過ぎると男女比は逆転する。理由は様々考えられるが、アンケート回答のインセンティブがポイント付与であることがその一つとして挙げられよう。

## 3. 関心領域マッピング

趣味や娯楽についての人々の関心領域として20項目を列挙し、対応分析によるマッピングを行った。各項目について興味関心の有無について複数回答を許して質問した。全ての回答者

連絡先: 森下 壮一郎, 東京大学人工物工学研究センター, 千葉県柏市柏の葉 5-1-5, E-mail: mori at race.u-tokyo.ac.jp

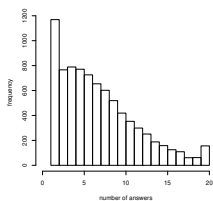


図 1: 回答数のヒストグラム

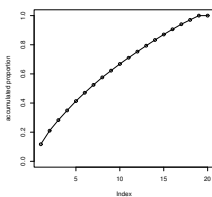


図 2: スコアの累積寄与率

について最低限一つの項目を挙げさせた．表 2 に関心領域項目と回答数を示す．

表 2: 関心領域項目とその回答数

関心領域項目	回答数	関心領域項目	回答数
ファッション	3,804	テレビ	3,586
ビューティー	2,751	ゲーム	3,365
経済・ビジネス	2,664	スポーツ	2,646
政治	1,657	車	1,911
音楽	4,913	ギャンブル	1,422
グルメ	3,900	旅行・温泉	4,151
インテリア・住居	2,708	恋愛	1,664
お稽古・スクール	1,339	健康・医療	2,769
芸能ニュース	2,872	映画	3,820
PC・インターネット	4,391	自然・環境	2,365

さらに図 1 に回答数のヒストグラムを示す．なお 156 人が 20 項目の全てに関心ありと答えている．これは、本調査で利用したようなポイント付与を行う調査サービスにおいては、ユーザの興味を事前に聞いてそれに応じたアンケートを行うことが多いので、なるべく多くの関心領域に興味があると回答した方がポイント獲得の機会を得るために有利であるという心理が回答者に働いたものと考えられる．

### 3.1 対応分析による関心領域マッピング

前述の理由により、20 項目の全てに関心ありと回答されているサンプルは回答の信憑性が疑わしく、また 19 項目に関心ありと回答されているサンプルについても 20 項目の全てに関心ありと答えるつもりで操作ミスをしたと考えられることから、以下の分析ではそれらを除いた  $N = 7957$  のサンプルを用いる．そして行に回答者、列に関心領域を列挙した 7,957 行 20 列の分割表を作成し、回答者と関心領域についての対応分析を行った．計算には統計処理ソフトウェアの一つである R を用いた [Ihaka 96]．対応分析は、対応関係のあるカテゴリカルデータの分析によく用いられる手法の一つであり、これにより分割表における行スコアと列スコアが得られる．それぞれのスコアは各項目間の近さを表す．各スコアにより張られる空間を重ねてプロットすることで、行と列の項目間の近さを観察することもできる．

図 2 にスコアの累積寄与率をしめす．これはスコアの次元に応じてデータの分布がどの程度再現されるかの指標になる．

スコアの次元が 16 次元のときに累積寄与率 0.9 を超えることから、データの分布を十分に表現するには 16 次元程度必要であることが分かる．しかしながら多次元の量を可視化することはできないので、スコアを 2 次元まで採用して平面上に分布をプロットした．図にその結果を示す．

横軸が第一番目のスコア、縦軸が第二番目のスコアである．

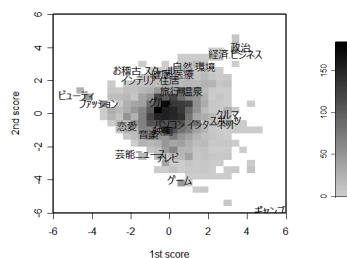


図 3: 関心領域の対応分析結果

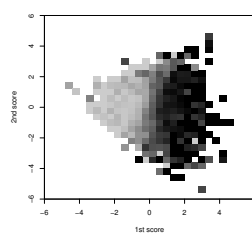


図 4: 男性のみの分布

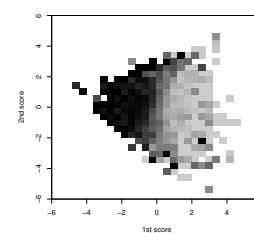


図 5: 女性のみの分布

グリッドは回答者の分布を表しており、原点に近いほど密度が高い．文字列は回答者の分布に対応する関心領域である．原点に近い位置に配された関心領域は多数の回答者が興味ありと答えており、原点から離れるほど興味ありと答える回答者が少数になる傾向にある．さらに、男性のみの分布と女性のみの分布をそれぞれ各階級において正規化して表したものを図 4 および図 5 に示す．男性の分布は右側に、女性の分布は左側に偏っている．すなわち、第一番目のスコアは男女差を表すものと解釈できる．また全体の分布を見ると、第二番目のスコアが大きいほど第一番目における分布が大きくなるような逆三角形の分布をしていることから、第二番目のスコアは男女差の分布の大きさを表すものと解釈できる．

## 4. おわりに

本稿では価値観推定のための個人のモデル化を目指し、ライフスタイル調査の結果に基づく対応分析により人々の関心領域のマッピングを行った．その結果、第一スコアとして性差を表す軸が、第二スコアとしては性差に依存するばらつきを表す軸が得られた．すなわち、人々の関心領域の分布は性差とそれに応じた分散により決まることが示唆された．

## 参考文献

[Benzécri 92] Benzécri, J.: *Correspondence Analysis Handbook*, CRC Press (1992)

[Ihaka 96] Ihaka, R. and Gentleman, R.: R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, Vol. 5, No. 3, pp. 299–314 (1996)

[Takenaka 08] Takenaka, T. and Ueda, K.: An Analysis of Service Studies toward Sustainable Value Creation, *International Journal of Sustainable Manufacturing*, Vol. 1, No. 1 (2008), in printing