撤回のお知らせ

撤回：数量化Ⅲ類におけるカテゴリー内積情報を用いたプロファイリング手法の提案

鈴木 誠, 吉川 大弘, 古橋 武

名古屋大学大学院工学研究科

日本感性工学会論文誌 Vol.11 No.1 (2012) pp.63-68

著者より同内容の論文が英文誌 International Journal of Affective Engineering Vol.12 No.2 p.89-94(2013)に掲載されているとの報告を受け、重複を確認いたしました。用語に誤りが見られることから、著者の要望により、先に掲載された本論文を撤回いたします。

日本感性工学会 論文審査編集委員会

Retraction of Duplicate Publication

Retraction: A Proposal of Profiling Method using Categorical Inner Products in Quantification Theory Type 3

Makoto SUZUKI, Tomohiro YOSHIKAWA and Takeshi FURUHASHI

Graduate School of Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

Transactions of Japan Society of Kansei Engineering, Vol. 11 No. 1 (2012) pp.63-68

Japan Society of Kansei Engineering
Editorial Board

原著論文

数量化Ⅲ類におけるカテゴリー内積情報を用いた プロファイリング手法の提案

鈴木 誠, 吉川 大弘, 古橋 武

名古屋大学大学院工学研究科

A Proposal of Profiling Method using Categorical Inner Products in Quantification Theory Type 3

Makoto SUZUKI, Tomohiro YOSHIKAWA and Takeshi FURUHASHI

Graduate School of Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

Abstract : In the field of marketing, it is an important issue to profile the users who will purchase new products. Correspondence analysis and quantification theory type 3 will be the popular methods for this profiling. Correspondence analysis can give the associations between groups and attributes, and then it can estimate the profiles of the users. However the deeper analysis of profiling is required because of the variety of their taste. This paper used the group and the attribute information that were converted into the Eigen space from the characteristic space, and aimed to evaluate the relations of the groups and attributes in detail by the independency of the coordinate axes in the Eigen space. The method to visualize the placement between groups and the inner product between a group and each attribute was shown by applying this proposal method to a test data. This paper also applied the proposed method to an actual questionnaire data on a user's interest, and the detailed profiling result for the relationship between the generation and their interests was shown.

Keywords : *Quantification Theory Type 3, Analysis of Questionnaire Data, Profiling*

1. は じ め に

マーケティング分野では、新規製品がどのようなプロファイルを持つユーザに受け入れられるか推定することが重要な課題であり、これまでもさまざまなプロファイリング手法の提案がなされている [1-4]。これらの手法に加え、年代や性別によるグループと、趣味の種類などアンケート回答から得られる属性を含むカテゴリーにおいて、カテゴリー相互の位置づけを明らかにしてグループのプロファイルを明確にする方法が求められている。このような課題に対応するため、従来用いられてきた有力な分析手段として対応分析（数量化Ⅲ類）が挙げられる [5-10]。この対応分析によって、グループと属性の間の位置づけを把握することが可能になり、ユーザがどのようなプロファイルを持っているか推定することができる。対応分析では、年齢などのグループ情報と趣味などの属性情報を同時に布置し、例えばユークリッド距離を比較して、より近いグループと属性を関連づけることによって、各グループに受け入れられる属性といったユーザプロファイルを求めることができる。

しかしながら、趣味やライフスタイルが多様化した現代では、グループと属性の関係が明確に把握できない場合があり、より詳細なプロファイルを得る手法が求められてきている。また複数のグループの中間に位置する属性が、これらのグループに近い関係にあるか、あるいはいずれのグループと

も関係が無い場合中間位置に配置されているかを明らかにする必要が生じている。

本論文では、アンケートデータに対する数量化Ⅲ類の解析において、特徴空間から固有空間に変換したグループ情報と属性情報を用いた、固有空間における座標軸の独立性に着目し、グループと属性からなる各カテゴリー間の関係を詳細に評価することを目指した。まず、固有空間におけるグループ情報からグループの相対的配置を求めるため、数量化Ⅲ類を適用してデータの特徴空間から固有空間に変換する方法を示した。さらにグループと属性の間の内積を求め、この値を調べることによって、より詳細なプロファイルを求める方法を示し、提案手法をテストデータに適用し、グループ間の配置およびグループと属性の間の内積値を可視化する方法を示した。最後に、年代で分けたグループのこだわりの領域のアンケート解析に本手法を用いた例を示した。

2. 数量化Ⅲ類

数量化Ⅲ類は林知己夫氏によって着想された解析手法である [5-7]。本論文では、各行がサンプル（サンプル数： m ）、各列が属性とグループからなるカテゴリー（カテゴリー数： n ）を表すものとして、 $m \times n$ のデータを用いてアルゴリズムを示す。

$$A = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \cdots & \sigma_{1n} \\ \sigma_{21} & \ddots & & & \vdots \\ \vdots & & \sigma_{ij} & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{m1} & \cdots & \cdots & \cdots & \sigma_{mn} \end{pmatrix} \quad (1)$$

ここで $\sigma_{ij} = \begin{cases} 1 & (\text{サンプル}i\text{がカテゴリ}j\text{に反応した時}) \\ 0 & (\text{サンプル}i\text{がカテゴリ}j\text{に反応しない時}) \end{cases}$

また、行要素および列要素の合計値から対角行列 B および C を用いる、

$$B = \begin{pmatrix} b_1 & & & 0 \\ & \ddots & & \\ & & b_i & \\ 0 & & & \ddots \\ & & & & b_m \end{pmatrix} \quad (2)$$

$$C = \begin{pmatrix} c_1 & & & 0 \\ & \ddots & & \\ & & c_j & \\ 0 & & & \ddots \\ & & & & c_n \end{pmatrix} \quad (3)$$

ここで、 $b_i = \sum_{j=1}^n \sigma_{ij}$, $c_j = \sum_{i=1}^m \sigma_{ij}$ である。

サンプルスコアを x 、カテゴリスコアを y で表す。それぞれ平均0、分散1の下で x と y の相関係数が最大になる条件から、固有方程式

$$Hl = \lambda^2 l \quad (4)$$

が求まる。

ここで $H = B^{-\frac{1}{2}} A C^{-1} A^T B^{-\frac{1}{2}}$ である。

この固有方程式より固有値 λ^2 と各固有値に対応した固有ベクトル l を求め、サンプルスコア x を算出することができる。固有値1を除いた h 番目の固有値 $\lambda^{2(h)}$ に対応する固有ベクトル $l^{(h)}$ を用いて、第 h 軸のサンプルスコア $x^{(h)}$ は

$$x^{(h)} = B^{-\frac{1}{2}} l^{(h)} \quad (5)$$

で表される。

さらに、本論文では固有空間における第 h 軸のカテゴリスコア $y^{(h)}$ の各座標軸に重み付けを行い

$$y^{(h)} = \frac{1}{\lambda^{(h)}} C^{-1} A^T x^{(h)} \times \lambda^{(h)} = C^{-1} A^T x^{(h)} \quad (6)$$

として解析を行った。

3. 提案手法

3.1 グループの布置

グループ情報を、円周方向の角度として表す。数量化Ⅲ類を行った後、 $\min(m,n)-1$ 次元の固有空間におけるカテゴリー情報から、グループ s とグループ t の間の角度差を

$$|\varphi_s - \varphi_t| = \cos^{-1} \left(\frac{\sum_{h=1}^{\min(m,n)-1} y_s^{(h)} y_t^{(h)}}{\sqrt{\sum_{h=1}^{\min(m,n)-1} y_s^{(h)2}} \sqrt{\sum_{h=1}^{\min(m,n)-1} y_t^{(h)2}}} \right) \quad (7)$$

で表し、

$$E = \sum_s \sum_t \left| |\varphi_s - \varphi_t| - |\theta_s - \theta_t| \right| \quad (8)$$

を用いて、固有空間における角度関係をできるだけ維持しながら、目的関数 E が最も小さくなるように2次元平面上での角度を求める。ここで $y_s^{(h)}$ は第 h 軸のグループ s の値、 θ_s はグループ s の2次元平面における角度である。

3.2 属性情報の一覧

複数の属性情報を一覧し、さらにより詳細の属性情報を把握できるように、2段階で表示を行う。まず、属性情報一覧では、グループ s に対する属性 r の内積値 I_{sr} は

$$I_{sr} = \sum_{h=1}^{\min(m,n)-1} y_s^{(h)} y_r^{(h)} \quad (9)$$

で表され、式 (9) を用いて属性 r の座標 (x_r, y_r) は

$$x_r = \frac{\sum_{s=1}^q I_{sr} \cos \theta_s}{q}, y_r = \frac{\sum_{s=1}^q I_{sr} \sin \theta_s}{q} \quad (10)$$

で求めることができる。ここで q はグループの数である。ただし式 (10) では、 $I_{sr} > 0$ のみ計算に用いる。

3.3 属性情報の詳細

さらに属性の詳細情報は、円周上に配置したグループで、グループと属性の内積値を径方向に配置する多角形状で表現する。

4. テストデータを用いた解析

表1のテストデータに提案手法を適用した。ここで1～15はサンプル、a～eおよび20～60はカテゴリーで、a～eは属性、20～60はグループを表す。

表1のテストデータに従来の数量化Ⅲ類を適用した例を図1に示す。この解析によって、例えば属性bがグループ30の近くに布置されることから、特性的に近いことが予測されるが、さらに他のグループとの関係や、グループ間の近さの情報が与えられることによって、より詳細な解析が可能となると考えられる。

提案手法を用いて解析した表1のテストデータの属性情報の一覧を図2に、属性情報の詳細として属性a～eの解析結果を図3(i) および(ii)に示す。図2は中心を0とし、3.1で得られたグループ間の角度関係に基づき、半径1の円周上にグループを配置して、式(10)で求めた座標に各属性を

配置している。また、図3 (i) および (ii) では、グループと属性の内積値は負の値もとるため、グループ情報を配置する円の内側に、半径がグループ情報の1/2となる円を配置した。この半径1/2の円を境にグループ情報の円との間は内積値が正、内側は内積値が負として示した。

図2ではa～eの属性が各グループに対してどのような位置づけとなるかを概観することができ、例えば属性bはグループ20とグループ30の中間で、グループ30に近い角度に配置されるため、属性bとグループ30は近い関係にあることがわかる。またグループ30, 40, 50が比較的近いこと、

表1 テストデータ

		カテゴリー									
		属性					グループ				
		a	b	c	d	e	20	30	40	50	60
サンプル	1	1	1	0	0	0	1	0	0	0	0
	2	1	0	0	0	0	1	0	0	0	0
	3	1	1	0	0	0	1	0	0	0	0
	4	0	1	1	0	0	0	1	0	0	0
	5	0	1	0	0	0	0	1	0	0	0
	6	0	1	1	0	0	0	1	0	0	0
	7	0	0	1	0	0	0	0	1	0	0
	8	0	0	1	0	0	0	0	1	0	0
	9	0	0	1	0	0	0	0	1	0	0
	10	0	0	1	1	0	0	0	0	1	0
	11	0	0	1	0	0	0	0	0	1	0
	12	0	0	1	1	0	0	0	0	1	0
	13	0	0	0	1	1	0	0	0	0	1
	14	0	0	0	1	1	0	0	0	0	1
	15	0	0	0	1	1	0	0	0	0	1

図1 従来方法による表示 (テストデータ)

図2 属性情報の一覧 (テストデータ)

(i) a～c

(ii) d, e

図3 属性情報の詳細 (テストデータ)

グループ20とグループ60とが遠く対極にあることなどが把握できる。さらに図3(i) および(ii) のように属性情報の詳細を用いてより詳しい特性を把握できる。図3(i) および(ii) で、例えば属性bはグループ30との内積値が大きく近い関係にあるが、グループ40～60に対して内積値は負で関連が少ないことがわかる。

5. アンケートデータの解析と考察

実際のアンケートデータの解析として、表2に示すようなこだわりの領域を質問するアンケートの回答に対して解析を行った。ただし実際には、本データは、2008年12月9日より12月19日までインターネット調査で得られた情報機器に関するアンケートの回答データであり、こだわりの領域の質問は本アンケートの一設問である。20代から60代の計5006名の回答から、空白を含む回答および矛盾を含む回答を除いた3803名の回答に対して解析を行った。

数量化Ⅲ類による結果を図4に示す。図4では各属性情報は1点に代表されるため、多数の属性情報を同時に一覧できる反面、対象となる属性情報が複数のグループに関わる状況の把握や、グループ毎の詳細な関係の把握が困難であった。

提案手法による属性情報の一覧を図5に示す。図5よりd(園芸・ガーデニング)が比較的高い年令で支持され、b(ゲーム)が比較的低い年令で支持されていることがわかる。属性の中でd(園芸・ガーデニング)、b(ゲーム)、q(Webサイト閲覧)、p(ブログ・SNS)、r(ドライブ・ツーリング)およびi(演劇・バレエ)に関する属性情報の詳細をそれぞれ図6(i)～(vi)に示す。提案手法による属性情報の詳細で、それぞれのグループに対する内積情報を示している。これによって、複数のグループの中間に位置する属性が、これらのグループに近い関係にあるか、あるいはいずれのグループとも関係が無い場合中間位置に配置されているか明らかにすることが可能になった。

図6(i)のd(園芸・ガーデニング)では20代、30代の興味が低く、50代、60代で関心が高くなる。一方、図6(ii)のb(ゲーム)は20代、30代の関心が高く、40代以上では急激に関心が低くなる。このように提案手法によって、年代

表2 アンケート内容

	あなたはどのような領域にこだわりをお持ちですか。
a	スポーツ
b	ゲーム
c	料理・アルコール(お酒)
d	園芸・ガーデニング
e	絵画・工芸品などの美術
f	華道・茶道・書道・日舞などの伝統芸能
g	写真撮影
h	書籍
i	演劇・バレエ
j	音楽
k	映画
l	ドラマ
m	俳優・タレント・アイドル・有名人
n	アニメ・マンガ
o	キャラクター
p	ブログ・SNS
q	Webサイト閲覧
r	ドライブ・ツーリング
s	アウトドア
t	旅行
u	ショッピング
v	免許・資格
w	株などの金融商品
x	ギャンブル

図4 従来手法による表示(アンケートデータ)

図5 属性情報の一覧(アンケートデータ)

(i) d (園芸・ガーデニング)

(iv) p (ブログ・SNS)

(ii) b (ゲーム)

(v) r (ドライブ・ツーリング)

(iii) q (Webサイト閲覧)

(vi) i (演劇・バレエ)

図6 属性情報の詳細の例 (アンケートデータ)

毎の変化を把握することが可能となった。図6(iii)のq(Webサイト閲覧)と図6(iv)p(ブログ・SNS)はこだわりを持つ年代が類似していると考えられがちであるが、提案手法によってq(Webサイト閲覧)が50代、60代にも受け入れられつつあるのに対して、p(ブログ・SNS)は若い世代が中心であることがわかる。

さらに図6(v)のr(ドライブ・ツーリング)は、提案手法によって全年代で関心が低い傾向があるものの、40代では比較的関心が高く、20代で特に低いことが明らかになった。図6(vi)のi(演劇・バレエ)では、r(ドライブ・ツーリング)とは異なり50代に比較的受け入れられていることがわかる。

6. お わ り に

本論文では、数量化Ⅲ類の解析によって求めた、固有空間におけるグループと属性からなるカテゴリー情報を用いたプロファイリング手法を提案した。グループ情報からグループ間の相対的配置を求め、さらにグループ情報と属性情報の内積値を可視化して表示する方法を示した。この内積情報の変化を解釈することによって、より明確なプロファイリングが可能になった。さらに提案手法を年代ごとに分けたグループがこだわる領域のアンケート解析に用いた例を示し、いくつかの結果例を示した。

提案手法を推進するため、可視化に適した表現方法および内積値の解釈については今後さらに検討を重ねる予定である。

参 考 文 献

- [1] 伊藤晃, 吉川大弘, 古橋武, 光松佐和子: 階層的可視化手法を用いたアソシエーション分析によるプロファイリング, 日本感性工学会論文誌, Vol.10, No.2, pp.205-212, 2011.
- [2] 上田尚一: クラスター分析-講座<情報をよむ統計学7>, 共立出版, 2003.
- [3] 上田尚一: 主成分分析-講座<情報をよむ統計学8>, 共立出版, 2003.
- [4] 松村真宏: 影響の普及モデルに基づくオンラインコミュニティ参加者のプロファイリング, 人工知能学会論文誌, Vol.18, No.4, pp.165-172, 2003.
- [5] 林知己夫: 林知己夫著作集<第3巻>質を測る-数量化理論, 勉誠出版, 2004.
- [6] 林知己夫: 多変量データの解析 [IV・完], 電子通信学会誌, Vol.68, No.7, pp.779-786, 1985.
- [7] 大隅昇: 対応分析法・数量化法Ⅲ類の考え方 <http://www.ai.u-hyogo.ac.jp/~arima/ca.pdf>
- [8] 君山由良: コレスポンデンス分析の利用法, データ分析研究所, 2005.
- [9] 齋藤, 豊田: コレスポンデンス分析における布置の精度, オペレーションズ・リサーチ, Vol.49, No.3, pp.168-173, 2004.
- [10] Donna L. Hoffman and George R. Franke: Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research, Journal of Marketing Research, Vol. XXIII, pp.213-227, 1986.

鈴木 誠 (学生会員)

昭和54年名古屋大学大学院結晶材料工学専攻博士課程前期過程修了。昭和56年1月ブラザー工業株式会社入社。現在に至る。平成22年4月より名古屋大学大学院工学研究科計算理工学博士課程後期課程に社会人として在籍。感性工学に関する研究に従事。電子情報通信学会の会員。

吉川 大弘 (正会員)

1997年名古屋大学大学院博士課程修了。同年カリフォルニア大学バークレー校ソフトウェア工学研究所客員研究員。1998年三重大学工学部助手。2005年名古屋大学大学院工学研究科COE特任准教授。2006年10月同研究科准教授。現在に至る。主としてソフトウェアとその応用に関する研究に従事。博士(工学)。

古橋 武 (正会員)

昭和60年名古屋大学大学院工学研究科博士後期課程電気系専攻修了。工学博士。平成16年名古屋大学大学院工学研究科計算理工学専攻教授。現在に至る。ソフトウェア工学, 感性工学に関する研究に従事。平成8年日本ファジィ学会論文賞受賞。IEEE, 日本知能情報ファジィ学会, 電気学会等の会員。