

スペクトラルクラスタリングを用いたアンケートデータ解析に関する一検討

稲垣 和人[†] 吉川 大弘[†] 古橋 武[†]

名古屋大学大学院 工学研究科[†]

1 はじめに

マーケティングにおいて、企業が市場調査を通して、自社の製品やサービスに対する顧客の需要や評価を把握することは極めて重要である。例えば、企業が新しい製品の開発をする際には、対象となる顧客の需要を理解した上で企画をし、また既製品に対する顧客の評価なども考慮して販売戦略が立てられる [1]。

このような市場調査の方法の1つがアンケート調査であり、評価対象に対する各質問項目に複数段階の評点を付けることで、回答者の対象に対する印象が数値化されたアンケートデータを得ることができる。得られたアンケートデータは一般的に、クラスター分析 [2] や、主成分分析 [3]、多次元尺度構成法 [4] などに代表される多変量解析手法を用いて解析される。しかしこれらのアプローチでは、解析結果に影響を与える可能性があるノイズとなる回答や、少数ではあるが有益な特徴を持った、いわゆるマイノリティを抽出することは難しい。そこで本稿では、スペクトラルクラスタリング [5] を用いることで、少数の特徴的な回答者群を抽出することを試みる。

2 スペクトラルクラスタリング

スペクトラルクラスタリングは、クラスタリングをグラフ分割の問題として解く手法である。このときグラフのノードにデータ、ノード間のエッジの重みにデータ間の類似度が対応する。このように表わされたグラフについて、枝切りを行うことで全体のグラフをいくつかのサブグラフに分割する。その際に、サブグラフ内のエッジが密になり、サブグラフ間のエッジが疎になるような評価関数を設定する。このための評価関数はいくつか提案されているが、ここでは代表的な $Ncut$ [5] を利用する。まず、グラフのノード集合 V を2つのサ

ブグラフ A と B に分けることを考える。あるノード u, v の間でのエッジの重みを $w(u, v)$ としたとき、サブグラフ A と B の類似度 $cut(A, B)$ を以下のように定義する。

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v) \quad (1)$$

このとき、評価関数 $Ncut$ は以下で表される。

$$Ncut(A, B) = \frac{cut(A, B)}{cut(A, V)} + \frac{cut(A, B)}{cut(B, V)} \quad (2)$$

この式を最小化することは、サブグラフ内の類似度を大きく、かつサブグラフ間の類似度を小さくすることに等しい。またこの最小化問題は、一般化固有値問題に帰着することが示されている。 W をデータ間の類似度行列、 D を W の次数を対角成分に持つ行列とすると、 $D^{-1}(D - W)$ の固有ベクトルがグラフの分割を与える。ただし最小固有値は0のため、2番目に小さな固有値に対する固有ベクトルを用い、ある値以上の要素値を持つノードをクラスタ A に、それより小さいノードをクラスタ B に対応させることでクラスタリングを行う。評価関数を用いてエッジのカット位置を求める方法もあるが、本稿では値が正ならばクラスタ A に、負ならばクラスタ B に対応させる簡易な方法を用いる。形成されたサブグラフのいずれかを再び2分割するという操作を、目的のクラスタ数が得られるまで、あるいはユーザの満足する分割が行えなくなるまで繰り返す。

3 実験

3.1 実験内容

実際のアンケートデータを用いて実験を行った。1014名の回答者に対して、次世代サービスに関するWebアンケートを行った。本アンケート調査では、各3つのサービス（アフターサービス、ユビキタス、リサイクル）についての曖昧な説明と具体的な説明の合わせて6つの説明を評価対象とし、評定尺度法により、10個の質問項目に対してそれぞれ5段階の評点 {1,2,3,4,5} で評価してもらった。

得られたアンケートデータに対して、2節で述べたスペクトラルクラスタリングを用いて回答者のクラス

A Study on Analysis of Questionnaire Data Using Spectral Clustering

[†]Kazuto Inagaki

[†]Tomohiro Yoshikawa

[†]Takeshi Furuhashi

[†]Graduated School of Engineering Nagoya University

タリングを行った。入力データは、各回答者を6対象×10質問の合計60次元のベクトルとして表わしたものをを用いた。また、回答者 a, b の回答データをそれぞれ $\mathbf{x}_a, \mathbf{x}_b$ としたとき、回答者 a, b 間の類似度 $w(a, b)$ を次式で定義した。

$$w(a, b) = \exp\left(-\frac{|\mathbf{x}_a - \mathbf{x}_b|^2}{\sigma^2}\right) \quad (3)$$

ただし、 $\sigma^2 = 10$ とし、2分割クラスタリングを6回実行した。

3.2 結果と考察

多次元尺度構成法による回答者の可視化結果を図1に示す。ただし、多次元尺度構成法における回答者 a, b 間の距離は、 $\mathbf{x}_a, \mathbf{x}_b$ 間のユークリッド距離を用いた。各点の番号はスペクトルクラスタリングにより形成されたクラスタ番号を表わす。また、各クラスタにおける平均評点を図2に示す。

図1より、中心付近の密集部分から外れた回答者が少数クラスタとして抽出されていることがわかる。図2より、それらの回答者の回答傾向は中心付近の回答者(クラスタ7)と大きく異なっていることがわかる。例えば、クラスタ2には、各評価対象に1または5のみを付けた、いわばノイズとなる回答者が、クラスタ4には、評価対象による印象の差がない回答者が、クラスタ6には、全体的に高い評点を付けた回答者が抽出されている。

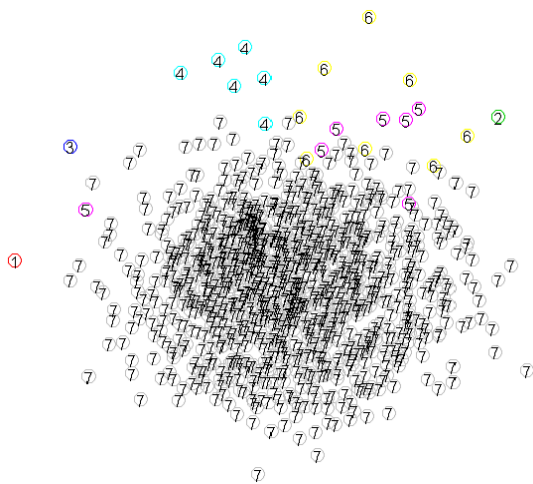


図1: 回答者の可視化結果

4 おわりに

本稿では、スペクトルクラスタリングを用いて、アンケートデータにおける回答者のクラスタリングを

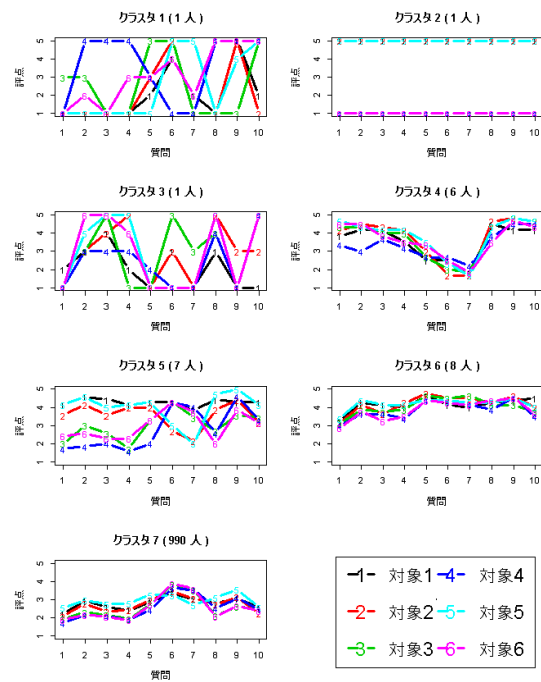


図2: 各クラスタの平均評点

行った。実際のWebアンケートデータに適用し、特徴的な回答傾向を持つ少数の回答者群が抽出されることを示した。今後の課題として、回答者間の類似度関数と得られる結果との関係性の解析や、クラスタ数に対する検討などが挙げられる。

参考文献

- [1] 木下祐介, 井上勝雄, 酒井正幸: 携帯電話機デザインの男女差の調査分析, 感性工学会研究論文集, Vol.7, No.3, pp.449-460(2008).
- [2] Michael R. Anderberg: クラスタ分析とその応用, 内田老鶴圃 (1988).
- [3] 田中義敏, 野間ロー由: 特許流通アドバイザーに求められる「マインド及び行動のスキル」に関する研究, Vol.24, No.1(2009).
- [4] 深見俊和, 吉川大弘, 古橋武, 原以起, 米田洋之: 可視化空間上でのインタラクティブクラスタリングによるマイノリティ発見に関する検討, 日本感性工学会論文誌, Vol.8, No.4, pp.1185-1191(2009).
- [5] J.Shi and J.Malik: Normalized cuts and image segmentation, IEEE Trans, Pattern Analysis and Machine Intelligence, Vol.22, N0.8, pp.888-905(2000)