

第5回異分野融合テキストマイニング研究会
於九州大学箱崎キャンパス旧工学部5号館(2016年5月15日)

ミックス法における テキストマイニングの活用

寺崎新一郎(九州大学)

はじめに

問題意識

- 手法としてのテキストマイニングは広く普及してきている一方、その活用の幅はいまだ限定的といえる。
- テキストマイニングでは、二次データの活用がほとんどであり、一次データの収集および分析はあまりみられない。

発表の骨子

- ミックス法の概観をレビューする。
- ミックス法における順次的探索的戦略とテキストマイニングを融合させる方略として、**MA-CAハイブリッド・アプローチ (M-CHA)**を紹介する。
- M-CHAの適用へ向けたリサーチデザインを併せて提案し、研究に活用する際のポイントを検討する。

本発表の流れ

1. データの収集および分析手法
2. 各手法が内包する課題
3. ミックス法とは何か、またその有用性
4. ミックス法のリサーチデザイン
5. 順次的探索的戦略
6. 順次的探索的戦略の応用：M-CHA
7. M-CHAにおける一次データの活用
8. 順次的探索的戦略におけるM-CHA概略図
9. 結語

データの収集および分析手法

データの収集および分析手法は、次の三つに弁別できる。

- **質的研究** (qualitative research)
- **量的研究** (quantitative research)
- **ミックス法** (mixed methods)

研究論文の大半は量的研究にもとづいており、次に質的研究が多くみられ、ミックス法を用いた研究は相対的に少ない。

ミックス法とは

- 量的および質的データを逐次的に収集し、分析をおこなう手法である。
- 異なる種類のデータを収斂させることを、**トライアングレーション** (triangulation) といい、ミックス法の説明過程で頻繁に用いられている。
- ミックス法の原型は、Campbell & Fiske (1959) にみられ、量的および質的データを同時に取り扱ったり、トライアングレーションさせたりする考え方が紹介されている (Jick, 1979)。

量的研究が多くみられる背景

- その「**客観性**」への信頼があると考えられる。
 - 諸現象を引き起こす「法則」や「要因」の探究は、社会科学と物理学などの自然科学とでは異質なものである（仲正, 2014）。
- ウェーバー（1998）は、社会科学において普遍的かつ汎用的な真理の存在は所与ではないという、量的研究に対する批判を投げかけている。
 - 人間が介在する限り、いかなる状況においても成り立つ方程式などない。

質的研究の課題

- 質的研究は、特定の社会状況、出来事、役割、集団、相互作用についての理解を目的としており (Locke et al., 2013)、基本的に**解釈的な手法**である。
- **個人である自己と研究者としての自己**が不可分なため、ウェーバーのいう「価値理念」に沿って厳密に質的データを収集、分析しなければ客観的な研究とはなり得ない。

※価値理念とは、「知るに値する事実を選択する究極の基準」(富永ら, 1998, p. 208-209)を指す。

ミックス法の有用性

- 量的および質的研究双方の限界を補完するミックス法も万能ではないが、**発見事実の精度**を高めたり、**新しい研究領域**を発展させたりするには有効な手法と考えられる(寺崎, 2016)。

- 1989年には**Field Methods** (SAGE publications) という、ミックス法を用いた研究やその方法自体を説明した論文を掲載する学術雑誌が発刊され、現在も多くの研究論文が掲載されている。

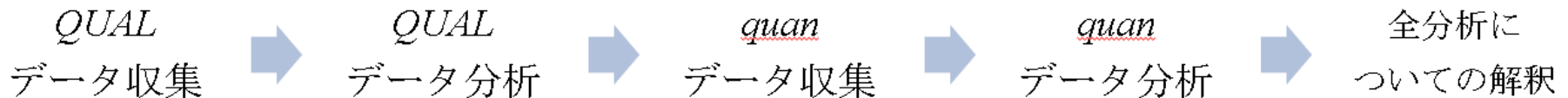


ミックス法のリサーチデザイン

- **順次的戦略** (sequential strategy) と並行的戦略 (concurrent strategy) の二つに大別でき、小分類として戦略別に三つのアプローチが存在している (クレスウェル, 2007)。
- 本発表では、テキストマイニングへの応用が期待できる、**順次的探索的戦略** (sequential exploratory strategy) を中心に説明する。

順次的探索的戦略

- 質的データの収集と分析の後、量的データのそれ
が実施される。
- まだあまり研究が進んでおらず、量的研究で用い
られる測定尺度が確立していないときなどに有効
な戦略である。



出典) クレスウェル (2007, p. 239)

順次的探索的戦略の適用例

- 消費者エスノセントリズムの測定尺度である **CETSCALE** (consumers' ethnocentric tendencies scale) を開発した Shimp & Sharma (1987) の研究。
- Shimp & Sharma (1987) は、消費者へのインタビュー調査、専門家による質問項目の吟味を経て (**質的段階**)、量的データの収集と分析によって当該尺度を開発する (**量的段階**) という、順次的探索的戦略を採用している。

順次的探索的戦略の応用

MA-CAハイブリッド・アプローチ (M-CHA)

- テキストマイニングおよびその形態論分析 (morphology analysis) によって導出された属性と水準にもとづき、コンジョイント分析をおこなう手法である (Yoon & Park, 2007; Yoon, 2008)。→ Morphology Analysis-Conjoint Analysis Hybrid Approach の略語。
- Lee & Bradrow (2007) も同時期に同様の手法を提案しており、今後にわかに注目を浴びていくものと考えられる。

コンジョイント分析とは

- コンジョイント分析は、**ニーズ調査**をする際に用いられ、複数の要因間の重要度の判別をおこなうことができる。
 - 人間が様々な選択行動をするとき、**複数の条件を比較検討**するという特徴がある。
 - 質問項目ごとに重要度を測定する場合、回答者はすべての項目に高い得点を付けることが多い。
 - コンジョイント分析を用いることで、具体的に条件提示をしながらニーズの把握ができる。
- 従来、新製品開発などに使われてきたが、教育や心理、福祉分野におけるニーズ調査にも応用可能と考えられる(真城知己, 2001)。

コンジョイント分析の例

Suh & McAvoy (2005) の研究

欧州、米国、日本からの観光客を対象に、ソウル旅行で重視する属性をコンジョイント分析によって明らかにした。

表 Suh & McAvoy (2005) で用いられた属性と水準

属性(要因)	水準
現地文化を体験する機会	多い
	最低限
食べ物	より伝統的な現地の食べ物
	あまり伝統的でない現地の食べ物
ショッピングの機会	多い
	最低限
宿泊地の立地	より伝統的な市場や観光地に近いホテル
	より現代的なショッピングセンターや観光地に近いホテル

出典) Suh & McAvoy (2005) をもとに寺崎が翻訳

参考)コンジョイント分析に必要なもの

- SPSS conjointを用いるのが一般的
- Rでも可能だが、コンジョイントカード作成の際に属性と水準のランダムイゼーションのみ調査会社に委託すると時間が節約できる
- なお、属性数は3以上、6以下が望ましいとされている(Smith, 1995)。水準数は2または3程度が一般的である。

M-CHAにおける一次データの活用

- 二次データではなく、一次データを用いる際は、データ収集に工夫が必要と考えられる。
- 一次データは、インタビューなどを通して収集し、テープ起こししたものを用いる。
 - 被験者間の相互作用により、データがリッチになると考えられるため、**グループインタビュー**が望ましい。
 - インタビューデータは親和図法+連関図法(全体構造の把握)、系統図法、GTAなどで分析される場合が多いが、テキストマイニングを活用することで、定量的に質的データを分析できる利点がある。

インタビューデータをテキストマイニングする場合、次のようなリサーチデザインが考えられる。

1. **半構造化インタビュー**へ向けて、予め質問項目やインタビューのフローを作成しておく。

→インタビューデータが散逸である場合、テキストマイニングが困難なため。

2. 任意の**尺度**を用いて、平均値 $\pm 1SD$ を目安にサンプル層別したのち、スコアがH群とL群に分けてグループインタビューを実施する。

→尺度の活用により、各群に特徴的な頻出語を精度高く抽出できる。

参考)よりデータをリッチにするために

知識と外部検索努力の間には**逆U字型**の関係がみられる (see Solomon et al., 2013) ことを念頭に、知識が中程度の被験者をインタビューに呼ぶ。

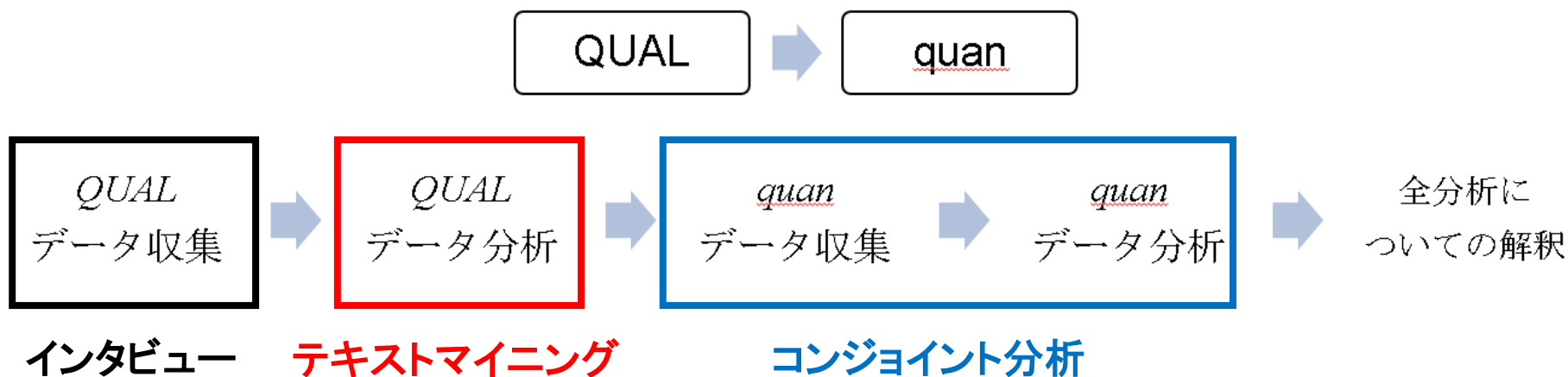
- 情報検索量がもっとも多いのは、知識が中程度の被験者となる。
- 何らかの尺度を用いて、知識が中程度の被験者をインタビューに呼ぶことを検討したい。

順次的探索的戦略におけるM-CHA概略図

M-CHA成功のカギは、インタビューにおいて頻出語を適切に抽出することであると考えられる。

ゆえに、インタビュー被験者の適切な抽出、インタビューガイドの精度が問われるであろう。

図 順次的探索的戦略におけるM-CHAの概略



出典) クレスウェル(2007)をもとに、著者作成

結語

- 本発表では、ミックス法の概要、そしてテキストマイニングに有用と考えられる、順次的探索的戦略について説明した。
- 次に、具体的なアプローチとして、M-CHAを紹介し、その適用へ向けたポイントを複数提案することで、応用研究への橋渡しをおこなった。
- 異分野融合をキーコンセプトとする大学学部(研究科)、研究機関などで活用が期待できるであろう。

参考文献

ウェーバー, M. (1998).『社会科学と社会政策にかかわる認識の「客観性」』, (富永祐治, 立野保男訳, 折原浩補訳). 岩波文庫. [原著: Weber, M. (1994). *Die "Objectivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis*].

クレスウェル, J. W. (2007).『研究デザインー質的・量的・そしてミックス法』, (操華子, 森岡崇訳). 日本看護協会出版会. [原著: Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches 2nd edition*, Sage Publications].

寺崎新一郎 (2016)「経営学におけるミックス法」九州大学ビジネス・スクール編『新たな事業価値の創造』QBS出版。

仲正昌樹 (2014).『マックス・ウェーバーを読む』. 講談社現代新書.

真城知己 (2001)『SPSSによるコンジョイント分析ー教育・心理・福祉分野での活用法』東京図書.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, 56(2), 81-105.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 602-611.

Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research*, 48(5), 881-894.

Locke, L. F., Spirduso, W. W., & Silverman, S. J. (2013). *Proposals That Work: A Guide for Planning Dissertations and Grant Proposals: A Guide for Planning Dissertations and Grant Proposals*. Sage Publications.

Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: Construction and validation of the CETSCALE. *Journal of Marketing Research*, 24(3), 280-289.

Smith, S. (1995). *Tourism Analysis: A handbook* (2nd ed.). Harlow: Longman.

Solomon, M., Russell-Bennett, R., & Previte, J. (2013). *Consumer Behavior* (10th ed.). Pearson Higher Education AU.

Suh, Y. K., & McAvoy, L. (2005). Preferences and trip expenditures—a conjoint analysis of visitors to Seoul, Korea. *Tourism Management*, 26, 325-333.

Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *Engineering Management, IEEE Transactions on Engineering Management*, 54(3), 588-599.

Yoon, B. (2008). On the development of a technology intelligence tool for identifying technology opportunity. *Expert Systems with Applications*, 35(1), 124-135.