

非定型自由記述法と定型自由記述法の比較

真柳麻誉美* 林 俊克** 平野広隆***

*女子栄養大学 食品学第一研究室

** (株)資生堂 CS 開発センター情報開発室 *** (株)アーキテクト

1. はじめに

自由記述型のテキストデータは、従来から各種調査での自由回答で収集されてきたが、適切な処理手順が確立、普及されていないことから、定量的な解析を行われることが少なかった。

特にインターネット環境の解放による蓄積型の自由記述データの増加も手伝って、自由記述データの客観的な集約記述方法として、定量的な解析への関心が急速に高まっている。

産業分野では、こういった「今まで捨てていたデータ」「自然に蓄積されたデータ」から何かが得られれば良いという発想だけではなく、より前向きな理由で自由記述型データの解析に大きな期待が寄せられている側面もある。

つまり、回答者の「自由な記述」を解析できれば、従来見落としていた要素を抽出、集約でき、無駄な情報を捨てることもない。お仕着せの無いより客観性の高い情報の分析。ここに新たな地平を見出そうと考えていると言って良い。積極的に自由記述型のデータを収集、解析しようという動きもまた活発化している。

「テキストマイニング」という語は盛んに使われてはいるが、指し示す範囲は極めて広い。一口に「自由記述型データと解析」言っても、データの内容だけでなく、形式、目的、解析法など様々である。

本報では、自由記述型データの「採取(調査)法」と「解析法」を整理、具体的な調査事例に基づいてその一部を比較検討する。また、筆者

らの目指す次世代テキストマイニングの形を論じる。

2. 自由記述データの分類

ここで議論する自由記述データは、下記の2つのタイプとする。

1. 非定型自由記述
2. 定型自由記述

1. 非定型自由記述は、内容以外の回答形式を回答者本人にゆだねる型の自由記述とし、一般に行われている「～についてご自由にお書き下さい」といった形式で採取されるアンケート項目の回答他、インタビュー記録や会話のログ等である。多くの蓄積型テキストデータがこれに当たる。

2. 定型自由記述は、内容の他、回答形式を指定する型の自由記述であり、回答者は形式に沿って該当する単語や短文を回答する。「カッコ()内に当てはまる言葉をお書き下さい」といった形式で採取された物がこれにあたる。具体的には「定義法」「文章完成法」「連想法」等がこの型である。蓄積型のデータではこの型のデータが少ない。

本報では、上記分類に沿って、特にアンケート形式の採取データを取り上げ、比較検討する。

3. 方法

1. 非定型自由記述、と、2. 「定義法」と「文章完成法」「連想法」による定型自由記述、さらに事前に定性調査を行って設問項目を吟味した定型選択にあたる7段階評点尺度による

3. SD法, の3種の形式で調査票を作成し, 女子栄養大学1年生544名を対象に調査を計画. 被験者に3種を無作為割付して調査を行った.

調査テーマは, 「女子大生における魅力的牛乳像の解明」であり, 具体的な調査内容は「普段飲んでいる牛乳」と「理想の牛乳」について.

理想の牛乳への回答で被験者の言葉によるニーズの探索を行い, さらに, 「普段飲んでいる牛乳」と「理想の牛乳」との乖離から, 各ニーズの充足度(一種の満足度)を探り, 付加すべき牛乳の魅力の解明するのが目的である.

調査は2001年5月25日および29日の両日を使用し, 学内教室にて実施した(集合調査法). 回答不備の著しいものは再配布し, 記入後に再回収した(実施日より2週間以内). 調査計画および回収結果は, 表1の通り.

表1 調査計画と回収結果

	計画数	回収数		
		回収	使用 不備	
1.非定型自由記述	182	回収	使用	181
			不備	0
		未回収		0
		未配布		1
2.定型自由記述	181	回収	使用	171
			不備	2
		未回収		2
		未配布		6
3.SD法	181	回収	使用	176
			不備	1
		未回収		2
		未配布		2
合計	544			544

*1不備は回収したが未回答により使用不能のもの

*2未回収は調査票を配布したが, 提出のなかったもの(再配布後の未提出含む).

*3未配布は欠席等により調査票自体を配布していないもの

4. 結果と考察

ここでは, 1.非定型自由記述と, 2.定型自由記述の比較を中心に行う.

まず, 実施における両手法の差異は, 特に回答に要する時間に現れる. 非定型自由記述の平均回答時間は, 約8.5分. 定型自由記述の平均

回答時間は約11.5分であり, 定型自由記述の負荷の方が高いことが伺われた.

非定型自由記述では, 事実や感想の列挙に留まる回答も多く, 事実と感想(結果)をつないだ因果を持つ表現が少ない. そのため, ニーズの抽出から具体的な手段を講じにくい傾向があった. しかし, 文章表現の他, 絵による多彩な表現も散見され, 被験者によっては自由な意見, 発想が盛り込まれやすいことがわかったが, そういった情報の処理は困難でもあった.

定型自由記述では, 回答に慣れを必要とし, 被験者の負荷も高く, 記述も少なくなる傾向があるが, 聞きたいことのみ抽出できるため効率的で, データの処理も簡便であった.

データの処理手順, 解析方法, 具体的に解析した結果の比較等の詳細は, 発表にて行なう.

5. まとめ

本報告では, 非定型自由記述と定型自由記述の調査比較を行い, それぞれの特徴が整理された. 定型自由記述データは非定型自由記述に比べ, 効率良く情報の収集が出来, データ処理の負荷が低いという長所があった. その反面, 関与が低い場合や不慣れな場合には, 被験者の回答負荷がやや高いことが判明した.

その他, 非定型自由記述型および定型自由記述型データの処理手順を整理し, 特に解析方法について検討の余地があることが示した. またこれにより, データの処理, 解析支援のために必要なツール, 機能についてが整理された.

参考文献

- 1) 大隈昇, Ludovic, L.: 調査における自由回答データの解析 - InfoMiner による探索的テキスト型データ解析 -, 統計数理, 48, 2, 339-376, (2000)