

## 評価・印象のテキスト・マイニング

文教大学大学院 情報学研究科 教授 岡野雅雄<sup>†</sup>

Masao Okano

### あらまし

メッセージについての評価・印象のテキスト・マイニングを行うためには、テキストの意味のカテゴリ化の研究の深化が求められている。その基礎理論となるのは、感情表現の心理学や、評価・印象の言語表現の研究である。

キーワード：テキスト・マイニング、感情表現、評価、印象

### 1. はじめに

情報発信をする際のメッセージを改善し、より効果的なものにするためには、受け手がメッセージに対していただいた評価や印象を知ることが必要である。また、現在はブログ・掲示板などの形でインターネット上に大量のテキストデータが存在する。これらの中にはさまざまな対象についての評価や印象が書き込まれている。そのために、さまざまな方法で評価・印象の測定方法が研究されている。ここでは、受け手の評価・印象をとらえるために評定尺度法を用いる方法について簡単に触れ、次いでテキスト・マイニングを用いる方法について述べたい。

### 2. 評価・印象をとらえるために評定尺度法を用いる方法

評定尺度を用いるのは心理学において発達してきた方法で、その代表的な例がSD法(Osgood,1957)である。SD法では、たとえば「明るい-暗い」、「温かい-冷たい」などの項目を20から25個程度並べて、測定したい対象の印象を評定してもらう。

ここで重要となるのは、どのような項目を用いるかである。実験を積み重ねて、妥当性や信頼性のある測定項目を構築してゆくことが必要となる。具体的には、先行研究で用いられている測定項目をすべて洗い出したものに妥当性の検証を行い、最適な測定尺度を構成する、という手続きを踏んで初めて、研究対象の測定を行うことができる。かつて筆者ら(岡野・浅川,2005a)は、テレビCMの測定尺度について、このような試みを行っているので、詳細は拙稿を参照されたい

ある程度、研究の蓄積のある分野ではSD法には利点が多いが、そうでない場合、つまりほとんど未開拓な領域の分野では事情が異なってくる。このような場合に有効と考えられるのが、テキスト・マイニングである。テキスト・

マイニングの有用性について、鈴木(2001)は、定性的な効果把握ができる、フレーミング効果がない、発想支援をサポートするなどを挙げている。この中で指摘されているように、テキスト・マイニングでは、あらかじめ評価項目を設定することなく、被調査者の生の声からボトムアップ的に結果を得ることが期待される。

### 3. 評価・印象をとらえるためのテキスト・マイニングとその課題

日本語の文章に対するテキスト・マイニングの最初の手順は、データとなる文の集合を形態素に分けることと、統語構造を解析し、係り受け構造を取り出すことにある。現在の技術水準では、この2点においては、多くの分析目的にとって十分に実用性高い精度の情報が得られる。

ただし、テキスト・マイニングを行う際に注意しなくてはならないのは、「ジップの法則」である。「ジップの法則」とは、言語学において発見された語彙分布の法則であり、テキストに含まれる語の出現頻度は、出現頻度順で上位の少数の語において高く、順位が低くなるにつれ急激に下がってゆくというものである。したがって、テキストマイニングの際に、出現頻度順に上位の語だけで切ってしまうと、多くの下位語の情報の取りこぼしが生じる可能性がある。

豊田(2005,p.28)は、「ジップの法則」へ対処し、テキストデータからより有効な知見を引き出すための方策として、データ分量の確保、データ範囲の限定と取り扱う内容の具体化、周辺情報の活用、の3点を指摘している。は、データ量を増やすことで、低頻度語のいわば底上げを行うことで取りこぼしを少なくするということである。は、回答となりうる文章の範囲にある程度の限定を加えることで、テキスト・マイニングの結果が広範囲に散ることを防ぐということである。たとえば、制約を加えないで回答してもらう場合と、ある程度の制約を加えた場合の違い

2009年3月31日受付

<sup>†</sup> 〒253-8550 神奈川県茅ヶ崎市行谷1100 okano@shonan.bunkyo.ac.jp  
Graduate School of Information and Communication, Bunkyo University

については、岡野・浅川(2005b)で検討を加えたが、制約付きの場合には、端的にキーワードが抽出されやすいことが認められた。は、テキスト以外のデータ、たとえば回答者の性別・年齢などの諸属性を同時に利用することである。

また、これ以外の重要な点として、類義語の処理がある。テキストには類義語ないし同義語と見なせるものが生じるのが常である。テキストに現れる低頻度語は、類義語としてまとめることで抽出できるものがある。また、比較的高頻度の語の場合でさえも、同一概念とみなせる語と統合すると順位がさらに上がる場合がしばしばある。

このように、類似した意味の語を括ることは、単に形式的に単語を抽出することを超えて、テキスト・マイニングで本来目指している「コンセプト」を抽出するために、必要な操作である。

だが、この段階において難題が生じる。言語学においては音素や形態素などの形式的な分析がまず発達し、意味の分析は困難を極め、発達が遅れたが、テキスト・マイニングにおいても同様に、意味を扱うことは先端的な課題となっている。現在、概念辞書を用いる方法や、単語の統計的な類似度を用いる方法などが用いられているが、形態素解析までの完成度と比べると、まだ試行的な段階と言える。この意味処理、具体的には類義語の設定・カテゴリー化をいかに行うかが、現在の重要な研究課題であると考えられる。

浅川・岡野(2009)では、その試みのひとつとして、タレントC に登場するタレントについての自由記述に表れる要素をテキスト・マイニングし、タレント自身の特性に対する評価、CMの表現・制作的要素に対する評価、およびタレント自身の特性と表現・制作的要素の適合性に関する内容の3タイプに分けたうえで、各回答がどのような視点に基づいているのかをさらに整理して「タレントを好きな理由」の仮説的分類体系をカテゴリー化している。

#### 4. 感情表現の研究の重要性

カテゴリー化は、単語の言語的な定義から一義的に決められるものではなく、調査目的によって必然的に変わるものである。そして、評価・印象のカテゴリーをいかに設定するかに関しては、感情に関する心理学や言語学が基礎的な理論を提供している。

テレビCMの評価・印象の研究に関してみると、これらの研究は、(i)心理学における基本情動説を応用した評価分析と、(ii)CM表現の評価・診断を目的とした評価の多次元的分析に分けられる(岡野・浅川,2005)。(i)は心理学における基本情動説を広告を視聴した場合の反応に応用した評価分析であり、広告反応を少数の基本情動に還元して把握しようとするものである。(ii)のCM表現の評価・診断を目的とした評価の多次元的分析は、米国の研究では、多くの測定項目を縮減して測定尺度を作成する形式がとられているものが多い。これらは、テキスト・マイニングで、評価・印象のカテゴリーを設定する際に有用である。

また、近年、感情表現の言語学的な研究が進んできており、これらの研究は、テキスト・マイニングの際にカテゴリー化を行うための立脚点として注目される。Kövecses & Palmer(1999)は、感情の言語学的研究を体系的に整理し、情緒の意味へのアプローチとして、ラベルによる方法、中核的意味(core meaning)による方法、次元による

(dimensional)方法、含意による(implicational)方法 プロトタイプ(prototype)による方法、社会的構成主義という、6種類を挙げている。また、基本情緒語彙(basic emotion terms)や、メタファー・メトニミーによる感情表現について触れており、これらは、評価・印象データをテキストから抽出する際に不可欠な理論となると思われる。

#### 5. むすび

先に引用した鈴木(2001)が「発想支援をサポートする」というメリットを指摘しているように、テキスト・マイニングは仮説検証的というよりは、問題発見的なものとして位置づけられる。いったん仮説が得られた後は、さまざまな条件のもとで繰り返し実施することが重要である。今後 Web上に大量に存在する自由記述の言語データを活用することを考えるとき(たとえば、ブログ等にかかれたタレントについてのコメントなど)、テキスト・マイニングを用いる手法は、さらに実証を積み重ねつつ開拓されてゆく必要がある。

#### 【文献】

- 浅川雅美・岡野雅雄.(2009). テレビCMに登場するタレントに対する態度を決定する要因の分析：自由記述のテキスト・マイニング. 『広告科学』, 50(1).
- 岡野雅雄・浅川雅美.(2005a). テレビCM 評価・印象の測定尺度の検討. 『情報研究』, 32,
- 岡野雅雄・浅川雅美.(2005b). テレビCMの視聴反応についてのテキストマイニング---制約なしと制約付きの自由記述の分析. 『生活科学』, 27
- 豊田裕貴.(2005). テキストマイニング入門. (上田隆穂, 戸谷圭子, 黒岩祥太, 豊田裕貴 編, 『テキストマイニングによるマーケティング調査』: 講談社 所収)
- 鈴木宏衛.(2001). 自然言語解析を用いた効果の把握.(仁科貞文編, 『広告効果論』: 電通. 所収)
- Kövecses, Z., & Palmer, G. B. (1999). Language and Emotion Concepts: What Experientialists and Social Constructionists Have in Common. In G. B. Palmer & D. J. Ochi (Eds.), *Language of Sentiment*. John Benjamins Publishing Company.
- Osgood, C. E. (1957). *The Measurement of Meaning*: University of Illinois Press.



おかの まさお  
岡野 雅雄

1957年生。筑波大学大学院博士課程  
文芸・言語研究科単位取得満期退学。  
1997年4月より文教大学情報学部  
に着任。2009年4月より大学院情報学  
研究科情報学専攻教授を兼任。心理  
言語学、記号論、文章心理学などが  
専門。本情報学研究科では「マルチ  
メディア・コンテンツ特論」を担当。