

## テキストマイニングを活用した クリティカルシンキング測定の試み

新潟医療福祉大学 健康スポーツ学科

後藤 康志

### 1 背景

情報の収集・整理を行う際に、得られた情報の真偽を批判的に捉える思考は批判的思考（クリティカルシンキング、以下CT）と呼ばれ、メディア・リテラシー育成にあたっては中心的な概念である。CTの測定についてはWatson-Graserテストなど様々な尺度が提案されており、選択肢によるものと自由記述によるものがある。自由記述は開かれた回答であることから膨大な情報を含んでいるが、その分析の難しさと煩瑣さが指摘されている。筆者はWeb情報における批判的思考に焦点化して自由記述のカテゴリー分析を用いたが（後藤2007）、データの持つ情報をすべて生かしたわけではない。

自由記述の分析で注目されているのがテキストマイニングである。テキストマイニングは、自由記述を最小単位である形態素にわけて対応分析を行うものであり、言葉同士の共通性や類似性を検討する手法である（藤井ら2005）。テキストマイニングを用いることで、対象者の特性に基づくCT測定が可能になると思われるが、そのような研究が蓄積されているとはいえない。

### 2 目的

CTの測定においてテキストマイニングを行うことにより、カテゴリー分析では明瞭にならなかった構成要素間の関係およびサンプルの特徴が検出できるかを検討する。

### 3 方法

- (1) 対象：対象は現職教員、大学生など53名である。
- (2) 自由記述データ：後藤(2007)による自由記述の問題を課した。あるダイエット食品のホームページに医学者が推薦文と、成功した3人の体験談が掲載されている。この情報の信頼性を確認するために必要な情報を、その情報を得るための方法を自由記述させるものである。
- (3) 分析：テキストマイニング用ソフトウェアとして日本電子計算製のWordMiner Ver1.100を用いた。文を形態素に分割する分かち書きとキーワード抽出を行ったのち、データのリファイン（同種の語を一つに統一する置き換え）を行った。「食品会社」「会社」「販売会社」など別の語で書かれていても文意から同じと思われる場合は「会社」に統一するなどである。得られた構成要素のうち頻度2以上のもの（閾値=2以上）を対象に対応分析を行った。抽出した15成分の累積寄与率は42.82%であり、それほど高くない。この点は課題で述べる。

### 4 結果と考察

#### (1) 構成要素のクラスタ化

従来、同一カテゴリーで扱ってきた内容が別の構成要素ク

ラスタとして見なせる可能性が示された。構成要素クラスタ1のメンバーシップは「OPAC、業績、勤め先、勤務、効用、実数、証拠、場所、精通、製薬会社、専門分野、存在、大学、電話、病院名、有無、領域」などである。データの原文では、推薦文を載せた医学者がその領域の専門家であるのかについて、専門分野や業績を勤務先に問い合わせたり、OPAC(Online Public Access Catalog)で確認したりする、という記述がある。成分スコアの2次元布置をみると、直近に構成要素クラスタ4がありそのメンバーシップは「医学書、医学的根拠、栄養、確かな、公表、身分、著作者、著者名、名」などであり、同様に医学者の信頼性の確認に関する項目である。カテゴリー分析ではこの両者は同じに扱ってきたが、別の意味がある可能性がある。この他にも構成要素クラスタ16「会社名、概要、企業年鑑、就職情報誌」のように企業情報の内容に特化したものも分離して布置されるなどカテゴリー分析では分離できなかったCTの構造が見えてきた。

#### (2) サンプルのクラスタ化

カテゴリー分析では同一カテゴリーに分類されるサンプルが、記述内容から見ると分類できる可能性がある。抽出したサンプルのクラスタはサンプル7クラスタであり、もっとも多いサンプルクラスタが44名、次に大きいサンプルクラスタが4名、あとは1名のクラスタが5である。別のクラスタに分けられる場合は、例えば前述の構成要素クラスタ16（企業に関する情報）に関する記述など他のサンプルが記述しない内容を記載した場合、分けられるようである。この場合、平面布置では類似した構成要素クラスタとサンプルクラスタが同一平面上の近い位置に布置され、データの全体的な特徴を把握することができる。

### 5 まとめと今後の課題

テキストマイニングを用いることでカテゴリー分析では明瞭にならなかった構成要素間の関係とサンプルの特徴が検出できる可能性が示唆された。

課題として2点を挙げる。第1に分析の範囲を絞って行うことである。今回扱ったデータでは情報の信頼性を確認するために必要な情報とその情報を得るための方法を同時に使った。これは同様のテキストマイニングを用いた分析（例えば藤井ら2005）に比べて広範囲すぎた可能性がある。比較的限定される方法に絞るなどして検討が必要である。第2にデータのリファインである。必要にして十分で、再現可能なデータのリファインのルールを決める必要がある。

### 6 参考文献

- 藤井美和・小杉孝司・李政元(2005) 福祉・心理・看護のテキストマイニング入門. 中央法規, 東京  
 後藤康志(2006) 学習者のWeb情報に対する「批判的な見方」の発達. 日本教育工学雑誌, 第30巻3号

謝辞：本研究の一部は、平成19年度新潟医療福祉大学研究助成金発展研究「テキストマイニングの自由記述解析によるクリティカルシンキング関連要因の探索的研究」により行われました。深く感謝申し上げます。