



読者の皆さまは「テキストマイニング」という言葉をご存知ですか？ マイニングとは「探掘」。よって文章の中に埋まっている宝を掘り出すことと言ったらいでしょうか。近年、食事処を決めるのにも、観光名所を探すのにも、電化製品を買うのにも、ネット上に書き込まれているレビューやクチコミを参考にされる方も多いでしょう。これらの信ぴょう性については今回のテーマではないのでひとまず横に置いて、善意で書き込まれたテキストには「お宝」が眠っているかもしれ

テキストマイニングのすすめ

テキストマイニングです。それでは一例をお見せします。研究室の学生が丁寧なデータを収集して分析してくれました。図表をご覧ください。これは2007年から2018年までのAKB6グループと坂道3グループ（ご存じない方は検索してください）のシングル楽曲すべての歌詞を分析ソフトに入力し、そこから使用頻度の高い単語を抽出し、そして発表された年と一緒に可視化したもの（対応分析の結果）です。その手順の詳細は省きますが、同じ曲中で用いられた単語ほど近くに配置されています。たとえば図の上部には「未来」と「LOVE」という語があり、近くに2009年と2012年が配置されているので、この二つの語はその年に発表された

た上部には「海」、「空」、「風」、「声」、「好き」といった名詞が並んでいて、初期のヒット曲「ポニーテールとシュシュ」や「ヘビローテーション」を思い出させます。一方下部には「気づく」、「愛す」、「変わる」、「知る」などグループのメンバーの思いが強くあふれてくるような歌詞が多用されています。

われわれは秋元氏の歌詞は確かに変化しているのだが、急激なイノベーションを与えているのではなく、ゆっくりとシフトしているのでは、と読み取りました。シングル曲それぞれを聞いてただけではその語感の流れに気付くことはなかったと思います。

上記の例は歌詞というテキストを用いましたが、研究室ではこの他にも名古屋の観光地の日英中国際クチコミ比較ですとか、日本でヒットした邦画評・洋画評比較等にテキストマイニングを適用して興味深い発見をしています。今回は文章の最小単位である単語を用いました。他にも語が持つ感情を掘り起こしたり、文節単位で意味を読み取ったりすることもできます。新聞記事や政治家の言葉から為替や株価の動向を予測するといった試みも行われています。

隠されたメッセージを

読み解く

ません。そのテキストを整理・分析し、結果を可視化して、法則性や隠れたメッセージを見つけ出すことが

智 茨木

曲に「よく使われていた」と判断します。さて図中の年データの位置をよく見ると上部から下部へ年代が移っていることにお気づきでしょうか。ま



名古屋市立大学大学院 経済学研究科准教授

いばらき・やまぐち オペレー ションズ・リサーチ、数理計画。京都大学大学院工学研究科博士 後期課程修了。工学博士。1965年生まれ。

