

語彙の極性分析を利用した自由回答からの重要項目の抽出と特徴分析

Extraction of Important Items and Feature Analysis using Semantic Orientations of Vocabulary from Free Answer

塩飽 朝美^{*1}, 椎名 広光^{*2}, 柳 貴久男^{*2}, 小林 伸行^{*3}
Asami SHIWAKU^{*1}, Hiromitsu SHIINA^{*2}, Kikuo YANAGI^{*2}, Nobuyuki KOBAYASHI^{*3}

^{*1} 岡山理科大学大学院総合情報研究科

^{*1} Graduate School of Informatics, Okayama University of Science

^{*2} 岡山理科大学

^{*2} Okayama University of Science

^{*2} 山陽学園大学

^{*3} Sanyo Gakuen University

Email: i15im01sa@ous.jp

あらまし：大学における講義の学生からの評価として，アンケートを実施されることがほとんどである．アンケートのうち自由回答方式での分析を，単語のポジティブ・ネガティブの極性評価と統計処理技術を利用することによって，学生が重要視している項目抽出や講義毎の特徴分析を行う．

キーワード：極性評価，自由記述解析，アンケート分析，講義評価

1. はじめに

現在，大学においてFD活動が義務付けられており，授業改善の一環として学生アンケートが実施されている．アンケートの方法は多様で，記名式無記名式，項目選択式，自由記述式など，それらを組み合わせたものが実施されている．アンケート分析自体は，さまざまな方法が行われており，自由記述を解析した事例も行われている^(1,2)．

本研究では，岡山理科大学総合情報学部情報科学科で実施された中間アンケートの自由回答の記述を対象として，その分析を単語のポジティブ・ネガティブの極性評価⁽³⁾と統計処理技術の主成分分析やデンドログラムによるクラスタリングから学生が重要視している項目抽出や教員ごとの特徴分析をについて述べる．

2. アンケートデータ

岡山理科大学総合情報学部情報科学科の 2014 年度春学期(4月～9月)の中間段階(15回中8回目の時期)に講義アンケートで実施された自由回答項目を利用した．調査した教員は16人中15人，講義の科目数は，40科目，アンケート回答数は1642個である．

3. 自由回答欄の解析手法

自由回答欄の解析手法として，①自由回答欄の記述(以下，コメント)のポジティブ・ネガティブを人手で評価しそれぞれを分離してから解析する場合と評価しないでまとめて解析する場合，②各単語に極性を付した場合と付さない場合に分けて解析を行っている．自由回答欄の解析の流れを，以下に示す．

(1) コメントを形態素⁽⁴⁾に分解し，人手によるコメント評価をポジティブ・ネガティブに分類している場

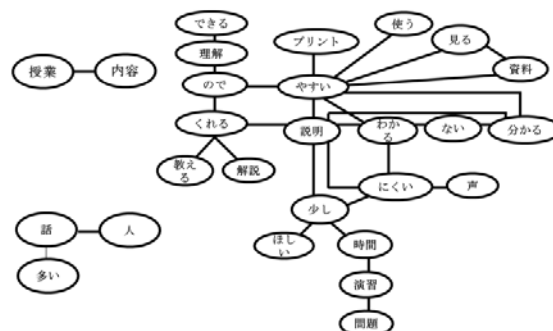


図 1 係り受けネットワーク(コメント評価無)

合は，ポジティブとネガティブに分けて単語集合を作る．語彙の極性値を利用して解析する場合は，コメントごとに単語の極性値の平均を求め，正であれば，コメントをポジティブ，負であればコメントをネガティブとして分類し，またそれぞれの単語集合を作る．コメントを評価しない場合は，1つの単語集合を作る．

(2) 頻度の多い6単語を単語集合から取り出し，その単語に係る単語を節とし，節間を辺でつなぎ，係り受けネットワークを作成する．コメントを評価しない場合の係り受けネットワークの例を図1に示す．

(3) 係り受けネットワークから出現回数の多い単語と係られている数が多い単語をアンケートの特徴語として取り出す．

(4) 特徴語が，教員ごとの自由回答に含まれているかの主成分分析を行う．教員の主成分は，各教員の特徴語の出現率から求める．

(5) 特徴語の主成分と教員の主成分の散布図を重ねて比較する．散布図は，特徴語と教員の主成分間の因子得点の組を座標とした散布図を作成している．散布図の黒色の数字は各教員に固有番号を付したもので，赤色の矢印は特徴語の各主成分による因子負荷量を座標に置き換えベクトルとして表現したもの

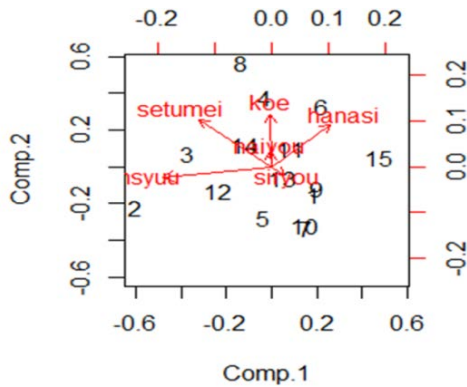


図 2 単語と教員の主成分分析

表 1 評価方法ごとの特徴語

コメント評価無	コメント評価有	極性評価
易い, 早い, 多い, 難しい, 良い, にくい, ほしい, できる, わかる, 教える, 見る, 使う, 分かる, 少し, プリント, 演習, 解説, 講義, 授業, 説明, 板書, 復習, 理解, 話, 問題, 課題, 黒板, 資料, 字, 声, 先生, 内容, 時間	易い(P), できる(P), 教える(P), 見る(P), 使う(P), 分かる(P), プリント(P), 演習(P), 授業(P), 説明(P), 復習(P), 理解(P), 話(P), 問題(P), 課題(P), 資料(P), 内容(P), にくい(N), ほしい(N), 分かる(N), 少し(N), 授業(N), 説明(N)	授業, プリント, 演習, 問題, 課題, 理解, 資料, 板書, 教え, 声, 講義, 良い, 解説, 黒板, 復習, 先生, 字

である。ベクトルの長さは因子負荷量の大きい特徴語, 数字(教員固有番号)と特徴語の位置関係は教員が反応している特徴語を図ることができる。図 2 は, コメントデータを第 1 主成分と第 2 主成分の分析例である。

(6)コメント全体で出現回数が 35 以上の単語を全教員の特徴語として抽出する。また, 各教員のコメントから出現回数が 5 以上かつその単語が教員全体で出現回数 1 であれば教員の特徴語として抽出する。

4. 調査結果

単語と教員の主成分分析(図 2)から, 2 番の教員は「演習」, 8 番の教員は「声」, 15 番の教員は「話」に関して特に評価を受けており, 学生は授業方法に関して「演習」, 教員に関して「声」と「話」に関して重要視していると考えられる。

また, コメント評価の有無, 極性評価による抽出される特徴語を, コメント全体と教員ごとに分けて表 1 と表 2 に示す。予めコメント評価した場合とそうでない場合では教員とその評価に関して結果に大きな差は見られなかった。また, 教員ごとの特徴として「モニター」「発表」といった授業方法に関する評価がされていることから各教員には授業における工夫が求められている。語彙の極性値を利用し, 極性値が 0 以上の場合をポジティブ単語と評価した際「楽しい」「明るい」といった授業の印象に関する単語が多く得られたが, 全極性値の平均を基準とした

表 2 教員ごとの特徴語

教員	特徴語		
	コメント評価無	コメント評価有	極性評価
1	質問	質問(P), 難しい(N)	難しい, 質問
2	確認	確認(P)	確認
3	せる, 少ない, 欲しい, きちんと, よく, 解答, 生徒, 答え	きちんと(P), 少ない(N), 欲しい(N), 教える(N), よく(N), 演習(N), 理解(N), 答え(N)	欲しい, きちんと, 生徒
4	づらい, 進む, 自慢, 宿題	良い(N), づらい(N), 自慢(N), 宿題(N), スクリーン(N), 先生(N)	自慢, 宿題
5	解く, モニター	多い(P), 解く(P), モニター(P)	解く, モニター
6	つく, 言う, 話す, 発表, 社会, 力	つく(P), 社会(P), 力(P), 1(P), 話す(N)	発表, 社会, 力
7	(なし)	(なし)	(なし)
8	何	何(N)	(なし)
9	について, デザイン	について(P), デザイン(P)	デザイン
10	詳しい, プロジェクター, 最後, 図, とき	詳しい(P), 見る(P), スクリーン(P), スライド(P), プロジェクター(P), 最後(P), 図(P), 資料(N), 時(N)	スライド, 最後, 図
11	とくに, 一つ	良い(P), 一つ(P)	(なし)
12	大きい, 追加, 練習	大きい(P), 追加(P), 練習(P), 解説(N), 課題(N)	教える, 追加, 練習
13	小さい, すぎる, もう少し, ゆっくり, 教え, 進行, 配付, 学生, 文字, 丁寧	ゆっくり(P), 配付(P), 先生(P), 丁寧(P), 小さい(N), すぎる(N), もう少し(N), スピード(N), 黒板(N)	早い, ゆっくり, 進行, 配付, 学生, 教え, 文字, 丁寧, 所
14	楽しい, おしえる	楽しい(P), おしえる(P)	楽しい
15	面白い, 聞ける, 教科書	面白い(P), 聞ける(P)	面白い, テスト, 教科書

*(なし)は特徴的な語が抽出されなかったことを表し, P,N はポジティブ評価を受けたコメントから抽出した単語を表す。

場合「方程式」「実習」といった授業内容に関する単語がポジティブ単語として抽出された。

5. 今後の課題

単語の極性評価では, コメント全体にネガティブな単語が多いため, コメント中の単語の極性値の平均では評価は難しい。今後は, 構文情報などを利用してより精度の高い極性評価を行う必要がある。また, ポジティブ・ネガティブは立場によって評価が相違するため, それらについて解析が必要である。

参考文献

- (1) 山西博之, "教育・研究のための自由記述アンケートデータ分析入門", 外国語教育メディア学会, メソドロジー研究部会 2010 年度報告集, pp110-124, (2010)
- (2) 岡山大学ダイバーシティ推進本部, 岡山大学の男女共同参画推進に関するアンケート調査, pp52-70, (2011),
- (3) 高橋大也, 乾孝司, 奥村学: "スピンモデルによる単語の感情極性抽出", 情報処理学会論文誌, 47(2), pp.627-637, (2006)
- (4) 工藤拓, 松本裕治: "チャンキングの段階適用による日本語係り受け解析", 情報処理学会論文誌, 43(6), pp.1834-1842, (2002)