

## アンケート自由回答文の複数人による評価を利用した教員・講義評価

## Lecture and Teacher Evaluation from Free Answers by Multiple People

塩飽 朝美<sup>\*1</sup>, 椎名 広光<sup>\*2</sup>  
Asami SHIWAKU<sup>\*1</sup>, Hiromitsu SHIINA<sup>\*2</sup>

<sup>\*1</sup>岡山理科大学大学院総合情報研究科

<sup>\*1</sup>Graduate School of Informatics, Okayama University of Science

<sup>\*2</sup>岡山理科大学

<sup>\*2</sup> Okayama University of Science

Email: i15im01sa@ous.jp

あらまし：本研究は、実際に行われた講義アンケートの自由回答文を対象に、一部の自由回答文を複数人で評価し、その評価を基にラベル無コメントの評価推定をそれぞれ行う。また、単語と自由回答文の評価を行い、その推定評価を互いに適用することで評価推定を繰り返し、全自由回答文の再評価を行う手法について述べる。そして、推定された自由回答文の評価から教員・講義評価を行った結果について述べる。

キーワード：多人数評価, 自由記述解析, 単語極性, 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。岡山理科大学総合情報学部情報科学科では, 独自にセメスターの中間で講義評価アンケートを実施しており, 自由記述のアンケートを実施している(2014年春学期, 調査教員15人, 講義の科目数40科目, アンケート回答数1642個)。

そこで本研究では, 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<sup>(1)</sup>を行い, アンケート内の単語リストから自由回答文を評価したうえで, 教員・講義の評価を行っている<sup>(2)</sup>。

自由回答文の評価手法については, アンケートの自由回答文に対する評価点数を取得していないため, 推定が必要となってきた。そこで自由回答文の一部を手で評価し, そこから単語の極性評価を推定し, すべての自由回答文の推定を行う手法を提案してきた。しかしながら, 自由回答文の評価は評価者によって相違し, それに連動して単語の極性評価もより異なるため, 複数人での自由回答文の評価が必要であると考えられる。

また, 複数人での評価の際に, 単語の極性評価が大きく割れるものがある。そこで単語の極性評価値の分散が大きいものと小さいものに分類し, 大きいものについては, 評価対象から外し上で自由回答文を評価, そして教員・講義評価を, 本研究では行っている。

## 2. 自由回答文の解析手法

自由回答文の評価は, 次の手順で行う(図1)。

### (1) コメントの人手による評価

(1-1) 教師データの作成(コメントの評価): 人手で「とても悪い」「悪い」「少し悪い」「少し良い」「良い」「とても良い」の6段階評価を行う。このコメント評価されたデータをシードとして利用する。

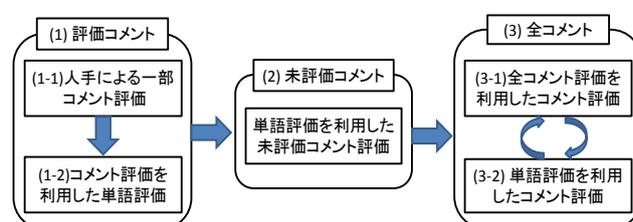


図1 コメントの評価手順

(1-2) 評価コメントを利用した単語評価推定: 評価コメントからコメントに含まれる単語の評価推定を行う。手順としては, 評価コメント中の各単語に付与された評価から単語ランク分布を作成し, 単語ランク分布から単語の評価を推定する。単語ランク分布の作成と単語の評価の推定の手順を以下に示す。

Step1: 評価コメントから名詞・動詞・形容詞の単語を抽出し, コメントの評価値をそれに含まれている単語の評価値として付与する。

Step2: すべての評価コメントとその中に現れる単語に対して Step1 を実施し, 単語ごとに付与された評価値の頻度から単語ランク分布を作成する。

Step3: 単語ランク分布から, 最大評価回数のランクを単語の評価とする。ランクの評価回数が, 同数の場合は低い評価とする。

(2) 未評価コメントの評価推定: 単語ランク分布から未評価コメントの評価推定を行う。評価コメントから作られた単語評価を直接使用し, それを利用して得られるコメント評価を再び単語評価に利用する。

### (3) 全コメントの評価推定

人手で評価したコメントとそれから推定された単語の評価, また未評価のコメントの評価を利用して, 全コメントの評価を推定する。推定手法としては, 次の(3-1)と(3-2)を全コメント評価の改善がなくなるまで繰り返す。

(3-1) 単語評価推定をもとに未評価コメントの評価

(3-2)未評価コメントを含む全単語の単語評価推定繰り返し停止後、コメントと単語の最終的な推定値とする。

### 3. コメント評価

前章の自由回答文の解析手法で、シードとなる自由回答文の評価を複数人場合、単語の極性評価が大きく割れるものがある。そこで単語の極性評価値の分散が大きいものと小さいものに分類し、大きいものについては、評価対象から外し上で自由回答文を評価、そして教員・講義評価を行う(図2)。

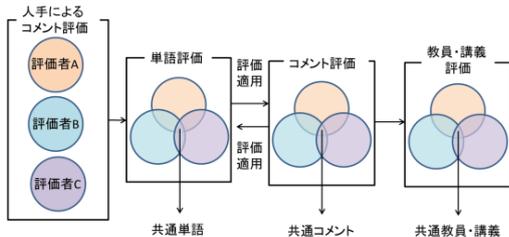


図2 複数人によるコメント評価手順

学生3人による評価を行った場合の単語の極性評価の相違を表1に一部示す。なお、標準偏差が1.5以上と評価された13単語については、評価対象から外して、コメント評価を行った(表2)。また、標準偏差が1.5以下の単語の評価が対象となる評価者のみ相違する場合、その単語を特徴語として抽出する。抽出した特徴語と共通単語、共通コメントの一部を表3に示す。

表1 単語評価の相違(平均ランク、標準偏差)

単語	単語推定ランク平均	単語ランク標準偏差
思う	3.67	2.52
テスト	2.33	2.31
見易い	5.33	1.15
あたらしい	1.00	0.00
ありがたい	5.00	0.00

表2 未評価コメントの推定評価

未評価コメント	個々の評価をそのまま利用した場合			相違の大きい単語を評価から外した場合		
	評価者1	評価者2	評価者3	評価者1	評価者2	評価者3
始めてみる方程式が...	1	5	1	1	1	1
1つ1つていねいに...	6	2	2	1	2	4

表3 推定した特徴語と共通単語・共通コメント

	評価者1	評価者2	評価者3
特徴語	いい, 見つめる, 足りる, 与える, 直す	もらえる, プロジェクト, 点, 強い	すごい, 行う, 進行, 配布, 意欲, 多く
共通単語	上げる, やる気, 感想, 話す, お願い, 担当		
共通コメント	マイクを使わないと聞こえない. 講義担当者...		
	配布資料があるとき用意してくれるのに授業...		

### 4. 教員・講義の評価結果

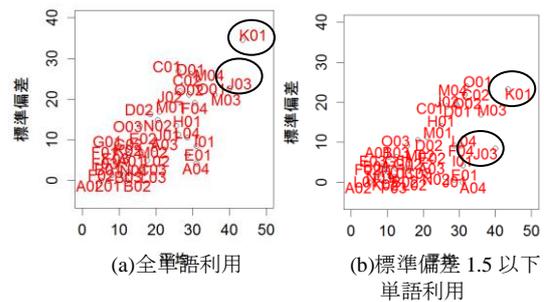


図3 評価者2 講義評価

表4 良い教員・講義

評価対象	全単語を利用した場合		標準偏差が1.5以下の単語を利用した場合	
	教員	講義	教員	講義
評価者1	なし	A01,B03,F02,L03	なし	A01,B03,F02,L03
評価者2	B, L	A01,B01,B03, C03,E01,F02, K02,L01,L02, L03	B	A01,B01,B03, C03,E01,F02, J01,L01,L03
評価者3	B	A01,B01,B03, C03,E01,F02, J01,L03	B	A01,B01,B03, C03,F02,J01, K02,L03

\*アルファベットは教員、数字は講義番号を表す

教員・講義ごとにコメント評価をまとめた散布図を図3に示す。また、コメント内の全単語を利用した場合と標準偏差が1.5以下の単語を利用した場合のコメント評価推定から教員・講義評価を行った結果を表4に示す。教員・講義評価の考察は、(1)平均と標準偏差の値が共に大きい講義はコメント数が多く、コメントの極性に偏りが見られる。(2)全単語を利用した場合と標準偏差が1.5以下の単語を利用した場合のコメント評価推定では、教員・講義評価に大きな相違は見られなかった。(3)評価者1に比べ、評価者2と評価者3はコメントをポジティブに評価しやすい傾向にあることから、良い教員・講義数が評価者1よりも多い。

### 5. 今後の課題

本研究では、複数人によるコメントの評価を利用してラベル無コメントの評価推定を行ったが、今後はさらに評価人数を増やすことでコメントの客観的な評価精度を向上させるための手法を検討する。また、学生の評価だけでなく教員のコメント評価を基にラベル無コメントの評価推定を行い、立場による評価の相違を評価する。

#### 参考文献

- (1) 高村, 乾, 奥村: "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).
- (2) 塩飽, 柳, 椎名, 小林: "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", 教育システム情報学会第40回全国大会, pp.215-216, (2015).