

単語ランクの所属確率を用いた講義アンケートの自由記述分析

塩飽 朝美¹ 小林 伸行² 椎名 広光³

¹ 岡山理科大学大学院 総合情報研究科 情報科学専攻

² 山陽学園大学 総合人間学部 生活心理学科

³ 岡山理科大学 総合情報学部 情報科学科

i15im01sa@ous.jp¹, koba_nob@sguc.ac.jp², shiina@mis.ous.ac.jp³

1 はじめに

体系的な大学教育及び大学院教育の充実をさせるため、教員の教育・研究指導能力の向上を図るFD活動が不可欠になっている [1]. FD 活動の具体的な活動内容として、教員相互の授業参観や全学レベルのセミナー・研修会への参加等が挙げられる [1]. その中でも、教員の教育活動に対する評価として、多くの大学では学生による講義アンケートを実施している. しかし、講義アンケートの分析を行うには、学生と教員における評価の相違が問題になると考えられる.

一方、評価文分類や評価文書分類に関する研究として、文書中に現れる単語をポジティブ・ネガティブによる極性評価する研究 [2] や、評価表現の単語出現比率等から文書全体の評価極性を推定する研究 [3], 正規分布を利用してコメントの評価クラス分類を行う研究 [4] 等が挙げられる.

本研究では、講義アンケートの自由回答文 (以下、コメント)[5]を対象に、学生と教員の立場が異なる評価者が人手で評価した一部コメント評価を基に、コメント評価の推定とコメントに含まれている単語のポジティブ・ネガティブの推定から、有用なコメントの抽出や教員の評価を行う. また、立場の異なる評価者間の違いについても調査を行っている. 本研究の提案手法の大きな特徴としては、評価ランクの所属確率 (単語ランク分布, コメントランク分布) を求めて単語とコメントの評価推定を行っている点にある (図 1).

2 コメント評価手法

2.1 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

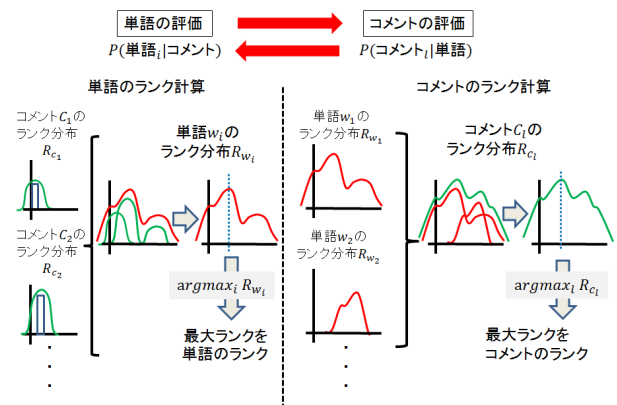


図 1: 単語とコメントの再帰的評価法

た. 調査した教員は 15 人, 講義の科目数は 41 科目, アンケート回答数は 1678 個である. なお, 講義アンケートのコメントは評価得点の取得をしていない.

2.2 解析手順の概要

コメント評価の解析手順を次に示す.

(1) 教師データの作成 (コメントの評価): 一部のコメントを人手で「とても悪い (ランク 1)」、「悪い (ランク 2)」、「少し悪い (ランク 3)」、「少し良い (ランク 4)」、「良い (ランク 5)」、「とても良い (ランク 6)」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.

(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、評価コメント中の各単語の評価ランク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.

(2-1) 評価コメントから名詞・動詞・形容詞・助動詞「ない」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ランクとする.

(2-2) すべての評価コメントとその中に現れる単語に

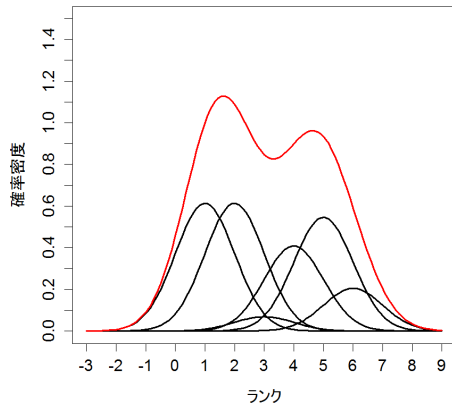


図 2: 単語「説明」の単語ランク分布

対して (2-1) を実施し、単語ごとの評価値の頻度から単語ランク分布を作成する。

(2-3) 単語ランク分布から、最大評価ランクを単語の評価とする。

(3) 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。人手による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。

(4) 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、次の (4-1) と (4-2) を全コメント評価の改善がなくなるまで繰り返す。

(4-1) 単語評価推定をもとに未評価コメントの評価推定

(4-2) 未評価コメントを含む全単語の単語評価推定
繰り返しが停止後、コメントと単語の最終的な推定値とする。

2.3 コメントと単語の再帰的評価

コメントと単語の評価手法に関して、出現頻度が最大である評価ランクをそれぞれの評価ランクとして決定する手法 [6] が考えられるが、この手法は他の評価ランクの影響を考慮していない。よって本研究では、単語とコメントの評価ランクの出現確率を正規分布に当てはめ、得られた混合正規分布から単語とコメントの評価推定を再帰的に行うことで各単語とコメントの評価ランクの出現頻度から作られる評価分布を推定する。評価手法について以下に述べる。

単語	コメントランク分布					
	1	2	3	4	5	6
分かる	1.41	1.24	1.11	1.03	1.16	1.05
にくい	1.50	1.31	1.12	1.04	1.00	1.04
分かりにくい	2.11	1.63	1.24	1.07	1.16	1.09

図 3: コメントランク分布

2.3.1 単語評価

(1) 対象となる単語 w を含むコメントの評価ランクを $i (= 1, \dots, M, M = 6)$ として評価ランクの出現ごとにランク i を中心とした $\mu_i (= i)$, 分散を σ^2 の正規分布を作成する。正規分布を次式で定義する。

$$\phi(x; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}$$

(2) 正規分布の再生性から全評価ランクの正規分布の結合を行い、単語の混合正規分布を作成し、評価ランク分布とする。単語「説明」を例に混合正規分布による評価ランクを図 2 に示す。混合正規分布の混合数を $M, M=6$, パラメータを $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_M)$, α_i をランク i に対する正規分布の重みとした混合正規分布を次式で定義する。

$$p(x; \alpha) = \sum_{i=1}^M \alpha_i \phi(x; \mu_i, \sigma^2),$$

$$\sum_{i=1}^M \alpha_i = 1.$$

(3) 混合正規分布の累積を 1 に正規化した各単語ランク分布から、最大値のランク R を単語評価とする。 k 番目の単語 w_k に対する評価ランクを次式で定義する。

$$R_{w_k} = \operatorname{argmax}_{i=1, \dots, M} (\alpha_i \cdot N(\mu_i, \sigma^2))$$

2.3.2 コメント評価

(1) 正規化した単語ランク分布の全ての値に 1 を加算する。

(2) コメント内に含まれる係り受け関係にある単語のランク分布同士を掛け合わせてコメントランク分布を作成する。コメントの「分かりにくい」を例にコメントランク分布を図 3 に示す。

(3) 各コメントランク分布から、最大値のランクをコメント評価とする。

2.4 評価ランク分布のパラメータ推定

各コメントの評価推定を行うには、人手によるコメント評価との推定差が小さい評価推定を行う必要がある。

平均ユークリッド距離

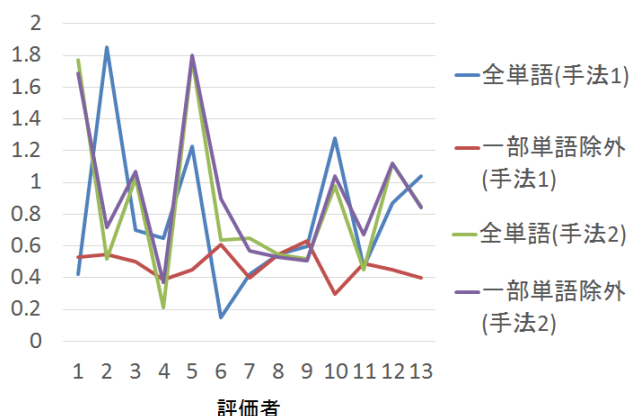


図 4: 各手法による評価コメントの平均ユークリッド距離

表 1: 単語の推定評価結果

単語	推定評価ランク (1~6)				平均	標準偏差
	学生 1	学生 2	学生 3	...		
できる	5	5	4	...	4.69	0.61
課題	1	2	2	...	1.92	0.62
丁寧	5	5	4	...	4.77	0.58
口癖	1	2	2	...	1.85	0.53
交流	1	2	3	...	3.23	1.67
話	1	2	2	...	2.00	0.78
楽しい	5	5	4	...	4.77	0.42

る。よって、本研究では次の 2 つの場合に分けて最急降下法によるパラメータ推定を行う。

(手法 1) 各評価ランクの正規分布に対する重みを等しいと考える場合、 $\alpha_i = \frac{1}{6}$ として、 σ^2 のみをパラメータとする。

(手法 2) 評価ランクごとに重みを付けた場合、各評価ランクの正規分布に対する重み α_i と分散 σ^2 をパラメータとする。

2.5 単語の評価揺れに対する処理

自由回答欄のコメント全 1678 件の内、コメントに含まれる単語数が多い上位 100 件のコメントを評価コメントとして学生 10 名と教員 3 名が人手で評価し、1578 件の未評価コメントの評価推定を行った。本研究で提案するコメントの解析手法は、シードとなるコメントの評価者が複数人である場合、単語の極性評価が大きく割れるものが存在する。そのため、単語の評価揺れに対する処理としてシードとなるコメント評価から得られた単語の評価推定ランクを基に Smirnov-Grubbs 検定 [7] により有意水準 5% の片側検定を行うことで

表 2: 未評価コメントの推定評価結果

未評価コメント	推定評価ランク			平均	標準偏差
	学生 1	学生 2	...		
演習が少ない	2	3	...	2.85	0.95
ペースが速い	2	2	...	2.15	0.53
演習が良い	5	4	...	4.62	0.62
説明が丁寧	5	5	...	4.77	0.58
実習がすくない	2	2	...	2.00	0.39
少人数	6	5	...	4.31	0.99
テンポがゆっくり	6	3	...	4.77	0.70
声が聞き取りにくい	1	2	...	1.77	0.42
字が見にくい	2	2	...	2.08	0.47

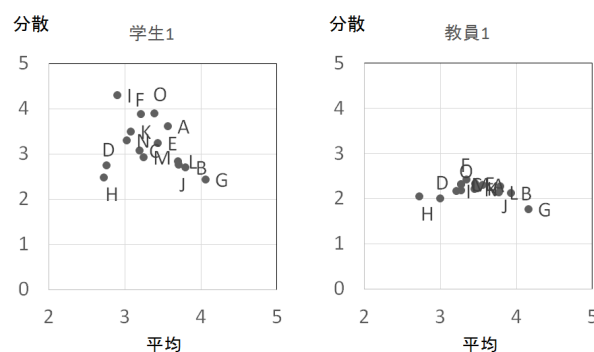


図 5: 教員評価

外れ値を求めた。そして、手法 1 の場合、単語極性の標準偏差が 1.1 以上の 14 単語を、手法 2 の場合、標準偏差が 1.14 以上の 47 単語を評価対象から除きコメント評価を行った。各手法による推定コメント評価と人手によるコメント評価の平均ユークリッド距離を図 4 に示す。

3 評価

3.1 コメントと単語の評価

手法 1 を用いて、評価者間の評価が揺れている単語を除外した場合の全単語の推定評価結果の一部を表 1 に、未評価コメントの推定評価結果の一部を表 2 に示す。

3.2 教員評価

前節のアルゴリズムで得られた全コメントの評価推定結果から、教員ごとにコメントの評価をまとめて教員評価を行う。手法 1 を用いて評価者間の評価が揺れている単語を除外した場合の学生 1 と教員 1 の教員評価を、各教員を A~O のアルファベットとして表した散布図として図 5 に示す。

表 3: 立場による単語とコメントの評価結果

評価者	各評価ランクの正規分布に対する重みが等しい場合		評価ランクごとに重みを付けた場合	
	単語評価の相違	コメント評価の相違	単語評価の相違	共通評価の相違
学生	配布 (P), 実行 (P), 実践 (P), 長い (N), 黒板 (N), 書く (N), プリント (P), 予習 (N)	社会に出て役に立つ (P), 細かい説明が足りない (N), 説明が少し難しい (P)	面白い (P), 役立つ (P), 要点 (P), 細かい (N), 発表 (N), テンポ (N), 企業 (P), レジュメ (P)	難しいけど面白い (P), 練習問題の量が少ない (N), プリントの配布 (P)
教員	細かい (P), 指摘 (P), 挨拶 (P), オーバー (N), 最高 (N), 場 (P), 文句 (N), 途切れる (N)	解説がわかりやすい (P), 配布物が多い (N), ねむくなる (N)	助かる (P), 進め方 (P), 形式 (P), 交流 (N), 簡単 (N), 配る (N), 振り返る (P), 黒板 (N)	話が深い (P), ねむくなる (N), 簡単です (N)
	共通極性評価単語	共通極性評価コメント	共通極性評価単語	共通極性評価コメント
全評価者	字 (N), にくい (N), 時間 (N), 丁寧 (P), 配布 (P), 面白い (P), 声 (N), 雑談 (N), 強調 (P), 実習 (P), 口調 (N), 私語 (N)	話が長い (N), 字が汚い (N), 難しいけど楽しい (P), 細かく教えてくれる (P)	字 (N), 課題 (N), 板書 (N), 丁寧 (P), 対応 (P), 情報 (P), 点 (P), 繰り返す (P), ミス (N), マイク (N), 自慢 (N), 速度 (N)	声が大きいの (N), 板書が少ない (N), 楽しい!! (P), 丁寧な説明 (P)

(P) はポジティブ評価, (N) はネガティブ評価

3.3 立場による評価の相違

評価ランク 1~3 をネガティブ, 4~6 をポジティブ評価として, 学生と教員の立場ごとに共通している極性評価と相違している極性評価結果となった単語とコメントの例を表 3 に示す。

4 まとめと今後の課題

本研究では, 評価ランクの所属確率を求めて講義アンケートのコメントとコメントに含まれる単語の評価推定を行った。各手法による調査結果として, (1) 手法 1 を用いて評価者間で評価が揺れている単語を評価対象から除外した場合, 他の手法に比べてコメントの推定評価と人手のコメント評価の平均ユークリッド距離が小さいことから, 提案する手法の中で最も人手に近いコメントの推定評価が出来たと考えられる。(2) 全評価者による単語とコメントの推定評価から, コメント中に声, 字, スピードに関する単語が出現するとネガティブ評価に, 復習, 説明の詳細さ, 社会に関する単語が出現するとポジティブ評価になりやすいと考えられる。また, 立場ごとの単語とコメントの推定評価から, 学生は「語尾」, 「口調」, 「口癖」等教員の話し方に関する単語や, 「自慢」, 「雑談」等話に関する単語をネガティブと評価している。よって, 学生は教員が講義中に講義内容と関係のない話をするのを好ましく思っていないと考えられる。

今後の課題としては, 単語の評価の種類と種類間の相関について調査したいと考えている。

参考文献

- [1] 文部科学省, <http://www.mext.go.jp/>
- [2] 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.
- [3] P.D. Turney, Thumbs up? thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002), pp.417424, 2002.
- [4] 三原隆義, 小林伸行, 椎名広光, 北川文夫, データ周辺の生起確率推定による機械学習によるコメントの評価クラス分類, 言語処理学会 第 21 回年次大会, 2015.
- [5] 岡山理科大学総合情報学部情報科学科中間アンケート, 2014.
- [6] 塩飽朝美, 椎名広光, 自由回答アンケートに対する再帰的評価手法を用いた講義アンケートの立場による相違評価, 電気・情報関連学会中国支部連合大会 第 67 回, 2016.
- [7] 石村貞夫, Desmond Allen, 劉 晨, すぐわかる統計用語の基礎知識, 東京図書, 2016.