

再帰的評価を利用した講義アンケートの 単語・コメント評価

塩飽 朝美^{1,a)} 椎名 広光^{2,b)} 小林 伸行^{3,c)}

概要：現在、各大学ではFD活動の一環として講義アンケートを実施し、教員や講義に対する学生の満足度等を調査することで教育や講義改善を図っている。本研究では、講義アンケートの自由回答欄のコメントを対象にコメントとそれに含まれる単語の評価推定を行う。コメントの評価方法として、一部コメントの複数人による評価に対してコメントと単語のそれぞれの評価推定を相互に再帰的に繰り返し、全コメントの評価推定を行う。また、個々人によるコメントと単語の推定評価の相違を評価する。

キーワード：多人数評価, 自由記述解析, 単語極性, 講義評価, 再帰的評価

Word and Comments Evaluation of Lecture questionnaire using Recursive Evaluation

ASAMI SHIWAKU^{1,a)} HIROMITSU SHIINA^{2,b)} NOBUYUKI KOBAYASHI^{3,c)}

Abstract: There is a call for an increase to education quality in response to FD (Faculty Development) activities becoming mandatory due to a revision of the University Establishment Standards in 2008. A survey on lectures was carried out as part of FD activities and used for improving lectures based on lecture evaluation and satisfaction from students. However, it is difficult to give an objective evaluation of this text data since these subjects of evaluation are wide-ranging, such as expectations towards lectures or opinions on teachers. An aggregation of the opinions gained through surveys also has its limits for manual assessment because of the heavy artificial costs and time costs required. Therefore, it is difficult to evaluate all comments given in a survey. This study uses actual open ended responses to surveys to describe a method for re-evaluating a comment by estimating the evaluation of words within comments from comment evaluations done through machine learning, as well as a method that uses a dictionary of emotional words to attach a polarity value to vocabulary and estimate the evaluation of the body of a comment from that value.

Keywords: Multi evaluation, Free questionair, Word polarity, Questinair analisys, Lecture evaluation Multiplayer evaluation, Word polarity, Questionnaire analysis, Lecture evaluation, Recursive Evaluation

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。しかし、これらテキストデータは講義に対する要望や教員に対する意見等、評価対象が多岐に渡ることから客観的評価が困難であると考えられる。また、アンケートによる意見集約は多大な人的コストや時間的コストが必要とされることから人手による評価には限界があるといえる。よって、アンケートにおいてコメ

¹ 岡山理科大学大学院総合情報研究科
Graduate school of Informatics, Okayama University of
Science, Ridaicho 1-1 Kitaku Okayama, Tokyo 700-0005,
Japan
² 岡山理科大学総合情報学部
³ 山陽学園大学総合人間部
^{a)} i15im01sa@ous.jp
^{b)} shiina@mis.ous.ac.jp
^{c)} koba_nob[at]sguc.ac.jp

ントを全て評価することは困難である。

近年では単語の評価表現や機械学習等による評価文分類や評価文書分類に関する研究が広く行われている [1] [2] [3]。その関連研究として、文書中に現れるポジティブ・ネガティブによる評価表現の単語出現比率から文書全体の評価極性を推定する研究 [5] や、評価情報の要素組を抽出し、評価極性を推定する研究 [6] 等が挙げられる。また、文書分類において機械学習は高い分類精度を達成しておりナイーブベイズやサポートベクトルマシン等を用いた評価文書分類の研究がある [7] [8]。

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。

単語のポジティブ・ネガティブの極性判定については、感情語 (Semantic Orientation) 辞書作成の関連研究がある [9]。本研究では、まず比較対象としてコメント内の各単語の評価に感情語辞書を適用することで、コメントの評価推定を行う。しかし、オープンデータの感情語辞書をそのまま利用すると全体評価が下がりすぎるきらいがある。そこで実際に行われた講義アンケートの自由回答欄のコメントを対象に、一部のコメントだけに評価を付け、残りのコメントの評価を推定する手法を適用する。コメントの評価については、コメントを評価したものからコメント内の単語の評価を推定し、コメントの再評価を行う再帰的評価手法を提案する。加えて、再帰的評価を行う中でのコメントから単語の評価、単語からコメントの評価のそれぞれの手法については、評価ランクの頻度を加算していく方法、正規化して乗算する手法について説明する。

また、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ている。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がある。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、単語の性質を調査している。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、本研究は行っている。

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15人、講義の科目数は41科目、アンケート回答

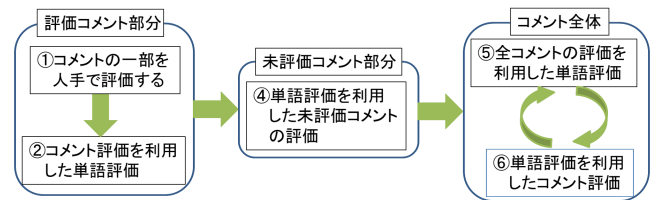


図 1 コメントの評価手順

Fig. 1 The procedure of comment

数は1678個である。

3. 自由回答文の解析手法

3.1 解析手順の概要

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。

(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

繰り返しが停止後、コメントと単語の最終的な推定値とする。

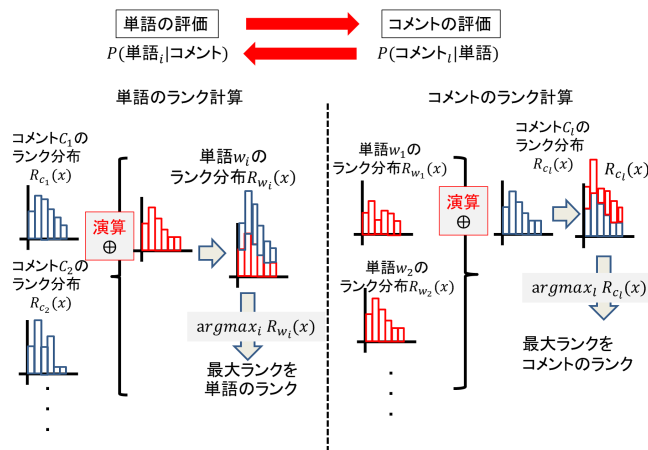


図 2 コメントと単語のランク再帰的評価

Fig. 2 Rank recursive evaluation of comments and words

3.2 コメントと単語の再帰的評価

前節では、評価コメントと未評価コメントに分け、未評価コメントの評価が定まるまで繰り返すようにコメント評価を行っていることを述べた。その中で、コメントの評価から単語の評価を行い、その逆の単語の評価からコメントの評価を行っている。本研究では、コメント評価から単語評価と単語評価からコメント評価を再帰的に行っている。コメント評価をもとに単語を評価の方法、その逆の単語の評価からコメントの評価の方法については、2種類の手法を提案する。

- (1) 評価ランクの頻度を合計する手法
- (2) 評価ランクをランク分布として、ランク分布同士を掛け合わせる手法

3.2.1 評価ランクの頻度を合計する手法

評価ランクの頻度を合計して、単語とコメントの評価推定を行う手順を以下に示す。

- (1) 単語評価
 - (1-1) 各単語ごとに自身を含むコメントのランク頻度を合計した単語ランク分布を作成する。
 - (1-2) 各単語の分布ごとに、最大値のランクを単語評価とする。
- (2) コメント評価
 - (2-1) コメントが含む単語の単語ランク頻度を合計したコメントランク頻度を作成する。
 - (2-2) 各コメントの分布ごとに、最大値のランクをコメント評価とする。

3.2.2 ランク分布同士を掛け合わせる手法

ランク分布同士を掛け合わせて、単語とコメントの評価推定を行う手順を以下に示す。

- (1) 単語評価
 - (1-1) 単語ごとに自身を含むコメントのランク頻度の合計

表 1 コメントの評価分布

Table 1 Comment evaluation of distribution

| コメントランク評価 | コメント数 |
|-----------|-------|
| とても悪い | 15 |
| 悪い | 21 |
| 少し悪い | 13 |
| 少し良い | 9 |
| 良い | 22 |
| とても良い | 20 |

表 2 単語評価

Table 2 Word evaluation

| | ランク頻度 | | ランク分布掛け合わせ | |
|-------|-------|------|------------|------|
| | 平均 | 標準偏差 | 平均 | 標準偏差 |
| 単語 | | | | |
| 思う | 3.67 | 2.52 | 3.33 | 2.08 |
| テスト | 2.33 | 2.31 | 2.33 | 2.31 |
| 見易い | 5.33 | 1.15 | 5.33 | 1.15 |
| あたらしい | 1.00 | 0.00 | 1.00 | 0.00 |
| ありがたい | 5.00 | 0.00 | 4.67 | 0.58 |

が確率 1 になるように正規化し、単語ランク分布を作成する。

- (1-2) 各単語の分布ごとに、最大値のランクを単語評価とする。
- (2) コメント評価
 - (2-1) 正規化した単語ランク分布の全ての値に 0.001 を加算する。
 - (2-2) コメント内に含まれ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。
 - (2-3) 各コメントの分布ごとに、最大値のランクをコメント評価とする。

4. コメントと単語の評価

2つの手法でのコメントと単語の評価結果について述べる。

自由回答欄のコメント 1678 個のうちコメントに含まれている単語数の上位 100 件をコメント評価を行ってから未評価コメント 1578 個を評価した。評価者 1 の評価コメント分布は表 1 に示す。

4.1 単語評価とコメント評価の評価例

表 1 のコメント評価をもとに、次の手法で単語とコメントを評価する。

- (1) 評価ランクの頻度を合計する手法
- (2) ランク分布同士を掛け合わせる手法

また、単語ランクの評価を表 2 に示す。

表 3 未評価コメントの評価
Table 3 Evaluation of unevaluated comments

| コメント | ランク 頻度 | ランク分布の 掛け合わせ |
|--|-----------|-----------------|
| 初めて見る方程式 | 1 | 1 |
| 1つ1ついいねいに... | 6 | 1 |
| 説明が分かりにくい部分がある, 常に 声の調子が一緒過ぎて頭に残りにくい | 3 | 1 |
| 字がきたない。説明がわかりづらい。 置いていかれる。 | 2 | 2 |
| こういう方向性の授業を求めている 人にはいいと思います。 | 1 | 1 |
| プロジェクターを使って最初に説明して から, 演習に入るところが良いと思う | 4 | 5 |
| 資料はちゃんと用意されているので, 予習, 復習がきちんと出来る | 5 | 5 |
| モニターとプリントが併用されて いてわかりやすい。 | 1 | 1 |

一方, 単語評価に対してコメントを構成する単語のリストに単語評価を適用し, コメント評価から単語評価の逆の単語評価からコメント評価を行う。それぞれの評価を表 3 に示す。

4.2 コメントと単語の評価の評価手法による違い

各手法による違いは次のとおりである。

- (1) ランク頻度の合計を行った場合, ネガティブコメント数よりもポジティブコメント数の方が多い。また, 評価者 1 はランク 1, 評価者 2 はランク 2, 評価者 3 はランク 4 の特徴語を多く持つ。
- (2) ランク分布の掛け合わせを行った場合, ネガティブコメント数とポジティブコメント数に大きな差はない。また, 評価者 1 はランク 1, 評価者 2 はランク 5, 評価者 3 はランク 4 の特徴語を多く持つ。

5. コメントと単語の評価の個人ごとの相違

前章のコメントと単語の評価では, 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 3)。

5.1 共通コメント評価法

学生 3 人による評価を行った場合の単語評価の違いを, 表 4 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

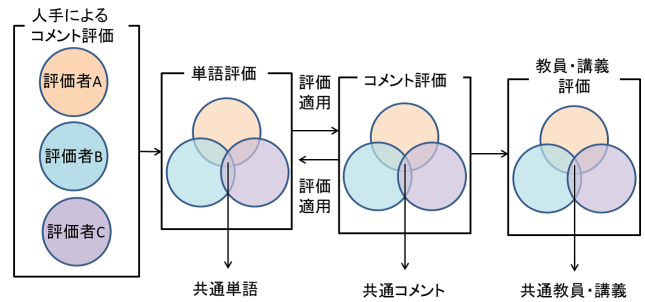


図 3 複数人によるコメント評価手順
Fig. 3 Procedure of comment evaluation by three person

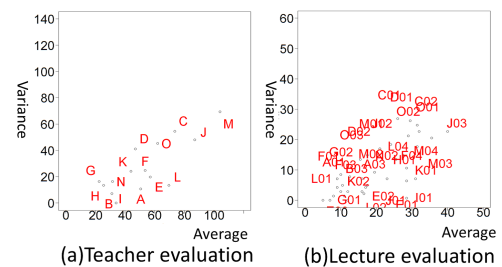


図 4 再帰的評価に散布図
Fig. 4 Scatter diagram using recursive evaluation

価を行っている。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も併せてを表 4 に示す。

5.2 単語評価とコメント評価

各手法による単語評価とコメント評価は, 次のとおりである。

- (1) ランク頻度の合計を用いた場合, 全評価者でポジティブ単語よりもネガティブ単語の方が多い。コメント評価として, 全評価者でネガティブコメント数よりもポジティブコメント数の方が多い。
- (2) ランク分布の掛け合わせを用いた場合, 評価者 1 と評価者 2 はポジティブ単語よりもネガティブ単語が多いのに対し, 評価者 3 はポジティブ単語が多い。コメント評価は, 評価者 1 は (1) 同様, ネガティブコメント数よりもポジティブコメント数が多いのに対し, 評価者 2 と評価者 3 はネガティブコメント数の方が多い。
- (3) 2つの手法とも, 単語とコメントの評価推定を繰り返すほど推定評価と人手による評価の平均ユークリッド距離は大きくなる。

表 4 未評価コメントの推定
 Table 4 Estimation of unevaluated comments

| 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 |
| 初めて見る方程式 | 1 | 5 | 1 | 1 | 1 | 1 |
| 1つ1ついいねいに.... | 6 | 2 | 2 | 1 | 2 | 4 |
| 課題や今日のやる内容などをプリン トに書いてあるので分かりやすい。 | 1 | 5 | 4 | 2 | 5 | 4 |
| 字がきたない。説明がわかりづらい。 置いていかれる。 | 2 | 2 | 2 | 2 | 1 | 2 |
| こういう方向性の授業を求めている 人にはいいと思います。 | 1 | 5 | 4 | 5 | 5 | 4 |
| たまに声が小さくて質問されている ことがわからない。 | 5 | 5 | 4 | 1 | 5 | 4 |
| 資料はちゃんと用意されているので、 予習、復習がきちんと出来る。 | 5 | 2 | 4 | 6 | 5 | 4 |
| モニターとプリントが併用されて いてわかりやすい。 | 1 | 5 | 4 | 6 | 5 | 4 |

表 5 推定した特徴語、共通単語、共通コメント
 Table 5 Estimated characteristic words, common words and common comments

| ランク頻度の合計 | | | |
|-------------|---|--|--|
| | 評価者 1 | 評価者 2 | 評価者 3 |
| 特徴語 | いい (1), 見つめる (6), 足りる (6), 与える (6), 直す (6), レポート (1), 指摘 (6), 間違い (1), 機会 (6), 担当 (3), 能力 (6), 内 (1), 場合 (1), 多い (1), | もらえる (5), プロジェクター (6), 見づらい (1), 残る (2), 大きい (2), わかる (1), 過ぎる (2), 改める (2), 含む (2), 面白い (5), 強い (2), 点 (5) | すごい (4), 行う (3), 進行 (3), 意欲 (4), 多く (4), 楽しい (4), やすい (4), できる (4), 示す (4), 資料 (4), 難しい (4), きこえる (2) |
| 共通語 | 上げる (4), やる気 (4), 感想 (4), 話す (2), お願い (2), 担当 (2), カラー (5), あたらしい (1), つらい (3), 鬱陶しい (1) | | |
| 共通語 コメント | マイクを使わないと聞こえない。講義担当者が最前列の学生とのみ話して講義を進めるのでよくわからない (2) 配布資料があるとき用意してくれるのに授業資料は事前に各自で印刷するという形式 (3) 進行スピードが早い, 何かモゴモゴして聞きづらい時がある, それ以前に他のクラスがやっていない応用に手を出すとか何を考えているのかと, プリントとか補助資料を出してみたらと (1) 指導に力が入りすぎて感情的になることがあるところ (3) スライドを使って説明してくれたりカラーの配布プリントをくれるのでとても学習するのに役立つありがたいです (5) | | |
| ランク分布の掛け合わせ | | | |
| | 評価者 1 | 評価者 2 | 評価者 3 |
| 特徴語 | 多い (1), シャベる (1), ずれる (2), 指摘 (6), 口癖 (1), ミス (1), スピーカー (2), 与える (6), 機会 (6), 笑い (1), 投げやり (1), うまい (5) | くわしい (1), 見づらい (1), 勉強 (5), づらい (1), 解説 (5), 統計 (5), バランス (3), まとめ (5), 見にくい (1) | 止まる (2), ありがたい (4), すごい (4), 計画 (3), 配る (1), 将来 (6), 要点 (4), 無駄 (2), 変わる (1) |
| 共通語 | にくい (2), 小さい (1), 早い (1), 板書 (2), 更新 (2), 退屈 (1), キーワード (2), スピード (1) | | |
| 共通語 コメント | 実力がない自分にとってはこの進行スピードは苦しい (1) 時間配分がおかしい, 進行速度が遅い (2) 記号が達筆で読めないことがある。(3) レジュメがいただけるのありがたいです (5) 全体への説明が雑で分かりにくい (2) | | |

表 6 教員評価
 Table 6 Teacher evaluation

| | 全単語を利用した場合 | | 標準偏差が 1.5 以下 | |
|-------|------------|--|--------------|---|
| | 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, A03,F02,J01, C03,F02,J01, K02,L03 |
| 評価者 3 | B | A01,B01,B03 C03,E01,F02 J01,L03 | B | A01,B01,B03 C03,F02,J01 K02,L03 |

6. 教員・講義評価

教員・講義ごとにコメント評価をまとめた散布図を図 4 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・講義評価として、

- (1) 平均と標準偏差の値が共に大きい教員・講義はコメント数が多く、コメントの極性に偏りがある。反対に、平均と標準偏差が小さな教員・講義はコメント数が少ないことから評価を行うことは困難であると考えられる。
- (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い。
- (3) ランク頻度の合計を行った場合とランク分布の掛け合わせを行った場合で、評価者 1 に教員・講義評価の相違はなかった。評価者 2 と評価者 3 はランク頻度の合計を行った場合よりもランク分布の掛け合わせを行った場合の方が、良い教員・講義数が評価者 1 よりも多い。
- (4) 標準偏差の大きな教員・講義は、評価者間のコメント評価に相違が多く存在する。

7. おわりに

本研究では、複数人による一部コメント評価を基にコメント内の単語と全コメントの評価推定を行った。各手法による調査結果として、(1) ラベル頻度の合計を行った場合とランク分布の掛け合わせを行った場合の 2 つの手法とも、単語評価として評価者 1 はランク 1、評価者 2 はランク 5、評価者 3 はランク 4 を持つ単語が最も多く存在する。これは、各評価者がコメントを人手で 6 段階評価した学習データの結果が大きく影響していると考えられる。また、コメント評価として評価者 1 はランク 1、評価者 2 はランク 5、評価者 3 はランク 4 の評価を持つコメントが最も多く存在する。このことから、評価者 1 はネガティブな評価、評価

者 2 と評価者 3 はポジティブな評価をする傾向にあると考えられる。また、評価推定を繰り返すほど各評価者のコメント評価は特定の評価値に評価が偏る傾向がある。評価者間の共通語・共通コメントとして、ネガティブな評価を持つ単語とコメントが多いことからポジティブな評価よりもネガティブな評価の方が評価者間で一致しやすいと考えられる。(2) ラベル頻度の合計を行う場合、評価者間で単語の極性評価値の分散が大きいのを除かずコメントの評価推定を行うよりも、除いた方が学習データとの推定評価の相違が少なかった。

今後の課題として、本研究では 3 名の学生による学習データを基に解析を行ったが、さらに調査人数を増やすことで個人と集団の評価の相違を調査し、より客観的な単語とコメントの評価精度を向上させるための手法を検討する。また、教員の立場からのコメント評価を調査することで学生と教員の立場による評価の相違を評価する。

謝辞 本研究は JSPS 科研費 16K01145 の助成を受けたものです。

参考文献

- [1] K. Nigam, A. McCallum, S. Thrun, T. Mitchell, Text Classification from Labeled and Unlabeled Documents using EM Machine Learning, 39 (2/3), pp.103-134, 2000.
- [2] 乾孝司, 奥村学: “テキストを対象とした評価情報の分析に関する研究動向”, 自然言語処理, Vol 13, No.3, pp201-241 (2006)
- [3] 鍛冶伸裕, 喜連川優: “文構造を考慮した評価文書分類のための確率モデル”, 電子情報通信学会第 17 回データ工学ワークショップ (2006)
- [4] 塩飽朝美, 貴久男, 椎名広光, 小林伸行, 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.
- [5] 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.417-424, 2002.
- [6] 鈴木泰裕, 高村大也, 奥村学: “Weblog を対象とした評価表現抽出”, 人工知能学会, セマンティックウェブとオンロジー研究会, SIG-SWO-A401-02, (2004)
- [7] I. Rish, An empirical study of the Naive Bayes classifier, IJCA 2001 workshop on empirical methods in artificial intelligence, Vol. 3, No. 22, pp.41-46, 2001.
- [8] V.N.Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
- [9] 高村大也, 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 (2002)