

Twitter と論文との自動対応付け

中橋 光[†] 難波 英嗣[‡] 竹澤 寿幸[†] 高須 淳宏[‡]

[†] 広島市立大学大学院情報科学研究科 〒731-3194 広島県広島市安佐南区大塚東三丁目 4 番 1 号

[‡] 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: [†] {nakahashi, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp, [‡] takasu@nii.ac.jp

あらまし 近年、学会議の参加者が論文(研究発表)に対するコメントを Twitter に投稿する機会が増えつつある。このようなツイートは、多様な観点から論文を理解するのに有益であると考えられる。そこで、我々は、ツイートと学術論文の自動対応付けの研究を行なっている。対応付けは 2 つのステップ、(1) 各ツイートと論文との対応付けおよび(2) ツイートの分類による有益なツイートの検出から構成される。提案手法の有効性を確認するため実験を行った。実験の結果、ツイートと論文の対応付けでは再現率 0.5080、精度 0.4853 を、ツイートの分類では、再現率 0.4533、精度 0.5782 を、それぞれ得た。

キーワード Twitter, 論文, DP マッチング, SVM

Automatic Alignment of Tweets with Research Papers

Hikaru NAKAHASHI[†] Hidetsugu NANBA[‡] Toshiyuki TAKEZAWA[†] and Atsuhiko TAKASU[‡]

[†] Graduate School of Information Sciences, Hiroshima City University

3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima 731-3194 Japan

[‡] National Institute of Informatics

Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: [†] {nakahashi, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp, [‡] takasu@nii.ac.jp

Abstract Recently, participants of academic conferences often post their comments for each research paper (presentation) to Twitter. Some of these comments are useful for understanding the papers from various viewpoints. Therefore, we have studied automatic alignment of tweets with research papers. The procedure consists of the following two steps: (1) alignment each tweet with a research paper and (2) classification of tweets for detecting useful ones. We conducted some experiments to confirm the effectiveness of our method. From the results, we obtained recall and precision scores of 0.5080 and 0.4853, respectively, for the alignment of tweets with research papers. We also obtained recall and precision scores of 0.4533 and 0.5782, respectively, for the classification of tweets.

Keyword Twitter, research paper, DP matching, SVM

1. はじめに

論文中で先行する著作を参照する習慣は、19 世紀に確立したと言われている。論文の参照は情報流通のひとつの形態であり、参照論文を分析することで、論文間の結びつきを見出したり、論文や研究者の評価を行ったりすることは、十分意義深いことであると考えられる。

論文間の参照関係はこれまで様々な目的に利用されている。そのひとつに、ある論文(被参照論文)を参照する複数の論文中で被参照論文について言及した箇所(参照箇所)を抽出し、まとめることで、被参照論文の読解支援をする、という試みがある[1]。しかし、この方法は、その論文が参照されるまでは適用できないという問題点がある。

そこで、本研究では Twitter に着目する。Twitter は、

近年、急速に普及が進んでいるコミュニケーションツールである。Twitter には、ユーザの日常生活に関するツイートだけでなく、学会等で発表された論文への意見や疑問を投げかけるツイートや論文の評価に関するも数多く投稿されている。学会で発表されている論文とこのようなツイートを対応付けることができれば、その論文に対するユーザの評価、意見、議論など論文からだけでは得ることができない多様な情報を得ることができると考えられる。

本研究では論文とツイートを自動的に対応付ける手法を提案する。Twitter はリアルタイム性に優れているため、学会終了直後にも評価等を知ることができる。ここで、ツイートの中には、意見や評判など論文の理解に役立つものの他に、発表の様子を実況中継したり、発表の開始時に論文表題と発表者名をツイートするといった、論文理解にはあまり貢献しないものも存在す

る。そこで、本研究では、ツイートをいくつかのタイプに自動分類し、有益なツイートの検出を目指す。

本論文の構成は以下のとおりである。2 節では関連研究を述べる。3 節では、Twitter と論文の自動対応付けとタイプ分類についての提案手法について述べる。4 節では、提案手法の有効性を調べるために行った実験について述べ、5 節で考察を行う。6 節で結論を述べる。

2. 関連研究

本節では、本研究に関連する研究について述べる。まず、対応付けに関する研究について述べる。越智ら[2]は学会での発表に対する聴講者のレスポンスを収集するために、Twitter と発表スライドとの対応付けを行っている。そして、レスポンスの収集のために、プレゼンシステムとモバイルシステムを開発している。プレゼンシステムでは、発表者がスライドのノート情報を管理サーバに送信することで、スライドのノート情報とスライド画像の URL が Twitter に自動投稿される。モバイルシステムでは、発表中のスライドが表示され、スライドに対するコメントとレスポンスの送信を行うことができる。この2つのシステムにより、スライド情報の送信とレスポンスの収集を行うことができる。発表論文のレスポンス情報を収集するという目的では、本研究と似ているが、越智らは Twitter を発表スライドと対応付けている点で本研究と異なる。

坂東[3]は、論文の重要度を測るために、Twitter などのソーシャルメディアと論文との対応付けを行っている。論文の重要度を測る従来手法には、論文間の引用関係を測る手法がある。論文の重要度を測る尺度として Garfield[4]の提案したインパクト・ファクタが代表的である。また、Yogatama ら[5]は、科学界の反響の予測のために、論文の引用数やダウンロード数を用いて予測を行っている。しかし、どちらの研究も、引用関係を調べるまでに時間を要するという問題点がある。そこで坂東は、従来手法の引用関係に代わる新たな手法として、Altmetric を提案している。Altmetric とは、ソーシャルメディアを活用して研究成果のインパクトを論文レベルでリアルタイムに計量化する新たな研究評価指標である。しかし、坂東の研究では、論文について言及しているツイート数が示されているが、論文への評価極性については考慮されていない。この問題を解決する手法の一つとして、本研究において対応付けされた論文に関するツイートを分析する手法が考えられる。この手法では、論文に対して肯定的なツイートと否定的なツイートを分類でき、より正確な論文の重要度を示すことができると考えられる。

次に、タイプ分類に関する研究について述べる。Twitter 上において評判情報を用いた研究として、Long ら[6]の研究がある。Long らは、Barbosa ら[7]の評価辞

書機能を用いて評価極性の判定を行っている。そして、Long らの研究では評価対象の拡張に着目している。例えば、“Microsoft”についての評価を表現するとき、“Microsoft”自体について評価するのではなく、“Microsoft technologies”のように、“Microsoft”における対象の何かしらの要素について言及することが多い。そこで、この要素を評価語の対象として選択する。本研究でも、この評判情報を用いてツイートのタイプ分類を行う。さらに本研究では、手掛かり語も使用することで、より精度の良い分類を実現する。

3. 提案手法

本節では、Twitter と論文との自動対応付けとツイートのタイプ分類についての手法を述べる。

本研究の流れとして、初めに、Twitter と論文との自動対応付けを行う。Twitter と論文を対応付ける手法については、3.1 節で説明を行う。

次に、ツイートには複数のタイプが存在するため、ツイートのタイプ分類を行う。タイプ分類手法については、3.2 節で述べる。

3.1. Twitter と論文との自動対応付け

本節では、学会の発表論文に関するツイートと学会で発表された論文を自動対応付けする手法について述べる。提案手法の流れを以下に示す。

- ① ツイートが投稿された時間と論文の発表時間を用いて、ツイートに対応する論文の候補を選定する。ツイートは、論文の発表中や発表後に投稿されることが多いため、本研究ではツイートの投稿された時間から 30 分前までに発表された論文を参照する。
- ② 対象のツイートと①で選定した論文間の類似度を算出する。類似性尺度として、DP マッチングとコサイン類似度を用いる。
- ③ ツイートに対して①で選定した論文の中から類似度が最も高い論文をそのツイートに対応付ける。

手順②において、ツイートは最大で 140 字という文字数制限があるため、論文との類似度を測る際、ツイートから得られる情報量が少ないため、正確な類似度が測れないという問題がある。また、1つの論文発表に対して、複数回ツイートするユーザが存在する場合がある。そこで本研究では、類似度を測る時に用いるツイートから、前後 20 分に投稿された同一のユーザのツイートを追加して類似度を測る手法を行う。本研究では、この手法を「ユーザ情報あり」と呼ぶ。

以下に、本研究で用いる DP マッチングとコサイン類似度について述べる。

DP マッチング

DP マッチングは、動的計画法を用いて2つのパター

ンの要素間の整列化を行い、類似度を測る手法である。類似度を求める際に重み付け(以下、ペナルティ)を行う。2つのパターンの要素間におけるペナルティの値が小さいほど、より類似していると判断している。

本研究では、DP マッチングの要素として形態素を用いる。類似度を測る計算方法を式(1)に示す。

$$DP = \sqrt{\frac{j}{\text{文 1 の形態素数}} \times \frac{j}{\text{文 2 の形態素数}}} \quad (1)$$

なお、j は文 1, 2 の形態素の一致した数である。本研究では、文 1 は論文、文 2 はツイートを示す。

コサイン類似度

コサイン類似度とは、2つのベクトル間の余弦の値を利用する方法である。1 に近ければ類似していることになる。2つのベクトルには、ツイートと論文の単語を用いる。文書ベクトルの重みには tf-idf を用いる。

3.2. ツイートのタイプ分類

本節では、タイプ分類の手法について述べる。本研究では、タイプ分類の手法として、機械学習に基づく手法を提案する。本研究の手順として、初めに、学会の発表論文に関するツイートを人手によりタイプ分類を行う。ツイートのタイプ分類についての詳細は 3.2.1 節で述べる。その後収集したツイートを機械学習により学会に関するツイートのタイプの自動分類を行う。ツイートのタイプの自動分類についての詳細は、3.2.2 節で述べる。

3.2.1. ツイートのタイプ分類

本研究では以下の4つのタイプに分類を行う。

1. type S (Sentiment)
2. type B (Broadcast)
3. type D (Discussion)
4. type O (Other)

本研究では、1つのツイートが複数タイプを持つ場合もある。それぞれのタイプについて述べる。

1: フォントとオノマトペの関係は面白いなあ

図 1 : type S のツイート例

図 1 のように、論文について「面白い」など評価しているツイートを type S に分類する。この type S のツイートの評価極性を分析することで、論文に対するユーザの評価を調べることができる。

1: 「ブログを中心とした観光情報の組織化」旅行ブログエントリ中に含まれるリンクを抽出する

図 2 : type B のツイート例

図 2 のように、発表している論文について実況しているツイートを type B に分類する。type B には論文の内容について実況しているツイートが多い。図 2 の場

合でも、論文に「旅行ブログエントリから観光情報リンクを抽出し分類する手法を提案する」という記述があり、論文の重要個所を要約してツイートしている。よって、type B のツイートを分析することにより論文の重要個所を容易に知ることができる。

1: 「ゆったり」「ゆっくり」みたいな語もオノマトペなんですか？
 2: @hijip 入力されたオノマトペは母音と子音に分解されて数値化されるので、どんなオノマトペであってもイケます！
 3: 「ゆったり」「ゆっくり」などは宿ホテルへの評価に含まれる特徴的な言葉なのではないか
 4: 関係しそう「共起情報抽出による日韓オノマトペの意味差の分析に向けて」
<http://t.co/HsoGL4Xh>

図 3 : type D のツイート例

図 3 の例 1 のように論文について質問しているツイート、例 2 のように質問に対して応答しているツイート、例 3 のように論文について提案しているツイート、例 4 のように論文について付加情報を示しているツイートを議論しているツイートとして type D に分類する。質問等に関するツイートを分析することで、学会での実際の議論の内容を容易に知ることができる。

1: RT @nanaya_sac オノマトペ満載のあのレシピ本は異端なのか。
 2: 徒歩 15 分の仕掛学セッションに来た。

図 4 : type O のツイート例

図 4 のように、学会の発表論文に関係のないツイートやリツイートを type O に分類する。本研究では、リツイートはユーザ自身のツイートではないので省く。

3.2.2. 機械学習によるツイートの自動分類

本研究では機械学習によりツイートタイプの自動分類を行う。機械学習に用いた素性を以下に述べる。評価表現辞書

評価表現辞書は、鍛冶ら[8]の大規模な評価文コーパスにより自動構築されたものであり、形容詞、形容詞句と評価極性値のペアが約 10,000 組登録されている。

評価表現辞書には、図 1 の例 1 の「面白い」など論文の評価に用いられる語が多く登録されている。本研究では、論文の評価に関するツイートを分類する際の手がかり語として、鍛冶らの評価表現辞書を用いる。

手がかり語

本研究では、手がかり語リスト作成のために加藤ら[9]の研究で用いられている情報利得を用いた。

本研究では、type S, type B, type D, type O の4つのタイプをクラスとして用いる。type B は実況についてのツイートなので、type S の評価語のような特定の

言語的特徴がないため手掛かり語を収集することは困難である。そのため本研究では、type S と type D について情報利得の値を求め、上位 100 件の単語を手掛かり語とする。収集した手掛かり語のツイートでの出現頻度を素性として用いる。

ツイートと論文の類似度

type B では、図 2 のツイート例 1 のように発表している論文の内容と類似しているツイートが多い。そこで、論文とツイートの類似度を素性として用いる。類似性尺度として DP マッチングを用いた。

文タイプ

ある論文に関して意見を述べる際、ツイートの文末には「方がいい」、「望ましい」などの表現が用いられることがある。そこで文末表現に着目して決定される文タイプ[10]を素性として用いる。

リツイートの有無

リツイートとは、Twitter において他のユーザのツイートを再投稿することである。本研究では、リツイートの有無を素性として用いる。

4. 実験

本節では、3 節で述べた提案手法の有効性を調べるための実験について述べる。4.1 節では、Twitter と論文との自動対応付けに用いるデータ、評価方法、および実験結果について述べる。4.2 節では、学会の発表論文に関するツイートのタイプ分類に用いるデータ、評価方法、および実験結果について述べる。

4.1. Twitter と論文との自動対応付け

実験に用いるデータと評価尺度

Twitter をハッシュタグごとにまとめられている Togetter¹から学会の発表論文についてのツイートを収集した。ハッシュタグとは、#記号と、半角英数字で構成される文字列のことである。ハッシュタグ付きのツイートは、検索画面などで一覧でき、同じイベントの参加者や同じ興味を持つ人の様々なツイートが収集しやすくなる。収集した学会の発表論文に関する 2,269 件のツイートデータのうち、4.2 節で述べるタイプ分類によって、type O に判定された学会の発表論文とは関係のないツイートを除いた全 783 件と、学会で発表された全 134 件の論文データを用いた。論文データの詳細を表 1 に示す。

表 1: 論文データ

学会名	データ数(件)
人工知能学会 2012	51
第 3 回楽天開発シンポジウム	7
言語処理学会 2012	29
地理情報システム学会 2011	47

評価尺度には、精度、再現率、F 値を用いた。

比較手法

本研究では、以下で述べる 4 種類の提案手法と 1 種類の比較手法で実験を行った。

● 提案手法

- ・ DP マッチング(ユーザ情報なし)：類似性尺度に DP マッチングを用いて類似度を測る。
- ・ コサイン類似度(ユーザ情報なし)：類似性尺度に コサイン類似度を用いて類似度を測る。
- ・ DP マッチング(ユーザ情報あり)：類似性尺度に DP マッチングを用い、類似度を測る際、類似度を測るときに用いるツイートから 20 分前後に投稿された、同一のユーザのツイート情報を用いる。
- ・ コサイン類似度(ユーザ情報あり)：類似性尺度に コサイン類似度を用い、類似度を測る際、類似度を測るときに用いるツイートから 20 分前後に投稿された、同一のユーザのツイート情報を用いる。

● ベースライン手法

- ・ ベースライン：ツイートの投稿時間と論文の発表時間を用いて、ツイートに対応付ける論文の候補を選定する。そして、論文の発表時間にあるツイートに対して対応付けを行う。

実験結果

Twitter と論文の対応付けの実験結果を表 2 に示す。

表 2: 対応付けの実験結果

	精度(%)	再現率(%)	F 値(%)
DP マッチング (ユーザ情報なし)	49.46	46.69	48.04
コサイン類似度 (ユーザ情報なし)	41.04	39.21	40.10
DP マッチング (ユーザ情報あり)	50.80	48.53	49.64
コサイン類似度 (ユーザ情報あり)	44.12	42.15	43.11
ベースライン	45.86	38.80	42.04

表 2 の結果から、比較手法と比べて、コサイン類似度(ユーザ情報なし)では若干低下したが、その他では、比較手法の F 値より高い値を得た。特に、DP(ユーザ情報あり)において、精度 4.94 ポイント、再現率 9.77 ポイント、F 値 7.60 ポイントの改善を行うことができた。

4.2. ツイートのタイプ分類

実験に用いるデータ

Togetter から収集した学会の発表論文に関する 2,269 件のツイートデータを用いる。ツイートデータの詳細を表 3 に示す。

¹ <http://togetter.com/>

表 3: ツイートデータ

学会名	ハッシュタグ	データ数(件)
人工知能学会 2012	#jsai2012	702
第 3 回楽天開発 シンポジウム	#rrds3	706
言語処理学会 2012	#nlp2012	505
地理情報システム 学会 2011	#jisa2011	356

次に、収集したツイートデータを、人手によりタイプ分類した結果を表 4 に示す。

表 4: ツイートのタイプごとの件数

分類タイプ	type S	type B	type D	type O
データ数(件)	274	348	304	1486

機械学習と評価尺度

機械学習には TinySVM を用いた。線形カーネルを使用し、2 分割交差検定を行った。評価尺度には精度、再現率、F 値を用いる。

比較手法

提案手法では、評価表現辞書、type S の手がかり語、type D の手がかり語、文タイプ、ツイートと論文の類似度、RT の有無、全単語の出現頻度を素性として機械学習を行う。ベースラインでは、全単語の出現頻度を素性として機械学習を行う。

実験結果

ツイートのタイプ分類の実験結果を表 5 に示す。

表 5: ツイートのタイプ分類の実験結果

	タイプ	精度(%)	再現率(%)	F 値(%)
ベース ライン	type S	60.04	47.02	52.74
	type B	63.59	49.96	55.96
	type D	51.06	34.83	41.41
	平均	58.23	43.93	50.04
提案 手法	type S	58.80	49.93	54.00
	type B	61.32	50.71	55.51
	type D	53.35	35.36	42.53
	平均	57.82	45.33	50.68

表 5 の結果から、比較手法に比べ、全タイプにおける平均の F 値が向上した。type D においては、精度 2.29 ポイント、再現率 0.57 ポイント向上した。

5. 考察

5.1. Twitter と論文との自動対応付け

次に、4.1 節の実験結果について考察する。システムが誤って抽出した例と抽出できなかった例について分析する。

システムが誤って抽出した例

人手では対応付けを行わなかったが、システムでは対応付けした例について考察する。例として、類似度を測る際、同じユーザの情報を追加した場合について

述べる。図 5 に同一ユーザのツイート例を示す。

1 : ハッシュタグだけでいい結果が出たら元も子もないな w #rrds3	2010/12/18 14:12:08
2 : ブログを中心とした観光情報の組織化 #rrds3	2010/12/18 14:12:59
3 : 旅行ブログエントリ中のリンクに注目。 #rrds3	2010/12/18 14:14:58

図 5: 同一ユーザのツイート例

図 5 より、文 1 では Twitter に関する論文について述べているのに対し、文 2 と 3 では、旅行ブログに関する論文について述べている。本研究では、同一論文について述べているツイートを情報として追加することを目的に同一ユーザのツイートの追加を行ったが、別の論文について述べているツイートが追加されている例が存在した。これは、図 5 の例では、旅行ブログに関する論文の発表が 14 時 10 分頃に開始されたが、前の論文発表について言及しているツイートもほぼ同時刻に投稿されたため、同じユーザのツイートとしてシステムが追加したと考えられる。これにより、類似度を測るときに用いるツイートとは別の論文について言及しているツイートが追加され、システムが誤って抽出したと考えられる。

システムが抽出できなかった例

次に、DP マッチングを用いた際に、人手では対応付けを行ったが、システムでは対応付けを行わなかった例について考察する。論文と類似度を比較すると、同一単語が存在するものの語順が異なるために、対応付けが行われなかった例が存在した。システムが対応付けを行わなかったツイートと論文の例を図 6 示す。

[ツイート]
「ゆったり」「ゆっくり」みたいな語もオノマトペなんですか？
[論文]
図 3 の頻度上位のオノマトペ表現は、大別して「しっくり」のように事物や行為の描写に用いられる場合と、「ゆっくり」などのように記述者の気持ちの記述に用いられる場合がある。
例えば「ゆったり」であれば「ベッド」や「部屋」などの設備が「ゆったりしていた」と用いられることがよくある。

図 6: 対応付けされなかったツイートと論文の例

図 6 のツイートと論文は、ともに下線部の「ゆっくり」や「ゆったり」という単語が出現しており、人手では類似していると判断した。しかし、DP マッチングでは語順を考慮するため、語順が異なる場合、類似度が小さくなるという問題点がある。その解決策としては、類似度を測る際、複数の類似性尺度を組み合わせ

て行うことで解決できると考えられる。

5.2. タイプ分類

4.2 節の実験結果について考察する。システムが誤って抽出した例と抽出できなかった例について分析する。例として type S の場合について考察する。

システムが誤って抽出した例

人手では分類を行わなかったが、システムでは分類した例について考察する。ツイートのタイプの判定誤りの原因として、type S において不要な手がかり語が存在したことがあげられる。

type S の手がかり語には、評価に関する語が多く存在する。しかし、情報利得により自動で収集したため、「幸せ」「えらい」といった論文の評価とは関係のない語も収集された。これにより、人手では type S でないと判定したツイートが、システムでは手がかり語が存在したため、type S と判定したと考えられる。

システムにより抽出できなかった例

次に、人手では分類を行ったが、システムでは分類を行わなかった例について考察する。ツイートのタイプの判定誤りの原因として、type S の手がかり語の不足があげられる。

type S の手がかり語として、「素晴らしい」や「面白い」といった、論文を評価する際に用いられる単語を素性として用いた。しかし、ツイートデータによっては、同じ「素晴らしい」という意味でも「スバラシイ」「すばらしい」など様々な記述が存在する。評価表現辞書では、「すばらしい」と「スバラシイ」、「素晴らしい」は、同じ単語の代表の表現である「素晴らしい」に変換されているため、「素晴らしい」しか登録されておらず、情報利得により収集した type S の手がかり語においても、「スバラシイ」は登録されていない。これにより、同じ意味でも表記の違う単語は評価語として扱われなかったため、抽出できなかったと考えられる。

6. おわりに

本研究では、Twitter と学会の発表論文を用いて、Twitter と論文との自動対応付けの手法と、学会に関するツイートのタイプを自動分類する手法を提案した。Twitter と論文との自動対応付けでは、類似度に基づく手法を提案した。また、類似度を測る際、同一ユーザの情報を追加した対応付けも行った。実験の結果、DP マッチングのユーザ情報を追加した場合で比較手法に比べて精度が 4.94 ポイント、再現率が 9.77 ポイント向上し提案手法の有効性を示した。タイプ分類では、type S、type B、type D、type O の 4 つのタイプに、機械学習を用いて自動分類を行った。実験の結果、提案手法が比較手法に比べて結果が向上し提案手法の有効性を示すことができた。

文 献

- [1] V. Qazvinian, D.R. Radev, S.M. Mohammad, B. Dorr, D. Zajic, M. Whidby, T. Moon, “Generating Extractive Summaries of Scientific Paradigms” *Journal of Artificial Intelligence Research*, Vol.46, pp.165-201, 2013.
- [2] 越智洋司, 中筋浩之, “プレゼンテーションと連動したモバイル型レスポンスシステムの開発” *ARG-WI2*, No.1, pp.11-14, 2012.
- [3] 坂東慶太, “Altmetrics の可能性 ソーシャルメディアを活用した研究評価指標” *情報管理*, Vol.55, No.9, pp.638-646, 2012.
- [4] E. Garfield, “Citation Indexes to Science: A New Dimension in Documentation Thought the Association of Ideas” *Science*, No.122, pp.108-111, 1995.
- [5] D. Yogatama, M. Heilman, B. O’Connor, C. Dyer, “Predicting a Scientific Community’s Response to an Article” *Proceedings of EMNLP ’11*, pp.594-604, 2011.
- [6] J. Long, Y. Mo, Z. Ming, L. Xiaohua, Z. Tiejun, “Target-dependent Twitter Sentiment Classification, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics” pp.151-160, 2011.
- [7] L. Barbosa, J. Fenq, “Robust Sentiment Detection on Twitter from Biased and Noisy Data” *Proceedings of COLING ’10*, pp.36-44, 2010.
- [8] 鍛冶伸裕, 喜連川優, “自動構築した評価文コーパスからの評価表現辞書の構築,” *日本データベース学会 Letters*, Vol.6, No.1, pp.41-44, 2007.
- [9] 加藤大智, NGUYENMANH CUONG, 橋本泰一, 横田治夫, “論文のラベル付きクラスタリングのための情報利得を用いたキーワード選定” *DEIM Forum 2012*, E10-1, 2012.
- [10] 横山憲司, 難波英嗣, 奥村学, “Support Vector Machine を用いた談話構造解析” *情報処理学会研究報告.自然言語処理*, 2003(23), pp.193-200, 2003.