

# SNS を用いた迅速な論文の重要度の評価

中橋 光† 難波 英嗣† 竹澤 寿幸†

†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 3-4-1

E-mail: †{nakahashi, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

**あらまし** 学術論文や学術雑誌の研究への影響や重要度の評価は、一般的に、インパクトファクターのような引用関係を用いて行われていた。しかし、これらの引用に基づく手法は、論文の重要度を評価するまでに時間を要するという問題がある。この問題を解決するために、本研究では、Twitter を用いて論文の重要度を評価する。近年、学術会議の参加者が、Twitter を利用して、発表論文に対して意見や評価などを投稿するケースが増えている。このようなツイートが多い発表論文は、その学術会議において強い影響力を持っていると考えられる。本研究は3つのステップ、(1) 有益なツイートの自動分類、(2) ツイートと論文との自動対応付け、(3) 有益なツイートに基づく論文の重要度の評価から構成される。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、ツイートの自動分類では、再現率 0.591, 精度 0.588 を、自動対応付けでは、再現率 0.483, 精度 0.525 を、論文の重要度の評価では、 $\epsilon$  値 0.236 を、それぞれ得た。

**キーワード** Twitter, 論文, 機械学習, DP マッチング, ランキング

## 1. はじめに

本研究では、迅速に論文の重要度を評価する手法を提案する。学術論文や学術雑誌の研究への影響や重要度の評価は、一般的に、インパクトファクター[1]のような引用関係を用いて行われていた。しかし、これらの引用に基づく手法は、論文の重要度を評価するまでに時間を要するという問題がある。例えば、ある年の学術雑誌の重要度を評価する際、インパクトファクターを用いた場合、過去2年間の被引用数から計算される。このため、インパクトファクターは、新しく発表された論文の重要度を評価することができない。この問題を解決するために、本研究では、SNS(ソーシャルネットワークサービス)を用いて論文の重要度を評価する。

本研究では、SNSの中でも特にTwitterに着目する。近年、学術会議に参加する研究者が、Twitter という媒体を利用して、発表論文に対して意見や評価、質問などを投稿するケースが増えている。このようなツイートが多い発表論文は、ユーザからの関心が高く、学術的に意味深いと考えられる。また、有益なツイートが多い発表論文は、その学術会議において強い影響力を持っていると考えられる。そこで本研究では、様々な学術会議で発表された論文に関する有益なツイートを収集し、それらを発表論文と対応付けることで、論文の重要度を評価する。

先行研究として、中橋ら[2]はツイートと論文を自動的に対応付けることで、論文の読解支援を行った。本研究では、対応付けにおいて論文の節との類似度を計測することで対応付けの性能を向上させ、さらに、論文の重要度を評価することで、迅速に重要論文を見つけることが可能になる。

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 を用いて論文の重要度を自動評価する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と結果についての考察を述べ、5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。論文の重要度を評価するために、これまでも数多くの手法が提案されている。Garfield[1]は、1節で述べたインパクトファクターに加えて引用半減期を提案している。引用半減期は、現在から遡って、学術雑誌に掲載された論文の被引用回数が被引用総数の中央値となるまでの年数により重要度を評価する。しかし、インパクトファクターと引用半減期は、論文の重要度を評価するために長時間を要するという問題点がある。

この問題を解決するために、いくつかの研究が行われている。最も初期の研究プロジェクトとして、KDD Cup 2003<sup>1</sup>がある。本プロジェクトでは、以下の2つの手法が行われた。

- (1) 被引用数予測 : 長い期間をかけてよく引用された論文の被引用数の変化を予測する。
- (2) ダウンロード数推定 : arXiv<sup>2</sup>にて、論文が出版後2ヶ月間にダウンロードされた数を推定する。

被引用数やダウンロード数を推定することにより、より迅速に論文の重要度を評価することが期待することができる。しかし、実験の結果あまりよい結果が得

<sup>1</sup> <http://www.cs.cornell.edu/projects/kddcup/>

<sup>2</sup> <http://arxiv.org>

られなかったことが報告されている<sup>3</sup>。Yogatamaら[3]は、異なるデータセットで同様の予測の実験を行ったが、実験の結果、あまりよい結果が得られなかった。また、本研究では、Yogatamaらの研究をベースラインにすることで、従来手法による論文のランキング結果と提案手法によるランキング結果を比較し、Twitterを用いた論文の重要度の評価の有効性を検証する実験を考えた。しかし、本研究では、学会の発表論文をデータとして用いており、データベースに論文が蓄積されていないため、被引用数やダウンロード数の計測が難しく比較実験を行うことはできない。また、Vaughanら[4]とKoushaら[5]は、Web上の文献からの引用であるWeb引用を用いる手法を提案した。Vaughanらは、論文についてのWebでのヒット数をもとに、Web引用の分析を行い、Koushaらは、Web引用に加え、Google Scholarを用いた論文の被引用数と論文のタイトルやオープンアクセス論文のURLのGoogle検索におけるヒット数をもとに、分析を行った。分析の結果、共にISI<sup>4</sup>による引用との間に強い相関があることを報告した。しかし、Web引用は、従来の引用関係よりは即時性に優れているものの、数を記録するまでに一定時間を要する。

そこで、近年、SNSを用いた論文の重要度を評価する手法に注目が集まり、いくつかの研究が行われている。Eysenbachら[6]は、Twitterに掲載されているツイートに基づき、Twitterを用いた引用の予測について分析した。結果、ツイートは論文発表後3日以内に高被引用論文を予測できる、Twitterは新しく発表された学術論文を見つけることができる、ツイート数は論文の重要度を測るうえで重要な尺度である、という結論を述べている。Wellerら[7]は、学会中に投稿された発表論文に関するツイートを収集し、論文へのURLを含むツイート数とリツイート数を用いて、論文の引用を分析する手法を提案した。

また、SNSを用いた論文の重要度を評価する手法の中でも、引用関係に変わる新たな研究指標である“Altmetrics<sup>5</sup>”が近年注目されている。Altmetricsとは、TwitterやFacebookなどのソーシャルメディア・プラットフォームでの言及数に基づき、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[8]。Altmetricsを計測する代表的なサービスとして、Altmetrics.com<sup>6</sup>やImpactStory<sup>7</sup>がある。また、日本語論文を対象としたサービスとし

ては、吉田[9]が開発したCeek.jp Altmetrics<sup>8</sup>がある。Ceek.jp Altmetricsでは、主に2つの機能を提供している。1つ目の機能は、学術文献をランキングする機能であり、これにより、現在旬な学術文献情報を提供することができる。2つ目の機能は、各文献の言及情報の詳細であり、これにより、いつ流行したのか、どれほどの期間流行したのかを容易に知ることができる。Zahediら[10]は、ImpactStoryと引用手法との比較を行い、Altmetricsのスコアが高い論文やジャーナルは重要度が高い可能性が大きいと結論づけた。Altmetricsは、ソーシャルメディア・プラットフォームの言及数に基づいて評価しているため、ツイートの内容について考慮されていない。本研究では、論文やプレゼンテーションへのツイート内容を分析し、ツイートが論文を評価するうえで有益であるかを分類し、有益であるツイートのみを用いて評価している点で異なる。また、Altmetricsは、全論文を評価の対象としているのに対し、本研究では、学会の発表論文を用いて学会ごとに論文を評価している点で異なる。

### 3. Twitterを用いた論文の重要度の自動評価

本節では、Twitterを用いて論文の重要度を自動評価する手法について述べる。本研究では、以下の3種類の手法を提案する。

- (1) 有益なツイートの自動分類
- (2) ツイートと論文との自動対応付け
- (3) 有益なツイートに基づく論文の重要度の評価

論文の重要度を評価する流れとして、まず有益なツイートを自動分類する。その後、有益なツイートと論文を自動対応付けし、対応づいた有益なツイート数に基づき論文の重要度を評価する。3.1節では、有益なツイートを自動分類する手法について述べる。3.2節では、ツイートと論文とを自動対応付けする手法について述べる。3.3節では、有益なツイートに基づき論文の重要度を評価する手法について述べる。

#### 3.1. 有益なツイートの自動分類

本節では、有益なツイートを自動分類する手法について述べる。ツイートの中には、意見や評判など論文の評価に役立つものの他に、発表の様子を実況中継する、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、論文の評価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、有益なツイートの自動分類を行う。3.1.1節では、有益なツイートの定義について述べる。3.1.2節では、手法の詳細について述べる。

<sup>3</sup> <http://www.cs.cornell.edu/projects/kddcup/download/KDDCup-Task3.ppt>

<sup>4</sup> Institute for Science Information, Thomson Reutersにより提供されている世界最大の学術文献引用索引

<sup>5</sup> <http://altmetrics.org/manifesto/>

<sup>6</sup> <http://www.altmetric.com/>

<sup>7</sup> <https://impactstory.org/>

<sup>8</sup> <http://altmetrics.ceek.jp/>

### 3.1.1. 有益なツイートの定義

有益または有益ではないツイート例を図 1 に示す。図 1 のように、論文を評価しているツイート(ツイート(1)), 論文について質問しているツイート(ツイート(2)), 質問への返信が含まれているとみなされるツイート(ツイート(3)), 論文に対して意見を述べているツイート(ツイート(4)), を有益なツイートとして分類する。また, URL は, 一般に関連論文や関連 Web サイトへのリンクが多いため, URL を含むツイートも有益なツイートとして分類する。

一方で, 論文とは関係のないツイート(ツイート(5)), 論文の書誌情報や研究発表について実況中継しているツイート(ツイート(6)), リツイート(“RT”から始まるツイート)(ツイート(7)), を有益ではないツイートとして分類する。

- |   |
|---|
| (有益なツイート)   |
| (1) フォントとオノマトペの関係は面白いなあ                                       |
| (2) 「ゆったり」「ゆっくり」みたいな語もオノマトペなんですか？                             |
| (3) @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! |
| (4) バックボーンとなる理論を組み上げることが必要だと思うのだがどうか。                         |
| (有益ではないツイート)  |
| (5) 徒歩 15 分の仕掛学セッションに来た。                                      |
| (6) 「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                |
| (7) RT @nanaya_sac オノマトペ満載のあのレシピ本は異端なのか。                      |

図 1: 有益または有益ではないツイート例

### 3.1.2. 有益なツイートの自動分類手法

本研究では, 機械学習により有益なツイートの自動分類を行う。機械学習には, SVM(Support Vector Machine)を用いる。機械学習に用いた素性を以下に述べる。

#### 評価表現辞書

本研究では, 各ツイート内の単語が評価表現辞書[11]に含まれる場合, 評価極性値を正規化した値を素性として用いる。この辞書には, 形容詞, 形容詞句と評価極性値のペアが約 10,000 組登録されており, 評価極性値とは, 形容詞・形容詞句の評価極性の強さを数値化したものである。

#### 手がかり語

本研究では, 素性として“各ツイート内の単語が収集した手掛かり語に含まれる出現頻度”を用いる。手がかり語には, 情報利得[12]の値が上位 100 件の単語を用いた。

#### ツイートと論文の類似度

実況中継しているツイートには, 図 1 のツイート(7)

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。これらのツイートは, その論文の評価にはあまり貢献しないツイートであると考えられる。そこで, 論文とツイート間の類似度を計測し, その値を素性として用いる。類似度の計算は, DP マッチング[13]を用いてツイートと論文の文を整列化し, それらの編集距離を測ることで求める。

#### 文タイプ

文タイプとは, 文末表現により決まる文のタイプのことであり, 横山ら[14]が人手で作成した対応規則を用いて決定された。論文に関して意見を述べる際, 図 1 のツイート(4)の「どうか」など, 特徴的な文末表現が用いられることが多い。そこで, 素性として“各ツイート内の単語が文タイプの文末表現に含まれるかどうか”を用いる。

上記に加え, URL の有無, ツイートの長さ, 論文のタイトルの有無, 顔文字の有無, 疑問符の有無, 全単語を素性として用いる。なお, 評価表現辞書, 手がかり語, 全単語において使用する単語は, MeCab<sup>9</sup>により分割された形態素とし, 品詞が名詞, 動詞, 形容詞であるものとする。

### 3.2. ツイートと論文との自動対応付け

対応付け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。ここで, 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. 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, 編集距離(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。

対応付けの例を図 2 に示す。図 2 では, 10 時 30 分に投稿されたツイートと 10 時から 10 時 20 分に発表されている論文 A, 10 時 20 分から 10 時 40 分に発表されている論文 B, 10 時 40 分から 11 時に発表されている論文 C を対応付けする。まず, 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ため, 論文 A と論文 B が対応付けの候補として選出される。その後, 類似度を算出し, 最も類似度が高い論文に対して対応付けするため, ツイートは類似度が 0.89 で最も高い論文 B に対応付けされる。

<sup>9</sup> <http://mecab.sourceforge.net>

表 1: データの詳細

学会名	ハッシュタグ	ツイート数	論文数
人工知能学会 2010	#jsai2010	458	17
人工知能学会 2012	#jsai2012	600	51
人工知能学会 2014	#jsai2014	246	4
第 3 回楽楽開発シンポジウム	#rrds3	706	7
言語処理学会 2012 第 18 回年次大会	#nlp2012	268	29
地理情報システム学会 2011	#gisa2011	296	47
第 1 回 Web インテリジェンスとインタラクション研究会	#sigwi2	195	11
第 2 回 Web インテリジェンスとインタラクション研究会	#sigwi2	241	16
第 3 回 Web インテリジェンスとインタラクション研究会	#sigwi2	75	8
第 4 回 Web インテリジェンスとインタラクション研究会	#sigwi2	436	11
第 4 回データ工学と情報マネジメントに関するフォーラム	#DEIM2012	754	63
第 5 回データ工学と情報マネジメントに関するフォーラム	#DEIM2013	295	15
第 6 回データ工学と情報マネジメントに関するフォーラム	#DEIM2014	155	9

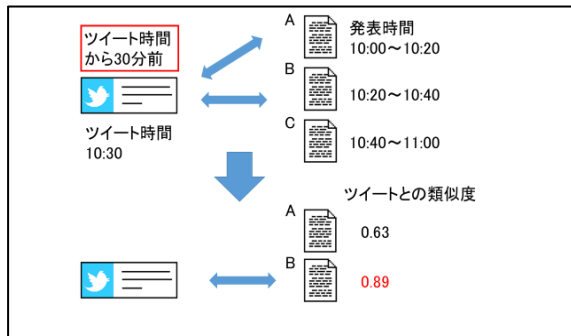


図 2: 対応付けの流れ

対応付けにおいて、以下の 2 点を考慮する。

#### ユーザ情報(User)

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、1 件のツイートからでは十分な情報が得られない可能性がある。そこで本研究では、類似度を測る際に、対象のツイートから、前後 20 分以内に投稿された同一ユーザによる全てのツイートを用いて類似度を測る。

#### 論文の分割(Sec)

ツイートには、論文全体について言及している場合や、手法や関連研究など論文のある一部分について言及している場合がある。特にツイートは最大で 140 字であるため、後者の場合が多い。そこで、ツイートと論文の節との類似度を測る。

### 3.3. 有益なツイートに基づく論文の重要度の評価

本研究では、各学会における重要論文を自動的に決定する手法として、単純な順位アプローチを用いている。有益なツイートに基づく論文の重要度を評価する手法の流れを以下に示す。

- 3.2 節で述べた手順により対応付けした結果を用いて、各論文の自動で対応づいた有益なツイート数を計測する。

- 有益なツイート数に基づき、各論文をランキングする。

## 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。Twitter を用いて論文の重要度を自動評価する手法として、以下の 3 種類の実験を行った。データセットについては 4.1 節、(1)の実験については 4.2 節、(2)の実験については 4.3 節、(4)の実験については 4.4 節で述べる。

- 有益なツイートの自動分類
- ツイートと論文との自動対応付け
- 有益なツイートに基づく論文の重要度の評価

### 4.1. データセット

本研究では、Togetter<sup>10</sup>から学会の発表論文についてのツイートを収集した。各学会には、特定のハッシュタグが存在するため、データには、各学会のハッシュタグ付きツイートを用いた。最終的に、本研究では、13学会で全 4,693 件のツイートを収集した。また、様々な学会で発表された全 291 件の論文を収集した。ツイートデータおよび論文データの詳細を表 1 に示す。また、3.1.1 節で述べた有益なツイートの基準に従って、収集したツイートを人手で分類した。その結果、4,693 件のツイートのうち 840 件のツイートが有益なものであった。

### 4.2. 有益なツイートの自動分類

#### 4.2.1. 実験方法

本節では、有益なツイートを自動分類するための実験について述べる。本研究では、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 値を用いた。データには、表 1 に示した 4,693 件のツイ

<sup>10</sup> <http://togetter.com/>

トと 291 件の論文を用いた。比較手法には、以下に示す提案手法とベースラインを用いた。なお、全単語に用いる単語として、動詞、形容詞、名詞を用いた。

- **提案手法**：機械学習の素性として、3.1.2 節で述べた素性を用いる。
- **ベースライン**：機械学習の素性として、全単語のみを用いる。

#### 4.2.2. 実験結果

実験結果を表 2 に示す。表 2 の結果から、ベースラインに比べ、F 値において 0.032 向上した。特に、再現率において 0.057 向上した。また、マクネマー検定によって提案手法とベースラインの結果を比較した結果、有意水準 0.01 において有意な差が得られた。これにより、提案手法の有効性を示すことができた。

表 2：有益なツイートの自動分類の実験結果

	精度	再現率	F 値
提案手法	<b>0.588</b>	<b>0.591</b>	<b>0.589</b>
ベースライン	0.581	0.534	0.557

#### 4.2.3. 考察

4.2.2 節の実験において、再現率は向上したものの、精度についてはあまり向上が見られなかった。そこで、誤って有益なツイートと分類してしまった例について考察を行う。誤って分類してしまった原因として、実況中継に関するツイートに、提案した手がかり語や疑問符、URL が含まれていたことが挙げられる。図 3 に、提案手法により有益と判断された実況中継に関するツイート例を示す。

(有益と判断された実況中継に関するツイート)

(1) A 会場。自治体職員が、「システムが使えた」というユーザーエクスペリエンスを獲得できれば定着化がすすむ。(佐藤)

(2) 大阪の人はオノマトペをより使うのではないかと? という仮説。

(3) 恐怖のデータ量。119 人×310 語×40 語  
<http://t.co/ScYjDOZe>

図 3：有益と判断された実況中継に関するツイート例

ツイート(1)は、本研究で提案した手がかり語「会場」が含まれているために、誤って有益と判断された。この「会場」は、評価や意見を述べる単語ではないため、不要な手がかり語である。本研究では、情報利得により自動で手がかり語を収集したため、このような不要な手がかり語が複数含まれている。この問題を解決するために、事前に不要語リストを作成しておくことなどが挙げられる。ツイート(2)は、疑問符が含まれているために、誤って有益と判断された。実況中継のツイートは、発表者のコメントをそのままツイートすることが多いため、このような疑問符を含むツイートが複数存在した。ツイート(3)は、URL が含まれているた

めに、誤って有益と判断された。実況中継のツイートには、特に Ustream<sup>11</sup> の URL が多く存在した。これは、学会によっては、学会の様子を Ustream で配信されることがあることが原因であると考えられる。

### 4.3. ツイートと論文との自動対応付け

#### 4.3.1. 実験方法

本節では、ツイートと論文との自動対応付けするための実験方法について述べる。評価尺度には精度、再現率、F 値を用いた。データには、人手で有益なツイートと判定した 840 件のツイートと 291 件の論文を用いた。比較手法には、以下に示す 8 種類の提案手法とベースラインを用いた。

##### 提案手法

- **DP**：類似性尺度として編集距離(DP)を用いる。
- **DP+User**：編集距離(DP)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、同一のユーザの全ツイートを用いる。
- **DP+Sec**：編集距離(DP)を用い、ツイートと論文の節との類似度を測る。
- **DP+User+Sec**：編集距離(DP)とユーザ情報を用い、ツイートと論文の節との類似度を測る。
- **Cos**：類似性尺度としてコサイン類似度を用いる。
- **Cos+User**：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、同一のユーザの全ツイートを用いる。
- **Cos+Sec**：コサイン類似度を用い、ツイートと論文の節との類似度を測る。
- **Cos+User+Sec**：コサイン類似度とユーザ情報を用い、ツイートと論文の節との類似度を測る。

##### ベースライン

- **ベースライン**：時間情報を用い、ツイートの投稿時間に発表している論文に対して対応付けする。

#### 4.3.2. 実験結果

実験結果を表 3 に示す。表 3 の結果から、編集距離(DP)を用いた全ての手法において、ベースラインより F 値が向上した。特に、DP+Sec 手法において F 値 0.092 向上した。また、マクネマー検定によって編集距離(DP)を用いた全ての手法とベースラインの結果を比較した結果、有意水準 0.01 において有意な差が得られた。これにより、編集距離(DP)を用いた DP, DP+User, DP+Sec, DP+User+Sec 手法の有効性を示すことができた。一方で、コサイン類似度を用いた手法では、Cos+User+Sec 手法においてはベースラインより F 値が向上したものの、その他の手法では、ベースライン

<sup>11</sup> <http://www.ustream.tv/>

より F 値が低下する結果となった。この結果から、コサイン類似度に比べ編集距離の方が、ツイートと論文の類似度を測る類似性尺度として優れていることが分かった。

表 3：ツイートと論文との自動対応付けの実験結果

	精度	再現率	F 値
DP	0.492	0.453	0.472
DP+User	0.477	0.438	0.456
DP+Sec	<b>0.525</b>	<b>0.483</b>	<b>0.503</b>
DP+User+Sec	0.514	0.473	0.493
Cos	0.406	0.375	0.390
Cos+User	0.392	0.361	0.396
Cos+Sec	0.413	0.380	0.396
Cos+User+Sec	0.461	0.425	0.442
ベースライン	0.463	0.370	0.411

### 4.3.3. 考察

誤って対応付けしてしまった例について分析する。誤って対応付けた原因として、同一セッションに類似した論文が存在していたことが挙げられる。図 4 に例を示す。

<p>[ツイート]            専門家が書いた<u>レシピ</u>は<u>オノマトペ</u>が少ない。。なるほど。</p>
<p>[論文 1]            2. <u>オノマトペ</u>によるランキング表示            オノマトペロリでは、料理<u>レシピ</u>と<u>オノマトペ</u>の適合度を求め、適合度を用いて以下のランキングを行っている。</p>
<p>[論文 2]            料理とスイーツの違いが現れる<u>オノマトペ</u>や、類似した<u>オノマトペ</u>でも微妙に関連する料理が異なることが分る、というような分析のヒントが得られた。</p>

図 4：誤って対応付けした例

図 4 の 2 つの論文やツイートはいずれもオノマトペに関する論文やツイートである。この二つの論文の内容は非常に類似しているおり、例えば「オノマトペ」といった表現が多く存在する。一方、ツイートにおける重要単語は「専門家」、「レシピ」、「オノマトペ」である。そのため、本研究の手法では、これらを正しく対応付けすることができなかつたと考えられる。この問題を解決するために、本研究では、ツイートの評価語と感嘆表現に着目する。例えば、学会発表中の論文に対するツイートは、「面白い」「素晴らしい」といった評価表現や「だなぁ」「そう」といった感嘆表現が用いられることが多い。そこで、ツイートが投稿時間に発表されている論文に対するツイートであるかどうかを識別することにより解決することができると考えら

れる。

## 4.4. 有益なツイートに基づく論文の重要度の評価

### 4.4.1. 実験方法

本節では、有益なツイートに基づく論文の重要度の評価についての実験方法について述べる。本研究では、Sidiropoulos ら [15] が用いた評価方法を適用した。Sidiropoulos らは、論文のランキング結果を評価する際に、学会における受賞論文を重要論文と仮定し、受賞論文の順位で評価を行った。そこで、本研究においても、各学会における受賞論文を重要論文と仮定し、受賞論文の順位を評価した。評価尺度には、式(1)に示す  $\epsilon$  という尺度を用いる [16]。この尺度は、質問応答システムの評価で一般的に用いられる MRR (Mean Reciprocal Rank) を拡張したものであり、各学会の受賞論文がランキング上位であれば、1 に近い評価値が得られる。

$$\epsilon = \frac{\sum_{i \in R_i} \frac{1}{i}}{\sum_{j \in \{1, 2, \dots, n\}} \frac{1}{j}} \quad (1)$$

ここで、 $n$  は各学会の各賞における受賞論文数、 $R$  は各学会の各賞における受賞論文の順位の集合、 $i$  は各学会の各賞における受賞論文の順位である。データには、極端に有益なツイート数が少ない学会、全ての受賞論文が一度もツイートされていない学会を除いた 11 学会に関する 796 件のツイートと 240 件の論文を用いた。比較手法には、以下に示す提案手法と 2 種類のベースラインを用いた。なお、3.3 節の手順 1 における対応付けの結果には、4.3.2 節の実験結果で最も良い結果である DP+Sec を用いた。

#### 提案手法

- **提案手法**：機械学習に 3.1.2 節で述べた素性を用いて有益なツイートの自動分類を行い、編集距離 (DP) を用いて論文の節との類似度を測る DP+Sec 手法で対応付けを行い、その結果を用いてランキングする。

#### ベースライン

- **ベースライン 1**：有益なツイートの自動分類を行わず、全ツイートを用いて対応付けを行い、その結果を用いてランキングする。
- **ベースライン 2**：機械学習に 3.1.2 節で述べた素性を用いて有益なツイートの自動分類を行い、ツイートの投稿時間と論文の発表時間を用いて対応付けを行い、その結果を用いてランキングする。

### 4.4.2. 実験結果

実験結果を表 4 に示す。表 4 の結果から、提案手法がベースラインを上回ることに成功した。これにより、本研究で提案した論文の重要度を評価する手法が、重

表 5：各学会の受賞論文の順位

	最優秀論文賞	優秀論文賞	若手奨励賞	萌芽研究賞	最優秀インタラクティブ賞	優秀インタラクティブ賞
人工知能学会 2010(17)		12 位 (1/13)				
人工知能学会 2012(51)		22 位 (1/9)				
第 3 回楽天開発シンポジウム (7)	5 位 (1/1)	5 位 (1/1)				
言語処理学会第 18 回年次大会 (31)	5 位 (1/1)	(0/4)	1 位, 15 位 (2/5)			
第 1 回 Web インテリジェンスとインタラクショナル研究会 (11)		9 位 (1/1)	(0/1)	5 位 (1/1)		
第 2 回 Web インテリジェンスとインタラクショナル研究会 (16)		11 位 (1/1)	14 位 (1/1)	1 位, 5 位 (2/2)		
第 3 回 Web インテリジェンスとインタラクショナル研究会 (8)		6 位 (1/1)	(0/1)	1 位 (1/1)		
第 4 回 Web インテリジェンスとインタラクショナル研究会 (11)		1 位 (1/1)	(0/1)	7 位 (1/1)		
第 4 回データ工学と情報マネジメントに関するフォーラム (63)	(0/1)	22 位 (1/1)	1 位, 30 位, 46 位 (3/15)		13 位, 46 位 (2/6)	(0/6)
第 5 回データ工学と情報マネジメントに関するフォーラム (15)	(0/1)	(0/1)	3 位, 5 位, 12 位 (3/30)		(0/2)	1 位 (1/13)
第 6 回データ工学と情報マネジメントに関するフォーラム (9)	(0/2)	1 位 (1/4)	(0/51)		(0/3)	(0/10)

要論文を見つけることに有効であることを示すことができた。

表 4：論文の重要度の評価の実験結果

	$\varepsilon$
提案手法	<b>0.236</b>
ベースライン 1	0.183
ベースライン 2	0.164

#### 4.4.3. 考察

各学会の受賞論文の順位を表 5 に示す。なお、表 5 において、斜線はその学会にその賞が存在しないもの、空欄は受賞論文に関するツイートが 1 件もなかったもの、ランキングが複数存在するものは、その賞が複数人存在するもの、1 列目の括弧は、その学会における論文数、2 列目以降の括弧は、その学会における賞の数を表す。

表 5 の結果より、低い順位になっているものも存在するが、比較的受賞論文がランキング上位に来る傾向があることが分かった。また、セッション数が一つで投稿論文数が少ない Web インテリジェンスとインタラクショナル研究会では、大多数の受賞論文が一度はツイートされ、ランキング上位に来る傾向があった。これにより、セッションが少ない学会では、全ての論文についてツイートが行われるため、正確なランキング結果が得られ、より Twitter を用いる論文の重要度の評価が有効であると考えられる。

一方で、一度もツイートされていない受賞論文が多く存在する例が数多く見られた。特に、第 5 回データ工学と情報マネジメントに関するフォーラムの若手奨励賞では、30 件の受賞論文のうち 3 件しかランクインされていなかった。これは、データ工学と情報マネジメントに関するフォーラムの若手奨励賞は、各セッ

ョンで受賞論文が決定しているが、ツイートが頻繁行われるセッションは限られていることが原因であると考えられる。また、投稿論文数が多い学会である、人工知能学会、データ工学と情報マネジメントに関するフォーラムでは、セッション数が多いこともあり、あまり良い結果が得られなかった。この結果から、セッション数や投稿論文数が多い学会では、ランキングの精度は高いものの、全ての受賞論文を網羅することは難しいと考えられる。

## 5. おわりに

本研究では、Twitter を用いた論文の重要度の自動評価を行った。そのために、まず、有益なツイートを自動検出するために、有益なツイートを自動分類する手法を提案した。有益なツイートの自動分類は、機械学習により行った。次に、ツイートと論文を自動的に対応付けする手法を提案した。ツイートと論文との自動対応付けは、ツイートと論文の類似度を測ることで行った。類似性尺度には、編集距離とコサイン類似度を用いた。最後に、各学会における重要論文を自動的に決定するために、有益なツイートに基づく論文の重要度の評価する手法を提案した。重要論文の自動決定は、単純な順位アプローチを用いて行った。3種類の手法の有効性を調べるために、291件の様々な学会の発表論文と4,693件の学会の発表論文に関するツイートを用いて、実験を行った。実験の結果、有益なツイートの自動分類では、再現率 0.591、精度 0.588 を、ツイートと論文との自動対応付けでは、DP+Sec 手法において再現率 0.483、精度 0.525 を、有益なツイートに基づく論文の重要度の評価では、MRR を拡張した尺度である  $\epsilon$  において 0.236 の値がそれぞれ得られ、提案手法の有効性を示せた。

## 参 考 文 献

- [1] Garfield, E., "Citation Indexes for Science: A New Dimension in Documentation Thought the Association of Ideas", *Science*, No.122, pp.108-111, 1955.
- [2] 中橋光, 難波英嗣, 竹澤寿幸, 高須淳宏, "Twitter と論文との自動対応付け", *電子情報通信学会技術報告*, データ工学, Vol.113, No.105, pp.65-70, 2013.
- [3] Yogatama, D., Heilman, M., O'Connor, B. and Dyer, C., "Predicting a Scientific Community's Response to an Article", *Proc. of EMNLP '11*, pp.594-604, 2011.
- [4] Vaughan, L. and Shaw, D., "Web Citation Data for Impact Assessment: A Comparison of Four Science Disciplines", *Journal of the American Society for Information Science and Technology*, pp.1075-1087, 2005.
- [5] Kousha, K. and Thelwall M., "Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis", *Journal of the American Society for Information Science and Technology*, Vol.58, Issue.7, pp.1055-1065, 2007.
- [6] Eysenbach, G., "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact", *Journal of Medical Internet Research*, Vol.13, No.4, 2011.
- [7] Weller, K., Dröge, E. and Puschmann, C., "Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences", *Proc. of the ESW2011 Workshop on Making Sense of Microposts*, pp.1-12, 2011.
- [8] 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.
- [9] 吉田光男, "計量書誌学の新たな挑戦 -国産オルトメトリクス計測サービスの開発-", *情報の科学と技術*, Vol.64, No.12, pp.501-507, 2014.
- [10] Zahedi, Z., Costas, R. and Wouters, P., "How well Developed are Altmetrics? A Cross-Disciplinary Analysis of the Presence of 'Alternative Metrics' in Scientific Publications", *Journal of the American Society for Information Science and Technology*, Vol.58, Issue.7, pp.1055-1065, 2007.
- [11] Kaji, N. and Kitsuregawa, M., "Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents", *Proc. of the Conference on Empirical Methods in Natural Language Processing*, pp.1075-1083, 2007.
- [12] 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", 第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012), 2012.
- [13] Su, K.-Y., Wu, M.-W. and Chang, J.-S., "A New Quantitative Quality Measure for Machine Translation Systems", *Proc. of the 14th Conference on Computational Linguistics*, Vol.2, pp.433-439, 1992.
- [14] 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.
- [15] Sidiropoulos, A. and Manolopoulos, Y., "A Citation-Based System to Assist Prize Awarding", *Proc. of the ACM SIGMOD Record*, Vol.34, Issue.4, pp.54-60, 2005.
- [16] 難波英嗣, 釜屋英昭, 竹澤寿幸, 奥村学, 新森昭宏, 谷川秀和, "論文用語の特許用語への自動変換", *情報処理学会論文誌*. データベース, Vol.2, No.1, pp.81-92, 2009.