

Quick Evaluation of Research Impacts at Conferences using SNS

Satoshi Fukuda, Hikaru Nakahashi,
Hidetsugu Nanba, Toshiyuki Takezawa

Graduate School of Information Sciences,
Hiroshima City University, Hiroshima, Japan

Background

- Traditionally, the research impact (importance) of academic journals has been evaluated using **citation relations**.

- Evaluation measure using citation relation

- Impact factor [Garfield 1955]

y Impact factor = A/B

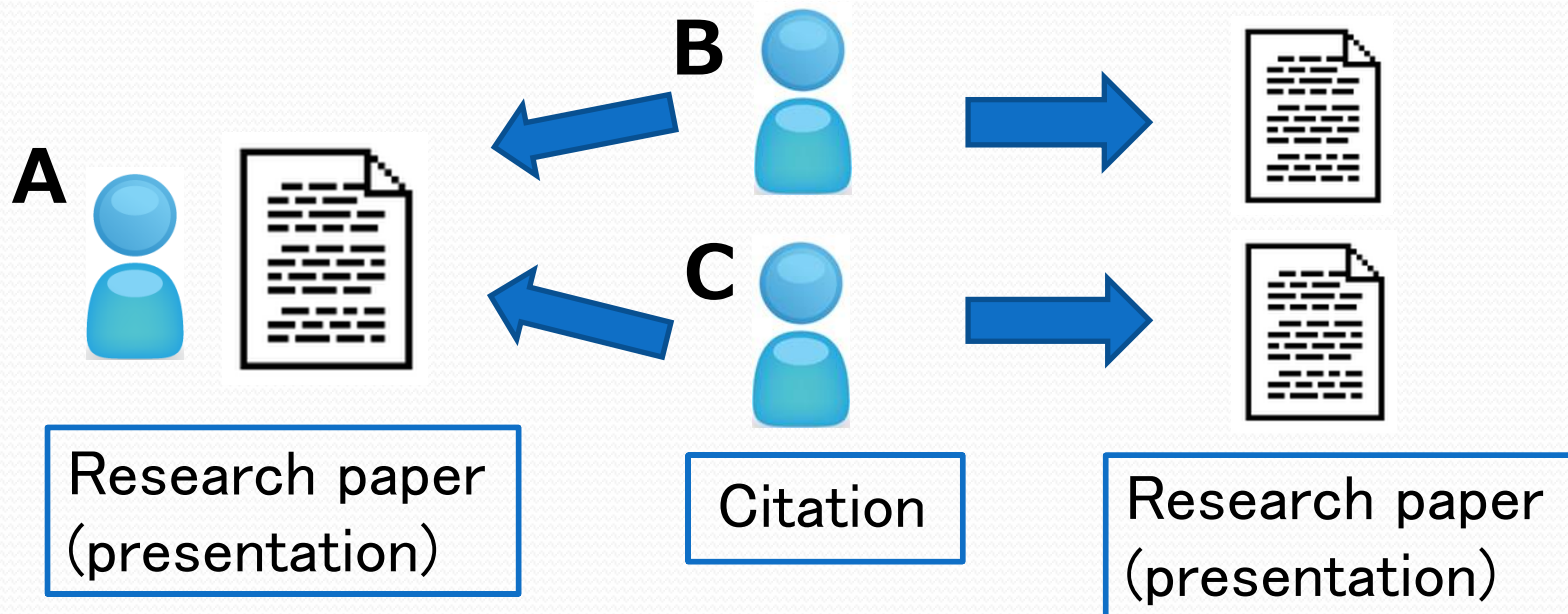
A: the number of times that all items published in that journal during the 2 preceding years were cited by publications in y .

B: the total number of citable items published by that journal during the 2 preceding years.

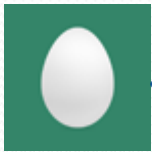
Background

Problem: Evaluation by citation-based methods requires long periods.

Impact factor: does not reflect the latest research impacts such as research papers presented in academic conference



Twitter



Relationship of onomatopoeia
and font is interesting.

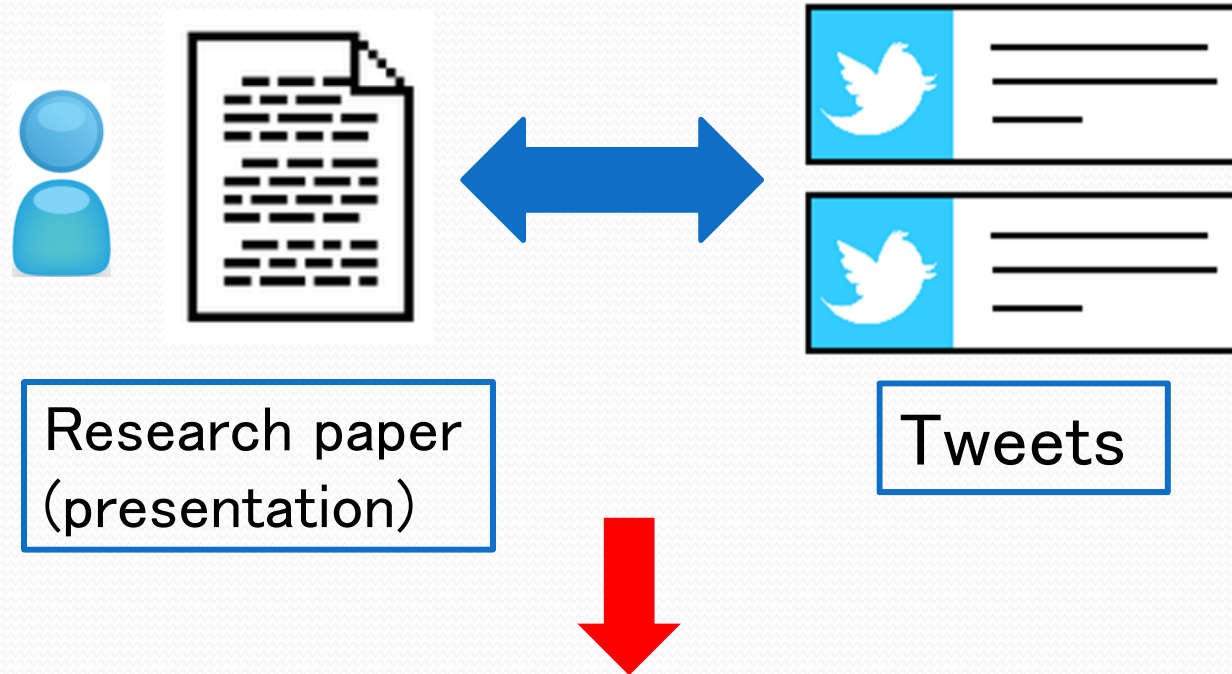
Have a strong
emotion



Sentiment,
opinion,
question etc.

Research papers (presentations) that have many comments are considered to be outstanding and to have a strong impact in the conference

Alignment of Tweets with Research paper



Evaluation of research impact

System Configuration

Detection of Valuable Tweets



Alignment of Tweets with Research Papers



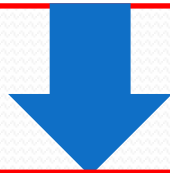
Calculation of Research Impact of Each Research Paper

System Configuration

Detection of Valuable Tweets



Alignment of Tweets with Research Papers



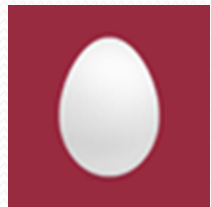
Calculation of Research Impact of Each Research Paper

Examples of Valuable Tweets



Relationship of onomatopoeia and font is interesting.

sentiment



I wonder whether some words such as “loose” and “slow” are characteristic ones that are included in the evaluation of hotel.

Opinion
(Comments)



Are terms like “loose” and “slowly” also onomatopoeia?

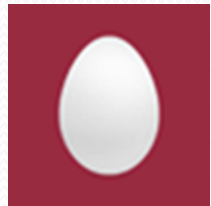
question

Examples of Valueless Tweets



I came to the shikakeology session by a 15-minute walk (> <).

Not related
research
presentation



“Automatic Organization of Travel Information”, they extract some links that are contained from travel blog entries.

Only
bibliographic
information



RT @nanaya_sac whether the recipe book of packed onomatopoeia is heresy.

RT tweet

Detection of Valuable Tweets

- We detect valuable tweets using machine learning.
 - Support Vector Machine (SVM)

We propose the following cue phrases as features for SVM to detect valuable tweets effectively.

1. Sentiment lexicons for sentiment analysis
2. Opinion phrase lists
3. Presence or absence of a question mark
4. The frequency of appearance of each word

sentiment

opinion

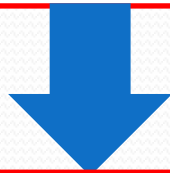
question

System Configuration

Detection of Valuable Tweets



Alignment of Tweets with Research Papers



Calculation of Research Impact of Each Research Paper

Alignment of Tweets with Research Papers

The procedure for alignment

1. Obtain candidate research papers corresponding to the tweet.
2. Calculate the similarity between a tweet and a candidate research paper using a similarity measure.
3. Align the tweet with the research paper that has the highest similarity.

Alignment of Tweets with Research Papers

1. Obtain candidate research papers corresponding to the tweet.
 - using the time of posting of each tweet and the time of presentation of research papers in the conference.

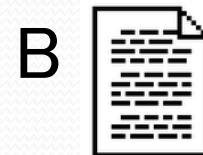
Refer to research papers presented up to 30 minutes before the posted tweet.



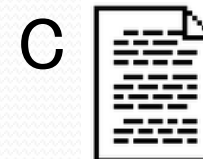
Posted tweet time: 10:30



Presentation time
10:00 ~ 10:20



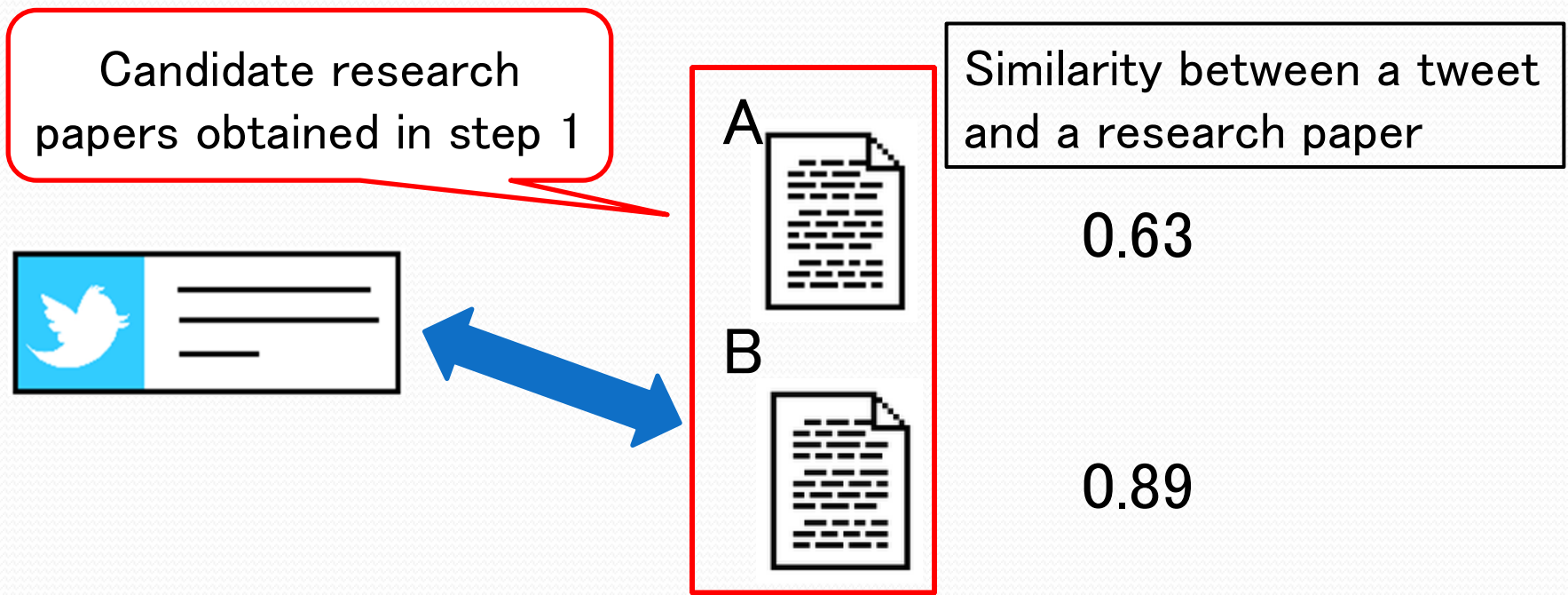
10:20 ~ 10:40



10:40 ~ 11:00

Alignment of Tweets with Research Papers

2. Calculate the similarity between a tweet and a candidate research paper using a similarity measure.



Similarity measure: DP (Dynamic Programming) matching

Alignment of Tweets with Research Papers

We consider the following problem in aligning tweets with research papers.

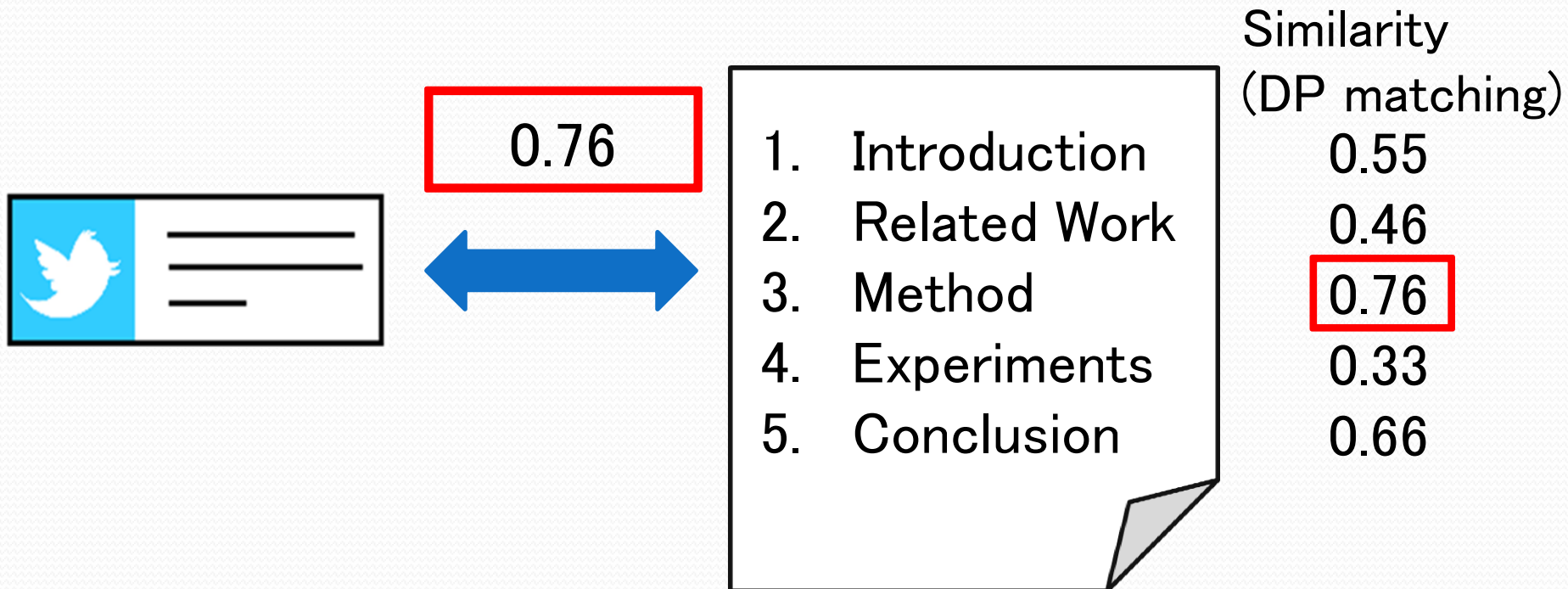
The maximum length of a tweet is 140 characters, and there may not be enough information in a tweet.

We take account of the following two points.

- **Sec:** Calculate the similarity between a tweet and a section in a research paper.

Alignment of Tweets with Research Papers

- **Sec:** Calculate the similarity between a tweet and a section in a research paper.

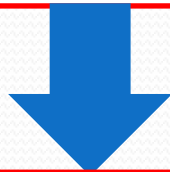


System Configuration

Detection of Valuable Tweets



Alignment of Tweets with Research Papers



Calculation of Research Impact of Each Research Paper

Calculation of Research Impact of Each Research Paper

To calculate of research impact, we use a simple ranking approach as follows.

⇒ Top-ranked research paper is more important.

(Step-1) Count the number of automatically aligned valuable tweets for each research paper.

(Step-2) Rank research papers by the number of aligned tweets.

Experiment

-Detection of Valuable Tweets-

- Dataset
 - 13 conferences and 291 research papers
 - 4,693 tweets
- Machine-learning package
 - TinySVM (linear kernel)
- Our method
 - use cue phrases (e.g. Sentiment lexicons for sentiment, Opinion phrase lists) as features
- Baseline
 - Not use cue phrases i.e. only use the frequency of appearance of each word

Experiment

-Detection of Valuable Tweets-

- Evaluation measure
 - precision, recall, F-measure
- Experimental Results

| | Precision | Recall | F-measure |
|------------|-----------|--------|-----------|
| Our method | 0.588 | 0.591 | 0.589 |
| Baseline | 0.581 | 0.534 | 0.557 |

McNemar test: **significance level of 0.01**
was obtained.

Experiment -Alignment of Tweets with Research Papers-

- Dataset
 - 13 conferences and 291 research papers
 - 840 valuable tweets manually judged by one annotator
- Evaluation measure
 - precision, recall, F-measure

Experiment -Alignment of Tweets with Research Papers-

Our method

- **DP** : Use DP matching as similarity measure for alignment of tweets with research papers.
- **DP+Sec** : Use **DP** to calculate a similarity measure for alignment of tweets with each section of research paper.

Baseline method

- **Baseline**: align all research papers within 20 minutes before and after with the target tweet.

Experiment -Alignment of Tweets with Research Papers-

- Experimental Results

| | Precision | Recall | F-measure |
|----------|--------------|--------------|--------------|
| DP | 0.492 | 0.453 | 0.472 |
| DP+Sec | 0.525 | 0.483 | 0.503 |
| Baseline | 0.463 | 0.370 | 0.411 |

McNemar test: **significance level of 0.01 was obtained (compared by DP+Sec with Baseline).**

Experiment – Calculation of Research Impact of Each Research paper–

- Evaluation method
 - We compared the output using our system with actual awards to research papers in each academic conference.
- Dataset
 - 11 conferences and 237 research papers
 - 796 tweets
 - excluding the conferences for which no tweets were aligned with any awarded research papers.
- Evaluation measure
 - Mean Reciprocal Rank (MRR)

Experiment – Calculation of Research Impact of Each Research paper–

Comparative method

Our method

- Count the number of tweets automatically aligned with the research papers using the DP + Sec method, and then sort research papers by the number of tweets.

Baseline method

- Count the number of tweets aligned with the research papers using the time information (posted time of tweet and presentation time of research paper).

Experiment – Calculation of Research Impact of Each Research paper–

| | MRR |
|------------|-------|
| Our method | 0.236 |
| Baseline | 0.183 |

Our system is useful for detecting research papers having higher research impacts.

Ranking of awarded research papers

| Japanese conference name | Best Paper Award | Paper Award | Young Paper Award |
|------------------------------------------------------------------|---------------------------|---------------|-------------------|
| Data Engineering and Information Management in 2012 (63) | (0/1) | (1/1) | 46 (3/15) |
| The Association for Natural Language Processing in 2012 (29) | Rank-5 (1/1) | (0/4) | Rank-1, 15 (2/5) |
| Web Intelligence and Interaction 1 st conference (11) | Number of research papers | Rank-9 (1/1) | (0/1) |
| Web Intelligence and Interaction 2 nd conference (16) | | Rank-11 (1/1) | Rank-14 (1/1) |
| Web Intelligence and Interaction 3 rd conference (8) | | Rank-6 (1/1) | (0/1) |

Award name

Number of awarded research papers

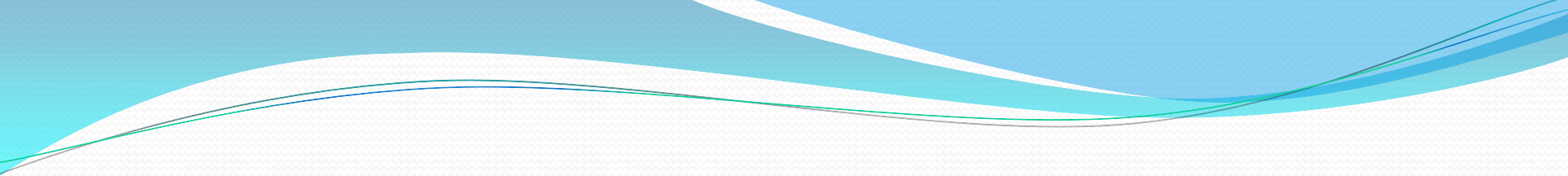
Number of research papers

Ranking of awarded research papers

| | Best Paper Award | Paper Award | Young Paper Award |
|------------------------------------------------------------------|------------------------|------------------|----------------------------------|
| Data Engineering and Information Management in 2012 (63) | (0/1) | Rank-22 (1/1) | Rank-1 , 30, 46 (3/15) |
| The Association for Natural Language Processing in 2012 (29) | Rank-5 (1/1) | (0/4) | Rank-1 , 15 (2/5) |
| Web Intelligence and Interaction 1 st conference (11) | | Rank-9 (1/1) | (0/1) |
| Web Intelligence and Interaction 2 nd conference (16) | | Rank-11 (1/1) | Rank-14 (1/1) |
| Web Intelligence and Interaction 3 rd conference (8) | | Rank-6 (1/1) | (0/1) |

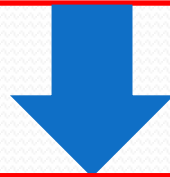
Conclusion

- We proposed a method for automatically evaluating research impact using Twitter.
 - Detection of valuable tweets
 - Alignment of tweets with research papers
 - Calculation of research impact of each research paper based on the number of aligned valuable tweets
- We conducted some experiments using tweets posted during several Japanese conferences and the research papers presented in them.
- Confirmed the effectiveness of our methods.

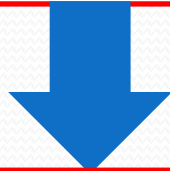


System Configuration

Detection of Valuable Tweets



Alignment of Tweets with Research Papers



Calculation of Research Impact of Each Research Paper

Alignment of Tweets with Research Papers

We consider the following problem in aligning tweets with research papers.

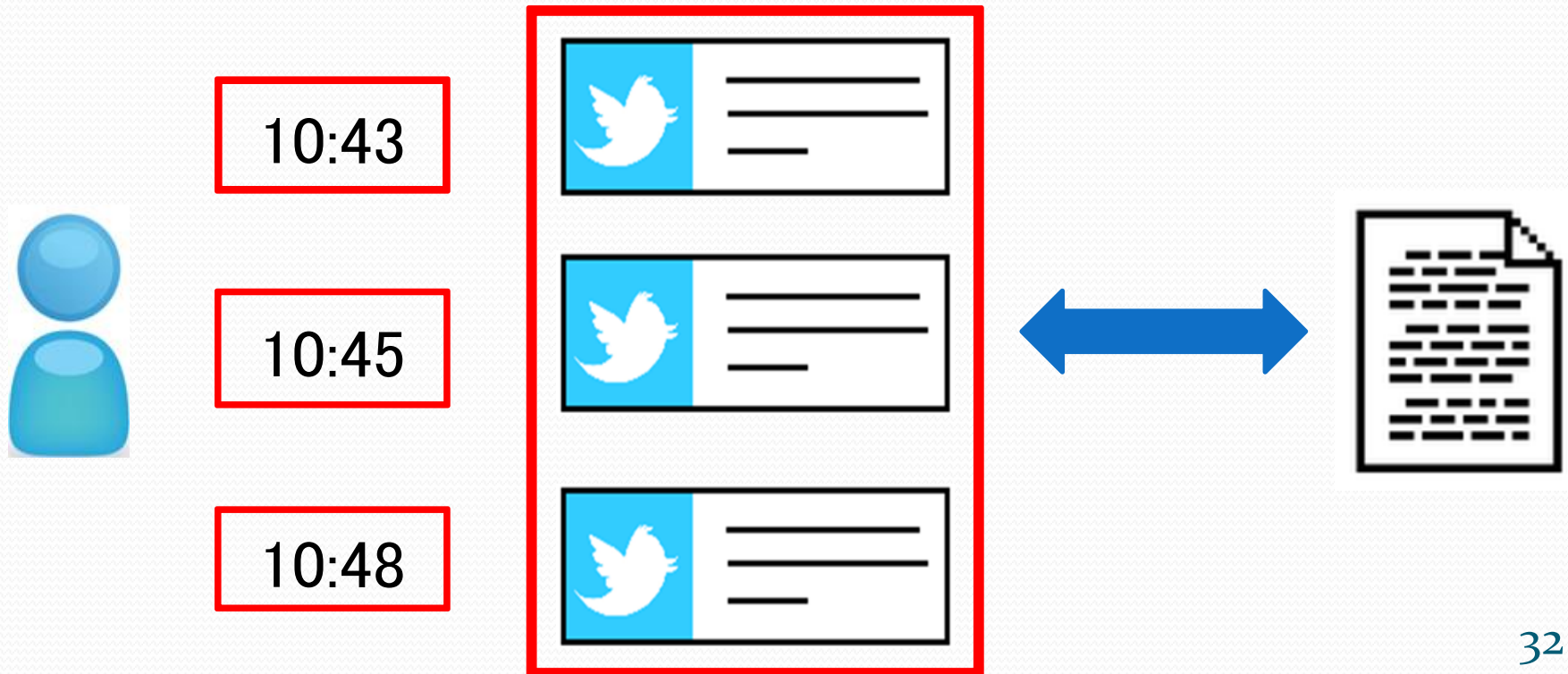
The maximum length of a tweet is 140 characters, and there may not be enough information in a tweet.

We take account of the following two points.

- **User:** Add some tweets posted by the same user.
- **Sec:** Calculate the similarity between a tweet and a section in a research paper.

Alignment of Tweets with Research Papers

- **User:** Add some tweets posted by the same user.
 - We use all tweets posted by the same user within 20 minutes before and after the target tweet.



Experiment -Alignment of Tweets with Research Papers-

Our method

- **DP** : Use DP matching as similarity measure for alignment of tweets with research papers.
- **DP+User** : Use DP with the same user's tweets within 20 minutes before and after the target tweet.
- **DP+Sec** : Use DP to calculate a similarity with a section in a research paper.
- **DP+User+Sec** : Use DP+User to calculate a similarity with a section in a research paper.

Baseline method

- **Baseline**: align all research papers within 20 minutes before and after with the target tweet.:

Experiment -Alignment of Tweets with Research Papers-

- Experimental Results

| | Precision | Recall | F-measure |
|-------------|--------------|--------------|--------------|
| DP | 0.492 | 0.453 | 0.472 |
| DP+User | 0.477 | 0.438 | 0.456 |
| DP+Sec | 0.525 | 0.483 | 0.503 |
| DP+User+Sec | 0.514 | 0.473 | 0.493 |
| Baseline | 0.463 | 0.370 | 0.411 |

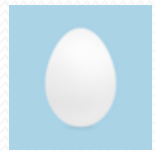
McNemar test: **significance level of 0.01 was obtained (compared by DP+Sec with Baseline).**

ご清聴ありがとうございました

Ranking of awarded research papers

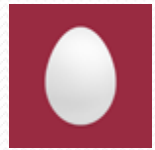
| | None of tweet for the awarded research paper | | Young Paper Award |
|------------------------------------------------------------------|----------------------------------------------|------------------|-----------------------|
| Data Engineering and Information Management in 2012 (63) | (0/1) | Rank-22 (1/1) | Rank-1, 30, 46 (3/15) |
| The Association for Natural Language Processing in 2012 | None of the award in the conference | 0/4) | Rank-1, 15 (2/5) |
| Web Intelligence and Interaction 1 st conference (11) | | Rank-9 (1/1) | (0/1) |
| Web Intelligence and Interaction 2 nd conference (16) | | Rank-11 (1/1) | Rank-14 (1/1) |
| Web Intelligence and Interaction 3 rd conference (8) | | Rank-6 (1/1) | (0/1) |

Automatic classification of Valuable Tweets



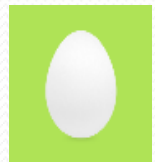
Relationship of onomatopoeia and font is interesting.

Sentiment



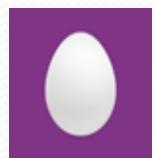
I wonder whether some words such as “loose” and “slow” are characteristic ones

Opinion



Are terms like “loose” and “slowly” also onomatopoeia?

Question



I came to the shikakeology session by a 15-minute walk (> <).

Not related presentation

DP (dynamic Programming) matching

The marshal two sentences by DP matching

S1 : [カスタマーレビュー][中][における][オノマトペ][の][調査]
S2 : [カスタマーレビュー][による][オノマトペ][*****][**][****]

Calculate the similarity between two sentences using edit distance

Edit distance

$$= \sqrt{\frac{j}{\text{The number of morpheme of sentence 1}} \times \frac{j}{\text{The number of morpheme of sentence 2}}}$$

j : The number of matched morpheme between two sentences

Experimental Results (using Cosine Similarity)

| | Precision | Recall | F-measure |
|----------|-----------|--------|-----------|
| DP+Sec | 0.525 | 0.483 | 0.503 |
| Cos | 0.406 | 0.375 | 0.390 |
| Cos+Sec | 0.413 | 0.380 | 0.396 |
| Baseline | 0.463 | 0.370 | 0.411 |

Experimental Result (using ROUGE)

| | Precision | Recall | F-measure |
|-----------|-----------|--------|-----------|
| DP+Sec | 0.525 | 0.483 | 0.503 |
| ROUGE | 0.367 | 0.333 | 0.349 |
| ROUGE+Sec | 0.375 | 0.389 | 0.382 |
| Baseline | 0.463 | 0.370 | 0.411 |

McNemar test

| | | Our method | |
|----------|-----------|------------|-----------|
| | | Correct | Incorrect |
| Baseline | Correct | a=191 | b=114 |
| | Incorrect | c=189 | d=289 |

Test statistic $T = \frac{(b-c)^2}{b+c} = 18.56$

$T = 18.56 > 6.634$ (Significant level: 1%)

受賞論文のランキング(人手)

| | 最優秀論文賞 | 優秀論文賞 | 若手奨励賞 | 萌芽研究賞 |
|--------------------------------|-------------|--------------|-----------------------|-----------------|
| 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) | 34位 (1/1) | 1位, 34位, 34位(3/15) | |
| 言語処理学会 第18回年次大会(31) | 3位 (1/1) | (0/4) | 7位, 14位 (2/5) | |
| 第1回Webインテリジェンスとインタラクション研究会(11) | | 5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ン研究会(16) | | 10位 (1/1) | 7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ン研究会(8) | | 4位 (1/1) | (0/1) | 1位 (1/1) |

受賞論文以外がランキング上位

- ネタ系論文

タイトル：音象徴の機械学習による再現最強の
ポケモンの生成

[ツイート]

これは面白そう

おもしろーい

やばい

ランキング1位の論文への反応

タイトル：Wikipediaの表記特徴を利用した別称コーパス生成ツールの開発

[ツイート]

別称コーパス生成ツール これは使えそう！
正式名称だと広告ばかりがひっかかり、略称だと意見がひっかかるというのは面白い
コーパスをみたい

今後の課題

- 引用手法との比較実験や相関分析
- Facebookやブログなど別のSNSも用いる
- 対応付けにおいて節との類似度を計測する際、時間情報を考慮させる

実験(論文の重要度の評価)

● 評価方法

論文をランキングし, ランキング結果を評価する際, 学会の受賞論文を重要論文として分析
[Sidiropoulos 2005]



本研究においても学会の受賞論文を重要論文とし, 受賞論文の順位を評価

機械学習に基づく手法

素性

- 1 : 評価表現辞書[Kaji 2007]
- 2 : ツイートと論文の類似度
- 3 : 文タイプ[横山 2003]
- 4 : URLの有無
- 5 : ツイートの長さ

機械学習に基づく手法

素性

6 : 論文のタイトルの有無

7 : 顔文字の有無

8 : 疑問符の有無

9 : 手がかり語(情報利得)

10 : 全単語(動詞, 名詞, 形容詞)

考察(ツイートと論文との自動対応付け)

- 同じセッションの論文は類似した研究が多い

[ツイート]

専門家が書いたレシピはオノマトペが少ない。。。

[論文1]

オノマトペによるランキング表示

オノマトペロリでは、料理レシピとオノマトペの適合度を求め、ランキングを行っている。

[論文2]

料理とスイーツの違いが現れるオノマトペや、類似したオノマトペでも微妙に関連する料理が異なる

考察(ツイートと論文との自動対応付け)

- 同じユーザの前後のツイートを追加する際、他の論文のツイートも追加してしまった場合がある

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ー中のリンクに注目. #rrds3 2010/12/18 14:14:58

他の論文
についての
ツイート

同一論文
についての
ツイート

考察(有益なツイートの自動分類)

- 誤って有益なツイートと判定された例

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

- ◆ 実況中継に関するツイート
- ◆ 情報利得による不要な手がかり語が存在した
- ◆ 不要な手がかり語リストの作成により解決

考察(有益なツイートの自動分類)

- 有益なツイートと判定できなかった例

フォントとオノマトペの関係はおもしろいなあ

- ◆ 表記揺れが存在した
- ◆ 形態素解析器JUMANを用いることで解決

人手により分類した結果

| | 分類数(件) |
|------------|--------|
| 有益なツイート | 840 |
| 有益ではないツイート | 3,853 |

ε の計算

順位が1位、5位

$$\text{MRR} = 1/2(1/1 + 1/5)$$

順位が1位5位、3位など複数存在

$$\varepsilon = \frac{\frac{1}{1} + 1/5}{\frac{1}{1} + 1/2}$$

$$\text{MRR} = 1/2(\varepsilon + 1/3)$$

評価表現辞書[Kaji 2007]

鍛冶らの研究の評価文コーパスにより、自動構築された辞書

辞書の例

難しい
面白い
厳しい
素晴らしい



論文の評価する
語句も多く含ま
れている

文タイプ[横山 2003]

横山らが人手で作成した、文末表現により決まる文のタイプのこと

文タイプの例

方がいい
必要だ
べきである
望ましい

論文を意見する
際の文末の表現
が多く含まれる



コサイン類似度

2つのベクトル間のコサイン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、論文とツイートの単語を用いる。単語の重みにはtf-idfを用いる。

$$\cos(\vec{p}, \vec{q}) = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| |\vec{q}|} = \frac{\vec{p}}{|\vec{p}|} \cdot \frac{\vec{q}}{|\vec{q}|} = \frac{\sum_{i=1}^{|\mathcal{V}|} p_i q_i}{\sqrt{\sum_{i=1}^{|\mathcal{V}|} p_i^2} \cdot \sqrt{\sum_{i=1}^{|\mathcal{V}|} q_i^2}}$$

\vec{p} : 論文の単語
 \vec{q} : ツイートの単語

情報利得

情報利得:単語のエントロピーの減少量を値にしたもの
→ 素性として有効な単語を得られる

有益

面白い
素晴らしい
楽しい
オノマトペ

有益でない

楽しい
オノマトペ
幸せ
感動

受賞論文のランキング

| | 最優秀論文賞 | 優秀論文賞 | 若手奨励賞 | 萌芽研究賞 |
|--------------------------------|--------|--------------|-----------------------|-----------------|
| 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) | |
| 人工知能学会2012(51) | | 22位 (1/9) | | |
| 第1回Webインテリジェンスとインタラクション研究会(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ン研究会(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ン研究会(8) | | 6位 (1/1) | (0/1) | 1位 (1/1) |