

## クラウドソーシングによる重要ニュース選択

## Breaking News Selection using Crowdsourcing

高濱 隆輔 \*1\*2      馬場 雪乃 \*1      清水 伸幸 \*3      藤田 澄男 \*3      鹿島 久嗣 \*1  
 Ryusuke Takahama      Yukino Baba      Nobuyuki Shimizu      Sumio Fujita      Hisashi Kashima

\*1 京都大学大学院 情報学研究科 知能情報学専攻

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

\*2 JST ERATO 河原林巨大グラフプロジェクト      \*3 ヤフー株式会社 Yahoo! JAPAN 研究所  
 JST, ERATO, Kawarabayashi Large Graph Project      Yahoo! JAPAN Research, Yahoo Japan Corporation

News sites should show users especially important news. Selection of important news is conducted by experts since criteria of importance are not obvious and algorithmic judgement is difficult. In order to lighten the work of experts, we propose the method that selects important news by using crowdsourcing. According to results of experiments, we could estimate the importance of news with high accuracy using the dataset collected by pairwise comparison. In addition, we applied the method that generates criteria of classification to this problem and confirmed it may create features which help to discover important news.

## 1. はじめに

様々なインターネットメディアによって日々多くのニュースが配信され続けている昨今、ニュースメディアはそれらの大量のニュースの中から特に重要なニュースを選び、ユーザに提示する必要がある。しかしながら、重要なニュースを機械的に判断するのは容易ではない。なぜなら、重要なニュースの選択基準は自明ではないので、特微量やルールを定義するのが難しいからである。例えば、Yahoo! JAPAN のトップページ \*1 には、特に重要と判断された 8 件のニュースが表示されているが、掲載するニュースの選択の大部分は **エディタ** と呼ばれる専門家が人手で行っている。1 日あたり約 4,000 件にも及ぶ配信ニュースのなかから重要なニュースを選択するエディタの負荷は大きい。

そこで我々は、クラウドソーシングを用いて重要ニュースの判別を行う手法を提案する。多数のクラウドワーカーからニュースに関する意見を収集し、これを統合することによってニュースの重要性を調べる。本手法が、例えば Yahoo! JAPAN トップページに掲載されるニュースの選択に用いられる場合、図 1 に示すようなシステムが構築されることになる。現行のシステムでは、掲載主から投稿された記事は全てエディタに提示され、エディタの意見をもとにトップページ掲載ニュースが選択されている。ここに、図 1 中では青く示されているような、クラウドソーシングによるニュースのフィルタリングを行う部分を挿入することにより、明らかに重要でないニュースをエディタに提示しないようにする。これによりトップページ掲載ニュース選択におけるエディタの負荷を軽減することができる。

我々はクラウドワーカーからの意見の収集法として **絶対評価** および **相対評価** を検討した。絶対評価に関する実験では、ワーカーに対して 1 つのニュースを提示し、そのニュースの掲載可否を問う。相対評価に関する実験では、ワーカーに対して 2 つのニュースを提示し、どちらのニュースを掲載すべきかを問う。これらの実験の結果、特に相対評価によって意見を収集した場

合にニュースの掲載可否を正しく判断できることを示した。

さらに、クラウドソーシングを用いて重要ニュースの判断基準を生成する方法についても検討した。Cheng と Bernstein は、相対評価タスクを用いて特徴を生成し、より高精度な分類を行うことができる分類器を構築するためのプラットフォームである Flock を提案している [Cheng 15]。我々は Flock における相対評価タスクを用いた判断理由の収集と、それら理由の統合による特徴の生成を重要ニュースの判断基準生成に適用し、重要ニュースの分類に有用な特徴が生成できる可能性があることを確認した。

本論文では、ニュースに 2 つのカテゴリを定義する: (i) **公共性ニュース** は、新聞や NHK ニュースで報道される傾向があるニュース群、(ii) **話題性ニュース** は、週刊誌やネットメディアで掲載される傾向があるニュース群である。本論文中で用いているデータセットにおける話題性・公共性の分類は全て筆者が行った。

## 2. 絶対評価による重要ニュース選択

まず、ニュースの重要性を単純にクラウドワーカーに尋ねた場合に、どの程度正しく重要ニュースを分類できるのかを確かめる。ここで、ニュースの分類精度に影響を及ぼしそうな要因として以下の 2 点が挙げられる: (i) クラウドワーカーに提示する質問文の形式、(ii) ワーカーの属性 (年齢/性別など)。以上の条件を変化させながら、重要ニュースの分類の精度を確認する実験を行う。

### 2.1 実験設定

Yahoo! ニュース \*2 より、まず公共性ニュースとして分類されたニュース記事群から、正例・負例をそれぞれ 50 件、計 100 件の記事を用意する。ここで、正例は実際に Yahoo! トップページに掲載されたニュース、負例は掲載されなかったニュースを指す。話題性ニュースについても同様に 100 件の記事を用意する。

このニュースセットに含まれる各ニュースに対し、Yahoo! ク

連絡先: 高濱 隆輔, 京都大学大学院 情報学研究科 知能情報学専攻, takahama@ml.ist.i.kyoto-u.ac.jp

\*1 <http://www.yahoo.co.jp/>

\*2 <http://news.yahoo.co.jp/>



図1: クラウドソーシングによる重要ニュース選択手法が Yahoo! JAPAN トップページに掲載されるニュースの選択に用いられる場合に構築されるシステムの例。現行のシステムでは、掲載主から投稿された記事は全てエディタに提示され、エディタの意見をもとにトップページ掲載ニュースが選択されている。ここに、図中で青く示されているような、クラウドソーシングによるニュースのフィルタリングを行う部分を実挿入することにより、明らかに重要でないニュースをエディタに提示しないようにする。これによりトップページ掲載ニュース選択におけるエディタの負荷を軽減することができる。

クラウドソーシング<sup>\*3</sup>を介して、クラウドワーカに以下の7つの質問のいずれかを問うタスクを依頼する：

- Q1. ヤフトップページには特に重要な8件のニュースしか表示されません。このニュースはヤフトップページに掲載すべきですか？
- Q2. このニュースはNHKニュースで報道すべきですか？
- Q3. あなたがニュースサイトを運営しているとして、このニュースをそのサイトに掲載しますか？
- Q4. あなたがこのニュースの見出しをニュースサイトで見たとして、あなたはこのニュースのリンクをクリックしますか？
- Q5. このニュースはヤフトップページを見る人にとって重要ですか？
- Q6. このニュースはあなたの周りの人にとって重要ですか？
- Q7. このニュースはあなたにとって重要ですか？

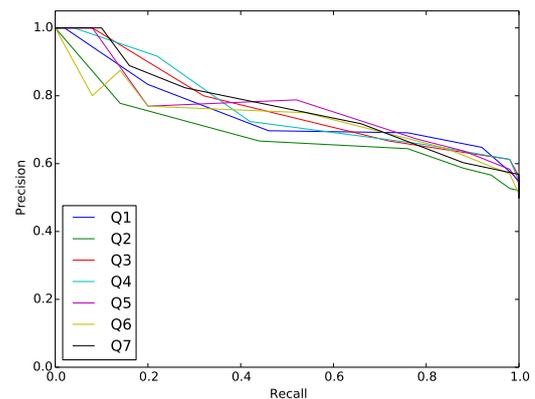
ワーカは各質問に対し、「はい」または「いいえ」の二択で回答する。また、各タスクは8属性のワーカに同数依頼される。8属性の内訳は、4種類の年齢(20代, 30代, 40代, 50代以上)および2種類の性別(男性, 女性)の組み合わせである。さらに、タスクを実行してもらう際に、同時に以下の属性もワーカに聞いておく：最終学歴(小/中学校, 高校/旧中, 短大/高専/専門学校, 大学, 大学院), 結婚歴(有り, 無し)。

## 2.2 実験結果

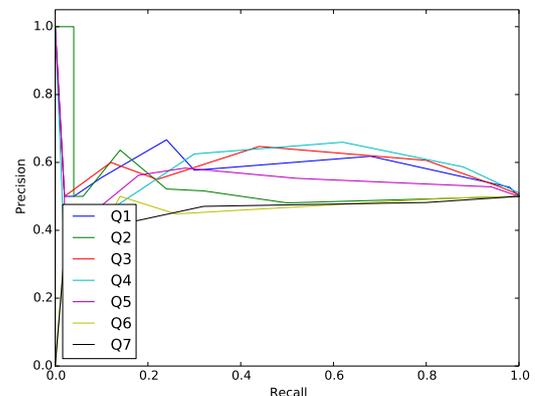
実験の結果を以下に述べる。まず、ワーカの属性や質問文による正答率の差はみられなかった。ここで、正答率とは、ニュース全体について、ワーカの意見の多数決で導かれた回答が正解と一致しているものの割合として計算する。各質問文について、属性を無視して全てのデータを用いて算出した正答率と、各属性のみのデータを用いて算出した正答率との差をフィッシャーの正確確率検定で検定したところ、どの質問文、属性の組に対しても有意水準1%で有意差が認められなかった。

次に、公共性ニュース、話題性ニュースそれぞれに関する Precision-Recall curve (PR curve) を図2に示す。複数のワーカの意見を統合する際に、何割以上のワーカが「はい」と答えていれば全体の意見を「はい」と統合するかの閾値を変化させることによって PR curve を描画した。図2より、話題性ニュースに比べて公共性ニュースの方が正例と負例を高精度に分類できていることが読み取れる。

続いて、各ニュースについて、「はい」と答えたワーカの割合、すなわち「掲載すべき」と判断したワーカの割合のヒストグラムを図3に示す。なお、先述の通り、各質問間の正答率には



(a) 公共性ニュース

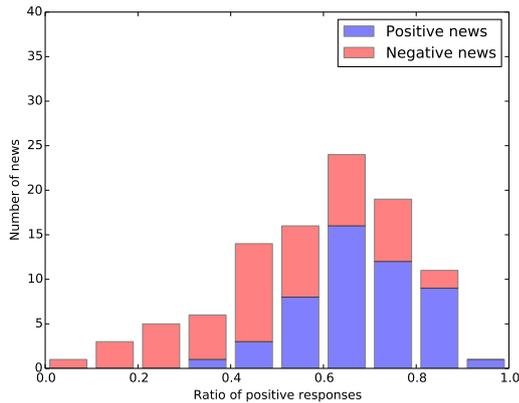


(b) 話題性ニュース

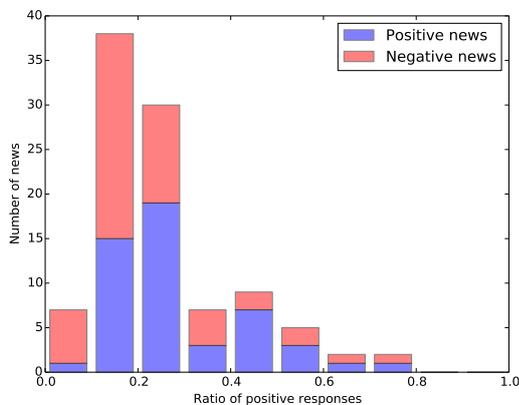
図2: Precision-Recall curve. Q1 から Q7 は、ワーカに問うた7つの質問の番号とそれぞれ対応している。話題性ニュースに比べて公共性ニュースの方が高精度に正負を分類できていることが読み取れる。

有意差が認められなかったため、1番の「ヤフトップページには特に重要な8件のニュースしか表示されません。このニュースはヤフトップページに掲載すべきですか？」という質問に対する回答を用いて図を描画した。図3より、話題性ニュースに比べ、公共性ニュースの方がワーカに「載せるべき」と強く判断される傾向があることが読み取れる。話題性ニュースと公共性ニュースの回答傾向に大きく違いがあるため、これらのニュースを区別せずにトップページ掲載の可否を判断することは難しいと結論づけることができる。

\*3 <http://crowdsourcing.yahoo.co.jp>



(a) 公共性ニュース



(b) 話題性ニュース

図3: 各ニュースについて「掲載すべき」と判断したワーカーの割合のヒストグラム。各質問の正答率に有意差が認められなかったため、Q1の回答結果のみを用いて図を描画した。公共性ニュースでは「掲載すべき」と判断される傾向があるのに対し、話題性ニュースでは逆の傾向があることが読み取れる。

### 3. 相対評価による重要ニュース選択

2章に示した通り、絶対評価による重要ニュース選択の実験では、話題性ニュースと公共性ニュースの回答傾向の違いや、質問文/ワーカー属性の精度への影響を調べることはできなかった。そのため、ここではより高精度な重要ニュース分類を行うため、相対評価によってワーカーから意見を収集する実験を行う。本論文では特に公共性ニュースに関する実験結果のみについて述べる。

#### 3.1 実験設定

Yahoo! ニュースより、公共性ニュースとして分類されたニュース記事群から、正例・負例をそれぞれ50件、計100件の記事を用意する。正例と負例からそれぞれ1件ずつ記事を選ぶことを繰り返し、50個のペアを作成する。

これらのペアに対し、Yahoo! クラウドソーシングを介して、クラウドワーカーに「ヤフトップページには特に重要な8件のニュースしか表示されません。ニュースAとニュースBのどちらをヤフトップページに掲載すべきですか?」と問うタスクを依頼する。ワーカーは2つのニュースAとBを提示され、どちらのニュースを掲載すべきかを回答する。

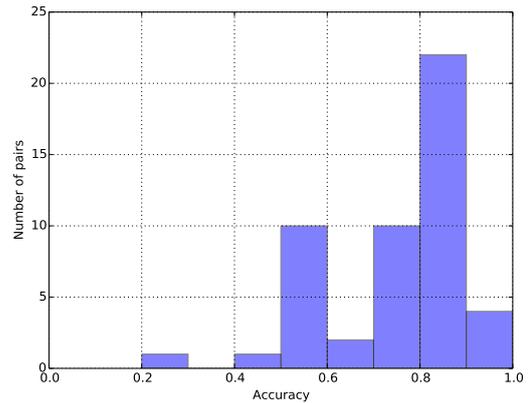


図4: 各ペアに関するワーカーの正答率のヒストグラム、公共性ニュースのみを対象にした実験結果である。ほとんどすべてのペアに対して正答率が0.5以上、つまり半分以上のワーカーが正しく正負を判別できていることが読み取れる。

### 3.2 実験結果

実験の結果として、各ペアに関するワーカーの正答率のヒストグラムを図4に示す。図4より、正答率が0.5を下回るものは2つのペアのみであり、50ペアのうち48ペアについては、多数決で正しく掲載の可否を決定できることがわかる。

以上より、公共性ニュースに関しては相対評価によるトップ掲載ニュース選択の正答率が十分に高いことを確認した。

## 4. 重要ニュース判断基準の生成

ここでは、クラウドソーシングを用いて重要ニュースの判断基準を生成する方法について検討する。重要なニュースの選択基準は自明ではないので、特徴量やルールを定義するのが難しく、そのため重要ニュースを機械的に判断するのも容易ではない。Cheng と Bernstein は、相対評価タスクを用いて特徴を生成し、より高精度な分類を行うことができる分類器を構築するためのプラットフォームである Flock を提案している [Cheng 15]。Flock では、機械的は得にくい特徴量をクラウドソーシングによって生成することで分類精度の向上を行っており、重要ニュースの判断にも応用が可能であると考えられるため、我々は Flock における相対評価タスクを用いた判断理由の収集と、それら理由の統合による特徴の生成を本問題に適用した。

Flock に基づいた、クラウドソーシングによる特徴生成および重要ニュースの判別が、例えば Yahoo! JAPAN トップページに掲載されるニュースの選択に用いられる場合、図5に示すようなシステムが構築されることになる。我々は、このシステムの一部を Yahoo! ニュースに適用した。

Yahoo! ニュースより、公共性ニュースとして分類されたニュース記事群から、正例・負例をそれぞれ50件、計100件の記事を用意する。正例と負例からそれぞれ1件ずつ記事を選ぶことを繰り返し、50個のペアを作成し、3章で用いたものと同様のニュースセットを得る。各ペアについて、ペアを構成する2つのニュースをそれぞれワーカーに提示し、正例をニュースA、負例をニュースBとする。ニュースAが正例である旨をワーカーに伝えた上で、「ニュースAがトップページに掲載された理由は何だと思いますか?」と問う。

これによって得られた理由群から、判断基準となりそうなものを著者が抽出した。まず、トップページ掲載に肯定的な基準には以下のようなものがある:

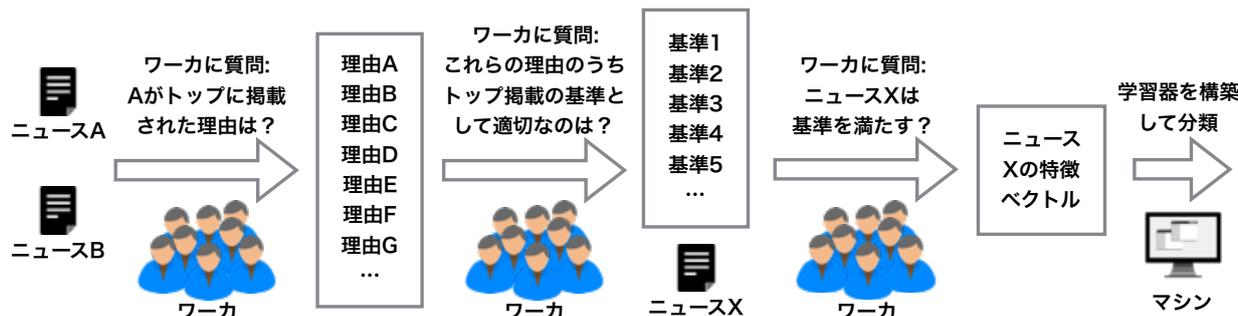


図 5: Flock に基づいた、クラウドソーシングによる特徴生成および重要ニュースの判別が Yahoo! JAPAN トップページに掲載されるニュースの選択に用いられる場合に構築されるシステムの例。クラウドワーカーに 2 つのニュース A と B を提示し、ニュース A が実際にトップページに掲載されたことを伝えた上でその理由を問う。次の段階では、集められた理由をクラウドワーカーに提示し、一般的な重要ニュース選択の基準を生成させる。さらに、各ニュースと各基準をクラウドワーカーに提示し、そのニュースがどの程度基準に合致しているかを問うことによって、基準に対応する特徴量を生成する。最後にこれら特徴量を用いて機械学習手法等によって重要ニュースか否かの分類を行う。

- 日本全国の人に関係がある
- 重大なニュースの続報である
- 学術的な価値がある
- 生活に密着した身近な話題である
- 人命・安全(事故・事件)に関わる
- 医療事件/不正事件/国家の安全保障に関わる
- アーカイブ記事として資料になり得る
- 国際社会での重要性がある
- 世界的に重要な人物についてのニュースである

続いて、否定的な基準には以下のようなものがある：

- ニュースの内容が古い・終了したイベントを扱っている
- 対象者が限定的である
- 過去に報道されたニュースの続報なので目新しさが無い
- 内容が専門的である
- ニュースというより広告である
- 内容に誤植がある

これらの理由のうち、「緊急性がある・旬な話題である」と「ニュースの内容が古い・終了したイベントを扱っている」、「日本全国の人に関係がある」と「対象者が限定的である」などは、あるニュースが片方を満たせば、他方は満たさない基準である。

一方で、「重大なニュースの続報である」と「過去に報道されたニュースの続報なので目新しさが無い」、「学術的な価値がある」と「内容が専門的である」などは、あるニュースが 2 つの基準を同時に満たすことが可能であり、あるニュースがトップページ掲載に適しているか否かが自明ではない。

以上のように、Flock に基づいて、重要ニュースの分類に有用な特徴が生成できる可能性があることを確認した。

## 5. おわりに

本論文では、クラウドソーシングを用いた重要ニュースの判別を目的として、絶対評価および相対評価によるトップ掲載ニュース選択に関する実験により、話題性ニュースと公共性ニュースでは回答傾向が大きく異なること、質問文/ワーカー属性の精度への影響に有意差は認められないこと、相対評価を用いると高精度に正負を判別できることを示した。また、相対評価タスクを用いて重要ニュースの判断基準を生成する手法である Flock [Cheng 15] を適用することにより、本問題でも分類に有用な特徴を生成できる可能性があることを確認した。

今後の課題として以下のような方向性が考えられる。第一に、2 章および 3 章で行った実験の意見統合手法に工夫を加える。

Whitehill らはワーカーの能力を考慮してラベルを推定する手法を提案し [Whitehill 09], Chen と Bennett は一対比較データを統合して順位を推定する Bradley-Terry モデル [Bradley 52] を、ワーカーの能力を考慮するよう拡張した手法を提案している [Chen 13]. これらを本問題で収集されたデータに適用することにより、より高精度な分類が行えると考えられる。

第二に、クラウドソーシングを用いて重要ニュースの判断基準を生成する方法を発展させる。4 章で Flock を本問題に適用した結果を述べたが、今回筆者が行ったワーカーから集められた理由を統合して判断基準を生成する部分をクラウドソーシングで代替したり、またどの程度基準を満たしているかをクラウドワーカーに問うことによって各ニュースの特徴量を生成する部分を実験したりといった課題が残っている。また、Zou らは、3 つのオブジェクトから似ている 2 つのオブジェクトを選択させ、それらが似ている理由を問うことによって特徴を生成する手法を提案しており [Zou 15], 加えて Flock ではワーカーの能力は考慮されていないため、Flock を拡張する様々な方向性があると考えられる。

## 参考文献

- [Bradley 52] Bradley, R. A. and Terry, M. E.: Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, in *Biometrika*, Vol. 39, pp. 324–345, Biometrika Trust (1952)
- [Chen 13] Chen, X. and Bennett, P. N.: Pairwise Ranking Aggregation in a Crowdsourced Setting, in *Proceedings of the sixth ACM International Conference on Web Search and Data Mining*, pp. 193–202 (2013)
- [Cheng 15] Cheng, J. and Bernstein, M. S.: Flock: Hybrid Crowd-Machine Learning Classifiers, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 600–611 (2015)
- [Whitehill 09] Whitehill, J., Ruvolo, P., Wu, T., Bergsma, J., and Movellan, J.: Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise, in *Advances in Neural Information Processing Systems*, pp. 2035–2043 (2009)
- [Zou 15] Zou, J. Y., Chaudhuri, K., and Kalai, A.: Crowdsourcing Feature Discovery via Adaptively Chosen Comparisons, in *Proceedings of the Third AAAI Conference on Human Computation and Crowdsourcing*, pp. 198–205 (2015)