

1L4-OS-09a-3in1

# クラウドソーシングによる重要ニュース選択

JSAI 2016, 2016/06/06 (Fri) 16:10-16:25

OS-9 ヒューマンコンピューテーションとクラウドソーシング (1)

**高濱 隆輔 (京都大学, JST ERATO)**

馬場 雪乃, 鹿島 久嗣 (京都大学)

清水 伸幸, 藤田 澄男 (Yahoo! JAPAN 研究所)

# 概要: クラウドソーシングによる重要ニュース選択に関する実験の結果、相対評価タスクの有効性を確認

---

- Yahoo! JAPAN のトップページに表示される重要ニュースは主に専門家による手作業で選択されており、選択候補となるニュースの数が大きくなると専門家の負荷が増大
- **専門家の負荷を減らすためにクラウドソーシングを利用したい**
  - ワーカーに直接「ニュースが重要かどうか」を問う
- 実験の結果、以下の事柄を確認した：
  - ニュースのタイプによってクラウドワーカーの意見の傾向が異なる
  - 絶対評価タスクより相対評価タスクのほうが高精度に重要ニュースを選択することができる
  - 相対評価タスクに関連する既存研究の手法を適用することで、クラウドソーシングから重要ニュース選択に有用な特徴を生成でき、それらの特徴による分類が機械学習による分類の精度に匹敵する

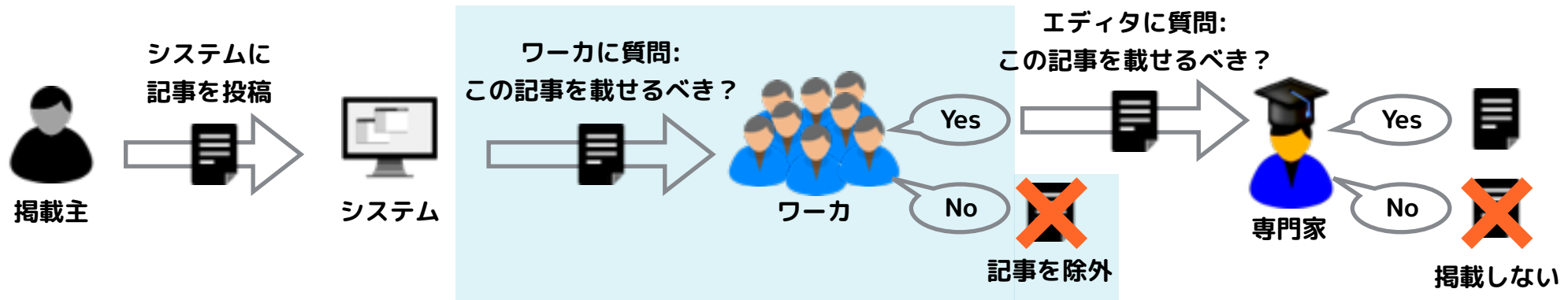
# 背景: Yahoo! JAPAN のトップページに掲載されるニュースは人力で選択されている

- Yahoo! JAPAN のトップページには特に重要と判断された8件のニュースが表示される
  - 現在は投稿されるニュースを専門家が全て見て重要なものを選択
  - ニュース数: 約 4,000 件 / 日, 通常30分おきに更新

The screenshot shows the Yahoo! JAPAN homepage. At the top, there are navigation links for 'Yahoo! BB きっず版 オフィス版', 'トラベル', 'ヤフオク!', 'ショッピング', 'YAHOO! JAPAN', 'My Yahoo!', 'ボックス', 'メール', and 'カテゴリ一覧 サイトの登録 無料ID活用'. Below these is a search bar with the text 'ウェブ 画像 動画 辞書 知恵袋 地図 リアルタイム 一覧' and a search button labeled '検索'. A banner below the search bar reads '本日「津波防災の日」もしもに備えよう' and 'ホテルオークラの家具がチャリティーに出品'. The main content area is divided into sections: '主なサービス' (Shopping, Y! Auction, LOHACO, Travel/Hotel, News, Weather, Sports, Finance, TV, GYAOL, Games), 'ニュース' (News), '経済' (Economy), 'スポーツ' (Sports), and 'その他' (Others). The 'ニュース' section is highlighted with an orange arrow pointing to the first item: '都構想再び 大阪知事選告示へ'. Other news items include 'いじめ調査ミス 重大事象3倍', '非正社員、初めて4割に上昇', '貴族、鳥二郎との訴訟和解', '自動運転車が公道で事故 愛知', 'FAの西武脇谷 巨人に復帰か', '高校サッカー劇的弾 世界絶賛', and 'お笑い芸人 ひったくり犯捕獲'.

# 問題設定: クラウドソーシングを用いて重要ニュースの 選択における専門家の負担を軽減したい

- 専門家の負担を軽減するため、以下のようなシステムを構築して重要ニュースの選択をサポートしたい:



# データセット: Yahoo! ニュースより、公共性ニュース、話題性ニュースそれぞれについて正例と負例を用意

---

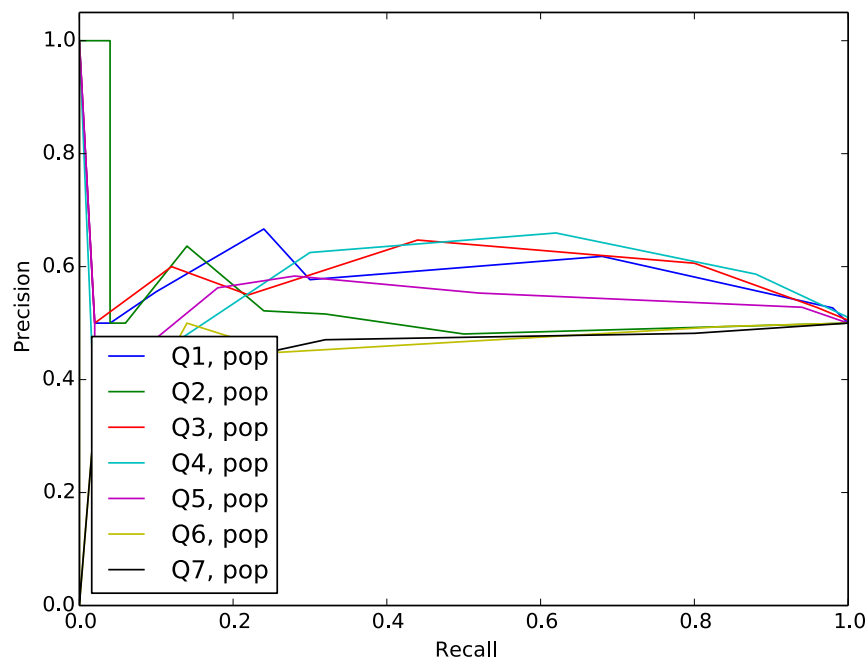
- Yahoo! ニュースにて配信された記事からデータセットを生成した
- ニュースの分類:
  - 公共性ニュース (100件): 新聞, NHK で報道されうる内容
  - 話題性ニュース (100件): 週刊誌, ネットメディアで報道されうる内容
- 各分類ごとに正例/負例を50件ずつ用意
  - 「正例」は実際に Yahoo! JAPAN のトップページに掲載されたニュース、「負例」はそうでないニュースを示す

# 絶対評価による重要ニュース選択:

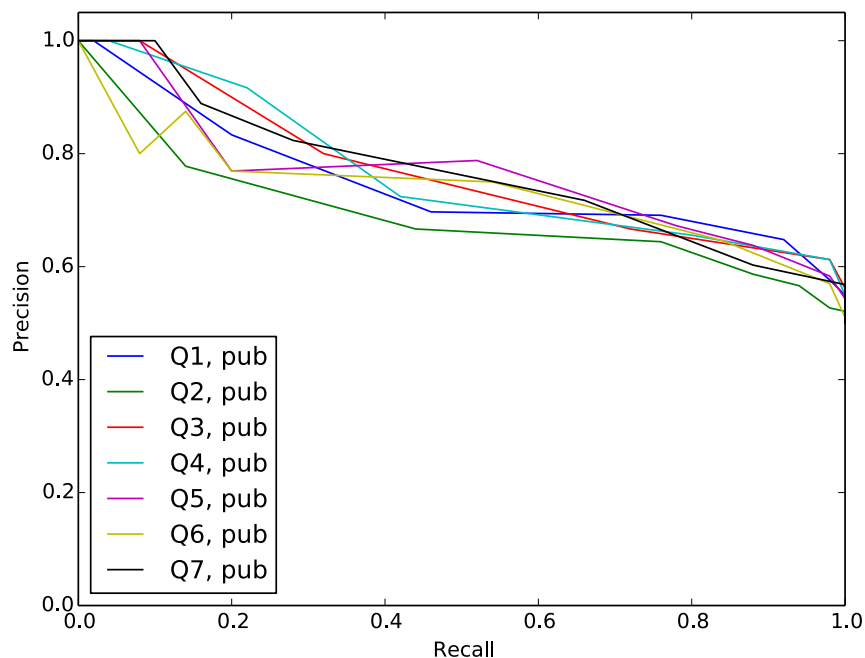
## 話題性ニュースより公共性ニュースのほうが高精度

- 「このニュースをトップページに載せるべき？」など、7通りの質問への回答をクラウドワークに依頼
- 多数決の閾値を変化させて Precision-Recall curve を描画
- 話題性ニュースより公共性ニュースのほうが高精度

話題性ニュース



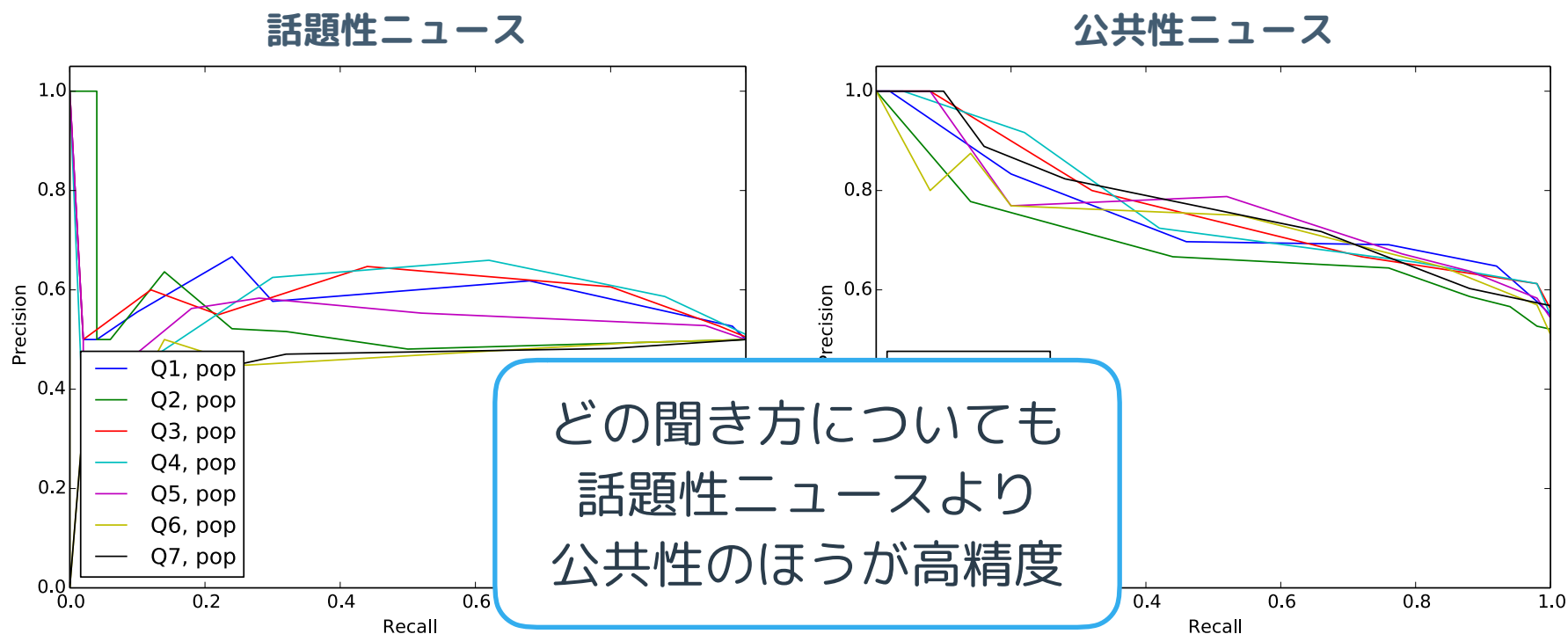
公共性ニュース



# 絶対評価による重要ニュース選択:

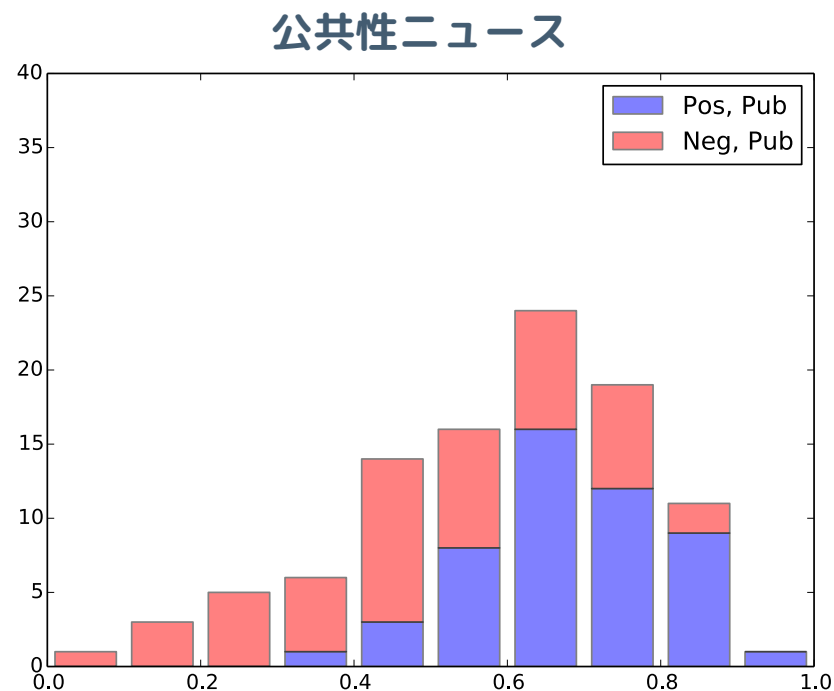
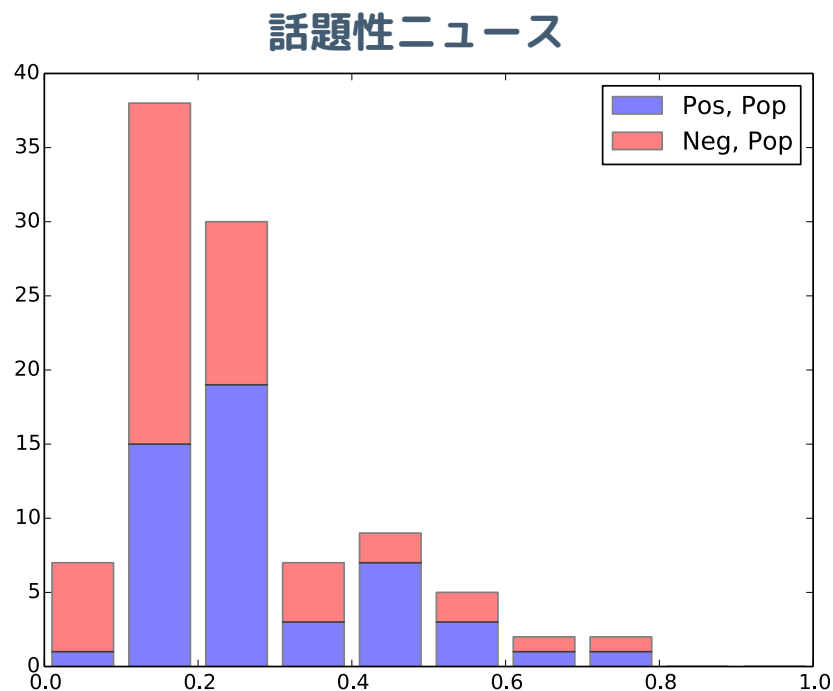
## 話題性ニュースより公共性ニュースのほうが高精度

- 「このニュースをトップページに載せるべき？」など、7通りの質問への回答をクラウドワークに依頼
- 多数決の閾値を変化させて Precision-Recall curve を描画
- 話題性ニュースより公共性ニュースのほうが高精度



# 絶対評価による重要ニュース選択: ワーカは公共性 ニュースの方をより載せるべきと判断しがち

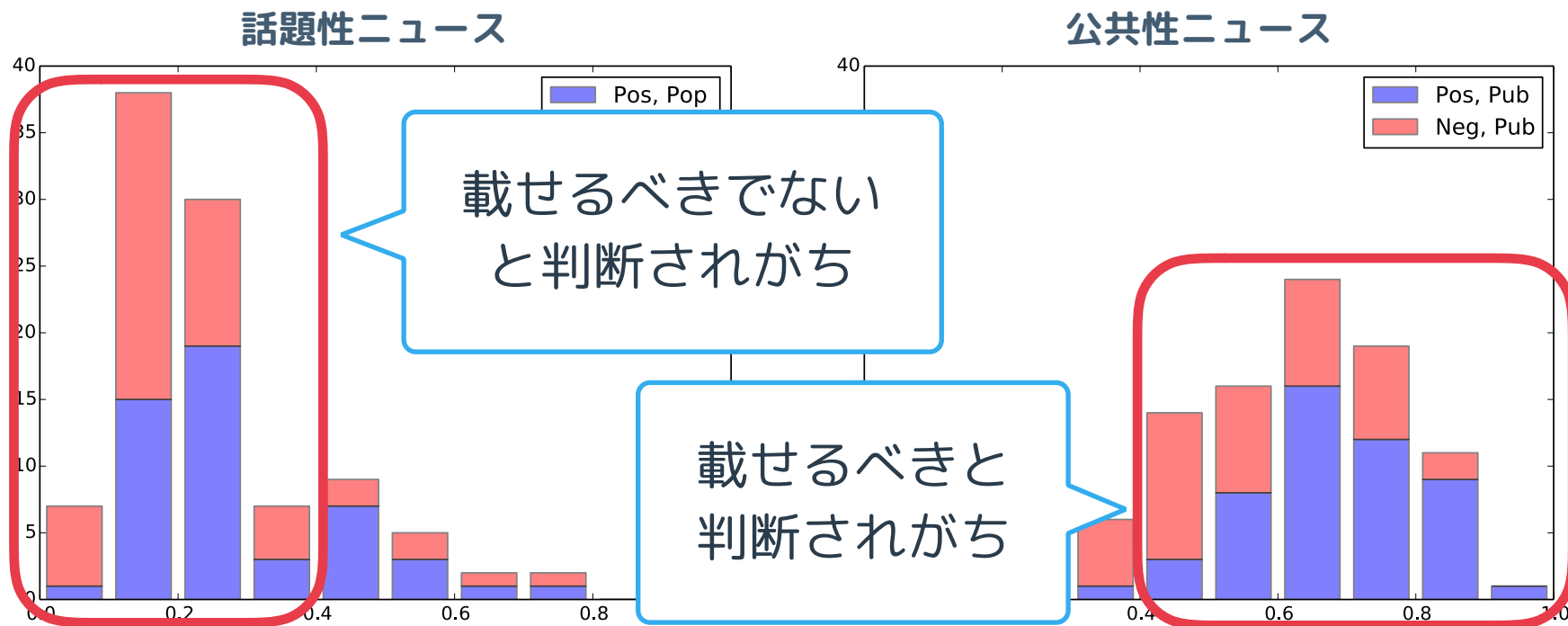
- 「載せるべき」と判断したワーカの割合を確率と解釈し、ヒストグラムをプロット
  - 公共性ニュースの方が載せるべきと判断されがちな傾向がある
  - 話題性・公共性ニュースを分けずに掲載の可否を決めるのは難しい





# 絶対評価による重要ニュース選択: ワーカは公共性 ニュースの方をより載せるべきと判断しがち

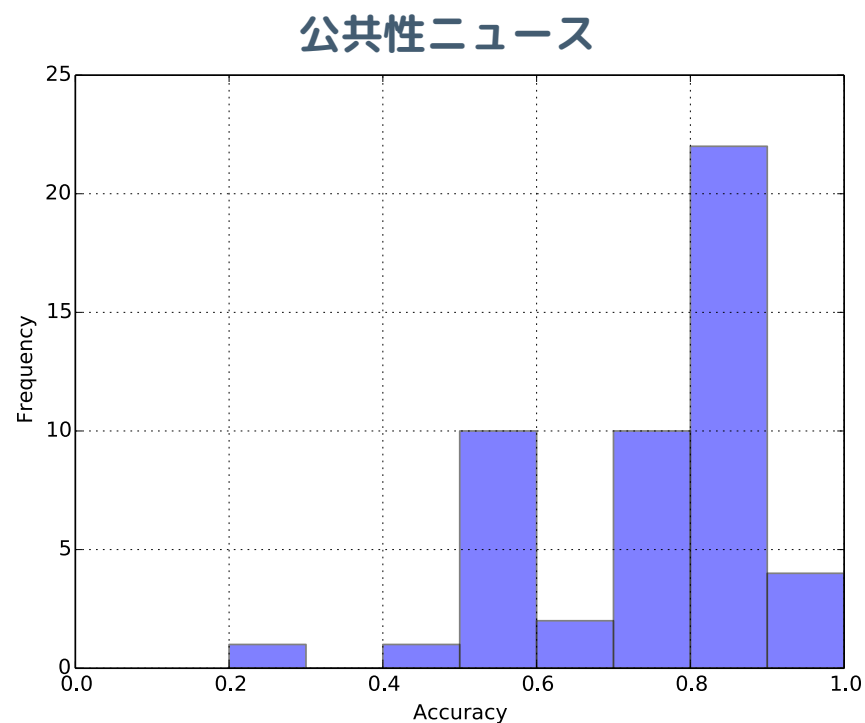
- 「載せるべき」と判断したワーカの割合を確率と解釈し、ヒストグラムをプロット
  - 公共性ニュースの方が載せるべきと判断されがちな傾向がある
  - 話題性・公共性ニュースを分けずに掲載の可否を決めるのは難しい



# 相対評価による重要ニュース選択:

かなり高精度に正負を判別できる

- 「どちらのニュースをトップページに載せるべき？」と質問
- 各ペアについての正答率に関するヒストグラムを描画
- 20人のワーカのうち18人が載せるべきものとして正例を選択していれば  $\text{Accuracy} = 0.9$
- Accuracy が 0.5 を下回るものは2つのみ

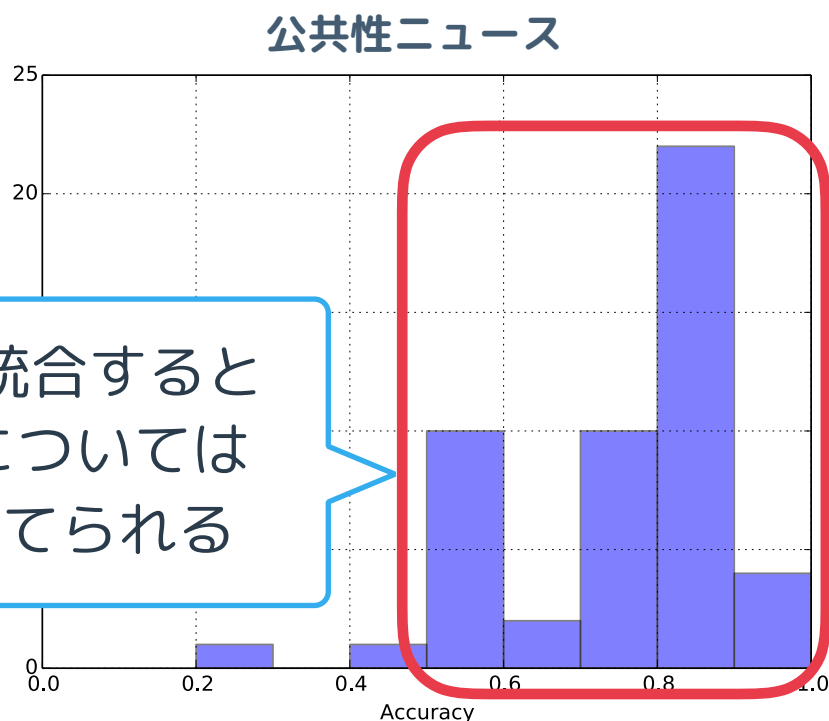


# 相対評価による重要ニュース選択:

かなり高精度に正負を判別できる

- 「どちらのニュースをトップページに載せるべき？」と質問
- 各ペアについての正答率に関するヒストグラムを描画
- 20人のワーカのうち18人が載せるべきものとして正例を選択していれば Accuracy = 0.9
- Accuracy が 0.5 を下回るものは2つのみ

相対評価の結果を統合すると  
50ペア中48ペアについては  
多数決で正解を当てられる



# Flock: Hybrid Crowd-Machine Learning Classifiers:

## 相対評価タスクによる特徴生成、機械学習による分類

- クラウドソーシングを用いて相対評価タスクから特徴を生成し、高精度な分類を行う分類器を作るプラットフォームに関する研究
  - ワーカに相対評価タスクを与え、比較結果を導いた理由を問う
  - 集まった理由を統合し、正負の判断基準を作る
  - 各オブジェクトが判断基準に一致しているか否かをワーカに問う
  - クラウドソーシングから生成された特徴量とその他の特徴量を用いて高精度な分類を行う
- 相対評価タスク → 特徴生成 というプロセスを本研究に適用する

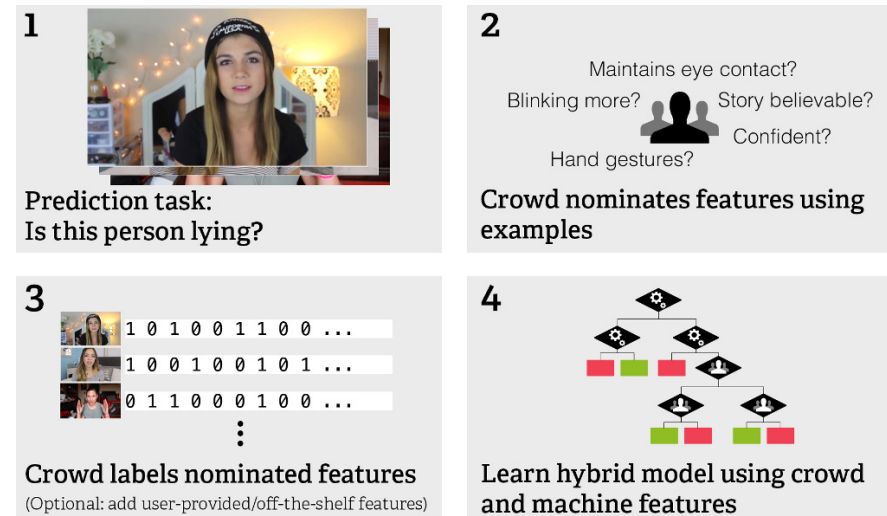


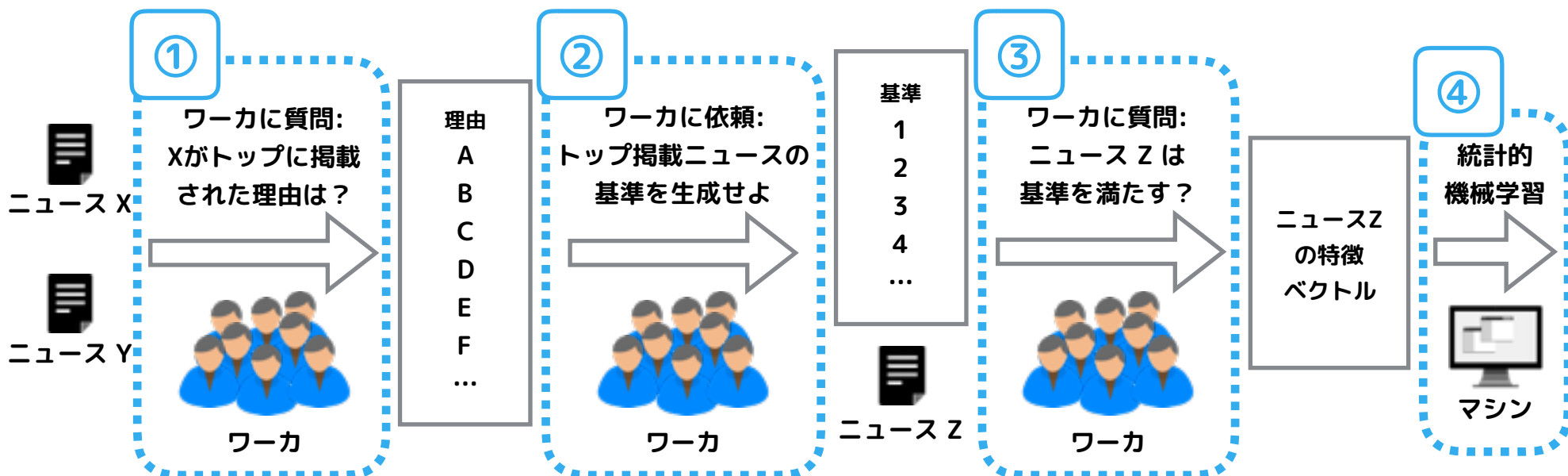
Figure 1. Flock is a hybrid crowd-machine learning platform that capitalizes on analogical encoding to guide crowds to nominate effective features, then uses machine learning techniques to aggregate their labels.

# Flock による重要ニュース選択:

## クラウドソーシングで特徴を生成、機械学習で予測



### ■ Flock のワークフローを本問題に適用

- ① ある記事が他方の記事より優れている理由を収集
- ② 理由を統合し、トップ掲載ニュースの基準を生成
- ③ 各記事の特徴ベクトルを生成
- ④ 機械学習手法で学習・予測 (positive/negative)



# Flock による重要ニュース選択: Flockによって生成された特徴のみによる予測でベースラインを上回る

- Logistic regression (LR), gradient boosting (GB) で予測
- Grid search でパラメータを探索
- 最高精度のものを下表に示す

Accuracy	話題性ニュース	公共性ニュース	
Flock	<b>0.760</b> (+/- 0.051)	<b>0.700</b> (+/- 0.059)	 : LR
Bag-of-Words	0.750 (+/- 0.078)	0.690 (+/- 0.052)	 : GB

- ベースライン: ニュース文章の Bag-of-Words で予測した結果
- Flockによって生成された特徴のみによる予測でベースラインに匹敵する精度を実現

# 結論: クラウドソーシングによる重要ニュース選択に関する実験の結果、相対評価タスクの有効性を確認

---

- 目的：クラウドソーシングによる重要ニュースの選択
- 実験により以下の事柄を示した：
  - ニュースのタイプによってクラウドワーカの意見の傾向が異なる
  - 絶対評価タスクより相対評価タスクのほうが高精度に重要ニュースを選択することができる
  - 相対評価タスクに関連する既存研究の手法を適用することで、クラウドソーシングから重要ニュース選択に有用な特徴を生成でき、それらの特徴による分類が機械学習による分類の精度に匹敵する
- 今後の課題
  - Bag-of-Words, Flock特徴を組み合わせた、より高精度な分類
  - Flock のワークフローに対する品質管理手法の適用