

# D-13

## ワーカーの確信度を用いたクラウドソーシングにおけるラベル統合

小山 聡(北海道大学), 馬場雪乃(東京大学), 櫻井祐子(九州大学), 鹿島久嗣(東京大学)

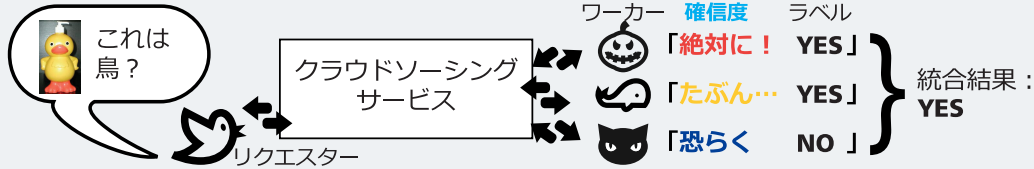
### 概要 ワーカーに答えさせた確信度情報を、クラウドソーシングのラベル統合に利用

問題：クラウドソーシングを用いて、人間にラベルづけをさせる

課題：ワーカーの能力・信頼性に差がある状況で、ラベルの正しさをどう保証するか

既存手法：同じアイテムに対して複数人にラベルづけさせ、統計的に統合 (= 真のラベルを推定)

本研究：ワーカーにラベルだけではなく「**確信度**」も回答させ、確信度情報も用いてラベルを統合



### 背景 既存研究は、ワーカーの能力等を考慮して真のラベルの推定を行う

**多数決**

どのワーカーも同等に扱う

**Dawid-Skeneモデル [Dawid&Skene 79]**

能力による重み  $\Delta$   $\Delta$   $\odot$

能力が高い人がつけたラベルほど正しいラベルだとみなす。ラベルづけ結果を利用して能力を推定。

$N$ : アイテム数,  $J$ : ワーカー数

$J_i$ : アイテム  $i$  にラベルづけしたワーカー集合

$t_i \in \{0, 1\}$ : アイテムの真のラベル

$y_{ij} \in \{0, 1\}$ : ワーカー  $j$  がつけたアイテム  $i$  のラベル

$p$ : アイテムの真のラベルが 1 である確率

$\alpha_t^{(j)}$ : ワーカー  $j$  が、真のラベルが  $t$  のときにラベル 1 をつける確率 (**能力**)

[Dawid&Skene 79] A. P. Dawid and A. M. Skene.: Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm, Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 28, No. 1 (1979)

### 提案手法 確信度をワーカーに聞く。ワーカーごとに異なる確信度パラメータを導入し既存のモデルを拡張

ラベルがどれだけ信頼できるのか、ワーカー本人が一番よくわかるはず...

**ワーカーごとに確信度の決め方が異なることを実データで確認**

- **自信過剰**なワーカー  
不正解でも高い確信度を報告
- **自信過小**なワーカー  
正解でも低い確信度を報告
- **適度な自信**をもつワーカー  
正しく自分の正解率を見積もることができる

**Worker Dependentモデル**

Dawid-Skeneモデルに以下を追加：

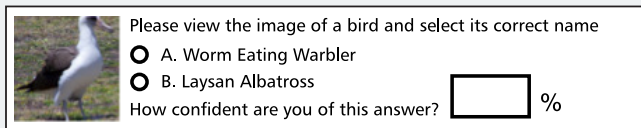
$c_{ij} \in \{0, 1\}$ : ワーカー  $j$  のアイテム  $i$  のラベルへの確信度

$\beta_{ty}^{(j)}$ : ワーカー  $j$  が、真のラベルが  $t$ 、与えたラベルが  $y$  のときに確信度 1 と答える確率

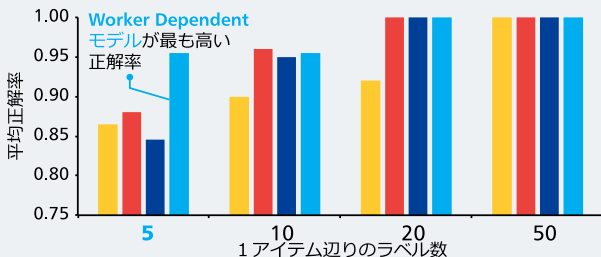
真のラベルを隠れ変数として、EMアルゴリズムで隠れ変数とモデルのパラメータを交互に推定

### 結果 1 アイテム辺りのラベル数が少ない場合には、提案手法がDawid-Skeneモデルよりも高い正解率

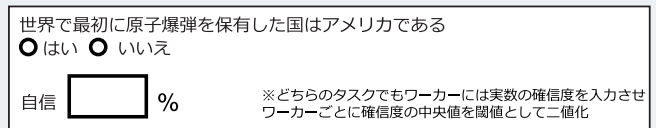
評価タスク 1：写真から鳥の名前を判定 (MTurk)



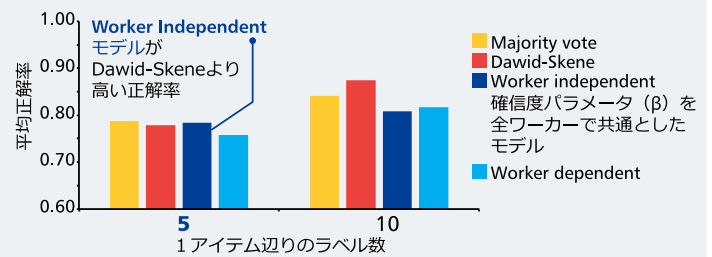
アイテム数：10、総ワーカー数：100  
1 アイテム辺りのラベル数：100



評価タスク 2：○×クイズに回答 (Lancers)



アイテム数：120、総ワーカー数：42  
1 アイテム辺りのラベル数：10



### 今後の展望 モデルの改良を検証・検討中

- 各問題ごとのクラス分布  $p_i$  をワーカーのラベルの割合から推定するモデル
- 確信度と真のラベルからワーカーのラベルが生成される ( $P(y_{ij}|c_{ij}, t_i)$ ) モデル
- 問題の難しさを考慮した既存モデル [Whitehill+ 09]に確信度を導入

[Whitehill+ 09] Whitehill, J., Ruvalo, 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 22 (2009)