

## 既存の教師データとクラウドソーシングを併用した教師付き学習

## Learning from Crowds and Experts

梶野 洸\*<sup>1</sup>    坪井 祐太\*<sup>2</sup>    佐藤 一誠\*<sup>3</sup>    鹿島 久嗣\*<sup>1</sup>  
 Hiroshi Kajino    Yuta Tsuboi    Issei Sato    Hisashi Kashima

\*<sup>1</sup> 東京大学大学院情報理工学系研究科  
 Graduate School of Information Science and Technology, The University of Tokyo

\*<sup>2</sup> IBM 東京基礎研究所  
 IBM Research - Tokyo

\*<sup>3</sup> 東京大学情報基盤センター  
 Information Technology Center, The University of Tokyo

Crowdsourcing services are often used to collect a large amount of labeled data for machine learning. Although they provide us an easy way to get labels at very low cost in a short period, they have serious limitations. One of them is the variable quality of the crowd-generated data. There have been many attempts to increase the reliability of crowd-generated data and the quality of classifiers obtained from such data. However, in these problem settings, relatively few researchers have tried using expert-generated data to achieve further improvements. In this paper, we extend three models that deal with the problem of learning from crowds to utilize ground truths: a latent class model, a personal classifier model, and a data-dependent error model. We evaluate these methods against two baseline methods on a real data set to demonstrate the effectiveness of combining crowd-generated data and expert-generated data.

## 1. 序論

機械学習は様々な分野で応用されはじめて久しいが、画像認識や自然言語処理などでは非常に大規模な訓練データを駆使する手法が提案され、訓練データを作成するコストは莫大なものとなっている。この問題を解決するため、機械学習のみならず自然言語処理 [Snow 08, Finin 10] やコンピュータビジョン [Whitehill 09, Welinder 10b, Welinder 10a] などの分野でクラウドソーシングを用いて訓練データを作成する試みが注目を集めている。この傾向は Amazon Mechanical Turk\*<sup>1</sup> (AMT) が 2005 年に登場して以来顕著に見られる。クラウドソーシングを用いて教師データを作る手法には次のような利点・欠点がある。利点として、データ作成にかかる金銭的・時間的コストを削減できるということが挙げられる。クラウドソーシングでは不特定多数の人 (ワーカーと呼ぶ) に小さい金額で仕事を依頼することができるためである。その点でクラウドソーシングは大規模な訓練データを作るのに適していると言える。一方、クラウドソーシングを行って得られるデータの信頼度がワーカーによって大きく異なるという欠点が指摘されている。ゆえにクラウドソーシングで訓練データを作成する際には、このようなワーカーの品質管理問題を解決することが不可欠となる。

このワーカーの品質管理問題に対して、2 つの問題設定がされている。1 つはデータから真のラベルを推定するという設定、もう 1 つはデータから識別器を構成するという設定である。前者の真のラベルの推定問題に対してはこれまでに多くの研究がなされている。提案された手法の多くは Repeated Labeling [Sheng 08] を元にした手法である。これは 1 つのデータに対して複数のワーカーにラベル付けを依頼し、得られた複数のラベルから真のラベルを推定する手法である。複数のラベルから真のラベルを推定する手法は様々な手法が提案されており、多数決を用いる方法や、真のラベルを潜在変数としてモデル化しワーカーの能力・問題の難しさなどもモデル化した上で

EM アルゴリズムを用いて推定する方法が用いられている。真のラベルを潜在変数としてモデルする手法は医師の診断の統合問題の文脈で提案されたのが始まりである [Dawid 79]。

後者の識別器推定問題は近年研究され始めた問題である。この問題に対する手法は 2 つに大別される。1 つは真のラベルも同時に推定する手法 [Raykar 10, Yan 10] で、もう 1 つは識別器のみを推定する手法 [Dekel 09, Wauthier 11, Kajino 12] である。機械学習では最終的に求めるべきものは識別能力の高い識別器であるため、このような問題設定は重要である。

以上の手法によりワーカーの品質管理問題は改善されるが、既存手法のほとんどが既存の教師データを併用することを考慮していないという問題がある。能力が高いと事前にわかっているワーカー (エキスパートと呼ぶ) がラベルを付けたデータを活用することで得られる識別器の性能は向上することが期待される。我々の知る限りでは、Tang ら [Tang 11] と Wauthier ら [Wauthier 11] のみこの問題を扱っている。

本論文では、クラウドソーシングを用いた学習法の中で代表的な 3 つのモデルである、識別器統合モデル [Kajino 12]、潜在ラベルモデル [Raykar 10]、データ依存潜在ラベルモデル [Yan 10] を拡張し、エキスパートによるデータを活用できる手法を提案する。実データを用いた実験でこれらの提案手法と 2 つの対抗手法を比較し、それぞれエキスパートによるデータを用いない場合と同等以上の性能となることを示す。

## 1.1 問題設定

本節では今回用いる記号と取り扱う問題を定義する。 $N$  個のデータが与えられ、それぞれのデータ  $i \in \{1, \dots, N\}$  に対して  $D$  次元特徴ベクトル  $\mathbf{x}_i \in \mathbb{R}^D$  が与えられる。その集合を  $\mathcal{X} = \{\mathbf{x}_i \mid i \in \{1, \dots, N\}\}$  とする。またデータ  $i$  に対して真のラベル  $y_i \in \{0, 1\}$  が存在すると仮定する (ただし一般的には観測されない)。このデータに  $J$  人のワーカーがクラウドソーシングを通じてラベルを付ける。ワーカー  $j \in \{1, \dots, J\}$  ( $j$  をワーカー ID と呼ぶ) がラベルを付けたデータの集合を  $\mathcal{I}_j \subseteq \{1, \dots, N\}$  とし、ワーカー  $j$  はそれぞれのデータ  $i \in \mathcal{I}_j$  に対してラベル  $y_{i,j} \in \{0, 1\}$  を付ける。このラベルは信頼度が低

連絡先: 梶野 洸, Hiroshi.Kajino@mist.i.u-tokyo.ac.jp

\*<sup>1</sup> <https://www.mturk.com/mturk/welcome>

く  $y_{i,j} = y_i$  となるとは限らない．ワーカー  $j$  から得られたラベルの集合を  $\mathcal{Y}_j = \{y_{i,j} \mid i \in \mathcal{I}_j\}$  とし，クラウドソーシングで得られたラベルの集合を  $\mathcal{Y} = \bigcup_{j=1}^J \mathcal{Y}_j$  とし，これらのラベルをクラウドラベルと呼ぶ．同様にデータ  $i$  に対してラベルを付けたワーカーの ID の集合を  $\mathcal{J}_i \subseteq \{1, \dots, J\}$  とし，データ  $i$  に付けられたラベルの集合を  $\mathcal{Y}_i = \{y_{i,j} \mid j \in \mathcal{J}_i\}$  とする．

本論文ではクラウドラベル以外に，信頼度の高いラベル「エキスパートラベル」を新たに導入する．ワーカーの他に信頼度の高いラベルを返すエキスパートが存在するとする．今回は常に真のラベルを返すエキスパートを考える．エキスパートのワーカー ID を  $j = 0$  とし，エキスパートがラベルを付けたデータの集合を  $\mathcal{I}_0 \subseteq \{1, \dots, N\}$  とし，それぞれのデータ  $i \in \mathcal{I}_0$  にラベル  $y_{i,0} (= y_i)$  を付けるとする．エキスパートから得られたラベルの集合を  $\mathcal{Y}_0 = \{y_{i,0} \mid i \in \mathcal{I}_0\}$  とする．

本論文では，訓練データ  $(\mathcal{X}, \mathcal{Y}, \mathcal{Y}_0)$  を入力として二値分類器  $f: \mathbb{R}^D \rightarrow \{0, 1\}$  を出力するという問題を考える．

## 2. 提案手法

本論文では識別器統合モデル [Kajino 12]，潜在ラベルモデル [Raykar 10]，データ依存潜在ラベルモデル [Yan 10] をそれぞれ拡張し，クラウドラベルとエキスパートラベルを併用した学習を実現する．潜在ラベルモデルとデータ依存潜在ラベルモデルは真のラベルを潜在変数としてモデル化しているためパラメタ推定問題は非凸最適化問題となるが，識別器統合モデルは真のラベルをモデル化しないことで凸最適化問題となる．

### 2.1 識別器統合モデル (Personal Classifier Model)

Kajino ら [Kajino 12] が提案した識別器統合モデルに従い各ワーカーのモデルを統合して真のモデルを推定する．

#### 2.1.1 モデル

真のモデルをパラメタ  $\mathbf{w}_0$  で表されるロジスティック回帰モデルとする：

$$p(y_i = 1 \mid \mathbf{x}_i, \mathbf{w}_0) = \sigma(\mathbf{w}_0^\top \mathbf{x}_i) = (1 + \exp(-\mathbf{w}_0^\top \mathbf{x}_i))^{-1}. \quad (1)$$

ここで  $\sigma$  をシグモイド関数とする．エキスパートラベル  $\mathcal{Y}_0$  はこのモデルに従って生成されるとする．またワーカー  $j$  のモデルをパラメタ  $\mathbf{w}_j$  で表されるロジスティック回帰モデルとする：

$$p(y_{i,j} = 1 \mid \mathbf{x}_i, \mathbf{w}_j) = \sigma(\mathbf{w}_j^\top \mathbf{x}_i) \quad (j \in \{1, \dots, J\}).$$

クラウドラベル  $\mathcal{Y}$  はこのモデルに従って生成されるとする．これらのモデルに対して，次のような関係を導入する．パラメタ  $\{\mathbf{w}_j\}_{j=1}^J$  は  $p(\mathbf{w}_j \mid \mathbf{w}_0, \lambda)$  に従って生成され， $\mathbf{w}_0$  は事前分布  $p(\mathbf{w}_0 \mid \eta)$  に従って生成されるとする ( $\lambda$  と  $\eta$  はハイパーパラメタ)．特にこれらは正規分布であるとする：

$$p(\mathbf{w}_0 \mid \eta) = \mathcal{N}(\mathbf{0}, \eta^{-1} \mathbf{I}), \quad p(\mathbf{w}_j \mid \mathbf{w}_0, \lambda) = \mathcal{N}(\mathbf{w}_0, \lambda^{-1} \mathbf{I}).$$

#### 2.1.2 パラメタ推定

訓練データが与えられた元での  $\mathbf{w}_0, \mathbf{W} = \{\mathbf{w}_j \mid j \in \{1, \dots, J\}\}$  の事後確率を最大化しパラメタ推定する． $\mathbf{w}_0, \mathbf{W}$  に関する負の対数事後確率  $F(\mathbf{w}_0, \mathbf{W})$  は，

$$\begin{aligned} F(\mathbf{w}_0, \mathbf{W}) = & - \sum_{j=0}^J \sum_{i \in \mathcal{I}_j} l(y_{i,j}, \sigma(\mathbf{w}_j^\top \mathbf{x}_i)) \\ & + \frac{\lambda}{2} \sum_{j=1}^J \|\mathbf{w}_j - \mathbf{w}_0\|^2 + \frac{1}{2} \eta \|\mathbf{w}_0\|^2, \end{aligned}$$

である．ここで  $l(s, t) = s \log t + (1-s) \log(1-t)$  とし，定数項は除いた．故にパラメタ推定問題は次の最適化問題となる：

$$\text{minimize } F(\mathbf{w}_0, \mathbf{W}) \text{ w.r.t. } \mathbf{w}_0 \text{ and } \mathbf{W}. \quad (2)$$

#### 2.1.3 アルゴリズム

最適化問題 (2) は  $\mathbf{w}_0$  を固定すると  $\mathbf{w}_j$  それぞれ独立に最適化できる．これを利用して，収束するまで  $\mathbf{w}_0$  と  $\mathbf{W}$  とを交互に最適化するアルゴリズムを提案する．これは Kajino ら [Kajino 12] によって提案されたアルゴリズムを元に行っているが， $\mathbf{w}_0$  に関する最適化が異なっていることに注意する．

### 2.2 潜在ラベルモデル (Latent Class Model)

Raykar ら [Raykar 10] が提案した潜在ラベルモデルを Tang ら [Tang 11] の手法を元に拡張する．Tang ら [Tang 11] のモデルでは特徴ベクトルを用いていない点で提案手法と異なる．

#### 2.2.1 モデル

真のモデルとしてロジスティックモデル (1) を仮定する．また各ワーカーは次のモデルでラベルを付けるとする：

$$\alpha_j = p(y_{i,j} = 1 \mid y_i = 1), \quad \beta_j = p(y_{i,j} = 0 \mid y_i = 0). \quad (3)$$

本論文ではエキスパートとしてワーカー 0 を導入し，そのパラメタを  $\alpha_0 = \beta_0 = 1$  と固定するという拡張を行った．

#### 2.2.2 パラメタ推定

モデルパラメタ  $\theta = \{\mathbf{w}_0, \{\alpha_j\}_{j=1}^J, \{\beta_j\}_{j=1}^J\}$  の推定は最尤推定で行った．尤度関数は  $p(\mathcal{Y}, \mathcal{Y}_0 \mid \mathcal{X}, \theta) = \prod_{i=1}^N [a_i p_i + b_i (1 - p_i)]$ ，と書ける．ここで  $p_i = \sigma(\mathbf{w}_0^\top \mathbf{x}_i)$ ， $a_i = \prod_{j \in \mathcal{J}_i} \alpha_j^{y_{i,j}} (1 - \alpha_j)^{1-y_{i,j}}$ ， $b_i = \prod_{j \in \mathcal{J}_i} \beta_j^{1-y_{i,j}} (1 - \beta_j)^{y_{i,j}}$  とおいた．

#### 2.2.3 アルゴリズム

$\theta$  の最尤推定量を EM アルゴリズムで近似的に求める．次に示す E ステップと M ステップを収束するまで繰り返す．E ステップ． $\mu_i = p(y_i = 1 \mid \mathcal{Y}_i, \mathbf{x}_i, \theta)$  を次のように更新：

$$\mu_i = \frac{a_i p_i}{a_i p_i + b_i (1 - p_i)}.$$

M ステップ． $\{\alpha_j\}_{j=1}^J$  と  $\{\beta_j\}_{j=1}^J$  を

$$\alpha_j = \frac{\sum_{i \in \mathcal{I}_j} \mu_i y_{i,j}}{\sum_{i \in \mathcal{I}_j} \mu_i}, \quad \beta_j = \frac{\sum_{i \in \mathcal{I}_j} (1 - \mu_i) (1 - y_{i,j})}{\sum_{i \in \mathcal{I}_j} (1 - \mu_i)},$$

と更新し， $\mathbf{w}_0$  を  $\mathbb{E}_{p(\{y_i\}_{i=1}^N \mid \mathcal{Y}, \mathcal{Y}_0, \mathcal{X}, \theta)} [\log p(\mathcal{Y}, \mathcal{Y}_0, \{y_i\}_{i=1}^N \mid \theta)]$  を最大化するように更新する．

パラメタの初期化は Raykar ら [Raykar 10] に従って， $\mu_i = \frac{1}{|\mathcal{J}_i|} \sum_{j \in \mathcal{J}_i} y_{i,j}$  とした．これは多数決を元に行っている．

### 2.3 データ依存潜在ラベルモデル (Data-Dependent Error Model)

2.2 節と同様にして Yan ら [Yan 10] が提案したデータ依存潜在ラベルモデルを拡張する．

#### 2.3.1 モデル

真のモデルとしてロジスティックモデル (1) を仮定する．Yan らはワーカーの能力がデータに依存して変化するモデルを提案した．すなわち，各ワーカーのラベル付けのモデルを

$$p(y_{i,j} \mid \mathbf{x}_i, y_i, \mathbf{u}_j) = (1 - \sigma(\mathbf{u}_j^\top \mathbf{x}_i))^{|y_{i,j} - y_i|} \sigma(\mathbf{u}_j^\top \mathbf{x}_i)^{1 - |y_{i,j} - y_i|}$$

とした．本論文ではエキスパートを次のようにモデル化した：

$$p(y_{i,0} \mid \mathbf{x}_i, y_i, \mathbf{u}_0) = 1 - |y_{i,0} - y_i|.$$

### 2.3.2 パラメタ推定

モデルパラメタの推定は最尤推定で行う．尤度関数は

$$p(\mathcal{Y}, \mathcal{Y}_0 | \mathcal{X}, \theta) = \prod_{j=0}^J \prod_{i \in \mathcal{I}_j} \sum_{y_i} p(y_{i,j} | \mathbf{x}_i, y_i, \mathbf{u}_j) p(y_i | \mathbf{x}_i, \mathbf{w}_0)$$

と書ける．ここで  $\theta = \{\mathbf{w}_0, \{\mathbf{u}_j\}_{j=1}^J\}$  とおいた．

### 2.3.3 アルゴリズム

$\theta$  の最尤推定量を EM アルゴリズムで近似的に求める．次に示す E ステップと M ステップを収束するまで繰り返す．E ステップ.  $i \notin \mathcal{I}_0$  に対して  $\mu_i = p(y_{i,0} = 1 | \mathcal{Y}_i, \mathbf{x}_i, \theta)$  を

$$\begin{aligned} \mu_i &\propto \sigma(\mathbf{w}_0^\top \mathbf{x}_i) \prod_{j \in \mathcal{J}_i} (1 - \sigma(\mathbf{u}_j^\top \mathbf{x}_i))^{1-y_{i,j}} \sigma(\mathbf{u}_j^\top \mathbf{x}_i)^{y_{i,j}}, \\ 1 - \mu_i &\propto (1 - \sigma(\mathbf{w}_0^\top \mathbf{x}_i)) \\ &\quad \prod_{j \in \mathcal{J}_i} (1 - \sigma(\mathbf{u}_j^\top \mathbf{x}_i))^{y_{i,j}} \sigma(\mathbf{u}_j^\top \mathbf{x}_i)^{1-y_{i,j}}, \end{aligned}$$

と更新し,  $i \in \mathcal{I}_0$  に対して  $\mu_i = y_{i,0}$  と更新する．

M ステップ.  $\sum_{j=1}^J \sum_{i \in \mathcal{I}_j} \mathbb{E}_{p(y_{i,0} | \mathbf{x}_i, \mathcal{Y}_i)} [\log p(y_{i,j}, y_{i,0} | \mathbf{x}_i)]$  を最大化して  $\theta$  を更新する．本論文では Yan ら [Yan 10] のアルゴリズムに従い L-BFGS 法を用いた．パラメタは Yan ら [Yan 10] に従い,  $\mathbf{w}_0 = \mathbf{0}$ ,  $\{\mathbf{u}_j\}_{j=1}^J$  はランダムに初期化した．

## 3. 実験

クラウドラベルとエキスパートラベルを組み合わせたことの有効性を示すために, 実データを用いた実験を行った．

### 3.1 データセット

固有表現抽出 (NER) の問題のデータセットを用いた．固有表現抽出は人名や地名などの固有表現を抽出する問題である．Finin ら [Finin 10] が AMT を用いて Twitter の文章に固有表現のラベルを付けたデータ<sup>\*2</sup>から訓練データ・テストデータを作った．このデータは通常の NER のデータとは異なり固有表現の境界情報は与えられていないため今回は各トークンが固有表現であるか ( $y_i = 1$ ) 否か ( $y_i = 0$ ) の二値分類問題とした．また Twitter で見られる「@(ユーザ名)<sup>\*3</sup>」という表現は固有表現抽出が容易なので, Ritter ら [Ritter 11] に倣い固有表現として取り扱わなかった．データ (トークン) 数は 212,720 個で, その内 8,107 個のデータにエキスパートがラベルを付けた．エキスパートラベルが付いているデータには他に 10 人以上のワーカーがラベルを付けており, それ以外のデータにはそれぞれ 2 人のワーカーがラベルを付けた．ワーカーは合計で 269 人である．それぞれのトークンに対して, Ritter ら [Ritter 11] と同様に 161,901 次元の特徴ベクトルを作成した．特徴空間の次元を削減するため 1 つのデータにしか使われない次元は用いなかった．

### 3.2 実験設定

訓練データ中のエキスパートラベル数  $|\mathcal{I}_0|$  を変化させて識別器の性能の変化を調べる．まず訓練データ・テストデータの作り方を説明する．エキスパートラベルが付いていないデータからは, 42 人のワーカーがラベルを付けたデータを 17,747 個選び訓練データとした．エキスパートラベルが付いているデータからは訓練データとテストデータを作った．エキスパートラ

ベルが付いているデータの中で, 上記の 17,747 個のデータと共通のワーカーがラベル付けを行った 2,790 個のデータを訓練データとし, 残りの 5,317 個をテストデータとして用いた．2,790 個のデータのうち, クラウドラベルは常に訓練データとし, 2,790 個のエキスパートラベルから  $|\mathcal{I}_0|$  個のエキスパートラベルをランダムに抜き出して訓練データとした． $|\mathcal{I}_0|$  は 0 から 2,750 まで 250 ずつ変化させた．

この訓練データを用いてそれぞれの手法で識別器を作り, テストデータで適合率, 再現率, F 値を計算して評価を行った．訓練データの構築・学習・評価を 10 回繰り返し, 適合率, 再現率, F 値の平均と標準偏差を計算した．

### 3.3 対抗手法

対抗手法として多数決法 (Majority Voting Method) と単純学習法 (All-in-One-Classifier Method) を用いる．多数決法はクラウドラベル  $\{y_{i,j} | j \in \mathcal{J}_i\}$  から多数決を用いて

$$y_i = \begin{cases} 1 & \text{if } \sum_{j \in \mathcal{J}_i} y_{i,j} > |\mathcal{J}_i|/2, \\ 0 & \text{if } \sum_{j \in \mathcal{J}_i} y_{i,j} < |\mathcal{J}_i|/2, \\ \text{random} & \text{otherwise.} \end{cases}$$

と真のラベル  $y_i$  を推定する手法である．エキスパートラベル  $y_{i,0}$  が付いているデータに対しては  $y_i = y_{i,0}$  とした．単純学習法はワーカー ID を用いずにすべてのデータから 1 つの識別器を学習する手法である．ここでエキスパートラベルは 1 人のワーカーが付けたラベルだと考えて学習を行った．

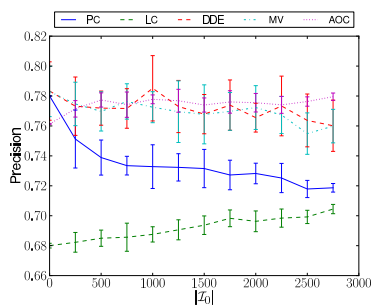
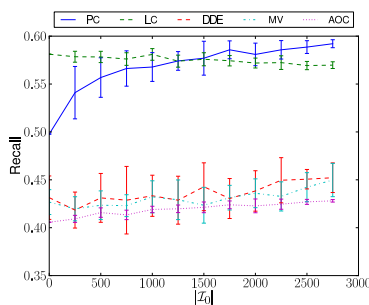
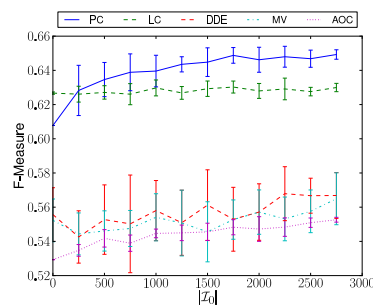
### 3.4 実験結果

それぞれの手法での適合率, 再現率, F 値の平均と標準偏差をそれぞれ図 1, 2, 3 にまとめた．識別器統合モデル, 潜在ラベルモデル, データ依存潜在ラベルモデル, 多数決法, 単純学習法はそれぞれ PC, LC, DDE, MV, AOC にあたる．これらの実験結果には以下の 3 つの傾向が見られた．(i) 識別器統合モデルと単純学習法では  $|\mathcal{I}_0|$  が増加するに従って F 値が大きくなった．(ii) データ依存潜在ラベルモデルと多数決法では  $|\mathcal{I}_0|$  が増加すると F 値も大きくなっているようだが他の手法と比べて分散が大きい．(iii) 潜在ラベルモデルの性能は F 値で測ると変化していない．

これらの観察から, 今回比較を行った手法は次のような特徴があることがわかる．(i) より, 識別器統合モデルと単純学習法はエキスパートラベルを活用できるモデルだが, F 値およびその変化量から, 識別器統合モデルの方がより効果的に用いることができると言える．(ii) より, データ依存潜在ラベルモデルと多数決法は不安定な手法であるといえる．1 つのデータに対して 2 つしかラベルが付けられていなかったため, 多数決法は不安定になったと考えられ, またデータ依存潜在ラベルモデルはパラメタ推定が非凸最適化問題となる上に初期値をランダムに設定したため, このような結果になったと考えられる．(iii) より, 潜在ラベルモデルはエキスパートラベルがない場合でもモデルパラメタをうまく推測できると言える．実際にエキスパートラベルがある場合とない場合とで推定されたパラメタを比較しても大きな違いは見られなかった．これは訓練データ数が多かったこと, 能力の低いワーカーが少なかったこと, 初期値を決定的に定めていることが原因であると考えられる．一方で Kajino ら [Kajino 12] は, データ数が少ない場合と能力の低いワーカーが多い場合との実験を行い, そのような場合にはうまく推定ができないことがあることを示している．

\*2 データセットは以下のサイトで入手可能である <http://sites.google.com/site/amtworkshop2010/data-1>

\*3 「@(ユーザ名)」はユーザに宛てた文章であることを表す．

図 1: 適合率と  $|I_0|$  とのグラフ .図 2: 再現率と  $|I_0|$  とのグラフ .図 3: F 値と  $|I_0|$  とのグラフ .

#### 4. 結論

本論文ではクラウドソーシングを用いた学習法の中で代表的な3つのモデルを拡張し、クラウドラベルとエキスパートラベルを併用した学習を実現した。識別器統合モデル [Kajino 12] では、真のモデルに対して訓練データを導入した。潜在ラベルモデル [Raykar 10] とデータ依存潜在ラベルモデルでは、Tangら [Tang 11] の手法を元にして拡張を行った。実データを用いた実験を行い、いずれの手法もエキスパートラベルを用いない場合と同等以上の性能となることを示した。

#### 謝辞

本研究の一部は最先端研究開発支援プログラム (FIRST) の助成を受けたものである。

#### 参考文献

- [Dawid 79] Dawid, A. P. and Skene, A. M.: Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm, *J. of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 20–28 (1979)
- [Dekel 09] Dekel, O. and Shamir, O.: Vox Populi: Collecting High-Quality Labels from a Crowd, in *Proc. of the 22nd Annu. Conf. on Learning Theory* (2009)
- [Finin 10] Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., and Dredze, M.: Annotating Named Entities in Twitter Data with Crowdsourcing, in *Proc. of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 80–88 (2010)
- [Kajino 12] Kajino, H., Tsuboi, Y., and Kashima, H.: A Convex Formulation for Learning from Crowds, in *Proc. of the 26th AAAI Conf. on Artificial Intelligence* (2012)
- [Raykar 10] Raykar, V. C., Yu, S., Zhao, L. H., Florin, C., Bogoni, L., and Moy, L.: Learning From Crowds, *J. of Machine Learning Research*, Vol. 11, pp. 1297–1322 (2010)
- [Ritter 11] Ritter, A., Clark, S., and Etzioni, O.: Named Entity Recognition in Tweets: An Experimental Study, in *Proc. of the 2011 Conf. on Empirical Methods in Natural Language Processing*, pp. 1524–1534 (2011)
- [Sheng 08] Sheng, V. S., Provost, F., and Ipeirotis, P. G.: Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers, in *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 614–622 (2008)
- [Snow 08] Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y.: Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks, in *Proc. of the Conf. on Empirical Methods in Natural Language Processing*, pp. 254–263 (2008)
- [Tang 11] Tang, W. and Lease, M.: Semi-Supervised Consensus Labeling for Crowdsourcing, in *ACM SIGIR Workshop on Crowdsourcing for Information Retrieval* (2011)
- [Wauthier 11] Wauthier, F. L. and Jordan, M. I.: Bayesian Bias Mitigation for Crowdsourcing, in *Advances in Neural Information Processing 24*, pp. 1800–1808 (2011)
- [Welinder 10a] Welinder, P., Branson, S., Belongie, S., and Perona, P.: The Multidimensional Wisdom of Crowds, in *Advances in Neural Information Processing Systems 23*, pp. 2424–2432 (2010)
- [Welinder 10b] Welinder, P. and Perona, P.: Online crowdsourcing: Rating annotators and obtaining cost-effective labels, in *Computer Vision and Pattern Recognition Workshops, CVPR Workshops, IEEE Computer Society Conf. on*, pp. 25–32 (2010)
- [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 22*, pp. 2035–2043 (2009)
- [Yan 10] Yan, Y., Rosales, R., Fung, G., Schmidt, M., Hermosillo, G., Bogoni, L., Moy, L., Dy, J., and Malvern, P.: Modeling annotator expertise: Learning when everybody knows a bit of something, in *Int'l Conf. on Artificial Intelligence and Statistics*, Vol. 9, pp. 932–939 (2010)