

## クラウドソーシングによるPOI収集のための品質管理手法

## A Method of Quality Control for Crowdsourced POI Collection

梶村 俊介\*<sup>1</sup>      馬場 雪乃\*<sup>2\*3</sup>      鹿島 久嗣\*<sup>4</sup>  
 Shunsuke Kajimura      Yukino Baba      Hisashi Kashima

\*<sup>1</sup>東京大学大学院情報理工学系研究科数理情報学専攻  
 Department of Mathematical Informatics, The University of Tokyo

\*<sup>2</sup>国立情報学研究所ビッグデータ数理国際研究センター  
 Global Research Center for Big Data Mathematics, National Institute of Informatics

\*<sup>3</sup>JST, ERATO, 河原林巨大グラフプロジェクト  
 JST, ERATO, Kawarabayashi Large Graph Project

\*<sup>4</sup>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, Kyoto University

Crowdsourcing, which enables us to ask an unspecified large number of people to complete some tasks via the Internet, is attracting increasing attention. Crowdsourcing is used in various business fields for collecting a large amount of data at low cost; however, it has a problem that there are large variations in the qualities of workers' products. In this paper, we consider POI (points of interest) collection tasks, which are tasks to collect positional information of specific spots that someone may find useful or interesting. In crowdsourced POI collection, workers neither necessarily provide correct spots nor provide exactly the same coordinates even if they mean a same spot. We propose a two-stage quality control method for POI collection tasks consisting of constraint exemplar clustering and reliability estimation, and demonstrate its effectiveness by experiments on an actual crowdsourcing service.

## 1. はじめに

クラウドソーシングにおいてはワーカーによる成果物の品質にばらつきが生じるため、品質のばらつきをいかに抑えるかという品質管理問題が盛んに研究されている。クラウドソーシングにおいてしばしば依頼されるタスクとして、真の回答が複数存在し、ワーカーができるだけ多くの回答を挙げる JPOI 収集が存在する。そこで、本研究では POI 収集に対する品質管理の手法を提案する。

クラウドソーシング (crowdsourcing) とは、インターネットを介して不特定多数のワーカー (crowd) に業務を外部委託 (outsourcing) する仕組みのことである。安価で大量のデータが取得できるクラウドソーシングは、近年インターネットの普及とともに様々なビジネス分野において利用が拡大している。適用分野は多岐に渡り、文書作成、デザイン作成の依頼等に利用され、また計算機科学分野においては人間には可能だが機械には難しい領域である言語処理、画像理解、音声認識等の分野において利用される。クラウドソーシングは不特定多数のワーカーに業務を依頼するという性質上、ワーカーの能力ややる気等によって得られる成果物の品質にばらつきが生じる。成果物の品質管理の方法としては、作業を複数の依頼人に委託して冗長化させることによって、得られた成果物を統計的に品質管理する手法が盛んに研究されている [4]。

クラウドソーシングにおけるはじめての統計的な品質管理手法は、複数医師の診断から適切と思われる診断結果を推定するモデル [1] を後にクラウドソーシングの品質管理に適用したものである。このモデルでは、複数の選択肢から一つを選ぶ択一選択型のタスクを扱っており、ワーカーの能力と真の回答

を同時推定している。このモデルに対して様々な拡張研究が行われている。新たな要素をパラメータとして導入しているものとして、Whitehill らが提案した、タスクの難易度をパラメータとして取り入れたモデル [9]、Welinder らが提案した、ワーカーの各単位タスクに対する得手不得手までを考慮したモデル [8] 等がある。また、タスクの種類を拡張した研究も行われており、Lin らはワーカーが実数で回答するタスクを対象にしたモデル [6] を提案している。

本研究では、タスクの種類を拡張して POI (point of interest) 収集にクラウドソーシングを適用することを考える。POI とは、地図上で誰かが便利あるいは興味があると思った特定の地点のことを指す。地図を利用するシーンは、不動産、飲食、旅行など様々な場面に広がっており、今後いっそう POI 収集の必要性が高まると考えられる。こうした POI 収集を自力で行うのは大変困難であるので外部委託することが一般的に考えられ、実際にクラウドソーシングを用いた例 [3] も存在する。POI 収集の具体例としては「東京都 23 区内にある 24 時間営業のセルフガソリンスタンドの列挙」、「京都市内にある寺院の列挙」などのタスクが考えられる。

## 2. 問題設定

本節では、本研究において扱ったクラウドソーシングによる POI 収集の問題設定を定義する。POI 収集においては条件を満たす正解が複数存在し、回答者はその問題に対して自分の思いっかかり回答を行える。このとき各ワーカーは自分の思いっ限り回答してよいが、重複回答は許されないものとする。このとき、ワーカー  $w_i$  ( $i = 1, \dots, W$ ) が列挙した回答集合を  $T_i$  とすると、全ワーカーによって列挙された回答集合は  $T_1 \cup T_2 \cup \dots \cup T_W$  と表すことができる。このとき、我々の目標はワーカーが挙げた回答集合  $\{T_i\}$  ( $i = 1, \dots, W$ ) を単に

連絡先: 梶村俊介, 東京大学大学院情報理工学系研究科数理情報学専攻, 6466555914@mail.ecc.u-tokyo.ac.jp

組み合わせたものよりもより精度の高い項目リスト  $\mathcal{P}$  を生成することである。つまり問題設定は、入力がワーカーが挙げた項目集合  $\{T_i\}$  ( $i = 1, \dots, W$ )、出力が項目リスト  $\mathcal{P}$  と定義される。

### 3. 提案手法

本研究では、クラウドソーシングによる POI 収集において、各ワーカーから挙げられた項目の重複からワーカーの信頼度を評価してより精度の高い項目リストを生成することを目標とする。項目リストの精度向上を目指す際、是正すべき不確定性を考える必要があるが、今回の POI 収集においては 2 種類の不確定性が存在すると考えられる。まず 1 つ目は、ワーカーは同一施設を指していたとしても厳密に同じ座標を指し示すとは限らないため、座標の数値的なゆらぎが発生するということがある。そして、2 つ目はワーカーが条件を満たす正しい施設を指しているとは限らないということである。

そこで、まずクラスタリングを行うことによって同じ項目を指している回答同士を同一クラスターに属させる。そのあとワーカーと列挙された項目の信頼度付けを行うことによりより精度の高い項目リストを生成する。ゆえに、品質管理手法全体の方針としては Algorithm 1 のように表される。

---

#### Algorithm 1 提案する品質管理手法

---

- 1: **Input:**  $\{T_i\}_{i=1, \dots, W}$
  - 2: **Output:**  $\mathcal{P}$
  - 3:  $Q \leftarrow$  クラスタリング ( $\{T_i\}$ )
  - 4:  $\mathcal{P} \leftarrow$  ワーカーと項目の信頼度付け ( $Q$ )
- 

#### 3.1 クラスタリング

今回の研究で用いるクラスタリング手法として、データ点上にクラスター中心をとる exemplar clustering を利用した。これは POI 収集のタスクにおいて、多くの場合に POI として指される項目同士の間点が POI として不適切な位置となり、データ点同士の間点を中心にとると意味のある解にならないためである。この問題は exemplar clustering を適用することによって解消される。

ただし、ワーカーは重複して項目を列挙しないという仮定より、同一ワーカーから挙げられた項目が同一クラスターに属してはいけないという cannot-link 制約を付加して制約付きクラスタリングを行う必要がある。また、例えばある建物の情報を保持していたとすると、その建物を指しているポイントは全て同一施設を指し、同一クラスターに属さなければならないという mut-link 制約も加えられる。従来の制約付きクラスタリングとして代表的な手法は COP-Kmeans [7] があるが、この手法はデータ点の間の位置にクラスター中心をとってしまうため本研究における手法としては不適切である。そこで、本研究において扱うクラスタリングは、データ点同士の非類似度を用いた凸最適化問題として定式化した手法である、Elhamifar らによる exemplar clustering [2] に制約を付加したものを用いた。

##### 3.1.1 exemplar clustering

Elhamifar らによると、 $N$  個のデータ点に対し、データ点同士の非負の非類似度行列を  $D = \{d_{ij}\}_{i,j=1, \dots, N}$ 、データ点  $i$  がデータ点  $j$  の代表点となる確率を示す変数  $z_{ij}$  を要素とする

行列を  $Z$ 、その第  $i$  行を  $\mathbf{z}_i^T$  とすると exemplar clustering は、

$$\begin{aligned} \min \quad & \sum_{j=1}^N \sum_{i=1}^N d_{ij} z_{ij} + \lambda \sum_{i=1}^N \|\mathbf{z}_i\|_q \\ \text{s.t.} \quad & \sum_{i=1}^N z_{ij} = 1, \forall j, \quad z_{ij} \geq 0, \forall i, j \end{aligned}$$

と  $l_q$  ノルムを用いて定義することができる。このとき、この最適化問題が凸性を保持するために  $q \in \{2, \infty\}$  とする。

##### 3.1.2 制約付き exemplar clustering

前節の exemplar clustering の定式化に cannot-link 制約、must-link 制約を付加する。まず、cannot-link 制約を付加するためには、同一ワーカーによって挙げられたデータ点の他点への帰属確率の割り当てにおいて、各点に対する帰属確率の和が 1 以下であるという制約を加えればよい。 $C$  に含まれる点集合のうちどの二つも同一クラスターに属してはいけないという cannot-link 制約を付加されているとし、 $C$  の集合を  $\mathcal{C}$  とする。データ点数を  $N$ 、 $C$  に含まれる点が点  $i$  を代表点とする帰属確率の総和を  $c_i$  として  $\mathbf{c} = (c_1, c_2, \dots, c_N)^T$ 、帰属確率行列を  $Z$  の第  $j$  列ベクトルを  $\mathbf{z}'_j$  とし  $Z = [\mathbf{z}'_1 \dots \mathbf{z}'_N]$  と書くと、cannot-link 制約は、 $\|\mathbf{c}\|_\infty = \left\| \sum_{j \in C} \mathbf{z}'_j \right\|_\infty \leq 1, \forall C \in \mathcal{C}$  として定式化できる。また、今回の提案手法であればこの must-link 制約も制約条件として加えることができる。cannot-link 制約の際と同様に、must-link 制約を  $M$  に含まれる点集合同士が同一クラスターに属さなければならないという条件として表すとし、 $M$  の集合を  $\mathcal{M}$  とすると、must-link 制約に含まれる点同士は同一の帰属確率ベクトルをもつはずであるから、 $\mathbf{z}'_i = \mathbf{z}'_j, \forall i, j \in M, \forall M \in \mathcal{M}$  という条件を付加すればよい。ゆえに、制約付き exemplar clustering は、

$$\begin{aligned} \min \quad & \sum_{j=1}^N \sum_{i=1}^N d_{ij} z_{ij} + \lambda \sum_{i=1}^N \|\mathbf{z}_i\|_q \\ \text{s.t.} \quad & \sum_{i=1}^N z_{ij} = 1, \forall j, \\ & z_{ij} \geq 0, \forall i, j, \\ & \left\| \sum_{j \in C} \mathbf{z}'_j \right\|_\infty \leq 1, \forall C \in \mathcal{C}, \\ & \mathbf{z}'_i = \mathbf{z}'_j, \forall i, j \in M, \forall M \in \mathcal{M} \end{aligned}$$

と定式化される。

##### 3.1.3 非類似度行列のスプース化

本研究においては、適当なカーネルを用いて類似度を定義した後、符号反転を行って非類似度を定義し、 $k$  近傍グラフを考えて除外される辺の非類似度を 0 とした。このようにグラフを再定義することで、非類似度行列をスプース化することができる。POI 収集においては扱うデータが座標であるため、今回はガウシアンカーネルを用いて類似度を定義した。

#### 3.2 信頼度付け

本節では、ワーカーから挙げられた項目リストは全て正解であるとは限らないため、ワーカーの信頼度付けを行ってより精度の高い項目リストを得ることを考える。信頼度の高いワーカーから挙げられた項目は数多く他ワーカーの列挙項目と一致する一方、信頼度の低いワーカーから挙げられた項目は他ワーカーのものとは一致しにくいと考えることができる。前節で、同

一項目に属する回答同士をクラスタリングしたことによる出力結果として点  $i$  が点  $j$  の代表点となる確率  $Z = \{z_{ij}\}$  を得ることで、この信頼度付けを行うことが可能になる。ここで列挙項目集合  $U$  は、 $\|z_i\| \neq 0$  なる条件を満たす  $i$  の集合と定める。この列挙項目集合とクラスタリング結果である各点の他点への帰属確率  $Z$  から、より精度の高い項目リスト  $\mathcal{P}$  を生成することを目指す。本研究では、ワーカーと項目への信頼度付けを行う際、Web ページにおける信頼度伝播のモデルである HITS [5] を修正して用いることとした。

### 3.2.1 HITS

HITS はリンク構造の解析によって authority と hub を決定するため、ページ  $i$  に authority の重み  $a_i$ 、hub の重み  $h_i$  を与え、これらの値を繰り返し処理によって更新してその収束値を求める。authority の重みが大きいページは、hub と認識されるページから多くリンクされており、hub の重みが大きいページは、authority と認識されるページにリンクに多くリンクしているという性質をもつとする。

このアルゴリズムは、各ページの authority、hub の重みを成分としてもつベクトルをそれぞれ  $\mathbf{a} = (a_1 \cdots a_n)^T$ 、 $\mathbf{h} = (h_1 \cdots h_n)^T$  と表し、ページ  $i$  からページ  $j$  へのリンクが存在する場合には  $l_{ij} = 1$  とし、それ以外は 0 とした隣接行列を  $L = [l_{ij}]$  とすると、authority、hub の重み更新を、 $\mathbf{a} = \alpha L^T \mathbf{h}$ 、 $\mathbf{h} = \beta L \mathbf{a}$  を定める。ただし、 $\alpha, \beta$  は適当な正規化定数とする。 $c$  を適当な正規化定数としてこれらをまとめると、 $\mathbf{a} = c L^T L \mathbf{a}$  となり、これは  $L^T L$  の最大固有値に対応する固有ベクトルを求める問題に帰着される。

### 3.2.2 修正 HITS

本研究においては、authority と列挙項目、hub とワーカーという類似性に着目して列挙項目とワーカーへの信頼度付けを行った。 $\mathbf{w}, \mathbf{s}$  をそれぞれワーカーの信頼度、列挙項目の信頼度を成分としたベクトルとし、クラスタリング後の列挙項目数を  $N'$  とする。今回のクラスタリング結果における  $Z$  の第  $i$  行を  $\mathbf{z}_i^T$  として  $\|z_i\| \neq 0$  となる  $i$  を用いれば、データ点  $i$  が列挙項目  $j \in U$  に帰属する確率を要素とする行列は、 $\tilde{Z} = [z_{i_1}, z_{i_2}, \dots, z_{i_{N'}}]$  と計算される。この  $\tilde{Z}$  の第  $i$  行を  $\tilde{z}_i^T$  とすれば、隣接行列  $L$  の第  $i$  行  $\mathbf{l}_i^T$  はワーカー  $i$  の列挙項目集合  $T_i$  を用いて、 $\mathbf{l}_i = \sum_{j \in T_i} \tilde{z}_j$  として求められる。この隣接行列  $L$ 、ワーカー、列挙項目それぞれの信頼度  $\mathbf{w}, \mathbf{s}$  により、本研究における信頼度付けモデルは HITS と同様に最大固有値問題として定式化できる。

更に本研究においては、ランダムに回答を行ったが偶然にも他のワーカーと回答が一致することによって信頼度が上がってしまう可能性を是正する。この現象が起きる確率はワーカーの回答数に比例した形で生じると考えられるので、ワーカーの回答数を  $\mathbf{n} = (n_1, \dots, n_W)^T$ 、 $\gamma$  を適当な係数としたとき、 $k_{ij} = \max((LL^T)_{ij} - \gamma n_i n_j, 0)$  として表される  $k_{ij}$  を要素とする  $K$  を  $L$  の代わりに用いることで解決する。

## 4. 評価実験

第 3 章で述べたクラウドソーシングによる POI 収集に対する品質管理手法を実データに対して適用した結果を示す。クラウドソーシングにおける POI 収集に対する品質管理の既存手法は未だ存在しないため、提案手法に対する比較対象としては得られた実データをそのまま適用したもの、第 3 章で提案したクラスタリング手法のみを適用したもの二つとした。また、本研究においてはクラスタリング手法における  $l_q$  ノルムを  $q = \infty$  とし線形計画問題を解いた。

### 4.1 データセットの概要

本実験においては、クラウドソーシングサービス Lancers を利用して実際の POI 収集を行った。ワーカー依頼するタスクにおいては、特定地域のみが赤枠で括られたグーグルマップの URL を掲載し、要求する POI の条件を満たす地点の緯度、経度をグーグルマップ上から抽出して入力させるものとした。各タスクの回答数の制限は基本的には設けず、ワーカーは自分の思いっただけ POI を入力できる形式とした。また、重複回答は行わないようにするよう注意書きを行った。それぞれのタスクの内容は表 1 に示した。

表 1: クラスタリング手法の比較

タスク	内容	回答数	ワーカー数
タスク 1	新橋駅周辺の公衆電話	133	5
タスク 2	高松駅周辺のうどん屋	63	7
タスク 3	新宿駅周辺の無料トイレ	82	11

### 4.2 正解の判定

正解リストを  $A = (a_1, a_2, \dots, a_n)$  と表したとき、適当な閾値  $d$  に対して項目  $v$  が正解であるとは、 $\exists a \in A, \|v - a\|_2 < d$  が成立することとし、これによって各項目が正解か否かの判定を行う。ただし、各正解項目に対して列挙項目は一つしか正解と判定できないものとし、2 つ以上の列挙項目が重複して同じ正解項目を指している場合は、そのうち一つのみが正解となるように定めた。正解データセットは、インターネット上で個人的に、あるいは企業としてそうした POI 収集を行っているサイトを正解とすることとした。具体的にはそれぞれ、公衆電話チズ\*1、食べログ\*2、トイレの三ツ星\*3における情報を正解データセットとした。

### 4.3 評価結果

#### 4.3.1 スпамワーカー数を変化させた場合

本研究において利用した Lancers は国内のクラウドソーシングサービスであり、報酬のみを目的としてランダムに回答を行うスパムワーカーの数が少なく品質管理を行わずともある程度の品質が期待できる。一方、海外のクラウドソーシングサービスである Amazon Mechanical Turk 等は Lancers に比べて信頼度の低いワーカーの数が多く、本研究において扱う POI 収集タスクを依頼するとランダムにプロットするスパムワーカーが数多く発生すると考えられる。そこで、そうしたスパムワーカーを人工的に発生させて評価実験を行うことにより、スパムワーカーが多くなった際のデータに対する品質管理手法の評価を行うこととした。

各タスクに対して、全データ数を回答したワーカー数で割ったワーカー当たりの平均回答数を  $\nu$  とし、人工スパムワーカーの回答数は平均  $\nu$  のポアソン分布によってサンプリングされた数  $x \sim Po(\nu)$  とした。上記の処理はスパムワーカーの数を定めた上で 20 回行い、それぞれに対して品質管理手法を適用し、その F 値の 95% 点を求めた。これをスパムワーカーの数を変化させながら行い、提案する品質管理手法の評価を行った。タスク 1 においてはスパムワーカーは 0 人～5 人、残り二つについてはスパムワーカーは 0 人～10 人と変化させた。修正 HITS

\*1 <http://www.telmap.net/>

\*2 <http://tabelog.com/>

\*3 <http://toilet.blog.shinobi.jp/>

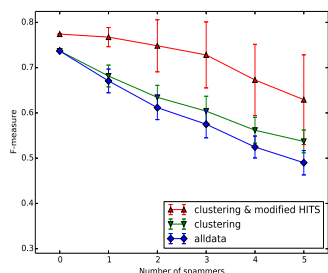


図 1: タスク 1

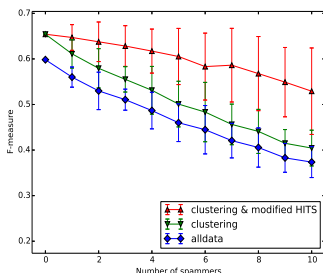


図 2: タスク 2

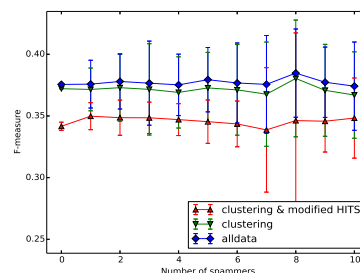


図 3: タスク 3

における係数  $\gamma$  は、データ点数  $N$  を用いて  $\gamma = 1/2N$  とし、正解判定に用いる距離閾値は  $d = 3 \times 10^{-4}$  として提案する品質管理手法の F 値を評価し、データをそのまま利用した場合、クラスタリングのみを行った場合との比較を行い、図 1~3 に示した。F 値は全て 95% 点までを表した。尚、クラスタリングにおける正規化パラメータ  $\lambda$  は  $\lambda = 0.1$ 、 $k$ -近傍グラフの  $k = 5$ 、ガウシアンカーネルの  $\sigma = 1 \times 10^{-3}$  とし、信頼度付けモデルにおいて列挙項目を項目リストに含めるか否かの判定に用いる閾値は、 $\epsilon = 1 \times 10^{-2}$  を用いた。評価結果を見ると、タスク 1 とタスク 2 においてはクラスタリング、提案手法が上手く機能していることがわかる。この 2 つのタスクにおいては、まずクラスタリングをすることにより F 値が向上し、さらに信頼度付けを行うことによって F 値がもう一度向上している。本研究で考えた 2 種類の不確定性は、同一項目を指す回答間にばらつきが存在すること、そしてワーカーが誤った項目を列挙してしまう可能性があることであるが、本研究における 2 段階の提案手法はこの 2 種類の不確定性を是正していることがわかる。またスパムワーカー数が増えるにつれ、提案手法と他の手法との F 値の差が広がっており、スパムワーカーに対しても頑健性を保持していることがグラフから読み取れる。

一方、タスク 3 では、クラスタリング、信頼度付けモデルが上手く機能しているとは言い難い。これは本研究で用いた信頼度付けの手法は信頼度の高い項目を列挙すればするほどワーカーの信頼度が高まる仕組みとなっており、信頼度の低い項目を列挙してもワーカーの信頼度が下がる要因とはならない。タスク 3 においては施設として適当そうに見える場所を地図上のみから判断して列挙したと思われるワーカーが存在している。そうした正答率の低いワーカーの信頼度が非常に高くなってしまい、正答率が高いが少ししかプロットしないワーカーの信頼度が下がってしまった事が原因である。しかし POI 収集問題においては、ワーカーが他人と一致しない項目を列挙しているのは他人が知らないから一致しないのか、あるいは本当に誤った回答をしているのかを区別することができない。ゆえに、本研究の信頼度付けの手段としては他人と一致しない項目を列挙していても信頼度降下のペナルティを課さないこととした。

## 5. 結論

本研究では、クラウドソーシングにおいて POI 収集を扱う際の品質管理手法を提案した。その際、データの品質のゆれの原因として、ワーカーが正しい項目を列挙するとは限らないこと、また同じ項目を指す回答を行うワーカー同士においても数値的なゆれがあること、この二つを考慮した。そこで、クラスタリングと信頼度付けの 2 段階の提案手法により品質管理を試みた。

実際のクラウドソーシングサービスを利用して評価実験を行ったところ、品質管理を行わずに得られたデータをそのまま利用する場合、クラスタリングのみを行った場合と比べて、提案した品質管理手法を適用した場合は精度の向上が見られた。これにより考慮した 2 種類の不確定性が 2 段階の提案手法により是正されていることがわかる。また人工スパムワーカーを発生させて評価を行った際は、品質管理を行わないとスパムワーカーの増加に伴ってデータ品質は下がってしまうが、品質管理を適用することによって精度の低下が抑えられることがわかった。

## 参考文献

- [1] A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied Statistics*, Vol. 28, No. 1, pp. 20–28, 1979.
- [2] E. Elhamifar, G. Sapiro, and R. Vidal. Finding exemplars from pairwise dissimilarities via simultaneous sparse recovery. In *NIPS*, pp. 19–27, 2012.
- [3] 東田圭介, 櫻木伸幸. クラウドソーシングを活用した POI の収集実験と課題. 2013 年度人工知能学会全国大会論文集, 2013.
- [4] 鹿島久嗣, 梶野洸. クラウドソーシングと機械学習. 人工知能学会誌, Vol. 27, No. 4, pp. 381–388, 2012.
- [5] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *SODA*, pp. 604–632, 1998.
- [6] C. Lin, Mausam, and D. Weld. Crowdsourcing control: moving beyond multiple choice. In *UAI*, pp. 491–500, 2012.
- [7] K. Wagstaff, S. Rogers, and S. Schroedl. Constrained k-means clustering with background knowledge. In *ICML*, pp. 577–584, 2001.
- [8] P. Welinder, S. Branson, S. Belongie, and P. Perona. The multidimensional wisdom of crowds. In *NIPS*, pp. 2424–2432, 2010.
- [9] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *NIPS*, pp. 2035–2043, 2009.