

No. 5 Crowdsourcing

馬場 雪乃（国立情報学研究所）

2014年7月20日 WWW 2014 勉強会

本セッション：クラウドソーシングの品質管理手法 特に「ワーカの選択方法」に着目

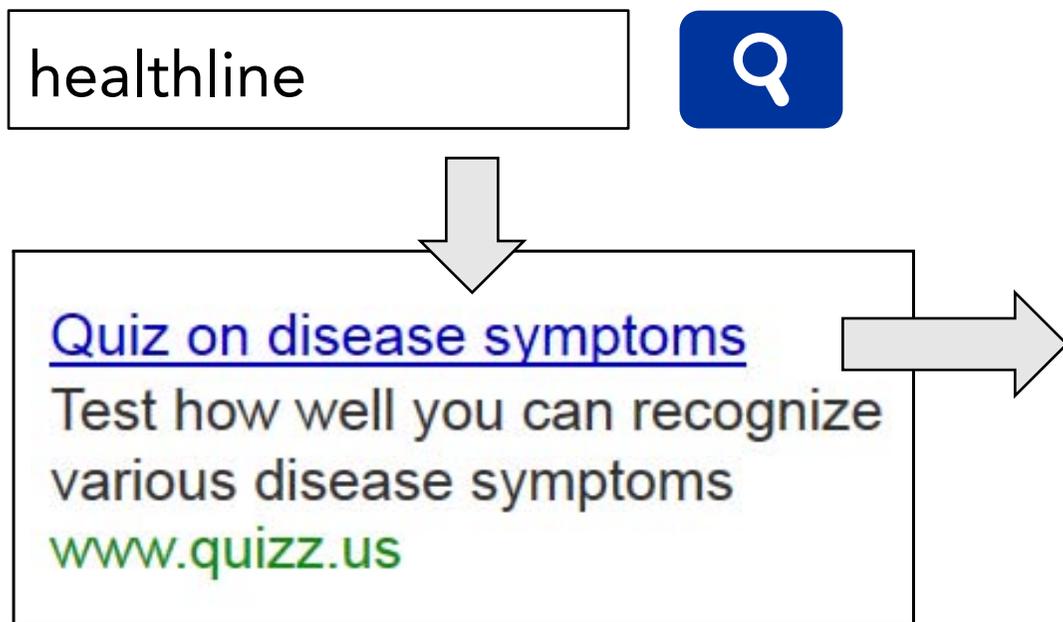
- 本セッションでのクラウドソーシング：
（答えを知りたい）多肢選択式問題を
人間に提示し回答させる仕組み
- 共通の課題
 - － 間違った回答をする人もいる
 - － できるだけ「正しい回答」を獲得したい
- 「同じ問題を複数人に聞く際の回答統合手法」
「正しい回答をしそうなワーカを選択手法」
が提案されている

1 本め概要：検索広告にクイズ形式タスクを埋め込む。
広告提示の仕組みを使い適切な人にタスクを割り当て

“Quiz: Targeted Crowdsourcing with a Billion (Potential) Users”

- NY大のP. IpeirotisがGoogle滞在時に実施
- 動機：Knowledge Graph充実のため
人間から知識を集めたい
- 課題：専門的知識は答えられる人間に限られる
(例：「モルジェロン病の症状はなんですか」)
- アプローチ：**Google広告にタスクを埋め込む**

1 本め概要：検索広告にクイズ形式タスクを埋め込む。 広告提示の仕組みを使い適切な人にタスクを割り当て



(元論文Fig. 3)



(元論文Fig. 2)

検索クエリに応じた問題へのリンクを広告に埋め込む。
ワーカの興味にあった問題を出せる
& 答えるかどうかはワーカに任せられる (**self-selection**)

手法：良いワーカの探索と活用のバランスを取るため MDPを利用し行動決定

- できるだけ能力が高いワーカの回答を獲得したい
- 2種類の問題を用意
 - 正解既知の問題：
ワーカの能力を推定するのに用いる（**探索**）
 - 正解未知の問題：
ワーカから知識を収集するのに用いる（**活用**）
- 各ワーカの回答状況に応じて、
探索・活用どちらのタスクを提示すべきかを
MDPで決める

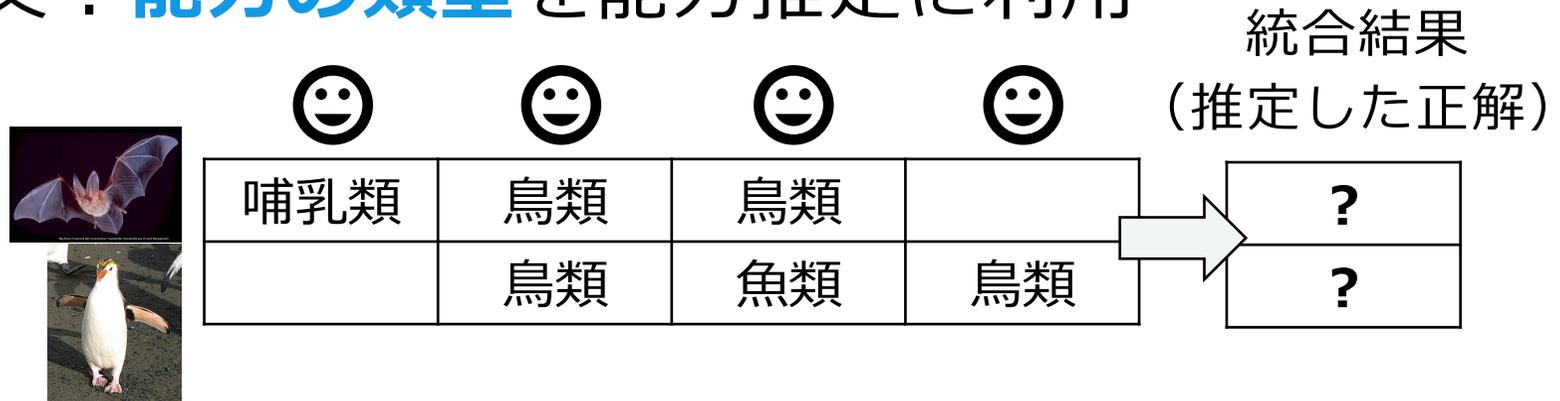
結果：99%の確率で正しい回答を、
一問あたり\$0.16で獲得することに成功

- あらかじめ用意した正解とワーカの回答を比較、
推定コストと正答率を評価（一つの回答が\$0.1、
一定の確信度になるまで複数人に聞く設定
→99%の確率で正答を得るのに一問\$0.16
- Targeting（検索クエリに応じた問題提示）と
Untargetingを比較
→targetingの方が高品質の回答が得られた
- 「能力が高いワーカほどたくさん回答する」傾向
→self-selectionにより高品質の結果が得られた

2本め概要：ワーカ能力の類型を考慮して 能力を推定し回答統合に利用

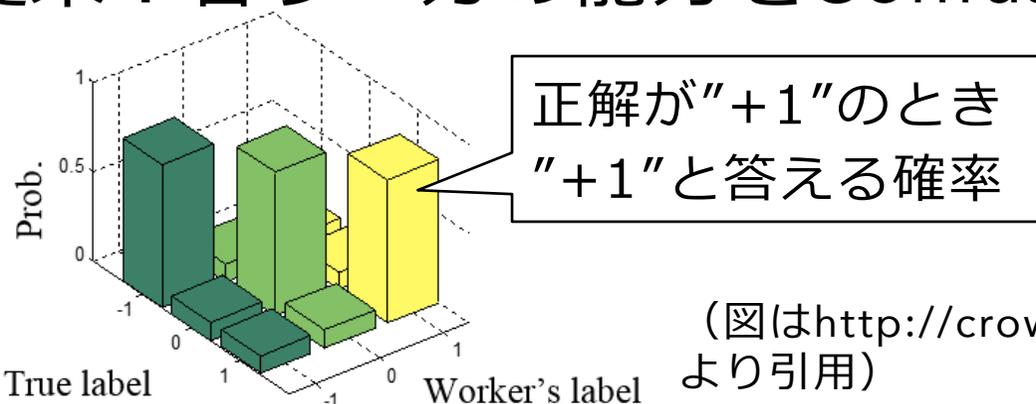
“Community-Based Bayesian Aggregation Models for Crowdsourcing ”

- 問題：同タスクを複数人に依頼する時の回答統合
- 従来：各ワーカ能力を考慮し回答統合
 - 正解推定（=統合） \Leftrightarrow 能力推定を交互に実施
- 本論文：能力の類型を能力推定に利用



手法：各ワーカーの能力は、能力の類型をベースにして生成されるとしてモデル化

- 従来：各ワーカーの能力をConfusion Matrixで表現



(図は<http://crowdresearch.org/blog/?p=8971>より引用)

- 本論文：能力の類型で複数ワーカーをまとめる

- ひとつの類型にひとつのコミュニティが対応
- 各ワーカーはいずれかのコミュニティに属する
- 各ワーカーのConfusion Matrixは、所属コミュニティのConfusion Matrixにノイズが加わって生成される

結果：特に回答数が少ないとき、 能力の類型を考慮した提案手法が有効

- 情報検索・NLPに関する4種類のタスクで実験
 - クエリ拡張結果の比較、検索結果の比較、感情分析、
成人コンテンツに関するクエリかどうかの判定
- 多数決、類型を考慮しない手法と比較
 - 特に獲得回答数が少ないときに提案手法が有効
- 解釈しやすい「類型」が得られた
 - 例：5段階の4, 5点ばかりを使う「コンサバ型」

3本め概要：

ワーカの属性を使って「良いワーカ」を発見

“The Wisdom of Minority: Discovering and Targeting the Right Group of Workers for Crowdsourcing”

- あるタスクにおけるワーカ的能力が**ワーカの属性（年齢、性別等）**に依存することはよくある
 - 例：「女子高生向け化粧品」に関するタスクなら「年齢 = 10代」「性別 = 女性」の能力が高そう
- **属性を使って良いワーカを選択**することで正しい回答の獲得を目指す

2つのアプローチ：「ワーカの属性から能力を推定」 「能力が高い人が満たす属性を推定」

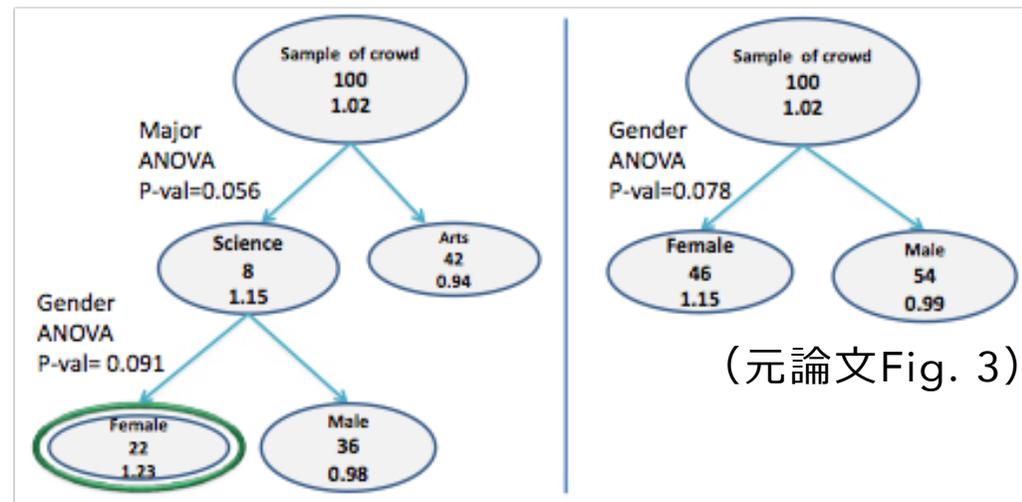
- 正解既知のタスクを用い一部のワーカの能力は推定済み
- ワーカ選択のアプローチを2つ提案

– **Bottom-up**：ワーカの属性から能力を推定

- 属性ベクトルの線形和で能力が決まるモデル
- 各次元の重みを学習

– **Top-down**： 能力が高い人が満たす 属性を推定

- 分散分析を用い
重要な属性を絞り込む



(元論文Fig. 3)

この例では“Major=Science,
Gender=Female”が重要属性となる

結果：属性によるワーカ選択後に回答統合する手法が複数のタスクで高精度達成

- 3タスクで実験：クイズ、語義判定、含意関係判定
- 提案手法：「良いワーカ」からの回答に絞った後、能力を考慮した回答統合手法を適用
- 多数決、ワーカ選択無しで回答統合、提案手法を比較
→**いずれのタスクでも提案手法が高精度。**
Bottom-upとTop-Downのどちらが良いかはタスクに依存
- クイズタスクでは“Major=Science”、
語義判定タスクでは“Major=Science”、
“Major=Engineering”が重要、“Education”はいずれでも重要ではない、と推定された