情報処理学会論文誌 Vol.56 No.3 1132–1140 (Mar. 2015)

推薦論文

ヒューマンコンピュテーションにおける非負値行列因子分解を用いたタスク割当て手法の提案

巻口 誉宗[1,a)] 並河 大地[1] 東 正造[1] 下村 道夫[1,2] 金丸 直義[1]

受付日 2014年10月4日, 採録日 2014年11月10日

概要：計算機の処理を人の処理能力によって補うヒューマンコンピュテーション（Human Computation）において，ワーカから高品質（高精度，高信頼）なデータを取得する品質管理手法（Quality Control）は重要課題の 1 つである．本研究では，品質管理手法の 1 つとして，タスクに対して適切なワーカを抽出する**タスク割当て手法**に注目する．提案手法としてワーカの過去のタスク回答結果をもとに非負値行列因子分解（Nonnegative Matrix Factorization）によるクラスタリングを行い，タスクに適したワーカ群（エキスパート群）を抽出する手法を提案する．そして固有表現ラベリングタスクの実データに提案手法を適用することで，提案手法が少ないテストタスク依頼数でエキスパート群を抽出できることを示す．

キーワード：ヒューマンコンピュテーション，クラウドソーシング，タスク割当て，非負値行列因子分解

A Task Allocation Method to Human Computation

Motohiro Makiguchi[1,a)] Daichi Namikawa[1] Shozo Azuma[1]
Michio Shimomura[1,2] Naoyoshi Kanamaru[1]

Received: October 4, 2014, Accepted: November 10, 2014

Abstract: Establishing a quality control methodology for obtaining a data set consisting of the results of tasks performed by workers is one of the most important problems to solve in Human Computation (HC). Our research proposes a new quality control methodology based on “efficient task allocation algorithm” which selects suitable workers for a given task. In concrete, the algorithm classify the workers into group by Nonnegative Matrix Factorization (NMF) clustering of the workers based on the result of the tasks performed by the workers in the past, aiming to reduce the cost of worker-sectioning process and easily select a large number of workers without sacrificing the quality. We have shown that the newly proposed algorithm successfully reduced the cost of worker-sectioning process for a given task, and can select a large number of suitable workers maintaining quality.

Keywords: human computation, crowdsourcing, task allocation, nonnegative matrix factorization (NMF)

1. はじめに

ヒューマンコンピュテーション（**Human Computation**，以下 **HC**）は，計算機技術のみで自動的に処理することが困難な問題を，人の処理能力によって解決しようとする手法 [1] である．計算機によって画像や音声，自然言語等のメディアを認識，加工するメディア処理技術（以下，エンジン）の中には，現時点ではサービス提供に必要な精度基準を満たしていないものや，さらなる精度向上のために大量の学習データを要するものがある．

こうしたエンジンを用いて提供されるサービスへ HC を

1 日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, Yokosuka, Kanagawa 239–0847, Japan
2 千葉工業大学（2014.4-）
Chiba Institute of Technology, Narashino, Chiba 275–0016, Japan
a) makiguchi.motohiro@lab.ntt.co.jp

本稿の内容は 2014 年 3 月のグループウェアとネットワークサービス研究発表会にて報告され，同研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

© 2015 Information Processing Society of Japan

適用することで，認識精度向上や認識対象の拡大等，サービスの品質向上が期待される．HC のサービス適用の一例として，ユーザが入力した画像を画像認識エンジンが自動的に認識し，関連情報を通知する「画像認識サービス」を基に説明する．このサービスにおいて，たとえばガラス製品のような光の反射，透過が起きやすいもの，布製品のような形状が変化しやすいもの等，エンジンが苦手とする処理対象がユーザから入力された場合，エンジンのみでは誤った結果がユーザに出力され，認識精度の低下につながる．こうした処理対象に対し，人がエンジンの認識結果の修正や補足，または認識処理自体の代行を作業者（以下，ワーカ）となって行うことで，エンジン単体と比べ認識精度の大幅な向上が期待できる．さらに人がエンジンの認識精度が低い処理対象に対して正解を付与したデータを生成することで，エンジンの教師データが増え，エンジンそのものの認識精度向上も期待できる．

従来，こうした仕組みを実現するためのワーカ確保の手段としては，ワーカを専属で確保し，作業（以下，タスク）を依頼することが一般的であったが，ワーカ集めの手間や報酬コストの高さから費用対効果が得られる用途が限定的であった．一方で，近年は低い報酬コストで大量のワーカにタスクを依頼できる手段として，Amazon Mechanical Turk [2]（以下，AMT）等に代表される**クラウドソーシング（Crowdsourcing）**が注目されている．

クラウドソーシングは，インターネットを用いることでワーカへタスクをアウトソーシングすることができる仕組みである．そのため，ワーカを専属で確保する場合と比較して低い報酬コストで必要量のタスクを容易に処理することが可能である．クラウドソーシングを活用した HC の一例として，AMT を用いた視覚障がい者への画像認識サービス [3] 等，従来の計算機技術のみでは実現が困難だったサービスが多数研究，実用化されている．

一方で，クラウドソーシングでは不特定多数のワーカにタスクを依頼することから，スパムワーカ等（後述）による処理結果の品質低下や，専門知識や特定のスキル等が必要とされる難易度の高いタスクの品質向上が困難であるといった問題がある．そのため，これまでのクラウドソーシングに関する HC おいては，大多数の一般的なワーカが正確に処理することが期待できるタスク（以下，**一般正解率の高いタスク**）を前提としたものが多く，適用先を限定する制約の 1 つになっている．

我々は HC を適用可能なサービスを拡大し，計算機を利用したサービスのさらなる品質向上や，計算機のみでは実現できない新サービスの創出を目指している．そのため，サービスのコスト削減の観点から低い報酬コストでタスクを処理できるクラウドソーシングの仕組みは欠かせない．一方で，今後の計算機技術の発展によって，人の処理能力による補足が必要なタスクも，より専門知識や特定のスキル等が必要なものに移行していくことが予想される．例として，翻訳等の専門知識が必要なタスク，写真内のオブジェクトのセグメント化等の繊細な作業が必要なタスク，および与えられた明確な規準に従う単語ラベリング等の高い再現性が必要なタスクは，一般的なワーカにとって難易度の高いタスクと考えられる．そこで我々は，クラウドソーシングにおいて，**一般正解率が低いタスク**に対しても高品質に処理できる手法を検討する．

本稿では，一般正解率の低いタスクの品質向上を実現する提案手法として，クラスタリングによるワーカの属性推定とテストタスクを組み合わせたタスク割当て手法を提案する．そして固有表現ラベリングタスクの実データを用いた解析結果から，提案手法によって少ないテストタスク依頼数でエキスパート群を抽出可能であることを示す．

2 章では，HC における品質低下を防ぐ品質管理手法について，先行研究のアプローチを述べる．3 章では品質管理手法の中でも，ワーカへのタスク割当て手法に注目し，従来手法の整理と提案手法について述べる．4 章では提案手法の有効性確認のために行った実データ収集実験と，解析手法について述べる．5 章では実験結果，6 章ではその考察を述べる．7 章では今後の課題，8 章ではまとめを述べる．

2. HC における品質管理手法

クラウドソーシングに限らず，人にタスクを依頼する場合にはその回答結果に様々なノイズが含まれる可能性があり，HC に関する研究では，ノイズを除去し，最終的により高品質な処理結果を得る**品質管理手法（Quality Control）**が重要な要素として多数研究されている．以下で概要を述べる．

まず本稿では，回答結果に含まれるノイズ源として，意図的に低品質な回答結果を返す「スパムワーカ」，タスク処理に対する知識やスキルが不足している「適性外ワーカ」，および外的・内的要因で本来のパフォーマンスを発揮できない「ヒューマンエラーワーカ」の 3 タイプのワーカを定義する．

これらのノイズを除去するための品質管理手法の先行研究は，主に以下の 2 つのアプローチから研究されている．

アプローチ 1

ワーカの回答結果解析による高品質な処理結果の算出．

アプローチ 2

ワーカの回答結果自体の品質向上．

アプローチ 1 は，回答結果からノイズを除去し，高品質な処理結果を算出するアプローチである．例として，同じタスクを複数のワーカに依頼し回答結果の冗長性を確保したうえで，複数のワーカで多数決をとる手法 [5] や，ワーカの属性を，多数決結果を正解と仮定して多数決結果と回答結果との一致度で推定し，優秀なワーカの多数決の重みを増加させる手法 [8]，正解の推定を行わずワーカの回答結

© 2015 Information Processing Society of Japan

果から直接機械学習の識別器を推定する手法 [9] 等が提案されている．

これらのうち，特に多数決を用いる手法は，一般正解率の高いタスクに対しては冗長化の効果によって高品質な処理結果が得られる．一方で，一般正解率の低いタスクに対しては，高品質な処理結果を算出するために必要なワーカ数が増加するほか，正解率が一定値より低い場合（たとえば 2 値選択式タスクにおいては正解率 50%以下），正解算出自体が困難である [5]．

アプローチ 2 は，タスクをワーカに依頼する前にタスク設計，報酬設計，ワーカ抽出等を工夫することでワーカの回答結果自体の品質を向上させるアプローチである．例として大規模なタスクを複数の小規模なタスクに分割する手法 [6] や，ワーカの自信の有無によって報酬額を選択させる手法 [7] 等が提案されている．なかでも，タスクの正解率が高いと想定されるワーカを抽出した後，タスクを割り当てる手法（以下，**タスク割当て手法**）は，タスク処理前にスパムワーカや適性外ワーカを除外でき，回答結果全体の品質向上が見込める．これにより冗長性削減による報酬コストや時間コストの削減に加え，適性のあるワーカを抽出することで一般正解率が低いタスクに対しても高品質な処理結果を算出できると考えられる．タスク割当て手法に関して，芦川ら [4] はタスク依頼者がワーカの情報を管理可能な Private Crowdsourcing System を提案し，ワーカのスキルや経験値に基づきタスクをマッチングすることで回答結果の品質が向上する可能性について述べている．

我々は，一般正解率の低いタスクに対して高品質な処理結果を得るために，適性ワーカの抽出を目的としたタスク割当てに注目する．次章では，タスクの割当て手法について既存手法を整理することで具体的な要件を定義し，提案手法を述べる．

3. タスク割当て手法

タスク割当ては，これから依頼したいタスク（以下，**新規タスク**）に対して高品質な回答結果を返すことが期待されるワーカを抽出する手法であり，新規タスクに適切なワーカを推定するワーカの属性推定が不可欠である．本稿では，属性推定を，人による属性推定と，計算機による属性推定の 2 パターンに分類し，それぞれのパターンのメリットとデメリットをまとめ，デメリットを解消するための要件を定義する．

3.1 人によるワーカの属性推定

本例として，ワーカに直接自己の属性を申告させる「自己申告」がある．代表的な自己申告の例として，リサーチ会社のスクリーニング等があげられる．自己申告のメリットは，ワーカのスキルや経験，得意分野や好み等の詳細な属性情報を容易に取得できる点等がある．一方で，スパムワーカ，ヒューマンエラーワーカ，および自らの能力を正しく把握していない適性外ワーカからは信頼できる申告結果が得られないというデメリットが考えられる．ワーカの自己申告の正確さについて，小山ら [10] はワーカが申告した「タスク処理結果への自信度」と実際の正解率の間にはばらつきがあることを報告している．この不確かさの解消のために，**ワーカの申告情報以外の情報を利用すること**を要件として定義する（**要件 1**）．

また，人によるワーカの属性推定の別例として，タスクを処理したワーカの評価を，タスクの依頼者や別のワーカが行う「他者評価」[11] がある．この手法は，属性を推定したいワーカが大量にいる場合，評価を依頼するタスク数が増加し，報酬コスト，時間コストが増加するといったデメリットがある．このデメリットを解消するため，**属性推定の際のタスク依頼数が十分に少ないこと**を要件として定義する（**要件 2**）．

3.2 計算機によるワーカの属性推定

本例として，ワーカの回答結果を用いた属性推定がある．Donmez らは IEThresh においてワーカの過去のタスク回答結果の多数決結果を正解とし，個々のワーカの正解率を推定する手法を提案している [12]．この手法は多数決を用いるため，一般正解率の低いタスクへの適用が困難であると考えられる．

また，ワーカの属性推定の別例としてテストタスクがあげられる．正解が既知のテストタスクをワーカに処理させ，回答結果の正解率を算出することで，スパムワーカ，適性外ワーカの除外に加え，そのタスクの正解率が高いワーカを抽出できる．テストタスクは一般正解率が低いタスクにおいて最も効果的なワーカの属性推定手法と考えられるが，デメリットとして，属性を推定したいワーカ全員にテストタスクを依頼する必要があり，ワーカ数の増加にともなって報酬コストや時間コストが増加する点がある（導かれる要件は要件 2 と等しい）．特に一般正解率の低いタスクにおいて，タスクを正確に処理できるワーカ（以下，エキスパートワーカ）をテストタスクによって抽出する場合，一般正解率が低いタスクほどエキスパートワーカの割合は減少していくため，テストタスクを依頼するワーカ数の母集団を増やす必要が生じる．

3.3 提案手法

ワーカの過去のタスク回答結果のみでは，一般正解率が低い新規タスクに対してワーカの適性を推定することは困難である．一方で，過去のタスク回答結果に基づいて，似た属性を持つワーカ群を推定することは，ワーカのタスク回答結果の正解不正解にかかわらず可能である．Gomes ら [14] は，客観的な正解が設定されていない画像のカテゴリ分類タスクを，ワーカの主観によって行わせ，その結

© 2015 Information Processing Society of Japan

果を基にクラスタリングすることで，似た画像分類を行うワーカ群の抽出が可能であることを示している．

そこで我々は，似た属性を持つワーカ群はタスクへの適性も似ていると考え，一部のワーカの適性をテストタスクによって推定することで，そのワーカが属するワーカ群の適性を判定する手法を提案する．

提案手法の手順を以下に示す．

1 過去のタスク回答結果に基づきワーカをクラスタリング．
2 クラスタの代表者に新規タスクのテストタスクを依頼．
3 テストタスクの正解率が高い代表者が属するクラスタを当該タスクの**エキスパート群**と推定し，エキスパート群から抽出したワーカに新規タスクを依頼．

提案手法では，ワーカの申告情報を用いず（要件 1），クラスタの代表者のみにテストタスクを依頼する．これにより，属性を推定したい全ワーカにテストタスクを依頼する場合と比較し，十分に少ないテストタスク依頼数（テストタスクを依頼するワーカ数）で適性のあるワーカ群を抽出できる（要件 2）と考えられる．

次章以降では，提案手法が一般正解率の低いタスクに対して，十分に少ないテストタスク依頼数で高品質な回答を行うエキスパート群の抽出が可能かどうかを，実データ解析によって検証する．

4. 実データ収集と解析手順

4.1 タスク設計

本稿では，実際に HC での処理が想定され，正解率によって回答結果の客観的な品質評価が行えるタスクとして**固有表現（Named Entity）**のラベリングタスクを初期検討の対象とした．固有表現とは，人名や地名等の固有名詞に加え，日付や数量表現を合わせた表現である．ラベリングされたデータは，形態素解析等のエンジンの教師データとして利用される．今回の実験では，関根 [15] の拡張固有表現階層を規準として過去に専属で確保したワーカがこの規準に従ってラベリングした結果をその単語の正解と定義する．

ワーカには，Wikipedia [16] からランダム抽出した単語を一覧で提示し，その単語について，固有表現の該当する分類（以下，クラス）を一覧から選択するよう依頼した．クラスの選択肢は，拡張固有表現階層における最上位階層 14 クラス（**表 1** 参照）とし，ワーカには拡張固有表現階層のリストとカテゴライズ例が記載されている Web サイトへのリンクを提示した．**図 1** にタスク処理画面の一部を示す．

このタスクは，ワーカの主観ではなく外部から与えられる明確な規準が正解となることから，当該タスクの単語やクラスが非日常的で難解なものである場合等，必ずしも多くのワーカが選ぶ多数派の回答結果が正解とならない．この点で，一般正解率の低下が予想されるタスクである．

表 1　拡張固有表現階層における最上位階層
Table 1　The top hierarchy in the named entity.

人名（Person）
神名（God）
組織名（Organizaton）
地名（Location）
施設名（Facility）
製品名（Product）
イベント名（Event）
自然物名（Natural Object）
病気名（Disease）
色名（Color）
時間（Timex）
期間（Periodx）
数値表現（Numex）
該当なし（Other）

以下の単語を見て、その単語がどのカテゴリ（分類）にあてはまるかをお選びください。 必須 ひとつだけ

※不明な固有名詞がありましたら、ネット検索等で調べてお答えください。

※各カテゴリの詳細、分類例はリンクを参考にして下さい。
（リンク先：「関根の拡張固有表現階層 -7.1.0-」）

回答方向	一斤	ケフェウス座	委員	アンモニア	美野山橋
人名（Person）	○	○	○	○	○
神名（God）	○	○	○	○	○
組織名(Organizaton)	○	○	○	○	○
地名(Location)	○	○	○	○	○
施設名(Facility)	○	○	○	○	○
製品名(Product)	○	○	○	○	○
イベント名(Event)	○	○	○	○	○
自然物名(Natural_Object)	○	○	○	○	○
病気名(Disease)	○	○	○	○	○
色名(Color)	○	○	○	○	○
時間(Timex)	○	○	○	○	○
期間(Periodx)	○	○	○	○	○
数値表現(Numex)	○	○	○	○	○
該当なし(Other)	○	○	○	○	○

図 1　固有表現ラベリングタスクの処理画面の一部
Fig. 1　The screen capture of the named entity task.

ワーカはリサーチ会社のモニタ 331 人で，78 単語に対して固有表現ラベリングを依頼した（ワーカ数 331，タスク数 78）．

4.2 クラスタリング手法

クラスタリング手法には，**非負値行列因子分解（Non-negative Matrix Factorization：以下 NMF）**を用いた [17]．代表的なクラスタリング手法である K-means は，すべての要素がいずれかのクラスタに強制的に分類されるハードクラスタリング手法である．一方で，NMF は，ワーカごとに与えられるそれぞれのクラスタへ属する強さ（以下，帰属度）を基に，1 人のワーカが複数のクラスタへ同時に所属することが可能なソフトクラスタリング手法である．この特性により，エキスパート群と判定したクラスタから，帰属度が高い順に任意の数のワーカを選択できるため，初期検討として NMF を用いた．

NMF は I 行 × J 列の行列 $\mathbf{X}$ を，I 行 × K 列の特徴行

© 2015 Information Processing Society of Japan

列 $\mathbf{T}$ と K 行 × J 列の特徴行列 $\mathbf{V}$ に分解し，$\mathbf{T}$ と $\mathbf{V}$ の積 $\mathbf{TV}$ と分解前の行列 $\mathbf{X}$ の距離 D が最小となるよう，$\mathbf{T}$ と $\mathbf{V}$ を最適化する手法である．

ワーカのタスク処理結果を行列 $\mathbf{X}$ に変換し NMF を適用することで，タスクとワーカのクラスタ $k \in \{1, \ldots, K\}$ への帰属度がそれぞれ特徴行列 $\mathbf{T}$ と $\mathbf{V}$ に表される．$\mathbf{TV}$ と $\mathbf{X}$ の距離 D の定義には，以下に示す一般化 Kullback-Leibler divergence（KL divergence）を用いた．

$$D*(\mathbf{X}, \mathbf{TV}) = \sum_{i=1}^{I} \sum_{j=1}^{J} d*(x_{ij}, \widehat{x_{ij}}) \tag{1}$$

$$d*(x_{ij}, t_i^T, v_j) = x_{ij} \log \frac{x_{ij}}{t_i^T v_j} - x_{ij} + \widehat{x_{ij}} \tag{2}$$

ここで $\widehat{x_{ij}} = t_i^T v_j$ であり，$\mathbf{T}$ と $\mathbf{V}$ の各要素 t_{ik}，v_{kj} は以下の更新式で最適化される．

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j \frac{x_{ij}}{\widehat{x_{ij}}} v_{kj}}{\sum_j v_{kj}}, \tag{3}$$

$$v_{kj} \leftarrow v_{kj} \frac{\sum_i \frac{x_{ij}}{\widehat{x_{ij}}} t_{ik}}{\sum_i t_{ik}} \tag{4}$$

4.3 タスク回答結果行列の作成

NMF の対象とする行列 $\mathbf{X}$ は，タスク回答結果から以下の手順で作成した．それぞれのタスク番号を $i \in \{0, \ldots, I\}$，ワーカの番号を $j \in \{0, \ldots, J\}$ とし，各タスクの回答結果を $[0, 1]$ の 2 値で表現する．すなわち，ワーカ j の i 番目のタスクへの回答結果ベクトルを $\mathbf{y}_i^j$ とする．たとえば固有表現ラベリングタスクの単語 i において，クラスの選択肢が「人名，地名，施設名」であり，ワーカ j が単語 i を「人名」とラベリングした場合，$\mathbf{y}_i^j = [1, 0, 0]$ と表現される．この回答結果ベクトルを結合し，タスク回答結果行列 $\mathbf{X}$ とする．

4.4 解析手順

解析はワーカ 331 人のタスク回答結果を評価対象とし（$N = 331$），以下の手順で行った．

手順 1. 学習タスク，テストタスク，評価タスクの分類

タスク回答結果をランダムに 3 等分し，それぞれ**学習タスク**，**テストタスク**，**評価タスク**とする．

手順 2. 学習タスク，テストタスクを用いたワーカ抽出

学習タスクとテストタスクのタスク回答結果を用いて，ワーカを n（$< N$）人抽出する．抽出方法は後述する．

手順 3. 評価タスクの正解率算出

手順 2 で抽出した n 人のワーカに対して，評価タスクの平均正解率を算出する．

学習タスクは，提案手法においてワーカの過去のタスク回答結果として利用するタスクである．また，テストタスクは新規タスクの正解率を測るための正解データが与えられたタスク，評価タスクは提案手法における新規タスクを意味する．このうち正解データを用いるのはテストタスク，評価タスクのみで学習タスクは正解データを必要としない．

学習タスク，テストタスク，評価タスクの分配はランダムに設定した．この乱数の違いによる偏りを排除するため，手順 1〜3 の試行を 1,000 回繰り返し，評価データの正解率の平均値を取得した．

4.5 ワーカの抽出手法

手順 2 において，全ワーカから n 人を抽出する手法は，提案手法に加え，比較手法として部分エキスパート抽出，多数派抽出，およびランダム抽出の 3 種類を定義する．以下にそれぞれの手法の詳細を示す．

提案手法（NMF）

全ワーカの学習タスクの回答結果から NMF によって K 個のクラスタを作成する．得られたクラスタ $k \in \{1 \ldots K\}$ から，帰属度が高い上位 m（$< N$）人のワーカを代表ワーカとする．この代表ワーカのテストタスクの正解率の平均値を算出し，この平均値が最も高いクラスタからクラスタの帰属度上位 n 人を抽出する．

部分エキスパート抽出（PartialExperts）

全ワーカからランダムに $(m \times K)$ 人のワーカを抽出し，テストタスクの正解率が高い順に上位 n 人を抽出する．提案手法のテストタスク依頼数を減らす効果を検証するための比較手法である．

多数派抽出（Majority）

全ワーカの学習タスクの回答結果の多数決をとり，多数決結果と学習タスクの回答結果との一致率が高いワーカを上位 n 人抽出する．多数派と同じ回答をしやすい傾向のワーカが抽出されるため，多数決結果が正解となるタスクでは最も正解率が高くなることが期待される．タスクの平均正解率による提案手法の効果を検証するための比較手法である．

ランダム抽出（Random）

全ワーカからランダムに n 人抽出する．タスク割当て手法で何も工夫しない場合と等しい．

それぞれの抽出手法において，テストタスクの依頼が必要なワーカ数（以下，テストタスク依頼ワーカ数）を**表 2** に示す．

表 2 各抽出手法におけるテストタスク依頼ワーカ数
Table 2 The number of workers that requested test task.

ワーカ抽出手法	テストタスクを依頼するワーカ数（人）
NMF	$m \times K$
PartialExperts	$m \times K$
Majority	0
Random	0

© 2015 Information Processing Society of Japan

提案手法において，クラスタを評価する代表ワーカの数 m が増加するとテストタスク依頼ワーカ数も増加する．今回の実験ではテストタスクの依頼数を可能な限り少なくするため，代表ワーカ数を多数決が実行できる最小人数の 3 人に設定した（$m = 3$）．

解析は，すべてのタスクを対象にワーカ抽出手法を比較する実験 1「全タスクを対象とするワーカ抽出手法の比較」と，タスクを平均正解率の高いタスク，低いタスクに 2 等分し，それぞれにおける抽出手法を比較する実験 2「タスクの正解率によるワーカ抽出手法の比較」を行った．実験 1 では，提案手法で必要なテストタスク依頼数の確認と，エキスパート群抽出効果の有効性検証を目的とする．実験 2 では，タスクの各平均正解率における最適なワーカ抽出手法の検証を目的とする．

5.　結果

5.1　実験 1：全タスクを対象とするワーカ抽出手法の比較

全ワーカ 331 人の全タスクに対する平均正解率は約 0.61（標準偏差 SD：0.14）であった．また，最も正解率の低いワーカの正解率は約 0.19，最も高い正解率のワーカは約 0.95 であった．

5.1.1　各ワーカ抽出手法における正解率比較

図 2 に，テストタスク依頼ワーカ数 18 人（提案手法におけるクラスタ数 $K = 6$）において，4 つの手法によって抽出したワーカ 3 人（$n = 3$）の平均正解率を示す．

このグラフから，提案手法（NMF：正解率約 0.86）が最も正解率が高く，ついで部分エキスパート抽出（PartialExperts：正解率約 0.80），多数派抽出（Majority：正解率約 0.67），ランダム抽出（Random：正解率約 0.61）である．

5.1.2　エキスパート群抽出効果

図 3 に，テストタスク依頼ワーカ数が 18 人の場合における提案手法（NMF）と部分エキスパート抽出（Random-Expert）において，抽出するワーカ数 n を 3〜17 まで増加させた際の n 人の平均正解率を示す．

このグラフから，提案手法では，抽出するワーカ数を増加させても平均正解率の減衰が部分エキスパート抽出と比較しても少なく，高い正解率を持ったワーカをより多く抽出できることが分かる．

5.1.3　テストタスク依頼数

図 4 に，提案手法と部分エキスパート抽出においてテストタスク依頼ワーカ数を変化させ，それぞれ抽出した 3 人（$n = 3$）の平均正解率を示す．また，図中の Experts は，全ワーカにテストタスクを依頼したと仮定し，テストタスク正解率上位 3 人の評価タスクの平均正解率を示している．

クラスタを評価する代表ワーカ数 m は 3 人に固定したため，提案手法では，テストタスク依頼ワーカ数の増加はクラスタ数 K の増加を意味する．また，部分エキスパート抽出では，テストタスク依頼ワーカ数は全ワーカからランダムに抽出する人数を意味する．

図 4 では，テストタスク依頼ワーカ数が同じ場合，部分エキスパート抽出が提案手法の平均正解率を超えるのはワーカ数が 66 人のときであった．提案手法ではテストタスク依頼ワーカ数，すなわちクラスタ数を増加させても正解率の大幅な変化はなく，テストタスク依頼ワーカ数 6 人

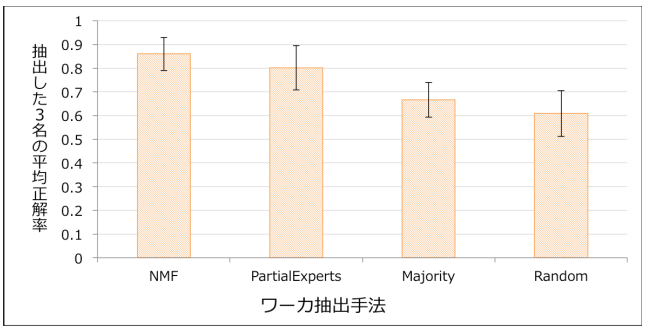

図 2　各手法における正解率（テストタスク依頼ワーカ数 18 人）
Fig. 2　The accuracy rate in each method.

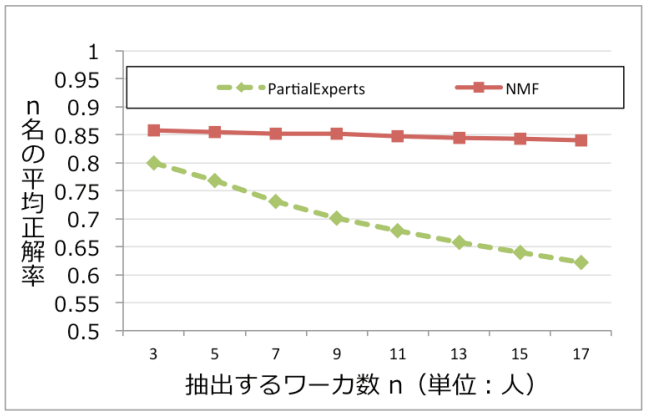

図 3　抽出するワーカ数 n を増加させた際の平均正解率（テストタスク依頼ワーカ数 18 人）
Fig. 3　The average accuracy rate increasing the extract worker number n.

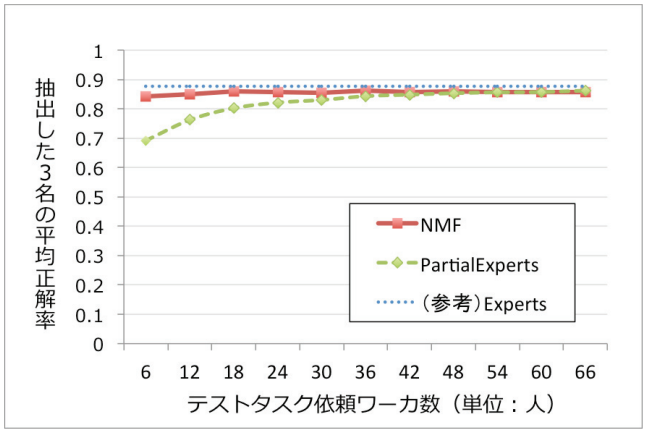

図 4　テストタスク依頼ワーカ数の増加における平均正解率
Fig. 4　The average accuracy rate increasing the number of test task workers.

© 2015 Information Processing Society of Japan

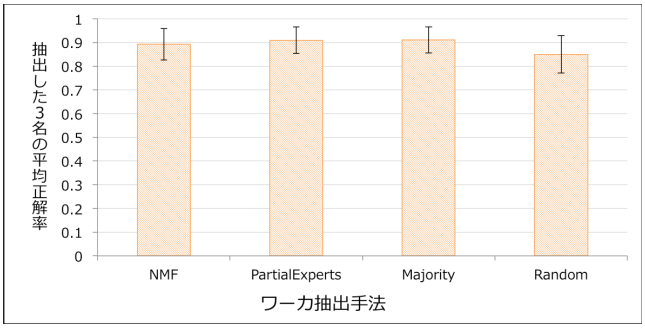

図 5　平均正解率 0.7 以上のタスクにおける各手法の正解率
Fig. 5　The accuracy rate of each method in the average accuracy rate 0.7 more.

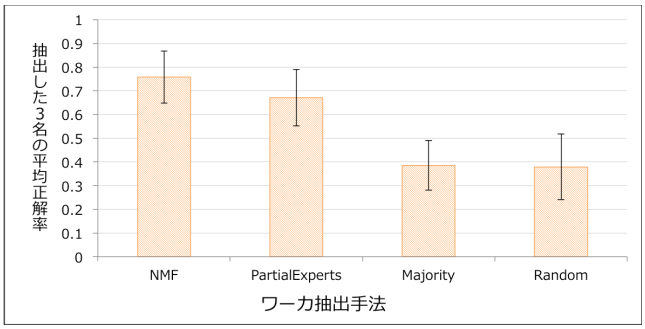

図 6　平均正解率 0.7 未満のタスクにおける各手法の正解率
Fig. 6　The accuracy rate of each method in the average accuracy rate under 0.7.

の段階で部分エキスパート抽出の 66 人の正解率とほぼ同程度の正解率を持つワーカを選出できていることから，テストタスク依頼数を 1/5〜1/10 程度に抑えられることが分かる．

以上の結果から，提案手法が十分に少ないテストタスク依頼数でエキスパート群を高効率に抽出できる点の基本的有効性が示された．

5.2　実験 2：タスクの正解率によるワーカ抽出手法の比較

全タスク 78 問を，ワーカの平均正解率に基づき 39 問ずつに 2 等分するしきい値は正解率 0.7 であった．ワーカの平均正解率が 0.7 を上回るタスク群を $Task_{High}$，それ以外のタスク群を $Task_{Low}$ とし，それぞれ前述の 4 つのワーカ抽出手法でワーカを抽出し，評価データの正解率を取得した（テストタスク依頼ワーカ数 18 人，試行回数 1,000）．

図 5 に，$Task_{High}$ における各手法の正解率を，**図 6** に，$Task_{Low}$，における各手法の正解率を示す．

図 5 から，正解率の高いタスクにおいては学習タスクの多数決結果を用いた多数派抽出（Majority）の正解率が提案手法（NMF）やランダム抽出（Random）よりも高くなることが分かる．一方，図 6 の正解率の低いタスクにおいては，多数派抽出の正解率はランダム抽出と同程度の低さである一方で，実験 1 の図 2 と同様，提案手法の有効性が確認できる．

このことから，正解率の高いタスクにおいては，テストタスクの依頼をいっさい必要とせずエキスパート群を抽出できる多数派抽出が優れており，正解率が低いタスクにおいては，少ないテストタスク依頼数でエキスパート群を抽出できる提案手法が優れていることが分かる．

6.　考察

6.1　NMF におけるクラスタ数について

NMF では，クラスタ数を指定する必要がある．今回のタスクでは，実験 1 で提案手法のクラスタ数を増加させた場合も正解率に大きな変動はなく，最も少ないクラスタ数 2 でもエキスパート群の抽出が可能であった．また，実験 2 では正解率が低いタスクほど提案手法の効果が明確であった．この要因として，NMF によるクラスタリング手法では，一般正解率が低いタスクほど，不正解のワーカの回答の分散が増加することや，特定の誤回答に回答が集中することで，高確率で正解を選択し続ける少数のエキスパート群がより特徴的なクラスタとして分離しやすくなるためと考えられる．

実用時では最適なクラスタ数を設定するため，テストタスク時にクラスタ数を 2 から除々に増加させ，クラスタの代表ワーカに依頼するテストタスクの正解率の変化量が一定のしきい値以下になったときにそのクラスタ数を選択する等の手法が有効と考えられる．

6.2　提案手法のコスト削減効果について

結果から，特に正解率の低いタスクに対して提案手法を適用することで，少ないテストタスク依頼数でエキスパート群を抽出できることを示した．提案手法によって新規タスクにおけるエキスパート群を事前に推定することで，回答結果の品質向上が期待できるため，従来より多数決における冗長性の削減が可能となる．このことから，提案手法は，ワーカ属性推定段階におけるタスク依頼数削減の効果だけではなく，新規タスク依頼時においてもタスク依頼数の削減，すなわち報酬コスト，時間コストの削減に効果があると考えられる．

また，今回は回答結果のみをクラスタリングに利用し，正解率が高いワーカの抽出を行ったが，タスクを完了するまでの回答速度もクラスタリングの情報に加えることで，タスクを正確なだけでなく，より早く回答できるエキスパート群を抽出し，新規タスクにおけるさらなる時間コストの削減を実現できる可能性が考えられる．

7.　今後の課題

7.1　新規タスクへの適用性向上

本稿では，学習タスクとテストタスク，評価タスクが同一の理想状態で提案手法の有効性を示した．一方で，実用時には学習タスクと評価タスクが完全に一致することは想

© 2015 Information Processing Society of Japan

定しにくく，過去のタスクとまったく異なるカテゴリ（固有表現ラベリングと画像のセグメント化等）の新規タスクに対してどの程度本手法が適用可能かを別途検証する必要がある．

新規タスクへの本手法の適用方法案としては，新規タスクを依頼する際のテストタスクの結果を用いて，過去のタスクと新規タスクを含めたタスクのクラスタリングを行い，似たタスクと判定された過去のタスクをワーカのクラスタの基とする手法や，すべてのワーカからランダムにテストタスクを依頼するワーカを選択し，最も成績の良かったワーカと類似するワーカ群を推定する手法が考えられる．

7.2 外部規準以外の高難易度タスクへの適用

今回の固有表現ラベリングタスクは，ワーカの主観や多数決結果等ではない明確な外部規準を正解としたことで，単純な多数決による正解算出が困難な高難易度タスクであった．一方でタスクの難易度が変動する要因としては，例として画像から特定のオブジェクトを見つけるタスクにおける画像サイズによる影響や，セグメント化タスクにおけるオブジェクトの複雑さ等，外部規準とは異なる要因も考えられる．今後はこのような要因でもたらされる一般正解率の低いタスクに対する提案手法の有効性を検証する必要がある．

7.3 タスク割当て手法の自動化

実験 2 の結果から，タスクの正解率の高さによって最適なワーカ抽出手法が変化することが分かる．このことから，正解率が高いタスクに対しては多数派抽出を用い，正解率が低いタスクに対しては提案手法を用いるといったワーカ抽出手法の切替えにより，新規タスクに対する適切なタスク割当てが行えると考えられる．

今後，新規タスクの難易度によらず高品質なタスク割当てを自動的に行える手法の確立を目指す．

8. まとめ

本稿では HC における品質管理手法としてワーカのタスク回答結果によるクラスタリングを用いたタスク割当て手法を提案し，固有表現ラベリングタスクに対する回答結果の実データを用い，少ないテストタスク依頼数でエキスパート群を抽出できることを示した．本研究はこれまでクラウドソーシングでの処理が困難だった一般正解率が低いタスクの品質を高めるものである．これによりクラウドソーシングの適用先拡大の足がかりとなることを期待する．

謝辞 本研究を進めるにあたり，ご教示をいただいた日本電信電話株式会社の木村昭悟氏，新井啓之氏，澤田 宏氏，貞光九月氏，片岡春乃氏に深謝いたします．

参考文献

[1] von Ahn, L.: Human Computation, Doctoral Thesis, UMI Order Number: AAI3205378, CMU (2005).

[2] Amazon Mechanical Turk, available from ⟨http://aws.amazon.com/jp/mturk⟩ (accessed 2014-02-13).

[3] Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S. and Yeh, T.: VizWiz: Nearly Real-time Answers to Visual Questions, *UIST 2010* (2010).

[4] 芦川将之，有賀康顕，宮村祐一：PrivateCrowdSourcing を用いた言語，音声資源の収集，人工知能学会全国大会，3M3-OS-07d-2 (2013).

[5] Sheng, V.S., Provost, F.J. and Ipeirotis, P.G.: Get another label? improving data quality and data mining using multiple, noisy labelers, *KDD 2008*, pp.614–622 (2008).

[6] 青木秀人，森嶋厚行，三津石智巳：クラウドソーシングによるデータ収集のためのタスク生成手法の提案，*deim2013*, pp.1–7 (2013).

[7] Sakurai, Y., Okimoto, T., Oka, M., Shinoda, M., and Yokoo, M.: Quality-Control Mechanism utilizing Worker's Confidence for Crowdsourced Task, *Proc. 12th International Conference on Autonomous Agents and Multiagent Systems* (*AAMAS 2013*) (2013).

[8] Dawid, A.P. and Skene, A.M.: Maximum Likelihood Estimation of Observer Wrror-Rates Using the EM Algorithm, *Journal of the Royal Statistical Society, Series C*, Vol.28, No.1, pp.20–28 (1979).

[9] 梶野 洸，坪井祐太，佐藤一誠，鹿島久嗣：エキスパートによる訓練データとクラウドソーシングで作成した訓練データからの教師付き学習，人工知能学会論文誌，Vol.28, No.3, pp.243–248 (2013).

[10] 小山 聡，馬場雪乃，櫻井祐子，鹿島久嗣：クラウドソーシングにおけるワーカーの確信度を用いた高精度なラベル統合，人工知能学会全国大会，2M5-OS-07b-2 (2013).

[11] 馬場雪乃，鹿島久嗣：非定型出力をもつクラウドソーシングタスクにおける成果物の統計的品質推定，人工知能学会全国大会，2M5-OS-07b-1 (2013).

[12] Donmez, P., Carbonell, J.G. and Schneider, J.: Efficiently learning the accuracy of labeling sources for selective sampling, *Proc. 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (*SIGKDD*) (2009).

[13] NTT コム リサーチ，入手先 ⟨http://research.nttcoms.com/service/qpolicy4.html/⟩（参照 2014-09-25）.

[14] Gomes, R., Welinder, P., Krause, A. and Perona, P.: Crowdclustering, *Advances in Neural Information Processing Systems* (2011).

[15] 関根の拡張固有表現階層 7.1.0，入手先 ⟨https://sites.google.com/site/extendednamedentityhierarchy/⟩（参照 2014-09-25）.

[16] Wikipedia, available from ⟨http://ja.wikipedia.org/⟩ (accessed 2014-09-25).

[17] 澤田 宏：非負値行列因子分解 NMF の基礎とデータ/信号解析への応用，入手先 ⟨http://www.kecl.ntt.co.jp/icl/signal/sawada/mypaper/829-833_9_02.pdf/⟩（参照 2014-02-13）.

推薦文

本稿は，計算機技術で自動的に処理することが困難な問題を，人の処理能力によって解決しようとする手法「ヒューマンコンピュテーション」において重要課題の 1 つである，

© 2015 Information Processing Society of Japan

少ない作業（タスク）依頼数で作業者（ワーカ）のタスク処理結果からより高品質なデータを取得するための，タスクに対し適切なワーカの選択を行うタスク割当て手法に関するものである．具体的には，ワーカの過去のタスク処理結果をもとに非負値行列因子分解（NMF）によるクラスタリングを行い，タスクに適したワーカを含むクラスタを抽出する手法を提案している．また，2 種類のタスクに対する処理結果の実データを用いて，提案手法によるタスク依頼数の削減と品質向上の実現可能性を示しており，有用な論文であるので，推薦する．

（グループウェアとネットワークサービス研究会主査
市村　哲）

巻口 誉宗　（正会員）

1988 年生．2010 年北海道大学工学部情報エレクトロニクス学科卒業．2012 年同大学大学院情報科学研究科博士前期課程修了．同年日本電信電話株式会社入社．現在，NTT サービスエボリューション研究所にてヒューマンコンピュテーション，クラウドソーシングの研究に従事．

並河 大地　（正会員）

1983 年生．2007 年鹿児島大学理学部卒業．2009 年同大学大学院博士前期課程修了．同年日本電信電話株式会社入社．現在，NTT サービスエボリューション研究所にて画像認識サービス等，サービス提供基盤に関する研究に従事．

東 正造

1994 年大阪大学基礎工学部情報工学科卒業．1996 年奈良先端科学技術大学院大学博士前期課程修了，同年日本電信電話株式会社に入社．自然言語処理技術，地図情報処理技術（GIS）等の研究に従事．現在，NTT サービスエボリューション研究所にて，センシング技術を活用した都市データアーカイブに関する研究開発に従事．電子情報通信学会会員．

下村 道夫

1988 年早稲田大学理工学部電子通信学科卒業，1993 年同大学大学院博士後期課程修了．同年日本電信電話株式会社入社．2012 年津田塾大学非常勤講師．2012 年より法政大学兼任講師．2014 年より千葉工業大学プロジェクトマネジメント学科教授．ICT サービスの研究開発等に従事．博士（工学），電子情報通信学会会員．

金丸 直義

1964 年生．1989 年東北大学大学院博士前期課程修了．1992 年同博士後期課程修了．同年日本電信電話株式会社入社．ロボティクス，ヒューマンインタフェース等の研究に従事．現在，NTT サービスエボリューション研究所プロジェクトマネージャ．日本ロボット学会，日本福祉のまちづくり学会各会員．

© 2015 Information Processing Society of Japan