

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 21 日現在

機関番号：12102

研究種目：基盤研究(A) (一般)

研究期間：2013～2016

課題番号：25240012

研究課題名(和文) データ中心型クラウドソーシングプラットフォームの高度化とその応用に関する研究

研究課題名(英文) Research on Advanced Data-centric Crowdsourcing Platforms and its Applications

研究代表者

森嶋 厚行 (MORISHIMA, Atsuyuki)

筑波大学・図書館情報メディア系・教授

研究者番号：70338309

交付決定額(研究期間全体)：(直接経費) 31,900,000円

研究成果の概要(和文)：本研究では、アカデミアで共同利用可能なオープン型の高度クラウドソーシングプラットフォームを構築し、その上で、各応用分野の研究者と協力してクラウドソーシングの基礎および応用研究を推進した。プラットフォームは50か国以上からの参加者により一日に数千のタスクが処理されるものとなり、基礎研究に関しては、クラウドソーシングのためのワークフローやタスクの設計支援等、応用研究に関しては、図書館領域をはじめとする様々な分野において成果が得られた。

研究成果の概要(英文)：In this research, we constructed an advanced open crowdsourcing platform that is open to academia and used it to conduct fundamental and applied research on crowdsourcing in collaboration with a variety of application-domain researchers. The platform has grown to have several thousands of tasks performed per day by workers from more than 50 countries. We obtained results on fundamental research such as better design of crowdsourcing workflows and tasks, and those on applied research in a variety of domains such as libraries.

研究分野：クラウドソーシングシステム，データベースシステム

キーワード：クラウドソーシング

1. 研究開始当初の背景

クラウドソーシングはコンピュータネットワーク等を通じて不特定多数の人々に仕事を委託することである。クラウドソーシングは様々な応用領域において、データを収集、処理するための新たなアプローチとして注目を集めており、多くのシステムや活用手法が提案されている。これらの開発を効率よく行うためには、個別のシステムを作り込むのではなく、共通の機能を提供するクラウドソーシングプラットフォームを利用する事が不可欠である。しかし、大規模な汎用クラウドソーシングプラットフォームは商用のクラウドソーシング市場が提供するものしか存在せず、また、それらの機能はあまりにも低レベルである。例えば、複雑な問題を解決するためのタスク生成支援機能、宣言的記述による処理の最適化機能、適切な人材へのタスクの割り当て機能などは提供されていない。

2. 研究の目的

本研究では、アカデミアで共同利用可能な高度クラウドソーシングプラットフォームの構築を行い、それを利用した研究を推進する。

研究期間内に、少なくとも基礎研究および応用研究を通じて高度クラウドソーシングプラットフォームが提供すべき機能のいくつかを明らかにし、プラットフォーム上で実装して、これらを利用した基礎・応用研究の成果を得る。具体的には、次のような問題に取り組む。Q1. これまで得られたクラウドソーシングの最新の科学的知見を踏まえた上で、複雑なクラウドソーシングシステムを正しく・迅速に開発するためには、プラットフォームおよび関連ツールはどのような機能を提供すればよいのか？Q2. クラウドソーシング分野の基礎研究の成果は各応用分野に適用可能か？Q3. 複雑なクラウドソーシングシステムに関与するクラウド(群衆)はどのように実際に動いており、これらを活用するためにはどうすればよいのか？

3. 研究の方法

アカデミアが内部にアクセス可能でありかつ容易に利用可能な高度クラウドソーシングプラットフォーム構築と、それを利用した基礎・応用研究の推進およびプラットフォーム開発へのフィードバックを並行して行う。基礎研究・応用研究・プラットフォーム構築を三つの柱とし、研究を推進する。研究期間の前半では、アカデミック・クラウドソーシングプラットフォームの構築に関する研究に重みを置く。後半は、それを利用したクラウドソーシング応用・基礎研究推進に焦点を当てる。さらに、研究期間後半には国際プラットフォーム構築研究も推進する。

具体的には次を並行して進めることにより、

研究を推進する。

- (1) 他の研究者と協力した、オープンなアカデミッククラウドソーシングプラットフォームの構築と運営
- (2) 応用分野の研究者と協力したアプリケーション開発を通じた、科学的知見の入手とプラットフォーム開発への反映
- (3) 宣言型処理、データ品質、開発効率化等の基礎研究。

基本的に、これらを並行して進め、基礎研究、プラットフォーム研究、応用研究、応用研究の成果を基礎研究やプラットフォーム研究に適用するというループを回すが、最初の2年と後半2年でフォーカスが異なる。最初の2年間は、アカデミッククラウドソーシングプラットフォーム構築に関わる研究に焦点を主に当てる。具体的には、プラットフォームに求められる高度機能や処理効率化のための基礎技術などである。後半の2年は、焦点をプラットフォーム構築そのものから、それを利用した基礎技術・応用技術の推進に焦点を当てる。

4. 研究成果

本研究の最大の成果は、最新の研究成果を反映させた高度機能を持つ研究システムでありながら、現実世界のクラウドソーシングプロジェクトで使われ続けるオープンプラットフォームを構築したことである。具体的には、前節であげた3つの項目に関してそれぞれ次のような研究成果をあげた。

- (1) 他の研究者と協力したアカデミッククラウドソーシングプラットフォームの構築と運営

本研究で構築・運営を行ったクラウドソーシングプラットフォーム Crowd4U は、我々の知る限り、研究者により構築と運営がされているプラットフォームでありながら、現実のクラウドソーシングプロジェクトで使われ続けている世界で唯一のクラウドソーシングプラットフォームである。関連する統計情報を図1に示す。最終年度には、50か国以上からの登録者によって一日に数千程度のタスクが処理されるプラットフォームに成長した。

#Current Projects	17
#Registered Tasks	57,206
#Registered Contributors	1,132
#Average Throughput	4,238 Tasks/day

図1. Crowd4Uの2017年2月の統計情報。行われるタスク数は平日を対象としたもの。

このプラットフォームには、これまでのクラウドソーシング研究の成果をいくつか実装している。例えば、世界中から集まるクラウドワーカーはスキルや言語など多様な属性を持つことに着目し、これらの人々が協調してタスクを行うためのチームを効率的に構築する仕

組みの設計を海外の研究者と協力して行い、実際に Crowd4U に実装した [5].

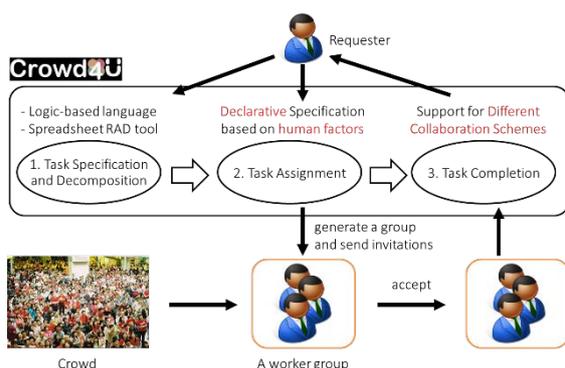


図 2. 宣言的ルールに従い、ワーカの多様な属性に応じてチームを自動的に構築

(2) 応用分野の研究者と協力したアプリケーション開発を通じた、科学的知見の入手とプラットフォーム開発への反映

いくつかの実際応用分野の研究者と協力してクラウドソーシングプロジェクトを行い、その成果をプラットフォームの開発に反映させた。特に、図書館領域においては専門家と協力して、国立国会図書館の書誌情報を対象とした書誌誤同定判定や、国立国会図書館デジタルコレクションのテキスト化などに活用し [5] [22]、その成果の一部の公開を行った。また、マイクロタスクのデザインに関して研究を行い書誌誤同定判定のための効率的なタスクデザインなどについて検証を行ったほか、マイクロタスクを PC 画面以外の環境に配布する仕組みを構築し、その場合の挙動と結果の品質を調査し、十分な品質が得られることを示した [1]。また、福島県双葉町のデジタルライブラリ構築プロジェクトにおいては、様々な言語・文化における画像へのタグ付けが要求されることから、タスク間で結果を動的にやり取りする仕組みを構築し、実際のプロジェクトにおいてもそれを用いて様々な言語のタスクを動的に生成し、17 か国語でのデータ入力を得た。また、クラウドソーシングによって学会のセッションプログラムの作成等の問題を解く手法を構築し、実際に国内の大規模学会のセッションプログラムの構築支援に利用した。

(3) 宣言型処理、データ品質、開発効率化等の基礎研究。

クラウドソーシングのためのプログラミング環境におけるワーカのインセンティブ構造の記述を支援する枠組みの有効性を理論的・実験的に示したほか、その枠組みを用いた人と計算機の処理をともに利用したソフトウェア記述の表現力に関する詳細な理論解析を行い、その限界を明らかにした [2]。具体的には、論理型言語に対してインセンティブ構造を表現するアスペクト記述を許可したプログラミング言語の表現力が、チューリングマシンの各

ステップでインタラクションを許したものと同等の表現力を持つことを示した。本枠組みを用いてインセンティブ構造を切り離して記述することにより人の行動の解析が容易になり、人による処理と計算機による処理の分担の割合を制御するなどの設計が容易となった (図 3)。

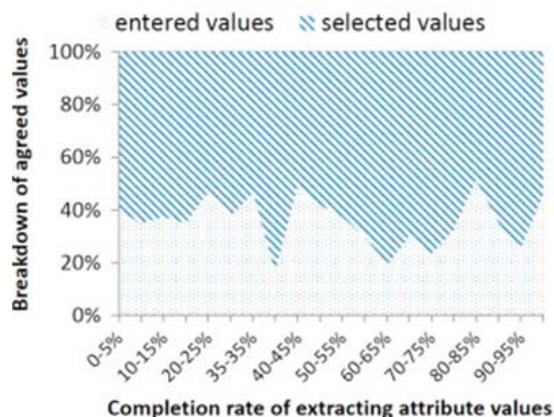


図 3. ツイートからの構造データ抽出を行うプログラムにおいて、人による抽出 (Entered values) と機械による抽出 (Selected values) を、作業初期 (x 軸左側) から機械による抽出が多くなるようにインセンティブを設定し、実際に実験したところ、理論通りの挙動を示した。

また、クラウドソーシングのためのマイクロタスクデザインに関しては、タスクの指示文を適切に記述することが困難であることが知られている。そこで、クラウドソーシングのタスク指示文自体をクラウドソースし、類似の意味を持つが表現が異なるタスク文面を集める手法を開発した。ここでの問題は、単純に指示文の書き換えをワーカに求めると、単語を置き換えたような代わり映えのしない指示文が集まる事である。そこで、タスクのデータと回答例を複数ワーカに提示し、指示文を推測させることによって多様な指示文のバリエーションを集める手法を提案した。また、実験によって、本手法によって集められた多くの指示文が、元の指示文と同様であるが全く同じではない指示文となっており、その中により良い指示文が含まれるケースがあることを示した [4]。

5. 主な発表論文等

[雑誌論文] (計 25 件)

[1] Kosetsu Ikeda, Ryota Hayashi, Kiyonori Nagasaki, Atsuyuki Morishima, “Human-assisted OCR of Japanese Books with Different kinds of Microtasks”,

- iConference 2017, 査読有, 5 pages ,
Wuhun ,China, March 24 ,2017.
- [2] Atsuyuki Morishima, Shun Fukusumi,
Hiroyuki Kitagawa, “CyLog/Game Aspect:
An Approach to Separation of Concerns in
Crowdsourced Data Management” ,
Information Systems, Elsevier, 査読有,
Vol.62, pp.170-184, December 2016.
- [3] Yuhki Shiraishi, Jianwei Zhang,
Daisuke Wakatsuki, Katsumi Kumai,
Atsuyuki Morishima, “Crowdsourced
Real-Time Captioning of Sign Language by
Deaf and Hard-of-Hearing People” iiWAS
'16 Proceedings of the 18th
International Conference on Information
Integration and Web-based Applications
and Services, 査読有, pp.54-63, November
28 - 30, 2016.
- [4] Ryota Hayashi, Nobuyuki Shimizu,
Atsuyuki Morishima, “Obtaining
Rephrased Microtask Questions from
Crowds” , the 8th International
Conference on Social Informatics
(SocInfo 2016) , 査読有, pp.323-336,
Bellevue, USA, November 14, 2016.
- [5] Atsuyuki Morishima, Kiyonori
Nagasaki, Kosetsu Ikeda, Takanori
Kawashima, “The Hondigi/L-Crowd Joint
Project: A Microtask-based Approach for
Transcribing Japanese Texts” ,
International Conference on Digital
Libraries 2016 (ICDL 2016), 査読無, pp.
108-115, India Habitat Centre, New Delhi,
Delhi, India, December 13-16, 2016.
- [6] Kosetsu Ikeda, Atsuyuki Morishima,
Habibur Rahman, Senjuti Basu Roy,
Saravanan Thirumuruganathan, Sihem Amer-
Yahia, Gautam Das. “Collaborative
Crowdsourcing with Crowd4U” . 42nd
International Conference on Very Large
Data Bases (VLDB2016), 査読有, pp.1497-
1500 New Delhi, India, September 5-9,
2016.
- [7] 権守健嗣, 森嶋厚行, 歳森敦, 北川博
之, 「クラウドソーシングデータ処理におけ
る一般化選択フィルタの利用と動的選択手
法」 日本データベース学会和文論文誌
Vol.14, Article No.8, 査読有, 8 pages,
2016年3月.
- [8] Emi Sakurai, Atsuyuki Morishima,
Kosetsu Ikeda, Nobutaka Suzuki,
“Bookshelf problem: a human-in-the-loop
approach for data grouping without
complete information” , Proc. of
iConference2016, 査読有, 10 pages,
Philadelphia, USA, March 21st, 2016.
- [9] Senan Kiryakos, Shigeo Sugimoto,
“A Linked Data Model to Aggregate
Serialized Manga from Multiple Data
Providers” , Proceedings of ICADL
2015(Springer Lecture Notes in Computer
Science 9469) , 査読有, pp.120-131,
December 5-9, 2016.
- [10] Nobuyuki Shimizu, Atsuyuki
Moirshima, Ryota Hayashi. “A
Crowdsourcing Method for Obtaining
Rephrased Questions” ,Proc. of Third
AAAI Conference on Human Computation and
Crowdsourcing (HCOMP2015)
WorkInProgress, 査読有, pp. 32-33, San
Diego, USA, November 8-11, 2015.
- [11] Li Chunqiu, Mitsuharu Nagamori,
Shigeo Sugimoto, “TEMPORAL
INTEROPERABILITY OF METADATA - AN
INTEROPERABILITY-BASED VIEW FOR
LONGEVITY OF METADATA -” , Proceedings
of the 6th International Conference on
Asia-Pacific Library and Information
Education and Practice(A-LIEP 2015), 査読
有, pp.212-222, October, 2015.

- [12] 森嶋厚行, 「クラウドソーシング - 新たな情報コンテンツ創造と社会デザインに向けて-」情報処理学会学会誌, 査読無, vol. 56, No. 10, pp. 978-981, 2015年10月.
- [13] Shun Fukusumi, Atsuyuki Morishima, Hiroyuki Kitagawa, “Game Aspect: An Approach to Separation of Concerns in Crowdsourced Data Management”, Proc. of 27th International Conference on Advanced Information Systems Engineering (CAiSE2015), 査読有, pp. 3-19, Stockholm, Sweden, June 10, 2015 (Best Paper Award).
- [14] 青木秀人, 森嶋厚行, 「クラウドソーシングによるデータ列挙のための分割統治手法の提案と評価」, 電子情報通信学会論文誌, 査読有, vol. J98-D, No. 5, pp. 760-773, 2015年5月.
- [15] Jan Askhoj, Shigeo Sugimoto, Mitsuharu Nagamori, “Development of a Manga Ontology by Extracting Instances from Web Resources and Bibliographic Records,” International of Metadata, Semantics and Ontologies, 査読有, Vol. 10, No. 1, pp. 1-11, March 2015.
- [16] 渡辺知恵美, 中村聡史, 「オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング」, 人工知能学会論文誌, 査読有, Vol. 30(2015), No. 1, p. 340-352, 2015年1月6月.
- [17] Atsuyuki Morishima, Sihem Amer-Yahia, Senjuti Basu Roy. “Crowd4U: An Initiative for Constructing an Open Academic Crowdsourcing Network”, Second AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2014) WorkInProgress, 査読有, pp. 50-51, Pittsburgh, USA, November 2-4, 2014.
- [18] Tetsuya Mihara, Yusuke Iwama, Mitsuharu Nagamori, Shigeo Sugimoto, “Development of a Manga Ontology by Extracting Instances from Web Resources and Bibliographic Records”, International Workshop on Global Collaboration of Information Schools WIS 2014, 査読有, 10 pages, November, 2014.
- [19] Atsuyuki Morishima, Shiori Tomita, Takanori Kawashima, Takashi Harada, Norihiko Uda, Sho Sato, Yukihiro Abematsu, “A Crowdsourcing Approach for Finding Misidentifications of Bibliographic Records,” Proc. of iConference 2014, 査読有, pp. 177-191, Berlin, Germany, March 6, 2014.
- [20] 森嶋厚行, 「ヒューマンコンピューテーションのためのプラットフォームとソフトウェア開発」, 人工知能学会誌, 査読無, Vol. 29, No. 1, pp. 12-18, 2014年1月.
- [21] 丹治寛佳, 森嶋厚行, 井ノ口宗成, 北川博之, 「Web情報を用いた竜巻経路推定支援のためのクラウドソーシング技術開発の試み」情報処理学会論文誌 データベース(TOD60), 査読有, vol. 6, No. 5, pp. 95-106, 2013年12月27日.
- [22] Atsuyuki Morishima, Takanori Kawashima, Takashi Harada, Norihiko Uda, Ikki Ohmukai, “L-Crowd: A Library Crowdsourcing Project by LIS and CS Researchers in Japan,” International Conference on Digital Library (ICDL2013), 査読無, pp. 40-47, New Delhi, Delhi, India, Nov. 28, 2013.
- [23] Hideto Aoki, Atsuyuki Morishima, “A Divide-and-Conquer Approach for Crowdsourced Data Enumeration,” The 5th International Conference on Social Informatics (SocInfo2013), 査読有, pp. 60-74, Kyoto, Nov. 25, 2013.
- [24] Kenji Gonnokami, Atsuyuki Morishima, Hiroyuki Kitagawa,

“Condition-Task-Store: A Declarative Abstraction for Microtask-based Complex Crowdsourcing,” The First VLDB Workshop on Databases and Crowdsourcing (DBCrowd2013), 査読有, pp. 20-25, Trento, Italy, Aug. 26-30, 2013.

[25] 権守健嗣, 森嶋厚行, 「クラウドソーシングシステム開発支援のための宣言的記述の提案」日本データベース学会論文誌, 査読有, Vol 12, No. 1, pp. 85-90, 2013年6月.

[学会発表] (計 49 件)

[図書] (計 2 件)

[26] Atsuyuki Morishima, Andreas Rauber, Chern Li Liew, Digital Libraries: Knowledge, Information, and Data in an Open Access Society - 18th International Conference on Asia-Pacific Digital Libraries, ICADL 2016, Tsukuba, Japan, December 7-9, 2016, Proceedings. Lecture Notes in Computer Science 10075, Springer 2016, ISBN 978-3-319-49303-9

[27] Atsuyuki Morishima, CyLog/Crowd4U: A Case Study of a Computing Platform for Cybernetic Dataspaces”, Handbook of Human Computation, Pietro Michelucci (eds.), Springer, pp. 561-572, November 2013 .

[産業財産権]

○出願状況 (計 1 件)

名称: 情報提供装置、情報提供方法および情報提供プログラム
発明者: 清水伸幸, 森嶋厚行, 丹治寛佳
権利者: ヤフー株式会社, 筑波大学
番号: 特願 2015-056469
出願年月日: 2015年3月19日
国内外の別: 国内

○取得状況 (計 1 件)

名称: 推定装置、推定方法および推定プログ

ラム

発明者: 清水伸幸, 森嶋厚行, 丹治寛佳
権利者: ヤフー株式会社, 筑波大学
番号: 5992951
出願年月日: 2014年4月28日
取得年月日: 2016年8月26日
国内外の別: 国内

[その他]

Crowd4U クラウドソーシングプラットフォーム: crowd4u.org

6. 研究組織

(1) 研究代表者

森嶋 厚行 (MORISHIMA, Atsuyuki)
筑波大学図書館情報メディア系・教授
研究者番号: 70338309

(2) 研究分担者

杉本 重雄 (SUGIMOTO, Shigeo)
筑波大学図書館情報メディア系・教授
研究者番号: 40154489

中村 聡史 (NAKAMURA, Satoshi)
明治大学総合数理学部・准教授
研究者番号: 50415858

池田 光雪 (Ikeda, Kosetsu)
千葉大学アカデミック・リンク・センター
・特任助教
研究者番号: 10779606

(3) 連携研究者

井ノ口 宗成 (INOBUCHI, Munenari)
静岡大学情報学部・講師
研究者番号: 90509944

宇陀 則彦 (UDA, Norihiko)
筑波大学図書館情報メディア系・准教授
研究者番号: 50261813

原田 隆史 (HARADA, Takashi)
同志社大学免許資格課程センター・教授
研究者番号: 30218648

渡辺 知恵美 (WATANABE, Chiemi)
筑波大学システム情報系・助教
研究者番号: 20362832

(4) 研究協力者

川島 隆徳 (国立国会図書館)
Sihem Amer-Yahia (CNRS, LIG, France)
Vilas Wuwongse (Mahidol University, Thailand)