



# 私のブックマーク

## クラウドソーシングとヒューマンコンピューテーション<sup>†1</sup>

馬場 雪乃 (国立情報学研究所 / JST, ERATO, 河原林巨大グラフプロジェクト)

### 1. はじめに

不特定多数の人に仕事を発注する仕組み「クラウドソーシング」の登場は、大量の人的資源の獲得を容易にした。コンピュータ科学の諸分野では、クラウドソーシングをより成熟した基盤にするために、品質管理やメカニズムデザイン、プラットフォーム設計など多様な観点で研究が進められている。また、クラウドソーシングの登場を受けて人工知能分野では、人間と計算機の協調問題解決を目指す「ヒューマンコンピューテーション」に関する研究が促進された。本稿では、以下五つの観点で情報源を紹介する。

- (1) クラウドソーシング・ヒューマンコンピューテーションと、関連する研究の概略
- (2) クラウドソーシング・ヒューマンコンピューテーションの最新研究動向
- (3) クラウドソーシング・ヒューマンコンピューテーションの活用事例
- (4) 研究でのクラウドソーシング活用
- (5) クラウドソーシング・ヒューマンコンピューテーションの研究資源

### 2. クラウドソーシング・ヒューマンコンピューテーションと、関連する研究の概略

- The Rise of Crowdsourcing [1]  
「クラウドソーシング」という言葉が最初に登場した米国 *Wired* 誌の記事。
- クラウドソーシング—みんなのパワーが世界を動かす [2]  
上記記事の執筆者 *Jeff Howe* による、クラウドソーシングの概念と事例紹介の書籍。
- Enrique Estellés-Arolas and Fernando González-Ladrón-de-Guevara: Towards an integrated crowdsourcing definition, *Journal of Information Science*, Vol. 38, No. 2 (2012) [3]  
クラウドソーシングの定義を調査した論文。
- Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton: The future of crowd works, *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW)* (2013) [4]  
クラウドソーシング研究のサーベイ論文。
- Luis von Ahn: Human Computation, Ph. D. Thesis, Carnegie Mellon University (2005) [5]  
ヒューマンコンピューテーションの生みの親である Luis von Ahn [6] の博士論文。
- Edith Law and Luis von Ahn: *Human Computation*, Morgan & Claypool Publishers (2011) [7]  
ヒューマンコンピューテーションの研究動向をまとめた書籍。
- Luis von Ahn and Edith Law: *Human Computation: Core Research Questions and State of the Art* (2011) [8]  
上記書籍と同様の内容のチュートリアル資料。
- Pietro Michelucci: *Handbook of Human Computation*, Springer (2013) [9]  
ヒューマンコンピューテーションに関する研究トピック紹介の書籍。
- *Crowdsourcing & Human Computation* [10]  
ペンシルベニア大学の Chris Callison-Burch [11] による、クラウドソーシングとヒューマンコンピューテーションの講義資料。

<sup>†1</sup> [http://www.ai-gakkai.or.jp/my-bookmark\\_vol130-no1](http://www.ai-gakkai.or.jp/my-bookmark_vol130-no1)

人工知能 Vol. 29 No. 1 (2014 年 1 月) では「ヒューマンコンピューテーションとクラウドソーシング」特集が生まれ、8 篇の解説記事が掲載されている。

- ・ 鹿島久嗣, 馬場雪乃: ヒューマンコンピューテーション概説
- ・ 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
- ・ 櫻井祐子, 松原繁夫: ヒューマンコンピューテーションのためのメカニズムデザイン
- ・ 小山 聡: ヒューマンコンピューテーションの品質管理
- ・ 水山 元: 予測市場とその周辺
- ・ 高木啓伸, 井床利生, 斉藤 新, 小林正朋: クラウドアクセシビリティ—クラウドソーシングによる障害者支援—
- ・ 後藤真孝, 吉井和佳, 中野倫靖, 緒方 淳: クラウドソーシングに基づくメディア処理サービス—能動的音楽鑑賞サービス Songle と音声情報検索サービス PodCastle—
- ・ 芦川将之, 池田朋男: クラウドソーシングを用いたアノテーション

人工知能学会誌 Vol. 27 No. 4 (2012 年 7 月) には以下の解説記事が掲載されている。

- ・ 鹿島久嗣, 梶野 洸: クラウドソーシングと機械学習

### 3. クラウドソーシング・ヒューマンコンピューテーションの最新研究動向

#### (1) blog

- ・ Follow the Crowd [12]  
クラウドソーシングの最新研究紹介 blog. 複数の研究者により運営されている。
- ・ A Computer Scientist in a Business School [13]  
クラウドソーシングの品質管理やプラットフォーム調査に関して多数の論文を発表している, ニューヨーク大学の Panos Ipeirotis [14] による blog.

#### (2) 国際会議

クラウドソーシングとヒューマンコンピューテーションに関する論文が発表される主要国際会議は以下である。

- ・ HCOMP [15]  
ヒューマンコンピューテーションとクラウドソーシングに関する国際会議. ワークショップとして 4 回開催された後, 2013 年より AAAI 主催の国際会議となった。
- ・ AAAI [16a], IJCAI [16b]  
人工知能に関する国際会議. ヒューマンコンピューテーションや, クラウドソーシングの品質管理についての論文が例年採択されている。
- ・ CHI [17]  
ヒューマンコンピュータインタラクションに関する国際会議. クラウドソーシングを利用したアプリケーションが多数発表されている。
- ・ UIST [18]  
ユーザインタフェースに関する国際会議. CHI と同様に, VizWiz [19] や SoyLent [20] など, クラウドソーシングを利用したアプリケーションが多数発表されている。
- ・ CSCW [21]  
Computer-Supported Cooperative Work と Social Computing に関する国際会議. Being a Turker [22] など, クラウドソーシングの調査論文が多数発表されている。
- ・ SIGMOD [23]  
データベースに関する国際会議. Crowd DB: Answering queries with crowdsourcing [24] など, ヒューマンコンピューテーションによるデータベース操作を扱った論文が多数発表されている。
- ・ WWW [25]  
Web に関する国際会議. 例年, クラウドソーシングのセッションが設けられ, 品質管理や報酬設定に関する論文が発表されている。

#### (3) ワークショップ

さまざまな国際会議で, クラウドソーシング・ヒューマンコンピューテーションを題材にしたワークショップが開催されている。そのうち, 2 回以上開催されているものを紹介する。

- Workshop on Crowdsourcing for Multimedia (CrowdMM) [26]  
マルチメディア研究におけるクラウドソーシング・ヒューマンコンピューテーション活用についてのワークショップ. ACM Multimedia [27] の併設で, 2012年より毎年開催されている.
- Workshop on Multimodal Crowd Sensing (CrowdSens) [28]  
クラウドソーシングを利用して実世界の情報を獲得する「クラウドセンシング」についてのワークショップ. 2012年に CIKM, 2014年に KDD の併設で開催されている.

ほかに, 機械学習分野の国際会議 ICML と NIPS では, 近年はクラウドソーシング・ヒューマンコンピューテーションに関連したワークショップが毎年併設されている. 特に, 最近3年間に開催されたものを紹介する.

- ICML'12 Workshop on Machine Learning in Human Computation & Crowdsourcing [29]
- ICML'13 Workshop on Machine Learning Meets Crowdsourcing [30]
- ICML'14 Workshop on Crowdsourcing and Human Computing [31]
- NIPS'12 Workshop on Human Computation for Science and Computational Sustainability [32]
- NIPS'13 Workshop on Crowdsourcing: Theory, Algorithms and Applications [33]
- NIPS'14 Workshop on Crowdsourcing and Machine Learning [34]

国内では人工知能学会全国大会において, オーガナイズドセッションが2013年と2014年に開催されている.

- オーガナイズドセッション「ヒューマンコンピューテーションとクラウドソーシング」[35]

#### (4) その他

- Crowdsourcing News, Events, and Resources [36]  
テキサス大学オースティン校の Matt Lease [37] が管理する, クラウドソーシング研究関連情報のまとめ.
- クラウドソーシング研究会 [38]  
国内クラウドソーシング研究者のコミュニティ.

#### 4. クラウドソーシング・ヒューマンコンピューテーションの活用事例

クラウドソーシング・ヒューマンコンピューテーションを活用したプロジェクトなどを紹介する.

- reCAPTCHA [39]  
「人間かロボットか」を判定する仕組み CAPTCHA を, 人手による書籍電子化に活用するプロジェクト.
- Foldit [40]  
タンパク質の構造予測問題をゲームにすることで人間に解決させる取組み.
- Jim Gray の検索 [41]  
海上で行方不明になった研究者 Jim Gray の検索プロジェクト. 「海域の衛星画像から Jim Gray のクルーザーを見つけるタスク」が Amazon Mechanical Turk で発注された.
- DARPA Network Challenge [42]  
米国内の10地点に設置された赤い風船を見つけ出すコンペティション. ソーシャルネットワーク活用やチームビルディングの手法を試すことを目的に開催された.
- Fast Adaptable Next-Generation Ground Vehicle (FANG) Mobility/Drivetrain Challenge [43]  
DARPA 主催の, クラウドソーシングを利用した軍用車両設計のコンペティション.
- Netflix Prize [44]  
オンライン DVD レンタルサービスを提供する Netflix が, 推薦アルゴリズム改善のために開催したコンペティション.
- Duolingo [45]  
言語教育プラットフォーム. 一定のスキルに到達した学習者に対して, 翻訳作業を発注する.

#### 5. 研究でのクラウドソーシング活用

##### (1) 汎用的なプラットフォーム (日本国外)

- Amazon Mechanical Turk [46]  
2005年にサービス開始された, 米国 Amazon 社運営のクラウドソーシングの老舗プラットフォーム. 画像分類や音声の書き起こしなどの小規模な仕事に特化している. 発注者になるには, 請求先として米国内住所が設定されているクレジットカードなどが必要である (2014年11月現在).

- CrowdFlower [47]  
Amazon Mechanical Turk と同種のプラットフォーム。発注時には, PayPal 経由で支払いが可能である (2014 年 11 月現在)。
  - oDesk [48]  
デザイン, ソフトウェア開発, 文書執筆などの比較的規模が大きい仕事を対象にしたプラットフォーム。
- (2) 汎用的なプラットフォーム (日本国内)
- ランサーズ [49]  
2008 年にサービス開始された, 日本における老舗クラウドソーシング・プラットフォーム。「タスク方式」と呼ばれる小規模な仕事と, 「プロジェクト方式」, 「コンペ方式」と呼ばれる比較的規模が大きい仕事, どちらも発注可能である。
  - クラウドワークス [50]  
ランサーズと同様に, ささまざまな種類の仕事に対応したプラットフォーム。
  - Yahoo! クラウドソーシング [51]  
小規模な仕事に特化したプラットフォーム。
  - Crowd4U [52]  
公共・学術利用を目的としたクラウドソーシング・プラットフォーム。ボランティアに対して仕事を発注することができる。
- (3) 特定の目的をもったプラットフォーム
- デザイン, ソフトウェア開発, 翻訳など, 特定の種類の仕事に特化したクラウドソーシング・プラットフォームも存在する。
- 99designs [53]  
ロゴや Web ページ, 名刺などのデザインに特化したプラットフォーム。
  - TopCoder [54]  
プログラミングコンテストの開催で有名な TopCoder は, ソフトウェア開発のクラウドソーシング・プラットフォームも提供している。
  - uTest [55]  
ソフトウェアのテストに特化したプラットフォーム。
  - Conyac [56]  
文章翻訳に特化したプラットフォーム。
- データ解析に特化したクラウドソーシング・プラットフォームも登場している。
- Kaggle [57]  
予測モデル構築のプラットフォーム。参加者を競わせ, 優れたモデルの獲得を目指す。
  - CrowdSolving [58], Deep Analytics [58]  
Kaggle と同様のプラットフォームの国内版。
  - ビッグデータ大学 [59]  
データ解析者の教育を目的とするプラットフォーム。
- (4) クラウドソーシング利用のノウハウ・ツール
- Guidelines for Academic Requesters [60]  
複数の研究者によって策定された, 学術目的でのクラウドソーシング利用のガイドライン。
  - Best Practices Guide for Requesters [61]  
Amazon Mechanical Turk の発注者を対象にした文書だが, その他のプラットフォームにも適用可能なコツが紹介されている。
  - TurKit [62]  
Amazon Mechanical Turk の API を Java, JavaScript で呼び出すためのライブラリ。

## 6. クラウドソーシング・ヒューマンコンピューテーションの研究資源

主に, 複数ワーカーの回答を統合することで品質を保証する「回答統合手法」の研究のために, データセットや共通タスクが整備されている。

- (1) ツール
  - ・ SQUARE [63]  
複数の回答統合手法が実装されており、ベースラインとして用いることができる。
- (2) データセット
  - ・ 回答統合手法用のデータセット [64]  
前述の SQUARE で用いられているデータセットの一覧。
  - ・ Amazon Mechanical Turk で収集された言語や音声のデータセット [65]  
NAACL'10 併設のワークショップ Creating Speech and Language Data With Amazon's Mechanical Turk [66] において収集されたデータセット。
- (3) 共通タスク
  - ・ CrowdScale [67]  
HCOMP'13 で実施された、回答統合に関する共通タスク。
  - ・ MediaEval Crowdsourcing Task [68]  
クラウドソーシングで獲得した、マルチメディア関連コンテンツに対するアノテーションの統合を目指す共通タスク。2013 年は、Flickr に投稿された写真へのアノテーション、2014 年はコメントへのアノテーションが対象とされた。
  - ・ TREC Crowdsourcing Track [69]  
情報検索分野で実施された共通タスク。文書とトピックの適合度のアノテーションをクラウドソーシングで精度良く収集する方法を競う。

## 7. おわりに

本稿では、クラウドソーシング・ヒューマンコンピューテーションに関する情報源を五つの観点に沿って紹介した。本研究分野拡大の一助となれば幸いである。

## 謝辞

本稿の執筆にあたり、クラウドソーシング研究会の皆様からご助言をいただきました。心より感謝申し上げます。

- [1] <http://archive.wired.com/wired/archive/14.06/crowds.html>
- [2] <http://www.amazon.co.jp/dp/4153200018>
- [3] <http://jis.sagepub.com/content/38/2/189.abstract>
- [4] <http://dl.acm.org/citation.cfm?id=2441923>
- [5] <http://reports-archive.adm.cs.cmu.edu/anon/2005/abstracts/05-193.html>
- [6] <https://www.cs.cmu.edu/~biglou/>
- [7] <http://www.morganclaypool.com/doi/abs/10.2200/S00371ED1V01Y201107AIM013>
- [8] <http://humancomputation.com/2011/Tutorial.html>
- [9] <http://www.springer.com/computer/ai/book/978-1-4614-8805-7>
- [10] <http://crowdsourcing-class.org/syllabus.html>
- [11] <http://www.cis.upenn.edu/~ccb/>
- [12] <http://crowdresearch.org/blog/>
- [13] <http://www.behind-the-enemy-lines.com/>
- [14] <http://people.stern.nyu.edu/panos/>
- [15] <http://www.humancomputation.com/2014/>
- [16a] <http://www.aaai.org/Conferences/AAAI/aaai15.php>
- [16b] <http://ijcai-15.org/>
- [17] <http://chi2015.acm.org/>
- [18] <http://www.acm.org/uist/uist2014/>
- [19] <http://www.vizwiz.org/>
- [20] <http://projects.csail.mit.edu/soylent/>
- [21] <http://cscw.acm.org/2015/>
- [22] <http://dl.acm.org/citation.cfm?id=2531602.2531663>

- [23] <http://www.sigmod2015.org/>
- [24] <http://dl.acm.org/citation.cfm?id=1989331>
- [25] <http://www.www2015.it/>
- [26] <http://www.crowdmm.org/>
- [27] <http://acmmm.org/2014/>
- [28] <http://sysrun.haifa.il.ibm.com/hrl/crowdsens2014/>
- [29] <https://crowdml12.wordpress.com/>
- [30] [http://www.ics.uci.edu/~qliu1/MLcrowd\\_ICML\\_workshop/](http://www.ics.uci.edu/~qliu1/MLcrowd_ICML_workshop/)
- [31] <http://crowdwisdom.cc/icml2014/index.html>
- [32] <http://www.cs.cornell.edu/~damoulas/Site/HCSCS.html>
- [33] [http://www.ics.uci.edu/~qliu1/nips13\\_workshop/](http://www.ics.uci.edu/~qliu1/nips13_workshop/)
- [34] <http://crowdwisdom.cc/nips2014/>
- [35] <https://sites.google.com/site/crowdsourcingresearch/jsai2014>
- [36] <http://ir.ischool.utexas.edu/crowd/>
- [37] <https://www.ischool.utexas.edu/~ml/>
- [38] <https://sites.google.com/site/crowdsourcingresearch/>
- [39] <https://www.google.com/recaptcha/intro/index.html>
- [40] <http://fold.it/portal/>
- [41] [http://archive.wired.com/techbiz/people/magazine/15-08/ff\\_jimgray?currentPage=all](http://archive.wired.com/techbiz/people/magazine/15-08/ff_jimgray?currentPage=all)
- [42] <http://archive.darpa.mil/networkchallenge/>
- [43] <http://www.darpa.mil/NewsEvents/Releases/2013/04/22.aspx>
- [44] <http://www.netflixprize.com/>
- [45] <https://www.duolingo.com/>
- [46] <https://www.mturk.com/mturk/welcome>
- [47] <http://www.crowdflower.com/>
- [48] <https://www.odesk.com/>
- [49] <http://www.lancers.jp/>
- [50] <http://crowdworks.jp/>
- [51] <http://crowdsourcing.yahoo.co.jp/>
- [52] <http://crowd4u.org/ja/>
- [53] <http://99designs.jp/>
- [54] <http://appirio.com/crowdsourcing-development-design-data-science/>
- [55] <http://www.utest.com/>
- [56] <https://conyac.cc/ja>
- [57] <https://www.kaggle.com/>
- [58] <https://crowdsolving.jp/>
- [59] <http://universityofbigdata.net/>
- [60] [http://wiki.wearedynamo.org/index.php/Guidelines\\_for\\_Academic\\_Requesters](http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters)
- [61] [http://mturkpublic.s3.amazonaws.com/docs/MTURK\\_BP.pdf](http://mturkpublic.s3.amazonaws.com/docs/MTURK_BP.pdf)
- [62] <http://groups.csail.mit.edu/uid/turkit/>
- [63] <http://ir.ischool.utexas.edu/square/>
- [64] <http://ir.ischool.utexas.edu/square/data.html>
- [65] <https://sites.google.com/site/amtworkshop2010/data-1>
- [66] <https://sites.google.com/site/amtworkshop2010/>
- [67] <http://www.crowdscale.org/shared-task>
- [68] <http://www.multimediaeval.org/mediaeval2014/crowd2014/>
- [69] <https://sites.google.com/site/treccrowd/>