

研究報告書

「(研究課題名)」

研究タイプ: 通常型

研究期間: 平成 22 年 10 月～平成 26 年 3 月

研究者: 森嶋厚行

1. 研究のねらい

本研究では、デジタルコンピュータと人の知を利用するハイブリッドコンピューティングを設計・実装するための“プログラミング言語”および“開発実行環境”を開発する。かつては、プログラミングの対象は計算機の動作だけであった。一方、近年は計算機では処理が困難な問題の処理を人間に委託するヒューマンコンピューテーションや、不特定多数の人に委託するクラウドソーシングが注目を集めている。今後人類が直面する困難な問題を解決するために、得意分野が異なる計算機と人間の知を融合して立ち向かう事が重要になると考えられる。

本研究で対象とするアプリケーションは次の性質を持つものである。(1) データの生成、入力、維持、検索などが主要な機能であり、かつ(2) 計算機だけでなく人による処理が重要な構成要素である。これらの性質を満たすアプリケーションは数多く存在するが、プログラムを体系的に記述するための枠組みはこれまで開発されていない。

これまでデータ中心型のプログラミングでは、SQL を代表とするデータベース言語が利用されていた。しかし、既存のデータベース言語では、そのセマンティクスで閉世界仮説(Closed World Assumption。必要な情報は全て計算機に格納されており、データベースが現実世界を完全に反映していると考えられること)を採用している。したがって、データベースには存在しないが人が知っているデータを扱うための処理は、データベース言語が扱う問題の範囲外になっており、上手く抽象化できず汎用言語で作り込みをせざるを得なかった。

そこで、本研究では、データベース言語を拡張して、閉世界仮説を前提としない意味を持つプログラミング言語を設計し、人と計算機の知の融合に関する研究を行う。また、これを実行するための開発実行環境を構築する。

本研究のチャレンジは二つの側面からとらえることができる。(1) 人と計算機の協調処理を記述し、プログラムし、議論するための宣言的なアブストラクションや関連する理論などの科学的知見の追求というサイエンス面でのチャレンジ。(2) 人と計算機の知の融合を実現するアプリケーションを迅速に開発・実行するための開発実行環境を構築するというソフトウェア工学的なチャレンジ。これらは本研究の両輪であり、どちらも平行して進めることに意義があると考えられる。

2. 研究成果

(1) 概要

本研究では、特にクラウドソーシングを想定し、人と計算機の処理を組み合わせた応用プログラムを記述するためのプログラミング言語 CyLog の設計およびその開発実行環境を構築し、並行してそれらを利用した理論・応用研究を行った。主な成果は次の通りである

【人の知の利用という視点からのプログラミング言語表現力の理論解析】理論研究における重要な貢献は、人による作業を構成要素として含むプログラミング言語の表現力の評価尺度として、プログラミング言語によって表現可能な(ゲーム理論的な意味での)ゲームのクラスを導入した事である。プログラミング言語が表現可能なゲームのクラスが大きいと言うことは、人の知を利用した処理の記述力が高いことを意味する。この評価尺度を利用して、既存のプログラミング言語、とくにクラウドソーシングを記述するための言語と比較して、CyLog がより表現力が高く、かつ宣言型記述が可能な言語である事を示した。

【クラウドソーシングにおける高次処理の可能性の開拓】既存のクラウドソーシング技術では、データの処理などを人に委託するものの、クラウドソーシングする問題自身の分割や処理の最適化といったより高次の処理は委託しないと言うことが一般的であった。本研究では、これらの高次処理をクラウドソーシングする枠組みを提案し、実験により特定の文脈においては有効なアプローチであることを示した。

【クラウドソーシングの研究インフラ構築】これまで、汎用のクラウドソーシングプラットフォームは商用のものが一般的であったが、CyLog プログラムの処理系を実装した非商用クラウドソーシングプラットフォーム Crowd4U を構築し、国内 20 大学の協力者を得て、大学間ネットワークを作った。これは、我々の知る限り、汎用クラウドソーシングエンジンを持つ世界初の大学間ネットワークである。

【応用領域専門家と協力したクラウドソーシング応用研究】構築したシステムを利用して、国立国会図書館の書誌情報のデータクリーニングや、竜巻災害時の状況把握のためのクラウドソーシング応用の研究を推進した。どちらも応用領域の専門家の協力を得て研究を進め、クラウドソーシングがこれらの領域で十分利用可能である事を示すことができた。

(2) 詳細

【研究テーマ A: 人と計算機を対象としたプログラミングの理論研究】

(1) 人と計算機の知の融合のためのプログラミング言語とその表現力の理論解析

CyLog は、データベース言語の一つである Datalog 風のプログラミング言語であり、人と計算機の知を融合したデータ集約型アプリケーションを簡潔に記述・実装可能である。CyLog では、Datalog の述語の一部が“オープン”であることを許可する。これは、その述語の評価が計算機ではなく人によって行われることを表すものであり、人と計算機の処理の融合を宣言的に記述することを可能にしている。しかし、そのような“オープンな”言語の問題は、プログラムのセマンティクス(実行時の結果)が明らかでなくなることである。本研究では、CyLog にデータ入力者へのフィードバックの仕組みを用意し、ゲーム理論を用いて「データ入力者が合理的であれば到達する結果」をプログラムのセマンティクスとすることを提案した。

本理論研究における最も重要な貢献は、人による作業を構成要素として含むプログラミング言語の表現力の評価尺度として、プログラミング言語によって表現可能な(ゲーム理論的な意味での)ゲームのクラスを導入した事である。さらに、現在一般的なマイクロタスク型クラウドソーシングプラットフォームで広く利用されているマイクロタスクテンプレートによるタスク記述の一般化であり、かつ CyLog と同等の表現力を持つ抽象表現 CTS ルールを提案した。また、既存のクラウドソーシング記述のための言語と比較して、CyLog が宣言的記述可能、チューリング完全、実装可能なゲームのクラスが大きい(多種多様である)という望ましい性質を満たすことを証明した(図 1)。

Measure	Declarative Description	Turing complete	Class of games
Abstraction / Framework			
MTurk alone	No	No	$\subseteq g_1$
CrowdForge	No	No	$\subseteq g_N$ (generally, $N = 3$)
CrowdDB / Deco / Qurk	Yes	No	$\subseteq g_N$
CTS Abstraction/CyLog	Yes	Yes	g_x
Imperative programming languages with MTurk API	No	Yes	g_x

図 1. クラウドソーシング記述言語の表現力の比較

(2) クラウドソーシングによる高次処理の可能性の開拓

既存のクラウドソーシング技術では、データの処理などを人に委託するものの、クラウドソーシングする問題自身の分割や処理の最適化といったより高次の処理は委託しないと言うことが一般的であった。本研究では、これらの高次処理をクラウドソーシングする枠組みを提案し、実験によって特定の文脈においては実現可能であることを示した。具体的には、データの結合演算における最適化のクラウドソーシング、および、データの列挙演算における問題分割のクラウドソーシングを行った。詳細は次の通りである。

- ・Crowdsourced Join Prefilter: クラウドソーシングを用いて、二つのデータ集合のマッチング(結合演算)を行う Crowdsourced Joins の研究を行った。特に、Crowdsourced Joins の処理の効率化そのものをクラウドソーシングする仕組みが可能であることを示した。具体的には、M 個のデータと N 個のデータの結合を行うためには、単純には $M \times N$ の比較タスクが必要であるが、そのタスク数を減らすための Join-Prefilter の発見をおこなうためのタスクをクラウドソーシングによって処理し、その増加分を差し引いても、最終的に必要なタスク数を大幅に削減可能であることを示した。

- ・Crowdsourced Data Enumeration: クラウドソーシングを用いて数多くのデータを列挙する演算において、分割統治を用いた手法を提案し、それにより再現率が向上することを理論的、実験的に示した。本手法は、問題領域に存在する階層構造をあらかじめ与えることにより、トップダウンに問題を分割するプロセスをクラウドソーシングを用いて行い、より小さな範囲のデータ列挙を行うタスクを多数生成することにより、最終結果の再現率の向上を図るものである。

【研究テーマ B: 開発実行環境の構築研究】

本テーマにおいては、単に CyLog の処理系を構築するだけでなく、クラウドソーシングの研究インフラの構築に向けた研究をおこなった。特に、国内 20 大学の協力者を得て、大学間ネットワーク Crowd4U を構築した。これは、我々の知る限り、汎用クラウドソーシングエンジンを持つ世界初の大学間ネットワークである。Crowd4U に関しては各種デモンストレーション発表を行ったほか、次に説明するとおり、Crowd4U を活用したクラウドソーシングの応用研究を行った。

【研究テーマ C: プログラミング言語・開発実行環境の応用研究】

本テーマにおいては、開発したプログラミング言語および開発実行環境を単に利用するというだけでなく、応用領域専門家と協力して、現実の問題を指向したクラウドソーシング応用研究を推進した。主なものは次の 2 つである。

・クラウドソーシングを用いた書誌情報のクリーニング: 国立国会図書館の書誌情報を利用し、機械的書誌同定を行った結果のうち同定の誤りの可能性があるものを対象として、間違いであるものを見つける「書誌誤同定判定タスク」をクラウドソーシングする研究を行った。様々なデザインのマイクロタスクを用意し、数少ないタスク数で精度の高い結果をもたらすようなタスクデザインを発見した。

・クラウドソーシングによる竜巻経路推定実験: 災害時に重要かつ困難な事は、災害状況の把握である。本研究では、2012 年に起きたつくば市での竜巻被害を題材に、クラウドソーシングによる竜巻経路推定の可能性を探る研究を行った。基本的なアイデアは、近年 Web に竜巻などの写真が直ちにアップロードされることに着目し、それらの写真と地図をマッチングするタスクを設計し、それらの結果から経路推定を行うものである。

3. 今後の展開

CyLog/Crowd4U を活用してより深い理論研究、より高度な応用研究を進めていきたい。

4. 評価

(1) 自己評価

研究成果は、ほぼ当初の研究のねらいに沿った形で得られたと考えられる。研究の方向性は提案時からそれほど悩むことなく、確信を持って進めることができたが、当初計画に記載したストーリーが埋まるよう幅広く研究を進めたため、全体的に、深い研究成果が得られたとは必ずしも言えない。しかし、当初描いた構想のほぼ全体を一周できたことにより、この領域におけるさらに深い研究の推進に必要な知見の入手や問題点の発見等につながったと考えられる。開発実行環境を、クラウドソーシングの研究インフラとして利用可能な大学ネットワークとして設計し、本格的な構築の第一歩を踏み出したことは、今後のクラウドソーシング研究の発展への貢献につながる可能性がある。また、応用研究に関しては、さきがけ研究者のつながり等も利用し、研究のための応用でなく、現実の問題を持つ領域でのクラウドソーシング応用に取り組むことによって、実利用可能性の検証と知見を得て、システム設計・実装などにもフィードバックすることができた。今後は、これまで得られた成果を踏まえ、より深い理論研究とより高度

な応用研究を進めていきたい。

(2) 研究総括評価

本研究では、クラウドソーシングを対象とするプログラミング言語 CyLog とプラットフォーム Crowd4U を考案・開発している。プログラミング言語 CyLog は、従来の論理型言語に人が評価する述語を導入したものである。論理型言語の問題分割能力をクラウドソーシングの高次処理に用いることができるのに加え、ゲーム理論を用いて言語の表現力を理論的に主張していることに新規性がある。一方、クラウドソーシングプラットフォーム Crowd4U は非商用で、大学や国立国会図書館などの公的機関の問題解決に利用が広がっている。また、災害時の状況把握への応用など、公共性の高い活動に適用が試みられている。今後の非商用分野での展開に期待したい。さらに、研究成果が最近刊行された当該分野の英文ハンドブックに掲載されていることも高く評価できる。近年、急展開したクラウドソーシング研究に日本の研究者が先鞭をつけた意義は大きい。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Atsuyuki Morishima, Shiori Tomita, Takanori Kawashima, Takeshi Harada, Norihiko Uda, Sho Sato, Yukihiro Abematsu. A Crowdsourcing Approach for Finding Misidentifications of Bibliographic Records. In iConference 2014 Proceedings (p. 177–191), 2014.
2. Atsuyuki Morishima, Takanori Kawashima, Takashi Harada, Norihiko Uda, Ikki Ohmukai. L-Crowd: A Library Crowdsourcing Project by LIS and CS Researchers in Japan. In Proceedings of International Conference on Digital Libraries 2013 (ICDL2013), pp. 40–47, New Delhi, 27–29 November, 2013. (Invited Paper)
3. 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) , pp.20–25, Aug 2013.
4. Atsuyuki Morishima, Norihide Shinagawa, Tomomi Mitsuishi, Hideto Aoki, Shun Fukusumi, “CyLog/Crowd4U: A Declarative Platform for Complex Data–centric Crowdsourcing,” PVLDB 5(12): 1918–1921 (2012), Istanbul, Turkey, Aug 27–31, 2012.
5. Atsuyuki Morishima, Norihide Shinagawa, Shoji Mochizuki. The Power of Integrated Abstraction for Data–centric Human/Machine Computations. First International Workshop on Searching and Integrating New Web Data Sources (VLDS2011), pp. 5–9, September, 2011

(2) 特許出願

研究期間累積件数：0 件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- ・ 受賞：2012 年度日本データベース学会論文賞(福角駿, 森嶋厚行, 品川徳秀, 北川博之, 「DB 抽象化とゲーム理論に基づくマイクロブログからの構造データ抽出 GWAP の開発」, 日本データベース学会論文誌, Vol 11, No. 1, pp. 19–24, 2012 年 6 月.)

- 著書: Atsuyuki Morishima, “CyLog/Crowd4U: A Case Study of a Computing Platform for Cybernetic Dataspaces,” Handbook of Human Computation, Pietro Michelucci (eds.), Springer, Nov. 2013 (Invited Chapter)
- 招待論文・講演: Atsuyuki Morishima, Takanori Kawashima, Takashi Harada, Norihiko Uda, Ikki Ohmukai. L-Crowd: A Library Crowdsourcing Project by LIS and CS Researchers in Japan. In Proceedings of International Conference on Digital Libraries 2013 (ICDL2013), pp. 40-47, New Delhi, 27-29 November, 2013.
- 招待講演: 森嶋厚行「Crowd4U/L-Crowd: アカデミアによる高度クラウドソーシングプラットフォームと図書館情報分野への応用」, 情報処理学会第 112 回情報基礎とアクセス技術研究会/第 45 回デジタル図書館ワークショップ合同研究会, 2013 年 9 月 26 日.