

5 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—

森嶋 厚行 (筑波大学)

古くて新しいクラウドソーシング

クラウドソーシングとは、問題解決の手段として不特定多数の人々に仕事を委託することである。Jeff HoweによってWiredの記事「The Rise of Crowdsourcing」が書かれたのは2006年、クラウドソーシングマーケットとして有名なAmazon Mechanical Turkがサービスを開始したのは2005年である。クラウドソーシングという言葉が出現してから、たかだか10年程度しか経っていない。

たくさんの人々の力を集めて問題を解決しようという試み自体は昔から行われてきた。たとえば、1938年から米国で行われた「Mathematical Tables Project」では450人以上の「Computer (計算する人)」が集められ、数多くの数表を作ったことが知られている。当時、「Computer」は機械計算機ではなく計算する人を指す用語であり、機械計算機は「Automatic Computer」と呼ばれていた。現在、計算する人間をわざわざ「Human Computer」と呼ぶことがあるが、時代の流れを感じさせる。

当時との決定的な違いは2つある。第1に、インターネットが登場したことにより、多くの人に仕事を委託するコストが劇的に変わったことである。第2に、計算機の普及により、これらのコーディネーションの自動化が可能になったことである。これらの違いにより、それ以前には考えられなかった「出来ること」の可能性が広がり、現在の盛り上がりにつながっている。

クラウドソーシングの強みと分類

このように近年身近になったクラウドソーシングであるが、その強みのいくつかを次に挙げる。

(A) 計算機より人間の方が得意なことがある

あくまでも現時点の話であるが、計算機より人間の方が得意なことがまだまだたくさんある。パターン認識はその1つであり、人間の方がより良い結果をもたらすことも多い。Webページで皆さんが一度は見たことがあると思われるCAPTCHAはこの性質を利用しており、Webにアクセスしてくる主体が人間かどうかを判定するいわゆるチューリングテストの1つとして、わざとゆがませた単語を見せ、正しい綴りを入力した主体を人間と判定する。reCAPTCHAはこのバリエーションであり、画像中の単語認識をクラウドソーシングするシステムである。ここでのポイントは、CAPTCHAで表示される単語の数が1つであるのに対し、2つの単語の入力をしなければならない点である。1つは人間か否かの判定に使うための単語、もう1つは、ドキュメントをスキャンしたOCR (Optical Character Recognition: 光学文字認識) の結果が疑わしい単語の画像であり、そこで入力されたテキストを、OCR結果の精度向上に利用する。

(B) 多くの人がいれば、その中にはあなたの仕事ができる人がいる

多くの人にアクセスできれば、その中には、あなたの問題を解決できる人がいる可能性がある。たとえば、広く使われているWeb上のQ&Aサービスはクラウドソーシングの1つであるが、投稿される



図-1 クラウドソーシングによるさまざまなコンテンツ創造の例。(左) Amazon Mechanical Turk を用いて集められたさまざまな 10,000 の羊の絵^{☆1}、(右) DARPA による地上車両の設計案のクラウドソーシング^{☆2}

ありとあらゆる種類の質問に対して回答が寄せられている。

(C) 多くの人がいれば、多様なやり方でやってくれる

Kaggle^{☆3} は、予測モデリング課題に特化したクラウドソーシングサービスであり、利用者がデータと課題の説明を登録すると、回答者が最適モデルの構築を競い合う。予測モデリング課題には多様なアプローチが考えられるため、さまざまな参加者が得られるクラウドソーシングとの相性が良い。これまで、NASA や大手企業と提携し、さまざまな問題に対して解決が行われてきた。

(D) 多くの人がいれば、皆で一斉にできる

もし、仕事がたくさん小さな仕事に分割できる種類の物であれば、多くの人がいれば一斉に並列に処理することができる。たとえば、広い範囲の航空写真に写っている範囲から、特定の形状を持つ建物を探す問題は、航空写真を 1 万要素に細かく分割し、1 万人で一度に探せば早く発見することができる。

クラウドソーシングにおける仕事は、必ずしも「仕事をしてください」と明確に依頼されて行っているとは限らず、仕事をしているという認識が弱い場合もある。先に挙げた reCAPTCHA はそのようなクラウドソーシングの例である。ほかの例としては、GWAP (Game With A Purpose) というコンセプトがよく知られている。これは、ユーザから見たときにはコンピュータゲームであるが、実は彼らがそのゲームをすることによって、何らかの問題の解決につながるというものである。ESP ゲームは有名な GWAP であり、複数のゲームプレイヤが、画面

に表示された画像に対して相手がどのようなキーワードを入力するか互いに推測する。ESP ゲームが解決したい問題は、画像のメタデータ入手である。

クラウドソーシングは、仕事の粒度やインセンティブ構造等に注目して分類することもある。代表的な形態としては、ソフトウェア開発のような大きな案件を委託する人を選ぶマッチング型もしくはプロジェクト型、課題に対して複数の提案を受け付け、最も優れている提案を選択するコンペティション型、短時間で終了する仕事の処理を委託するマイクロタスク型などがある。

クラウドソーシングによる新しい情報コンテンツ創造

クラウドソーシングによる情報コンテンツの可能性については、Wikipedia を代表とする辞典的コンテンツをはじめ、さまざまな試みが行われている (図-1)。アートの分野では、オンライン、オフラインにこだわらずクラウドソーシングを用いたさまざまな試みが行われている。たとえば、The Sheep Market はクラウドソーシングの強み (C) (D) を生かした作品と言える。このパターンはアート系に多い。また、クラウドソーシングを用いて各種設計を行う試みも行われている。DARPA では FANG Challenge と称した取り組みの一環で、

☆1 <http://www.thesheepmarket.com>

☆2 <http://www.darpa.mil/uploadedImages/Content/NewsEvents/Releases/2013/reviseddrivetrain.jpg> (2015 年 5 月 17 日現在)

☆3 <https://www.kaggle.com/>

新しいソリューションスペース

新しい社会のデザインスペース



図-2 クラウドソーシングのインパクト—クラウドソーシング“大陸”の出現—

Fast, Adaptable, Next-Generation Ground Vehicle (FANG) と呼ばれる陸上車両の設計を公募し、2013年4月に3人組チームの提案を採択し100万ドルの賞金を与えた。これは、クラウドソーシングの強み (A) (B) (C) を特に生かした例であるとともに、コンペティション型のクラウドソーシングである。このパターンのクラウドソーシングでは、車や靴のデザイン、小説など、多種多様な取り組みがこれまで行われている。今後は、クラウドソーシングの強みをよりさまざまに組み合わせた情報コンテンツ創造の試みが増大すると予想される。うまくやれば、昔からあるリレー小説のように皆で少しずつ担当して作られた作品が、プロレベルの品質を持てるかもしれない。

クラウドソーシングのインパクト

クラウドソーシングのインパクトは、大きく分けると、新しいソリューションスペースの出現と、新しい社会のデザインスペースの出現に集約できる。特に、クラウドソーシングが人とかかわるコンセプトであることから、社会に対しても大きなインパクトがあることが1つの特徴である。クラウドソーシングにより新しく現れた、このような広大なスペースを持ついわば「クラウドソーシング大陸」をどのように開拓していくかが、この分野の重要な研究課題と言える (図-2)。

■ 新しいソリューションスペース

まず、クラウドソーシングは、情報処理のための新しい問題解決の手段を提供する。これまでは、ソフトウェアによる問題解決や、専任の人を雇用して人手で行うという2つが主要な手段であった。これらは、ソフトウェアがすでにある場合や、専任の人がすで

にいる場合には有効なアプローチと言えるが、ソフトウェア開発や人の雇用コストを含めたトータルコストで比較した場合には、クラウドソーシングを用いた方が、より早くより低コストで問題解決できる場合が数多くあることが認識されるようになってきた。

■ 新しい社会のデザインスペース

クラウドソーシングは人間がかかわるコンセプトのため、新しい社会のデザインに影響を及ぼす。地方在住の個人がオンラインの仕事で生計を立てるための仕組みの構築は、技術以外の要因もあるが、クラウドソーシング技術応用の重要な項目の1つであろう。ネット上でボランティアをする「マイクロボランティア」が容易にできる環境も整ってきた。また、クラウドソーシングで上手く仕事を発注する過程は、作業記述の明確化を行うことにほかならない。作業知識を形式知として扱う技術の発展は、流動性の高い組織や社会の実現につながる可能性がある。

✦ ビッグデータとクラウドソーシング

特にビッグデータの利活用に関しては、先に述べたクラウドソーシングの強みと関係して、次のような活用が考えられる。

■ ビッグデータ解析

Kaggle の例で挙げたように、クラウドの多様な観点を利用することは、データ解析に関して本質的である。ビッグデータの解析過程の一部にクラウドソーシングを導入することは自然なアプローチであ

ろう。

■ データ収集とクリーニング

既存のビッグデータを補完する手段として、人手でしか収集しにくいデータをクラウドソーシングで収集することはこれまでも行われてきた。また、データのクリーニングは、対象データ以外の知識や常識などが必要な場合も多く、その作業の一部をクラウドソーシングで行うことは品質向上に大きく寄与する。

■ 人工知能の教育

ビッグデータの処理や解析をすべて人手でやるのは現実的でないため、機械による人工知能でできることがあれば、それにやらせるべきである。人工知能の能動学習における教師としてクラウドを活用することにより人工知能の効率的な教育が可能と考えられる。

研究対象としてのクラウドソーシング

これまでのクラウドソーシング研究の流れは大きく分けて2つある。第1に、クラウドソーシングの可能性を追求する流れである。さまざまな領域・コミュニティで、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を用いて、機械だけではできなかったアプリケーションや問題解決ができたという事例が発表されている。たとえば、アルゴリズムだけでは難しい旅程の決定の問題をクラウドソーシングで行う研究や、Webサイト構築などのプロジェクトの必要要員をクラウドソーシングだけでリクルートし、プロジェクトを短時間・低コストで遂行できることを示した研究などがある。

第2に、クラウドソーシング結果の品質向上やアルゴリズムなど、効率良く品質の高い結果を得るための基礎研究がある。ここでは、ゲーム理論や機械学習などを応用し、タスク結果のみならず、ワーカ

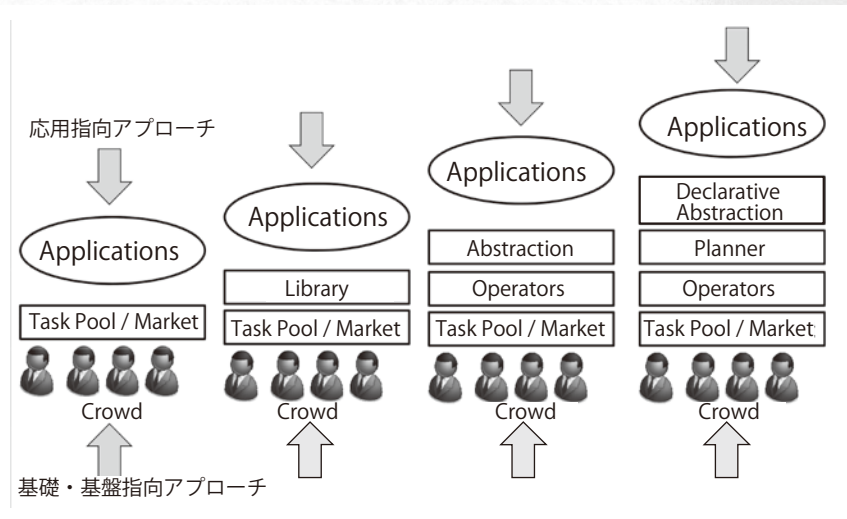


図-3 クラウドソーシングシステム・アーキテクチャの発展過程（左から右に進む）。まずは応用指向研究と基礎研究が行われ、ライブラリから始まって徐々にミドルレイヤの研究が重要になる。究極的にはデータベース管理システムのような宣言型の汎用クラウドソーシングプラットフォームの登場が期待される

やタスクそのものの評価や品質向上に関する研究が行われている。一般に、クラウドソーシングを行うシステムは、次の4つの問題に取り組む必要があり、基礎研究の多くはこれらに関連する。(1) 仕事をする人をどのようにリクルートするか、(2) 彼らに何をしてもらうか、(3) どのように入力を組み合わせるか、(4) 彼らをどのように評価するか。

現時点では、第1の流れと第2の流れに関しても、まだまだ取り組むべき問題が数多く残っており、当面この2つの観点からの研究は続く予想される。さらに、今後はこれらの流れの成果を統合し、基礎オペレータの上でアプリケーションを動かすためのミドルウェアや総合的なアーキテクチャの構築に向けて動いていくと予想される(図-3)。そのための鍵は、クラウドソーシング応用を記述するためのアブストラクションと基本演算子等の同定である。このような研究は、データベースコミュニティ等で一部行われているが適用範囲は限定されており、これからの発展が期待される分野である。

(2015年5月17日受付)

森嶋 厚行 (正会員) mori@slis.tsukuba.ac.jp
筑波大学知的コミュニティ基盤研究センター/図書館情報メディア系教授。多くの研究者とともに2011年よりCrowd4Uクラウドソーシングプラットフォームを運用し、さまざまなクラウドソーシングプロジェクトにかかわる。