

データ指向クラウドソーシングにおける網羅演算とその処理

青木 秀人

近年、群衆の知や力を利用することによってデータ収集・整理・処理等のデータ指向処理を行うクラウドソーシングが広く行われている。本論文では、このようなデータ指向処理のクラウドソーシングにおいて重要な演算である「網羅演算」を提案し、これを効率的に処理する仕組みについて議論する。

本論文で提案する網羅演算とは、「ある条件を満たす全ての項目を集めること」を指す。このような網羅演算は幅広い応用を持つ基礎的な演算である。例えば、Wikipedia などの見出し語や記述項目の列挙や、食べログなどの店舗リストやメニュー収集などが挙げられる。このようなアプリケーションで網羅演算を効率的に行う一つの手法は、網羅の範囲を複数の同値類に分割することである。例えば、つくば市のレストランの網羅を効率化する際には、レストランの網羅を行う範囲をいくつかの地域に細分化し、その地域ごとに網羅を行うといった手法が考えられる。

しかし、一般的に網羅の範囲の分割には、その網羅対象に関する固有の知識（ドメイン知識）が必要であるが、システム開発者が適切なドメイン知識を持っているとは限らない。したがって、本論文では、網羅対象の入力だけでなく、この網羅範囲の分割自体もクラウドソーシングによって行う網羅演算の処理手法も提案する。

本論文では、まず網羅演算を提案し、次にその処理手法の提案を行う。さらに手法の実装について説明し、最後に処理手法の評価を行うための予備的な実験について説明する。この予備実験においては、演算処理手法に関する仮説を立て、実験結果に基づく検証を行った。その結果、いくつかの仮説に関しては成立する可能性が示唆されたが、必ずしも十分な裏付けは得られていない。しかし、短い時間では網羅の範囲の分割をクラウドソーシングするためのコストの負の影響が大きいことがわかるなど、いくつかの興味深い洞察が得られた。

(指導教員 森嶋厚行)