

文脈を用いた Human-powered Join 処理方式

中村 優太

近年、計算機では処理が困難な問題の解決に人の力を利用するヒューマンコンピューテーションに注目が集まっている。ヒューマンコンピューテーションが有効な、計算機と比較して人間ならば容易に解決が可能な問題として、文脈を用いた結合処理が挙げられる。

文脈を用いた結合処理とは、ある組合せの結合に文脈の情報を必要とする処理のことを指す。例えば、人物の実体同定が挙げられる。仮に“田中 太郎”と“Taro Tanaka”という人物が同一人物かどうかを作業者に判別してもらうことを考える。人物の名前の情報のみでは、同姓同名の可能性があるため、同一人物かどうか判別ができない。そこで、両者が執筆した論文のタイトルや所属の同定を先に行ってもらい、もし執筆した論文や所属が同じ場合、同一人物と考えられる。つまり、執筆した論文や所属の情報を先に文脈として結合させることによって、同定が可能となっている。このように文脈を用いた結合処理は、周辺の情報の結合を行い、文脈を取り揃えることで初めて処理が可能となる。

こういった特徴を持つため、文脈を用いた結合処理では、必要な文脈が揃う前に結合を行おうとしても結合できず、処理が無駄になってしまう。よって、文脈を用いる結合処理問題の効率的な解決には処理を行う順序が重要となる。

そこで、文脈を用いた結合処理を効率的に行うために、タスク処理の順序を制御する方式を **PuzzleJoin** 方式として提案する。まず、結合に必要な文脈が揃っている組合せから、結合処理を行うための文脈を含むタスクを生成する。このタスクを人に行ってもらい、結合結果を得る。この結合結果は他の組合せの文脈として利用される。以上の処理を繰り返すことで、問題を解決する。結合に必要な文脈が揃っているかどうかを判別する関数と、組合せに対して文脈として提示される情報を返す関数を用いる。

提案手法の有用性を確認するために、シミュレーションを行った。航空写真データの組合せの結合処理に対し、ランダムな順番で行う単純な方式と **PuzzleJoin** 方式での 2 通りの方式でシミュレーションを行い、比較した。シミュレーション結果から、単純な方式では、この文脈が必要な結合の数だけ処理が無駄になっているのに対し、**PuzzleJoin** 方式では無駄となるタスクが発生しないことが確認できた。

(指導教員 森嶋厚行)