

## 論文の内容の要旨

論文題目	マイクロタスク型クラウドソーシングを用いた大規模データ処理における精度向上手法ならびにシステム開発と運用に関する研究
学位申請者	芦川 将之

### 第1章：序論

クラウドソーシングはCrowd(群衆)+Sourcing(調達)の造語であり、「企業、組織が、自社もしくはアウトソースの人材により実施していた業務を、よりオープンかつ不特定多数のCrowd(群衆)から人材を集め実施すること」と定義されている。このようなクラウドソーシング技術は、大規模データの解析や構築などを低コストで行うことが可能であり様々な分野や用途で利用されている。しかしその特性上、処理速度の速さや低コストの利点に対して処理結果の精度においては専門家による処理よりも劣るため問題視されており、様々な精度向上手法が研究されている。

本研究では適切な作業員(ワーカー)のマネジメントと育成を行うことで、安易な解雇を行うことなく高精度なデータ処理結果を高速かつ低コストで取得する手法の開発を目的とする。

### 第2章：関連研究

本章では、本研究で取り扱うクラウドソーシング関連の研究に関して述べる。まずは既存のクラウドソーシングに関する概要を説明する。そして、クラウドソーシング全体に関する研究に関して説明し、次にクラウドソーシングの精度向上手法に関する研究に関して説明し、さらにクラウドソーシングにおいてワーカーの学習や機械学習を用いた研究に関して説明する。

### 第3章：プライベートクラウドソーシングシステムの構築

本章では、本研究で開発を行った独自のマイクロタスク型クラウドソーシングに関して述べる。まずはクラウドソーシングにおけるタスクに関して説明し、システムの構築、及び既存のサービスとの比較に関して説明する。

### 第4章：ワーカーのフィルタリングによる精度向上手法の提案

PCSS の環境を構築しただけでは得られるタスクの処理結果の精度は低いため、

研究データとして使用するには十分ではない。本章ではPCSSにおけるワーカーフィルタリングによるタスク処理結果の精度向上手法に関して提案する。PCSSにおけるワーカーのフィルタリングは事前フィルタリング、動的フィルタリング、結果フィルタリング、推測フィルタリングという4つの独自のフィルタリングの組み合わせで行われており、それぞれの手法に関して説明する。

#### 第5章：ワーカーのフィルタリングによる精度向上手法の評価及び考察

本章ではPCSSを用いた語彙収集における4章で紹介したワーカーのフィルタリングによる精度向上効果に関して述べる。PCSS上で実装したフィルタリングを用いて語彙収集において精度が32.4ポイント上昇していることを確認した。

#### 第6章：ワーカーの段階的学習による精度向上手法の提案

本章ではマイクロタスク型クラウドソーシングにおける段階的学習手法に関して述べる。まずはマイクロタスク型クラウドソーシングにおけるワーカーの育成及び学習の必要性に関して述べ、そのための手法であるタスクグループのタスクカテゴリ分類と、タスクカテゴリ間の関係性の解析に関して提案する。

#### 第7章：ワーカーの段階的学習による精度向上手法の評価及び考察

本章では6章で述べた段階的学習手法の効果を確認するための実験とその結果から段階的学習手法の有効性に関して考察する。低品質な結果の多いタスクに対して学習タスクの算出を行ったところ、9種類のタスクに対して合計31種類の学習タスクを導出することが出来た。また、この導出された学習タスクを用いて低品質なワーカーに学習させ、改善効果を測定したところ平均7.8ポイントの改善効果が確認できた。比較対象として決定木でも学習タスクを導出したが、ベイジアンネットワークを用いて導出した学習タスクよりも効果が低いことが確認できた。

#### 第8章：ワーカーのフィルタリング及び段階的学習の事例紹介

本章ではPCSSを用いて知識処理研究に必要な語彙を収集した事例について述べる。125億文のWebテキストをクロールして収集し、得られたWebテキストをPCSSで処理する事によって14万語の語彙を獲得することが出来た。

#### 第9章：結論

本研究では研究データを安価、高速かつ高精度に作成するためのマイクロタスク型クラウドソーシングに関する精度向上手法の提案を行った。独自の精度向上手法を用いるためにプライベートなクラウドソーシングシステム(PCSS)を構築し、システム内に様々な精度向上手法を組み込むことで既存のマイクロタスク型クラウドソーシングサービスでは困難なシステム面からの精度向上を実現することが出来た。今後はワーカーのモチベーションコントロール、セキュリティ向上への検討が望まれる。

## 論文審査の結果の要旨

学位申請者氏名 芦川 将之

審査委員主査 大須賀 昭彦

委員 南 泰浩

委員 植野 真臣

委員 田原 康之

委員 川村 隆浩

近年の研究に用いられるデータはビッグデータと称される大量なデータであることが多い。従来の人手による作業では巨大データを扱うにはコスト、速度の面から難しくなっており、このような問題に対してクラウドソーシングを用いることによる解決が検討されている。しかし既存のクラウドソーシングでは精度の面で充分ではないという問題がある。本論文では作業者（ワーカー）のマネジメントと育成を行うことでクラウドソーシングの精度向上を行う手法の開発を目的としている。

第2章では、本論文で取り扱うクラウドソーシングに関して網羅的に説明されている。また、クラウドソーシングシステム、精度向上手法、ワーカーの学習や機械学習の適用に関する既存研究に関して説明されている。

第3章では、申請者が主導して開発し運用を行っている、本論文の提案手法を実装しているプライベートなクラウドソーシングシステム（PCSS）の構築手法と運用に関して説明されている。また既存のサービスに関する比較についても説明されている。以上により、PCSSは実用性の高い大規模なクラウドソーシングシステムであることが明らかとなっている。

第4章では、本研究で提案する手法の一つであるワーカーに対するフィルタリングによる精度向上手法について説明している。ワーカーの成長過程、行動履歴に応じた事前フィルタリング、動的フィルタリング、結果フィルタリング、推測フィルタリングという4つのフィルタリングを段階的に適用するという、独創的な手法を提案している。

第5章では、第4章で提案したワーカーのフィルタリング手法を第3章で説明したPCSS上で構築し、知識処理研究に必要な語彙収集を行う過程で得られた精度改善効果に関して説明している。実データを用いてフィルタリング手法適用前と適用後の比較検証が行われており、フィルタリング手法により有用性が明確に高まる程度の精度改善を実現していることも確認されている。

第6章では、ワーカーの段階的学習手法に関して説明している。本手法ではワーカーの育成及び学習の必要性に関して説明されており、クラウドソーシングシステムで実現するための手法として、タスクグループのタスクカテゴリ分類と、タスクカテゴリ間の関係性の解析を提案している。特に関係性の解析は、ベイジアンネットワークを適用した独創性の高い手法となっている。

第7章では、第6章で提案した段階的学習手法を第3章で説明したPCSS上で検証している。PCSSの運用から得られたワーカーの行動履歴に対してベイジアンネットワークを用いて解析することで学習タスクを導出している。また、得られた学習タスクを用いて実際にワーカーに学習させる評価実験をPCSS上でっており、学習効果があることが確認されている。さらに、比較実験として決定木を用いて学習タスクを導出し、同様にワーカーに学習させる比較実験を行っており、ベイジアンネットワークで得られた学習タスクよりも学習効果が低いことが確認されている。以上により提案手法は、クラウドソーシングにおけるワーカーのスキル向上に多大な効果があることが示されている。

第8章では、第4章、第6章で提案した精度改善手法を第3章で説明したPCSS上で実装し、知識処理研究に必要な語彙収集作業で実際に運用したケースに関して説明している。この実際の運用により知識処理研究に必要な精度を持った語彙が収集できたことが確認されており、提案手法がクラウドソーシングの実用性向上に大いに貢献できることが示されている。

第9章では、提案手法と評価結果をまとめ、研究データの解析にクラウドソーシングを用いるにあたって本研究の提案手法が精度改善の有効な手段となることが期待できることを示している。

以上で述べてきたように、本研究成果はオリジナリティに富み、実用面でも高い価値を持つものである。よって、本論文は博士(工学)の学位論文として十分な価値を有するものと認める。

以上