

論文 / 著書情報
Article / Book Information

題目(和文)	群衆の叡智からの統計的学習
Title(English)	Statistical Learning from the Wisdom of Crowds
著者(和文)	ZHANGHAO
Author(English)	Hao Zhang
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10239号, 授与年月日:2016年3月26日, 学位の種別:課程博士, 審査員:杉山 将,徳永 健伸,篠田 浩一,村田 剛志,石田 貴士
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10239号, Conferred date:2016/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	Hao Zhang		
論文審査 審査員		氏名	職名		氏名	職名
	主査	杉山 将	連携教授		石田 貴士	准教授
	審査員	徳永 健伸	教授	審査員		
		篠田 浩一	教授			
村田 剛志		准教授				

論文審査の要旨 (2000 字程度)

本論文は「Statistical Learning from the Wisdom of Crowds (群衆の叡智からの統計的学習)」と題し、英文4章から成っている。

第1章「Introduction」では、群衆の叡智に基づく統計的学習の枠組み、および、本論文の全体構成を示している。統計的学習の分野では、インターネットを通じて多数の非専門家（ワーカーとよぶ）から、データ（タスクとよぶ）に対するラベルを安価に収集するクラウドソーシングとよばれる枠組みが、近年注目を集めている。クラウドソーシングによって集められたデータからの統計的学習では、限られた予算のもとで、どのワーカーにどのタスクのラベル付けを依頼するかを決めるワーカー・タスク選択問題が重要な研究課題である。このワーカー・タスク選択問題に対して、これまでに様々な手法が提案され、その有効性が実証されてきた。一方、最近のクラウドソーシングでは、複数の分野のタスクが混在する複雑な場面が増加しつつある。従来ワーカー・タスク選択手法は単一分野のタスクを対象として設計されていたため、複数分野のタスクが存在する場合には性能が大幅に低下してしまうという問題があった。このような背景のもと、本論文の目的は、タスクが複数の分野からなる状況でも優れた性能を発揮する、新しいワーカー・タスク選択手法を提案することであると述べている。

第2章「Worker Selection in Crowdsourcing」では、与えられたタスクのラベル付けに相応しいワーカーを選択する Bandit-Based Task Assignment (BBTA, バンディットに基づくタスク割り当て) アルゴリズムを提案している。BBTA では、まず、全てのワーカーに少数の同じタスクを与え、各ワーカーの信頼度の初期推定を行う。そして、その後、各ワーカーの信頼度の更新とワーカーのタスクへの割り当てを逐次的に行っていく。BBTA は、クラウドソーシングにおけるワーカーの選択問題を、多腕バンディット（多数のスロットマシン）問題とよばれる数学的なオンライン意思決定問題として捉えるところに特徴がある。この数学的に厳密な定式化のもと、BBTA アルゴリズムによって得られるラベルと、全てのワーカーによる多数決との差を表すリグレットとよばれる指標の理論的上界を明らかにしている。最後に、計算機シミュレーションを通して、タスクが複数の分野からなる状況において、BBTA によるワーカー選択の精度が、従来法よりも優れていることを示している。また、タスクが単一分野に限定されている場合でも、最も良い従来法と同程度の性能が BBTA によって得られることを実験的に示している。

第3章「Task Selection in Crowdsourcing」では、BBTA アルゴリズムにおいて、どのような順番でタスクをワーカーに提示するのが良いかを論じている。クラウドソーシングにおけるタスクの選択は、標準的な教師付き学習における能動学習と同様な問題とみなすことができる。そこで、まず、能動学習で用いられている主要なタスク選択規準である信頼性規準、マージン規準、情報密度規準を概観している。そして、これらの規準を BBTA に適用し、計算機シミュレーションによって、信頼性規準と情報密度規準の性能が良いことを示している。更に、情報密度規準はパラメータの調整が容易でないことを実験的に示し、調整パラメータを含まない信頼性規準が実用的に最も有用であると結論づけている。

第4章「Conclusions and Future Work」では、本論文の成果を総括し、今後の課題を述べている。

以上を要するに本論文は、クラウドソーシングに基づく統計的学習技術の精度を向上させるものであり、工学上、および、工業上貢献するところが大きい。よって我々は、本論文が博士（工学）の学位論文として十分価値あるものと認める。