

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Discriminative Methods with Imperfect Supervision in Machine Learning
著者(和文)	NiuGang
Author(English)	Gang Niu
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9330号, 授与年月日:2013年9月25日, 学位の種類:課程博士, 審査員:杉山 将,佐藤 泰介,秋山 泰,篠田 浩一,瀬々 潤,工藤 一 浩
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9330号, Conferred date:2013/9/25, 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	Gang Niu	
論文審査 審査員		氏名	職名		
	主査	杉山将	准教授	瀬々潤	准教授
	審査員	佐藤泰介	教授		
		秋山泰	教授		
篠田浩一		教授			

論文審査の要旨 (2000 字程度)

本論文は「Discriminative Methods with Imperfect Supervision in Machine Learning」と題し、英文5章から成っている。

第1章「Introduction」では、研究の背景および本論文の全体構成を示している。機械学習の目的は、教師付きデータからその背後に潜む入出力規則を獲得することである。画像の認識やウェブページの分類問題では、画像やウェブページなどの入力データは自動的に収集できるが、画像に含まれる物体の種類やウェブページが属するカテゴリなどの教師情報は人手で付与する必要がある。そのため、大量の教師付きデータを準備することは困難であり、ビッグデータ時代にむけ新たな学習手法が必要とされている。このような背景のもと、本論文では不完全な教師情報に基づく学習手法を提案している。具体的には、教師データが全く与えられない教師なしのクラスタリング、入力データ対の類似性・非類似性に関する情報のみを用いる弱教師付きの計量学習、少数の教師付きデータと多数の教師なしデータを用いる半教師付き学習のパターン認識の三つの手法を提案している。

第2章「Maximum Volume Clustering」では、教師なし学習問題の一種であるクラスタリングについて論じている。これまでに様々なクラスタリング手法が開発されてきたが、従来のクラスタリング研究では、有限標本に対する安定性およびクラスタリング誤差評価という二つの重要な理論的性質が議論されてこなかった。そこで、半教師付き学習の理論研究で用いられている体積という概念を教師なし学習に拡張し、体積最大化原理に基づく新しいクラスタリングの枠組みを提案した。そして、マージン最大化クラスタリング、スペクトラル・クラスタリングなどの主要な従来法が特別な場合としてこの枠組に含まれることを示した。更に、体積最大化クラスタリングの具体的なアルゴリズムを構成し、このアルゴリズムが有限標本に対する安定性を有すること、すなわち、全ての局所的最適解が同じクラスタリング解を与えること、および、漸近的にクラスタリング誤差を最小にすることを理論的に証明した。最後に、提案法の有効性を数値実験により示した。

第3章「Information-theoretic Semi-supervised Metric Learning」では、類似性・非類似性ラベルを用いる弱教師付きの計量学習について論じている。未知の入力データに対する出力を正しく予測できる汎化能力は、学習したい入出力規則の滑らかさによりもたらされるが、入出力規則の滑らかさは入力空間の距離計量の選び方に依存する。この距離計量をデータに基づいて最適化することが計量学習の目的である。従来の計量学習法では、入力データ対に対して類似性・非類似性の情報が付与された弱教師付きデータと、ラベルが全く付与されていない大量の教師なしデータ間の類似性を用いていた。しかし、画像認識などの超多クラス分類問題では、ほとんどの入力データ対が非類似の関係をもつことから、従来法では教師なしデータのほんの僅かの情報しか活用できていなかった。そこで、エントロピーを用いた正則化を行うことにより教師なしデータ間の非類似性も活用できる計量学習法を提案し、その有効性を計算機実験により示した。

第4章「Squared-loss Mutual Information Regularization」では、半教師付きのパターン認識問題を論じている。相互情報量を用いた従来法は多クラス同時学習が可能であるなど優れた性質を持っているが、学習規準が非凸であるため最適化が困難であった。そこで、二乗損失相互情報量を用いた新しい手法を開発し、緩やかな条件のもとで最適化問題が凸になること、および、最適解が解析的に計算できることを示した。そして、提案手法の有効性を大規模な計算機実験により実証した。

第5章「Conclusions and Future Work」では、本論文で得られた成果を総括するとともに、残された課題や更なる研究の発展性について述べている。

以上を要するに本論文は、機械学習分野における不完全な教師情報に基づく新しい学習手法を提案するものであり、工学上、及び、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。