

論文 / 著書情報  
Article / Book Information

題目(和文)	
Title(English)	Learning under Class-Balance Change: Distribution Matching via Direct Divergence Estimation
著者(和文)	MARTHINUSC.DUPLES
Author(English)	Marthinus Christoffel du Plessis
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9558号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:杉山 将,佐藤 泰介,徳永 健伸,村田 剛志,瀬々 潤
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9558号, Conferred date:2014/3/26, 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

(博士課程)

## 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	Marthinus Christoffel du Plessis		
論文審査 審査員		氏名	職名		氏名	職名
	主査	杉山将	准教授	審査員	瀬々潤	准教授
	審査員	佐藤泰介	教授			
		徳永健伸	教授			
村田剛志		准教授				

### 論文審査の要旨 (2000 字程度)

本論文は「Learning under Class-Balance Change: Distribution Matching via Direct Divergence Estimation」と題し、英文7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。教師付き学習の目的は入出力が組になった訓練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り、通常の教師付き学習の理論では訓練データとテストデータが同じ確率分布に従うことが仮定されている。しかし、その仮定が実データでは成り立たないことがあるため、何らかの補正を行う必要がある。本論文の目的は、クラスのバランスが変化する状況に対応できる適応的分類学習法を提案することであると述べている。

第2章「Divergences」では、クラスバランスの推定において重要な働きをする確率分布間の距離尺度を紹介している。具体的には、まず  $f$  距離とよばれる確率密度関数の比に基づく一般的な距離尺度のクラスを導入し、カルバック・ライブラー距離、ピアソン距離などの代表的な距離尺度を示している。次に、確率密度関数の差を用いた距離のクラスを導入し、 $L_p$  距離をその代表例として紹介している。

第3章「Estimation of Divergences」では、第2章で紹介した距離尺度をデータから推定する方法を紹介している。具体的には、二乗誤差規準を用いたピアソン距離と  $L_2$  距離の推定法を導入するとともに、フェンシユル双対を用いた  $f$  距離や  $L_p$  距離の統一的な推定理論を紹介している。

第4章「Semi-Supervised Class-Prior Estimation」では、テストデータのクラスバランスの半教師付き推定法を提案している。提案法では、訓練データに含まれる各クラスのデータの確率分布を混合し、ラベルなしのテストデータに適合することによってテストデータのクラスバランスを推定する。この確率分布の適合を、カルバック・ライブラー距離やピアソン距離のもとで、それぞれの確率分布を推定することなく直接的に実行することが提案法の特徴であり、確率分布の推定を明示的に行う従来法と比べて高い推定精度が得られることを実験的に示している。また、 $L_2$  距離を用いた手法も同時に提案し、その有用性も実験的に示している。

第5章「Labeling Data Differing by Class Balance」では、ラベルなしのデータセットが二つ与えられる状況を考え、それらをもとに分類器を学習する「ラベル付け」という問題を論じている。これは一見不可能に思える問題であるが、二つのデータセットのクラスバランスが異なる場合、クラス事前確率が等確率の状況に対する分類器を学習することが可能であることを理論的に示している。この考え方は、確率密度の差の絶対値を推定することにより実現することができ、具体的なアルゴリズムを構成するとともに、その有効性を理論的および実験的に示している。

第6章「Prior Estimation in Positive-Only Labeled Data」では、正のクラスの標本と、(正と負のクラスの標本が混ざった) ラベルなしの標本が与えられる状況でのクラスバランス推定問題を論じている。第4章で提案しているピアソン距離に基づく確率分布適合の手法を拡張することにより、正のクラスの標本とラベルなしの標本だけからクラスバランスの推定が行える手法を提案し、その有効性を理論的および実験的に示している。

第7章「Conclusions and Future Work」では、本論文の成果を総括するとともに、今後の課題を述べている。

以上を要するに本論文は、機械学習分野においてクラスバランスが変化する状況下での分類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。