

論文の内容の要旨及び論文審査の結果の要旨の公表

学位規則第 8 条に基づき、論文の内容の要旨及び論文審査の結果の要旨を公表する。

○氏名	NGUYEN MANH HIEN (ぐえん まいん ひえん)
○学位の種類	博士 (工学)
○授与番号	甲 第 882 号
○授与年月日	2013 年 3 月 31 日
○学位授与の要件	本学学位規程第 18 条第 1 項 学位規則第 4 条第 1 項
○学位論文の題名	Learning Algorithms for Classification in Imbalanced Data Environments (不均衡データ環境における分類のための学習アルゴリズム)
○審査委員	(主査) 亀井 且有 (立命館大学情報理工学部教授) 西川 郁子 (立命館大学情報理工学部教授) ラック ターウォンマット (立命館大学情報理工学部教授) 小柳 滋 (立命館大学情報理工学部教授)

<論文の内容の要旨>

従来の機械学習は対象となるデータが均衡したクラス分布であることを仮定しているため、不均衡データに対しては十分満足できる結果を出すことができない。本学位論文では、静的環境および動的環境におけるクラス不均衡問題の幾つかの解法について述べている。

静的環境におけるクラス不均衡問題では、分類器の 1 つである SVM (Support Vector Machines) の決定境界をより効率よく絞り込むために補間と外挿の両手法を用いて作られた少数クラスデータを適切に配置する新しいオーバーサンプリング手法を提案し、分類研究分野では世界標準となっている UCI ベンチマークデータを用いた実験結果より、提案手法の有効性を示した。

動的環境におけるクラス不均衡問題では、以下の 3 点について述べている。

1. 不均衡データストリームから逐次学習する 2 つのオンラインサンプリング手法を提案し、UCI ベンチマークデータを用いた実験結果より、それらの有効性を示した。
2. サンプリング手法の比較実験にもとづきアンダーサンプリング手法がオーバーサンプリング手法より優れていることを実験的に示した。
3. 過去の少数クラスデータの再利用にもとづいてコンセプトドリフトとクラス不均衡の両方を有するデータストリームを効果的に処理する 2 手法を提案した。提案手法の有効性を検証するために人工的に発生させたコンセプトドリフトを有する UCI ベンチマークデータ

および実世界データを用いた実験結果より、提案手法が従来手法よりも優れていることを確認した。

<論文審査の結果の要旨>

本論文は、静的環境および動的環境におけるクラス不均衡問題の解法について述べたものである。本論文は以下の点について評価できる。

1. 静的環境において、SVM の決定境界をより効率よく絞り込むために補間と外挿の両手法を用いて作られた少数クラスデータを適切に配置する新しい手法 (BOS) を提案し、UCI ベンチマークデータを用いた実験結果より、提案手法の有効性を確認した。
2. 動的環境において、SVM とオーバーサンプリング手法の 1 つである SMOTE (Synthetic Minority Over-sampling Technique) を組み合わせた手法 (OSVM+Smt) を提案し、UCI ベンチマークデータを用いた実験結果より、提案手法の有効性を確認した。
3. 動的環境において、データストリームの多数クラスデータの一部から逐次学習するアンダーサンプリング手法と分類器の 1 つである Naïve Bayes を組み合わせた手法 (IOL) を提案し、UCI ベンチマークデータを用いた実験結果より、提案手法の有効性を確認した。
4. 動的環境において、サンプリング手法の比較実験より、アンダーサンプリング手法の 1 つである MRUS (Multiple Random Under-Sampling) 手法が他のアンダーサンプリング手法やオーバーサンプリング手法より優れていることを実験的に示した。
5. 動的環境において、AUC-Weighted Ensemble Model を用いて逐次学習をする手法 (INS) を提案し、人工的に発生させたコンセプトドリフトを有するデータストリームを用いた実験結果より、提案手法の有効性を確認した。
6. 動的環境において、過去の少数クラスデータを適応的に再利用する手法 (Adaptive Data Reuse) を提案し、UCI ベンチマークデータをもとに人工的に発生させたコンセプトドリフトを有するデータストリームおよび実世界データであるスラッシュドットを用いた実験結果より、提案手法の有効性を確認した。

本論文の審査に関して、2013 年 2 月 8 日 (金) 17 時 00 分~18 時 30 分、クリエーションコア 4 階知能情報学科会議室において公聴会を開催し、学位申請者による論文要旨の説明の後、審査委員は学位申請者 NGUYEN MANH HIEN に対する口頭試問を行った。各審査委員および公聴会参加者より、前提にしているクラス不均衡について、パラメータ決定について、コンセプトドリフトの妥当性について、などの質問がなされたが、いずれの質問に対しても学位申請者の回答は適切なものであった。よって、以上の論文審査と公聴会での口頭試問結果を踏まえ、本論文は博士の学位に値する論文であると判断した。

<試験または学力確認の結果の要旨>

本論文の主査は、学位申請者と本学大学院理工学研究科総合理工学専攻博士課程後期課程在学期間中に、研究指導を通じ、日常的に研究討論を行ってきた。また、本論文提出後、

主査および副査はそれぞれの立場から論文の内容について評価を行った。

学位申請者は、本学学位規程第 18 条第 1 項該当者であり、論文内容および公聴会での質疑応答を通して、学位申請者が十分な学識を有し、博士学位に相応しい学力を有していることを確認した。また、学位申請者は工学的な面においても学術的な面においても国際的に評価される研究を行っており、2009 年 9 月に開催された 5th International Workshop on Computational Intelligence and Applications において IEEE SMC Hiroshima Chapter' s Best Presentation Award を獲得するなど、量的ならびに質的に優れた研究業績により後期課程 2 年 6 ヶ月在学での修了が適当と判断した。

以上の諸点を総合し、学位申請者に対し、本学学位規程第 18 条第 1 項に基づいて、「博士（工学 立命館大学）」の学位を授与することが適当であると判断する。