

Title	勾配ブースティングの高度化と応用に関する研究
Sub Title	A study on the advancement and application of gradient boosting
Author	鈴木, 秀男(Suzuki, Hideo)
Publisher	慶應義塾大学
Publication year	2021
Jtitle	学事振興資金研究成果実績報告書 (2020.)
JaLC DOI	
Abstract	<p>転移学習の一分野に、ドメイン適応と呼ばれるものがある。ドメイン適応とは、予測対象である転移先のデータ (Target Domain) と予測対象でない転移元のデータ (Source Domain) の分布は異なるが、同一のタスクを想定する位置づけのことである。その中でも、サンプル数が少ないラベル付きTarget Domainとサンプル数が多いラベル付きSource Domainを用いて学習を行う教師ありドメイン適応では、TrAdaBoostやTrBaggなどのアンサンブル学習をベースとした手法が提案されている。</p> <p>本研究では、多重加法的回帰木 (Multiple Additive Regression Trees; MART) にニューラルネットのドロップアウトの概念を導入した手法であるDARTをドメイン適応に応用した、新たな手法TrDARTを提案した。TrDARTは既存手法と比べて2つの利点がある。1つ目は、アンサンブル学習の中でも高い予測性能を期待できる勾配ブースティング木を用いることによる予測性能の向上である。2つ目は、Source DomainとTarget Domainの学習を分離させることで、Target Domainに対する学習時間を大幅に短縮できる点である。複数の公開データセットを用いて、提案手法と既存手法を比較した結果、TrDARTは他の手法と比べて安定して高い予測精度を記録した。特に、回帰問題のデータセットに対して優れた予測精度を得ることができた。また、データサイズの大きいデータセットでのTarget Domainに対する学習時間を大幅に削減することができた。これらの結果により、予測性能の向上と学習時間の削減という2つの面での提案手法の有効性が示された。</p> <p>One of the methods in transfer learning is Domain adaptation. This method assumes that the same task, although the distribution of target domain data and source domain data are different. Particularly, as for supervised domain adaptation, which use a small sample size of labeled target domain and a large sample size of labeled source domain, there are several methods such as TrAdaBoost and TrBagg based on ensemble learning.</p> <p>In this study, TrDART is proposed, which is regarded as a methods of supervised domain adaptation using ensemble learning similarly existing methods. This method has two algorithms: applying Dropout to an existing tree learned using source domain and applying DART to new trees learned using target domain. TrDART has two advantages over existing methods. The first is to improve prediction performance by using a gradient boosting decision tree. The second is to reduce learning time for the target domain by separating the learning of source domain and target domain.</p> <p>TrDART obtained high prediction score stably through results of comparing the proposed method with existing methods using multiple open datasets. Particularly, excellent prediction accuracy was obtained for the regression problem datasets. In addition, we succeeded in significantly reducing the training time for target domain in large size datasets. These results show the two effectiveness of our proposed method: improvement of prediction performance and reduction of learning time.</p>
Notes	
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=2020000008-20200107

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

研究代表者	所属	理工学部	職名	教授	補助額	300 (A) 千円
	氏名	鈴木 秀男	氏名 (英語)	Hideo Suzuki		
研究課題 (日本語)						
勾配ブースティングの高度化と応用に関する研究						
研究課題 (英訳)						
A Study on the Advancement and Application of Gradient Boosting						
1. 研究成果実績の概要						
<p>転移学習の一分野に、ドメイン適応と呼ばれるものがある。ドメイン適応とは、予測対象である転移先のデータ(Target Domain)と予測対象でない転移元のデータ(Source Domain)の分布は異なるが、同一のタスクを想定する位置づけのことである。その中でも、サンプル数が少ないラベル付き Target Domain とサンプル数が多いラベル付き Source Domain を用いて学習を行う教師ありドメイン適応では、TrAdaBoost や TrBagg などのアンサンブル学習をベースとした手法が提案されている。</p> <p>本研究では、多重加法的回帰木(Multiple Additive Regression Trees; MART)にニューラルネットのドロップアウトの概念を導入した手法である DART をドメイン適応に応用した、新たな手法 TrDART を提案した。TrDART は既存手法と比べて2つの利点がある。1つ目は、アンサンブル学習の中でも高い予測性能を期待できる勾配ブースティング木を用いることによる予測性能の向上である。2つ目は、Source Domain と Target Domain の学習を分離させることで、Target Domain に対する学習時間を大幅に短縮できる点である。複数の公開データセットを用いて、提案手法と既存手法を比較した結果、TrDART は他の手法と比べて安定して高い予測精度を記録した。特に、回帰問題のデータセットに対して優れた予測精度を得ることができた。また、データサイズの大きいデータセットでの Target Domain に対する学習時間を大幅に削減することができた。これらの結果により、予測性能の向上と学習時間の削減という2つの面での提案手法の有効性が示された。</p>						
2. 研究成果実績の概要 (英訳)						
<p>One of the methods in transfer learning is Domain adaptation. This method assumes that the same task, although the distribution of target domain data and source domain data are different. Particularly, as for supervised domain adaptation, which use a small sample size of labeled target domain and a large sample size of labeled source domain, there are several methods such as TrAdaBoost and TrBagg based on ensemble learning.</p> <p>In this study, TrDART is proposed, which is regarded as a methods of supervised domain adaptation using ensemble learning similarly existing methods. This method has two algorithms: applying Dropout to an existing tree learned using source domain and applying DART to new trees learned using target domain. TrDART has two advantages over existing methods. The first is to improve prediction performance by using a gradient boosting decision tree. The second is to reduce learning time for the target domain by separating the learning of source domain and target domain.</p> <p>TrDART obtained high prediction score stably through results of comparing the proposed method with existing methods using multiple open datasets. Particularly, excellent prediction accuracy was obtained for the regression problem datasets. In addition, we succeeded in significantly reducing the training time for target domain in large size datasets. These results show the two effectiveness of our proposed method: improvement of prediction performance and reduction of learning time.</p>						
3. 本研究課題に関する発表						
発表者氏名 (著者・講演者)	発表課題名 (著書名・演題)	発表学術誌名 (著書発行所・講演学会)	学術誌発行年月 (著書発行年月・講演年月)			
柴田頼仁, 岩本大輝, 多賀友哉, 小森晴菜, 鈴木秀男	異常分布を考慮した タクシーの需要予測	日本経営工学会 2020 年秋季大会	2020 年 10 月 25 日			
柴田 頼仁, 鈴木 秀男	DART を用いたドメイン適応手法 TrDART の紹介	第 23 回情報論的学習理論ワークショップ	2020 年 11 月 23 日~26 日			
柴田 頼仁, 岩本 大輝, 小森 晴菜, 多賀 友哉, 鈴木 秀男	複数ドメインからの転移を想定した回帰手法とタクシーの 需要予測: Multi Source Two-stage TrAdaBoost.R2 の提案	オペレーションズ・リサーチ, Vol.66, No.3, 2021	2021 年 2 月 1 日			