

Department of Complexity Science and Engineering  
Graduate School of Frontier Sciences  
The University of Tokyo  
東京大学大学院新領域創成科学研究科  
複雑理工学専攻

2019  
平成30年度  
Master's Thesis  
修士論文

**Unsupervised Domain Adaptation Based  
on Source-guided Discrepancy**  
(転移元データのラベルを活用した分布間距離に基づく教師無しドメイン適応)

Supervisor: Lecturer Issei Sato  
指導教員：佐藤 一誠 講師

2019年3月8日提出  
Submitted March 8, 2019  
キーワード：統計的機械学習，ドメイン適応  
Keywords: Statistical machine learning, Domain adaptation

**Seiichi Kuroki**  
黒木 誠一  
47-176104

# Abstract

*Unsupervised domain adaptation* is the problem setting where data generating distributions in the source and target domains are different and labels in the target domain are unavailable. An important question in unsupervised domain adaptation is how to measure the difference or *discrepancy*, between the source and target domains. Existing discrepancy measures for unsupervised domain adaptation either require high computation costs or have no theoretical guarantee. To mitigate these problems, this thesis proposes a novel discrepancy measure called *source-guided discrepancy (S-disc)*, which exploits labels in the source domain unlike the existing ones. As a consequence, S-disc can be computed efficiently with a finite-sample convergence guarantee. In addition, it is shown that S-disc can provide a tighter generalization error bound than the one based on an existing discrepancy measure. Finally, experimental results demonstrate the advantages of S-disc over the existing discrepancy measures.