## Semi-supervised learning with density-ratio estimation

Masanori Kawakita • Takafumi Kanamori

Received: 17 April 2012 / Accepted: 14 January 2013 / Published online: 9 February 2013 © The Author(s) 2013

Abstract In this paper we study statistical properties of semi-supervised learning, which is considered to be an important problem in the field of machine learning. In standard supervised learning only labeled data is observed, and classification and regression problems are formalized as supervised learning. On the other hand, in semi-supervised learning, unlabeled data is also obtained in addition to labeled data. Hence, the ability to exploit unlabeled data is important to improve prediction accuracy in semi-supervised learning. This problem is regarded as a semiparametric estimation problem with missing data. Under discriminative probabilistic models, it was considered that unlabeled data is useless to improve the estimation accuracy. Recently, the weighted estimator using unlabeled data achieves a better prediction accuracy compared to the learning method using only labeled data, especially when the discriminative probabilistic model is misspecified. That is, improvement under the semiparametric model with missing data is possible when the semiparametric model is misspecified. In this paper, we apply the density-ratio estimator to obtain the weight function in semi-supervised learning. Our approach is advantageous because the proposed estimator does not require well-specified probabilistic models for the probability of the unlabeled data. Based on statistical asymptotic theory, we prove that the estimation accuracy of our method outperforms supervised learning using only labeled data. Some numerical experiments present the usefulness of our methods.

Keywords Semi-supervised learning · Density ratio · Semiparametric model · Missing data

Editor: Xiaojin Zhu.

M. Kawakita

T. Kanamori (🖂)

Department of Computer Science and Mathematical Informatics, Nagoya University, Furocho, Chikusaku, Nagoya 464-8603, Japan e-mail: kanamori@is.nagoya-u.ac.jp

Department of Informatics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan e-mail: kawakita@inf.kyushu-u.ac.jp