論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Density ratio approaches to multi-label classification and outlier detection
著者(和文)	Hyunha Nam
Author(English)	Hyunha Nam
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10011号, 授与年月日:2015年9月25日, 学位の種別:課程博士, 審査員:杉山　将,秋山　泰,篠田　浩一,村田　剛志,藤井　敦
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10011号, Conferred date:2015/9/25, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	Computer Science 専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of (Engineering)
学生氏名： Student's Name	Hyunha Nam	指導教員（主）： Academic Advisor(main)	Sugiyama Masashi
		指導教員（副）： Academic Advisor(sub)	Koichi Shinoda

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

The ratio of two probability density functions appears in various machine learning tasks such as conditional density estimation, outlier detection, non-stationarity adaptation, dimensionality reduction, independent component analysis, clustering and classification. Therefore the problem of estimating the density ratio is attracting a great deal of attention these days. Recently, direct density ratio estimation methods have been proposed and shown to be effective for many machine learning problems. The key idea of the direct density ratio approach is that the ratio is directly estimated so that difficult density estimation is avoided. There are still two major open topics to improve the density ratio approach in terms of accuracy and computational efficiency. This thesis focuses on them.

Firstly, we investigate a computationally efficient solution for more general problem setting. Multi-label classification allows a sample to belong to multiple classes simultaneously, which is often the case in real-world applications such as text categorization and image annotation. In multi-label scenarios, taking into account correlations among multiple labels can boost the classification accuracy. However, this makes classifier training more challenging because handling multiple labels induces a high-dimensional optimization problem. We propose a scalable multi-label method based on least-square approach to density ratio estimation. Through experiments, we show the usefulness of our proposed method.

Secondly, we investigate a deep model for density ratio estimation. So far, parametric and non-parametric direct density ratio estimators with various loss functions have been developed, and the kernel least-squares method was demonstrated to be highly useful both in terms of accuracy and computational efficiency. On the other hand, recent study in pattern recognition exhibited that deep architectures such as a convolutional neural network can significantly outperform kernel methods. We propose to use the convolutional neural network in density ratio estimation, and experimentally show that the proposed method tends to outperform the kernel-based method in outlier detection tasks in images.

Given the encouraging experimental results of the proposed methods, we conclude that the proposed density ratio approaches with a deep model and multi-label problem setting are successful and worth a further study in the future.

The thesis consists of 6 chapters. Chapter 1 introduce overview of thesis. In Chapter 2, we introduce multi-label classification. Section 2.1 gives a brief introduction of the single-label

classification and the multi-class classification. Section 2.2 reviews the multi-task classification and its applications. Section 2.3 shows an overview of multi-label classification and its application. Various transformation methods of the multi-label classification problem are introduced.

Chapter 3 covers our work on density ratio estimation for multi-label classification. Section 3.1 reviews the least-squared probabilistic classification (LSPC) for a single-label dataset. Section 3.2 shows the LSPC for multi-task learning. In Section 3.3, we extend multi-task LSPC to multi-label setup in a computationally efficient manner. We give the experiment results comparing with existing methods.

Subsequently, Chapters 4 and 5 are devoted to deep models for density ratio estimation. In Chapter 4, we present a deep learning method. Section 4.1 introduces deep learning. We review deep learning algorithms according to several criteria in Section 4.2. Especially, Section 4.2.1 and Section 4.2 explain deep autoencoders and convolutional neural networks (CNN).

In Chapter 5, we present density ratio estimation using deep models and demonstrate the performance of our proposed method in inlier-based outlier detection. In Section 5.1, we explain related studies and an overview of the proposed method. In Section 5.2, we introduce least-squares importance fitting (LSIF) that uses the squared-loss to fit a density ratio model to data. It is using a linear model. Section 5.3 derives how to apply the uLSIF criterion to the multilayer perceptron (MLP). We propose the uLSIF based on a deep CNN model in Section 5.4. Section 5.4.1 introduces the CNN model and Section 5.4.2 formulates the problem of training our method. In Section 5.5, we describe the inlier-based outlier detection and experimental setup. Then the proposed CNN-based uLSIF is compared with the kernel-based uLSIF and the kernel-based KLIEP in experiments of inlier-based outlier detection.

Finally, we summarize this thesis and show future works in Chapter 6.