

两种半监督多类水下目标识别算法的比较

杜方键, 杨宏晖

(西北工业大学航海学院, 陕西西安 710072)

摘要: 基于半监督学习理论的水下目标识别系统能够从未知类别测试集中识别出已学习类别测试样本, 并拒判未学习类别测试样本。描述并讨论了两种基于支持向量数据描述(Support Vector Data Description, SVDD)的半监督水下目标识别算法: 半监督 SVDD 与半监督 SVDD 集成。利用四类实测水下目标样本进行实验, 训练样本为三类已知类别样本, 测试样本为四类样本, 包含一类未学习类别样本。对两种算法实验结果进行比较, 表明半监督 SVDD 集成算法比半监督 SVDD 算法能更好地识别已学习类别测试样本, 并能有效拒判未学习类别测试样本, 不足之处为时间消耗与过程复杂程度比半监督 SVDD 算法高。

关键词: 半监督; 水下目标识别; 类别测试样本; 支持向量数据描述; 分类器集成

中图分类号: TB556

文献标识码: A

文章编号: 1000-3630(2014)-01-0010-04

DOI 编码: 10.3969/j.issn1000-3630.2014.01.003

Comparison of two semi-supervised multiclass underwater target recognition algorithm

DU Fang-jian, YANG Hong-hui

(School of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China)

Abstract: The underwater target recognition system based on the theory of supervised learning can recognize class learned underwater targets' samples and reject class unlearned underwater targets' samples from class unknown testing set. This paper describes and discusses two semi-supervised underwater target recognition algorithms based on Support Vector Data Description(SVDD): Semi-supervised SVDD algorithm(SS-SVDD) and Semi-supervised SVDD ensemble algorithm(SS-SVDDE). Four different classes of underwater targets' samples are used in the experiment, the training samples are 3-classes samples, the testing samples are 4-classes samples which contain one class of class unlearned samples. By comparing the two algorithms' experimental results, it is found that SS-SVDDE algorithm can recognize class learned testing samples better than SS-SVDD algorithm and effectively reject class unlearned testing samples, but SS-SVDDE algorithm's time consumption and process complexity are higher than SS-SVDD algorithm.

Key words: semi-supervised; underwater target recognition; class test samples; Support Vector Data Description (SVDD); classifier ensemble

0 引言

基于有监督学习理论的水下目标识别系统^[1,2]利用已知类别样本训练识别系统, 该系统在测试时, 会将没有学习过的类别的样本归于学习过的某类类别, 不能实现拒判^[3]。水下目标类别的多样性以及类别数目的不断增加使得有监督水下目标识别的系统不能有效地识别水下目标, 因此构建可以实现拒判的半监督^[4]多类水下目标识别系统具有实际意义。半监督学习理论的水下目标识别系统能够从未知类别测试集中识别出已学习类别测试样本, 并

拒判未学习类别测试样本。

支持向量数据描述^[5-8](Support Vector Data Description, SVDD)是一种典型的单类分类器, 通过建立包围目标类的超球来拒绝非目标类数据, 具有很好的数据描述能力, 能够较好地提供每类数据分布的紧凑描述, 对高维数据表现出良好的性能。训练时只需要一类样本, 有比较快的训练速度, 可以用于解决多类的分类问题, 通过构造多个数据描述来达到多类分类的目的。因此利用 SVDD 分类器构建半监督多类水下目标识别系统成为水下目标识别研究的一个重要方向。

本文将描述并讨论两种基于 SVDD 的半监督水下目标识别算法: “半监督 SVDD 算法”(Semi-supervised SVDD algorithm, SS-SVDD)与“半监督 SVDD 集成算法”(Semi-supervised SVDD ensemble

收稿日期: 2012-09-02; 修回日期: 2012-12-26

作者简介: 杜方键(1987-), 男, 河南固始人, 硕士研究生, 研究方向为模式识别与声信号信息处理。

通讯作者: 杨宏晖, E-mail: hhyang@nwpu.edu.cn

algorithm, SS-SVDDE)。利用两种算法构建目标识别系统，对两种算法识别性能进行比较。利用四类实测水下目标样本进行实验，训练样本为三类已知类别样本，测试样本为四类样本，包含一类未学习类别样本。实验结果表明 SS-SVDDE 算法比 SS-SVDD 算法能更好地识别已学习类别测试样本，并能有效拒判未学习类别测试样本，但时间消耗与过程复杂程度比 SS-SVDD 算法高。

1 SS-SVDD 算法

SS-SVDD 算法的基本原理：对水下目标中已知类别的所有训练样本集利用 SVDD 算法训练一个分类性能高的单类分类器，该单类分类器能有效地识别已学习类别的测试样本并拒判未学习类别的测试样本。用该单类分类器测试未知类别测试样本，如果样本被该单类分类器拒绝，则判定未知类别测试样本不属于已学习目标；如果样本被该单类分类器接受，则判定未知类别测试样本属于已学习目标之一，利用已学习目标的训练样本集训练多类分类器，测试未知类别测试样本，可以最终判定测试样本的类别归属。

具体算法流程如图 1 所示。

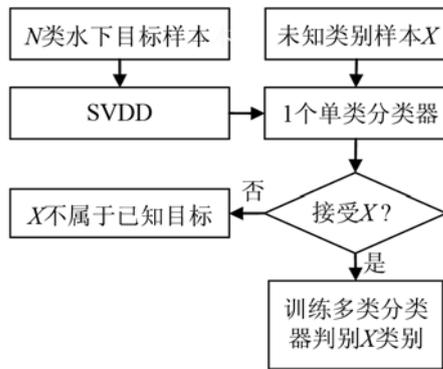


图 1 SS-SVDD 算法流程
Fig.1 Flow chart of SS-SVDD algorithm

算法步骤如下：

(1) 训练阶段

用 N 类类别已知训练样本集全体作为训练集训练 1 个 SVDD 单类分类器。该单类分类器能够识别已学习类别的样本，拒判未学习类别的样本。

(2) 测试阶段

用训练得到的单类分类器测试未知类别样本 X 。根据测试出现的两种结果来判别 X 的类别：

① 若测试结果显示 X 被单类分类器拒绝，则判别 X 不属于已学习目标，实现水下目标样本的拒判；

② 若测试结果显示 X 被单类分类器接受，则判定未知类别测试样本属于已学习目标之一，可以利用已知的 N 类水下目标训练样本集训练多类分类器判别 X 的类别。

2 SS-SVDDE 算法

SS-SVDDE 算法的基本原理：对水下目标中已知类别的每类训练样本集利用 SVDD 算法训练分类性能高的单类分类器，每个单类分类器能有效识别目标类样本和拒绝非目标类样本。用各单类分类器构成分类器集成系统，测试未知类别测试样本，如果样本被全部单类分类器拒绝，则判定未知类别测试样本不属于已学习目标；如果样本只被一个单类分类器接受，则判定未知类别测试样本属于该类已学习目标；如果样本被多个单类分类器接受，则判定未知类别测试样本属于识别正确率最大的一类已学习目标。

具体算法流程如图 2 所示。

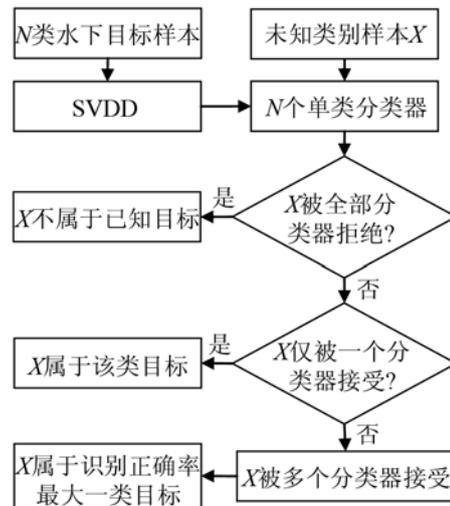


图 2 SS-SVDDE 算法流程
Fig.2 Flow chart of SS-SVDDE algorithm

算法步骤如下：

(1) 训练阶段

对 N 类类别已知训练样本集分别训练 SVDD 单类分类器，共获得 N 个单类分类器。每个单类分类器能够有效识别目标类样本和拒绝非目标类样本。

(2) 测试阶段

用训练得到的 N 个单类分类器构成分类器集成系统，测试未知类别样本 X ，共得到 N 个测试结果，综合 N 个测试结果，给出集成系统的最终判别结果。分以下三种情况给出最后的判别结果：

① 若测试结果显示 X 被全部单类分类器拒绝，

则判别 X 不属于已学习目标, 实现水下目标的拒判;

② 若测试结果显示 X 只被一个单类分类器接受, 则判别 X 属于该类已学习目标;

③ 若测试结果显示 X 被多个单类分类器接受, 则将 X 判为识别正确率最大的一类已学习目标。

3 实验结果与分析

3.1 实验数据

本实验所用数据集为某实测水声数据集: 共分为四类目标, 1920 个样本, 每个样本由 71 个特征描述, 每类样本个数为 480。依照采集的先后顺序将四类实测水下目标标记为 A、B、C、D。

3.2 SS-SVDD 与 SS-SVDDE 算法实验

将四类实测水下目标样本集 A、B、C、D 取出三类作为已知类别样本集, 余下一类作为未学习类别样本集。见表 1 所示。

实验过程如下:

第一步:

假设已知类别为 A、B、C, 剩余 D 类为未学习类别。将已知类别样本集每类随机分成两部分: 训练集(A_{tr} , B_{tr} , C_{tr})和测试集(A_{ts} , B_{ts} , C_{ts}), 测试集为训练集大小的 1/3。将 D 类样本全体作为测试集。

第二步:

(1) SS-SVDD 算法: 用三类已知类别训练样本集(A_{tr} , B_{tr} , C_{tr})全部样本训练一个 SVDD 单类分类器, 用该单类分类器测试未知类别测试样本(A_{ts} , B_{ts} , C_{ts} , D)。

(2) SS-SVDDE 算法: 用三类已知类别训练样本集(A_{tr} , B_{tr} , C_{tr})分别训练三个 SVDD 单类分类器, 构成半监督多类水下目标识别系统, 用识别系统测试未知类别测试样本(A_{ts} , B_{ts} , C_{ts} , D)。

第三步:

(1) SS-SVDD 算法: 用单类分类器测试未知类别测试样本(A_{ts} , B_{ts} , C_{ts} , D), 获得识别正确率。

① 若测试结果显示未知类别测试样本被单类分类器拒绝, 则未知类别测试样本不属于已学习三类目标中的任何一类, 实现拒判;

② 若测试结果显示未知类别测试样本被单类分类器接受, 则利用已学习训练样本集(A_{tr} , B_{tr} , C_{tr})训练多类分类器测试未知类别测试样本, 判定其类别。

(2) SS-SVDDE 算法: 用训练得到的三个单类分类器测试未知类别测试样本(A_{ts} , B_{ts} , C_{ts} , D),

获得识别正确率。

① 若测试结果显示未知类别测试样本被全部单类分类器拒绝, 则未知类别测试样本不属于已学习三类目标中的任何一类, 实现拒判;

② 若测试结果显示未知类别测试样本只被一个单类分类器接受, 则未知类别测试样本属于该类;

③ 若测试结果显示未知类别测试样本被多个单类分类器接受, 则未知类别测试样本属于识别正确率最大的一类。

表 1 为 SS-SVDD 与 SS-SVDDE 算法实验结果, 表 1 中的测试集识别正确率分为两部分:

(1) 训练集与测试集同类: 测试集识别正确率为分类器正确识别目标类样本的识别正确率;

(2) 训练集与测试集不同类: 测试集识别正确率为分类器正确拒绝非目标类样本的正确拒绝率。

对表 1 中 SS-SVDD 算法实验结果进行分析:

第一组, 单类分类器不能准确识别已学习类别测试样本, 主要原因是: 为了保证 SVDD 单类分类器的推广能力^[9], 允许部分目标样本点落在超球外, 而 C 类训练样本由于远离主样本群 A、B, 造成对 C 类训练样本的错误拒绝率高, 导致对 C 类测试样本的识别正确率低。

第二组, 与第一组同样的原因导致对 D 类测试样本的识别正确率低。

第三组, 单类分类器不能准确拒判未学习类别测试样本, 且对已学习类别测试样本的识别正确率不高。主要原因是: A、C、D 类样本群间距离远, 分类器分界面包括大片空白区域, 使得对 B 类测试样本的识别正确率低; C、D 类训练样本由于较多的落在超球外, 导致对 C、D 类测试样本的识别正确率低。

第四组, 与第三组同样的原因导致单类分类器不能准确拒判未学习类别测试样本, 且对已学习类别测试样本的识别正确率不高。

对表 1 中 SS-SVDDE 算法实验结果进行分析:

从实验结果中可以看出, 每类单类分类器能准确地识别目标类测试样本, 拒绝非目标类测试样本。以第一组为例, D 类测试样本被全部单类分类器拒绝; 根据判别规则, 由三个单类分类器构成的分类器集成系统实现对 A、B、C 三类已学习类别测试样本的准确识别, 并实现对 D 类未学习类别测试样本的拒判。

从四组实验结果可以看出, SS-SVDDE 算法能有效地识别已学习类别测试样本并拒判未学习类别测试样本, 与 SS-SVDD 算法相比, 实验结果更加准确可信。然而由于 SS-SVDDE 算法需要对每

表 1 SS-SVDD 与 SS-SVDDE 算法实验结果
Table 1 The experimental results of SS-SVDD algorithm and SS-SVDDE algorithm

	训练集	(SS-SVDD)测试集识别正确率%				(SS-SVDDE)测试集识别正确率%			
		A _{ts}	B _{ts}	C _{ts}	D	A _{ts}	B _{ts}	C _{ts}	D
第一组	A _{tr}	90.00	83.33	24.17	100	80.00	93.33	93.33	100
	B _{tr}					92.50	79.17	90.00	100
	C _{tr}					100	100	72.50	100
	D _{tr}								
第二组	训练集	测试集识别正确率%				测试集识别正确率%			
	A _{tr}	A _{ts}	B _{ts}	D _{ts}	C	A _{ts}	B _{ts}	D _{ts}	C
	B _{tr}	90.00	91.67	24.17	76.88	80.00	93.33	100	93.13
	D _{tr}					92.50	79.17	100	90.42
第三组	训练集	测试集识别正确率%				测试集识别正确率%			
	A _{tr}	A _{ts}	C _{ts}	D _{ts}	B	A _{ts}	C _{ts}	D _{ts}	B
	C _{tr}	99.17	65.00	58.33	20.00	80.00	93.33	100	95.00
	D _{tr}					100	72.50	100	100
第四组	训练集	测试集识别正确率%				测试集识别正确率%			
	B _{tr}	B _{ts}	C _{ts}	D _{ts}	A	B _{ts}	C _{ts}	D _{ts}	A
	C _{tr}	84.17	29.17	61.67	31.67	79.17	90.00	100	93.75
	D _{tr}					100	72.50	100	99.79

类已知类别样本训练单类分类器，与 SS-SVDD 算法相比，时间消耗大，过程复杂。对于已知类别较多，数据样本较多，样本群间距离较小的识别问题，SS-SVDD 算法时效性更优；对于已知类别较少，数据样本较少的识别问题，SS-SVDDE 算法性能更优。需要根据不同的实际问题选择合适的算法。

4 总结

本文描述了两种基于 SVDD 的半监督水下目标识别算法：SS-SVDD 算法与 SS-SVDDE 算法。对两种算法识别性能进行比较，结果表明 SS-SVDDE 算法比 SS-SVDD 算法能更好地识别已学习类别测试样本，拒判未学习类别测试样本，对于已知类别较少，数据样本较少的识别问题，SS-SVDDE 算法性能更优。SS-SVDD 算法时间消耗及过程复杂程度低，对于已知类别较多，数据样本较多，样本群间距离较小的识别问题，SS-SVDD 算法时效性更优。

参 考 文 献

[1] 杨宏晖, 孙进才. 基于支持向量机集成的水下目标自动识别系统[J]. 测控计算, 2006, 25(12): 14-16.
YANG Honghui, SUN Jincal. Automatic classification of multi-class underwater targets by SVM ensembles[J]. Measurement and control technology, 2006, 25(12): 14-16.

[2] 杨宏晖, 孙进才, 袁骏. 基于支持向量机和遗传算法的水下目标特征选择算法[J]. 西北工业大学学报, 2005, 23(4): 512-515.
YANG Honghui, SUN Jincal, YUAN Jun. A new method for feature selection for underwater acoustic targets[J]. Journal of Northwestern Polytechnical University, 2005, 23(4): 512-515.

[3] 徐引玲. KSVDD 及其在拒识判别中的应用[J]. 计算机工程, 2010, 36(19): 195-197.
XU Yinling. KSVDD and its application in rejection determination [J]. Computer Engineering, 2010, 36(19): 195-197.

[4] 赵莹. 半监督支持向量机学习算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2010: 23-29.
ZHAO Ying. Research on semi-supervised support vector machine learning algorithms[D]. Harbin: Harbin Engineering University, 2010: 23-29.

[5] David M J Tax. Support vector data description[J]. Machine Learning, 2004, 54: 45-66.

[6] Sakla W, Chan A, Ji J, et al. An SVDD-based algorithm for target detection in hyperspectral imagery[J]. Geoscience and Remote Sensing Letters, 2011, 8(2): 384-388.

[7] 朱孝开. 基于核方法的图像目标识别技术研究[D]. 长沙: 国防科学技术大学研究生院, 2009: 85-102.
ZHU Xiaokai. Kernel methods for image object recognition[D]. Changsha: Graduate School of National University of Defense Technology, 2009: 85-102.

[8] 刘晟, 朱玉全, 孙金津. 基于核空间相对密度的 SVDD 多类分类算法[J]. 计算机应用研究, 2010, 27(5): 1694-1696.
LIU Sheng, ZHU Yuquan, SUN Jinjin. SVDD multiclass classification algorithm based on relative density in kernel space [J]. Application Research of Computer, 2010, 27(5): 1694-1696.

[9] 朱孝开, 杨德贵. 基于推广能力测度的多类 SVDD 模式识别方法[J]. 电子学报, 2009, 37(3): 464-469.
ZHU Xiaokai, YANG Degui. Multi-class support vector domain description for pattern recognition based on a measure of expansibility[J]. Acta Electronica Sinica, 2009, 37(3): 464-469.