Deep Anomaly Detection and Search via Reinforcement Learning

Chao Chen, Dawei Wang, Feng Mao, Zongzhang Zhang, Yang Yu
National Key Lab for Novel Software Technology, Nanjing University, China
Alibaba Inc., China
181840013@smail.nju.edu.cn, david.wdw@alibaba-inc.com, maofeng.mf@alibaba-inc.com,
zzzhang@nju.edu.cn, yuy@nju.edu.cn

目录 CONTENTS

一、摘要

1、摘要

半监督异常检测（AD）是一种数据挖掘任务，旨在从部分标记的数据集中学习特征以帮助检测异常值。现有的半监督AD方法大多数都存在对标记数据的充分利用和对无标记数据的探索不足。

为了解决这些问题，提出了深度异常检测和搜索（DADS），它应用强化学习（RL）来平衡利用和探索。

在训练过程中，智能体通过分层结构的数据集搜索可能的异常，并使用搜索到的异常来提高性能，这本质上借鉴了集成学习的思想。

二、介绍

1、介绍

在经典的半监督AD任务中，训练集通常包含测试集中所有类别的异常。在现实场景中，传入的异常会出现以前未见过的分布。这需要 AD 方法具有更强的挖掘未标记数据集的能力。

所以 AD 方法需要对受污染的未标记数据具有鲁棒性，能够利用已知的异常来高效地探索受污染的未标记数据，还需要有效地利用标记的已知异常。

2、基于RL的AD

最近提出的一种基于 RL 的 AD 方法称为 DPLAN ， 它使用 RL 设置来根据标记的异常探索未标记的样本。
DPLAN基于距离的搜索很容易陷入局部最优，面临过拟合和对未标记数据集污染的鲁棒性弱等问题。

RL将成为AD的强大工具，因为它在平衡短期奖励和长期奖励（也称为平衡开发和探索）方面具有很强的优势。

提出了一种称为 DADS 的方法，它将 RL 有效地应用于 AD。借鉴了集成的思想，具有分层异常搜索机制的环境。

三、算法

1、数据集

异常数据集Da：仅包含异常的小数据集。
无标签数据集Du：包含异常数据和正常数据的相对较大的数据集。

根据以上数据集找到一个异常评分函数 $f(\cdot): D \mapsto R$，使得 $\phi(s_i) > \phi(s_j)$，其中 s_i 为异常， s_j为正常。

2、DADS示意图

基于SAC的代理和异常搜索环境。

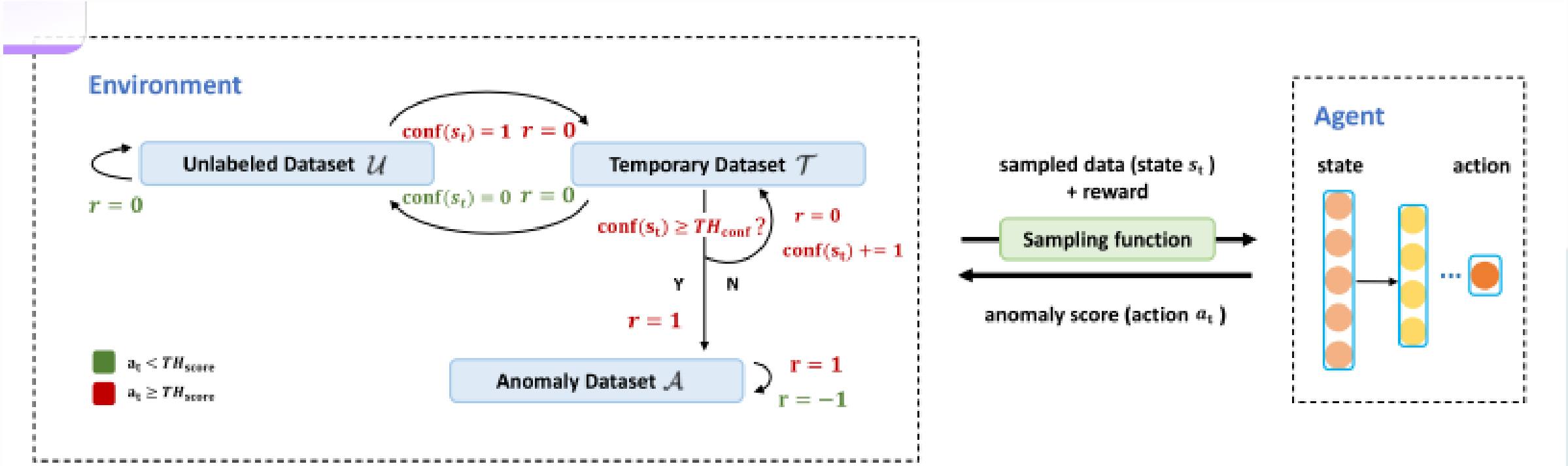


Figure 1: An illustration of our method DADS. See text for details.

3、训练数据转化强化学习数据流环境

将训练数据集转换为明确定义强化学习关键组件的数据流环境。

状态空间：是整个训练数据集 D = {Da, Du}，其中在时间步 t 采样的每个$s_t \in D$都是一个状态。
动作空间：是[0, 1]范围内的连续空间，其值对应于输入数据的异常分数。
环境：有自己的内部数据集，每个训练步骤都会采样单个数据作为状态。当接收到智能体的动作时，其内部数据集会根据一系列规则对其内部数据集进行调整并计算奖励。
代理：将当前数据作为输入，利用其内部网络计算相应的动作（输入数据的异常分数），并将其返回给环境。

4、基于SAC的代理

强化学习算法SAC，它在RL的原始目标上增加了一个额外的熵项，即最大化期望累积奖励，因此其目标函数变为：

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\tau_\pi} \gamma^t \left[r(s_t,a_t) + \alpha \mathcal{H}\left(\pi(\cdot \mid s_t)\right) \right]$$

其中 π 表示为策略网络，(st, at) 是策略 π 生成的轨迹 Tπ 内的状态-动作对，$\mathcal{H}(\pi(. \mid s_t))$ 是评估策略 π 不确定性的熵项，α > 0 是权衡系数。

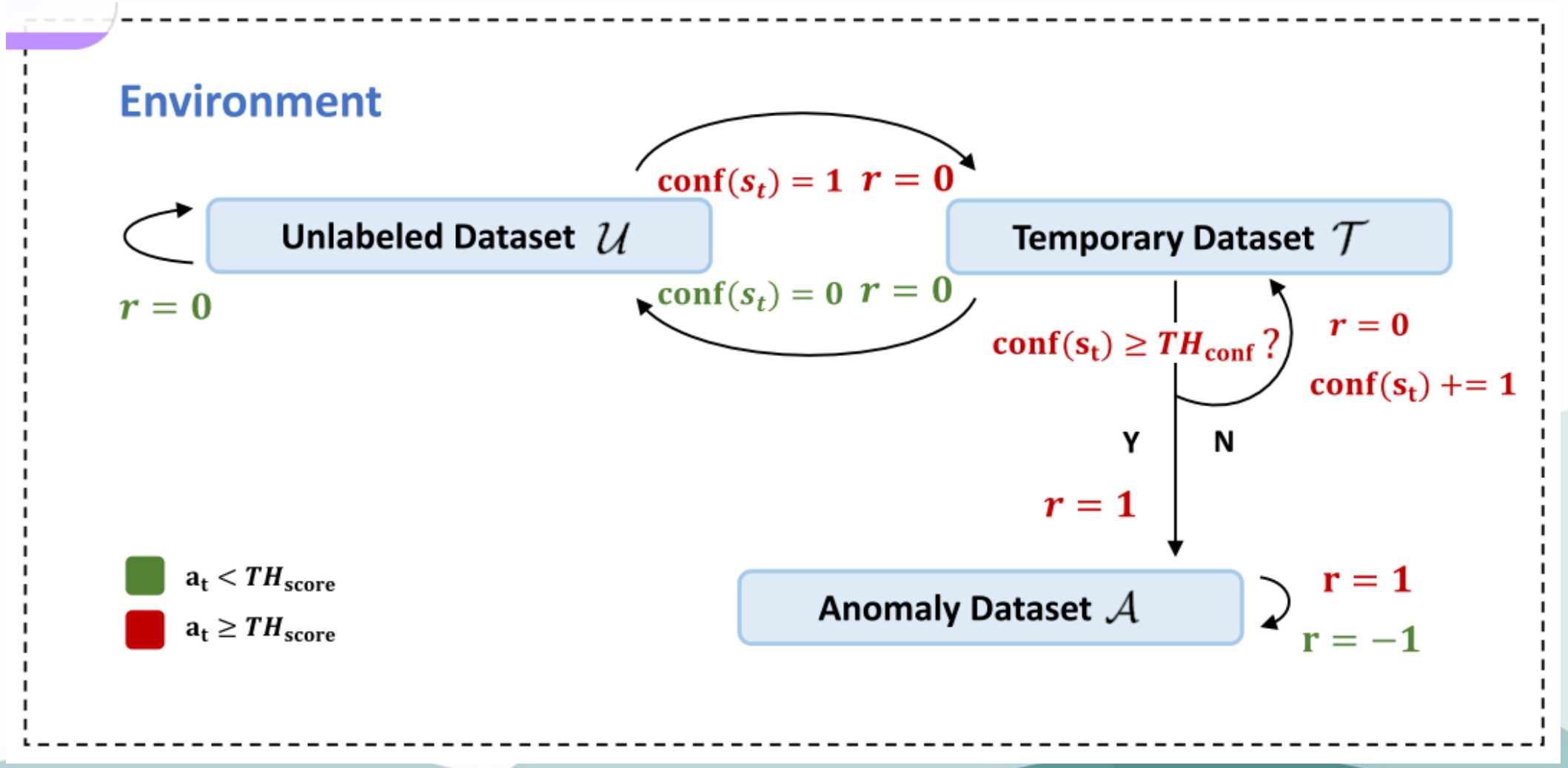
SAC 中目标函数的熵项对于 DADS 的异常搜索有很大帮助，因为它鼓励智能体采取不同的行动，从而发现更多可能的异常。这种熵正则化与 DADS 的采样功能一起，共同构成了 DADS 的搜索机制。

5、异常搜索环境

DADS的环境设计分为以下三个部分：

- 专为异常搜索而定制的分层数据集
- 基于集成的采样功能
- 异常检测奖励功能

5、异常搜索环境


Environment
$\mathrm{conf}(s_t) = 1 \quad r = 0$
Unlabeled Dataset $\mathcal{U}$
Temporary Dataset $\mathcal{T}$
$\mathrm{conf}(s_t) = 0 \quad r = 0$
$r = 0$
$\mathrm{conf}(s_t) \geq TH_{\mathrm{conf}}$?
$r = 0$
$\mathrm{conf}(s_t)$ += 1
Y
N
$r = 1$
$a_t < TH_{\mathrm{score}}$
$a_t \geq TH_{\mathrm{score}}$
Anomaly Dataset $\mathcal{A}$
r = 1
r = −1

6、专为异常搜索而定制的分层数据集

训练数据集由 Da 和 Du 组成。
将整个训练数据集分为三个互连的数据集：异常数据集A、临时数据集T和未标记数据集U。A 初始化为 Da，T 初始化为 ∅，U 初始化为 Du。
两个阈值超参数，分别是THscore和Thconf。

$$\texttt{Trans}(s_t, a_t) = \begin{cases} \mathcal{A} & s_t \in \mathcal{A} \\ \mathcal{A} & a_t \geq TH_{\text{score}}, s_t \in \mathcal{T}, \\ & \quad \texttt{conf}(s_t) \geq TH_{\text{conf}} \\ \mathcal{T} & a_t \geq TH_{\text{score}}, s_t \in \mathcal{U} \\ \mathcal{U} & a_t < TH_{\text{score}}, s_t \in \mathcal{U} \cup \mathcal{T} \end{cases}$$

7、基于集成的采样功能

环境从其内部数据集{A,T,U}的采样策略可以概括为两个阶段。
在第一阶段，环境根据预定义的概率分布从三个内部数据集中选择某个数据集。
在第二阶段，环境从第一阶段选择的数据集中采样数据。如果第一阶段选择的数据集在{A,T}中，只使用随机采样。如果在第一阶段选择了U，为了更好地探索U中的异常，借鉴集成学习的思想，设计了异常偏向的采样策略。

7、基于集成的采样功能

选择三种无监督 AD 算法，即隔离森林、基于直方图的异常值评分和 One-Class SVM。定义了一个随机分数生成器，给出 [0, 1] 范围内的分数。

每次环境都会首先从U中采样一定大小的一批数据S，然后根据预定义的概率分布（例如均匀分布）选择上述三种无监督方法中的一种（包括随即分数生成器）表示为UNSUP(x)。

用它来计算S中每个数据的异常分数，最后选择异常分数最高的数据。

$$s_t = \begin{cases} \mathrm{U}(\mathcal{C}) & \mathcal{C} \in \{\mathcal{A}, \mathcal{T}\} \\ \arg\max_{s \in S} \mathrm{UNSUP}(s) & \mathcal{C} = \mathcal{U} \end{cases}$$

8、异常检测奖励功能

环境的奖励函数是基于监督奖励和无监督奖励设计的。对于从A中采样的st，如果正确分类st，它将获得正奖励，否则将受到惩罚。对于从T中采样的st，如果conf(st)≥THconf并且分类为异常，则环境会将st移动到A并给予正奖励；否则，奖励为0。对于从U中采样的st，为了使代理能够借助无监督方法学习数据分布，环境将使用隔离森林给予无监督奖励，写为IForest(st)。

$$r(s_t|a_t) = \begin{cases} 1 & s_t \in \mathcal{A}, a_t \geq TH_{\text{score}} \\ -1 & s_t \in \mathcal{A}, a_t < TH_{\text{score}} \\ \texttt{IForest}(s_t) & s_t \in \mathcal{U} \\ 1 & s_t \in \mathcal{T}, a_t \geq TH_{\text{score}}, \\ & \quad \texttt{conf}(s_t) \geq TH_{\text{conf}} \\ 0 & \text{else} \end{cases}$$

四、评估

1、数据集

5 个单异常类数据集和 3 个多异常类数据集。
前5个数据集用于场景1的实验。后3个数据集用于场景2的实验。
如果异常类别已知，则它将同时出现在异常数据集中和未标记数据集中，否则仅存在于未标记数据集中。

Table S3: Datasets with single anomaly class.

Dataset			Normal Class	Anomaly Class
Name	N	D		
annthyroid	7200	6	6666	534(7.4%)
cardio	1831	21	1655	176(9.6%)
satellite	6435	36	4399	2036(31.6%)
satimage2	5803	36	5732	71(1.2%)
thyroid	3772	6	3679	93(2.5%)

Table S4: Datasets with multiple anomaly classes.

Dataset			Normal Class		Anomaly Class		
Name	Size	Dim	Name	Size	Name	Size	Known
annthyroid	3772	21	normal	3488	hypothyroid subnormal	93(2.5%) 191(5.1%)	N Y
cardio	2126	21	normal	1655	suspect pathologic	295(13.9%) 176(8.28%)	Y N
har	10299	561	walking, sitting standing, laying	7349	upstairs downstairs	1544(15.0%) 1406(13.7%)	Y N

2、评估指标

设置两个可调参数，即异常率和污染率。异常比率对应于已知异常与总异常的比率（在场景2中，总异常指的是已知异常类别的异常）；污染率对应于未标记数据中异常的百分比。
我们将每个数据集分层分为训练集、验证集和测试集，分别占总数据集的 60%、20% 和 20%。训练数据集是根据上述两个可调参数：异常率和污染率手动生成的，由Da和Du组成。
我们选择受试者工作特征曲线下面积（AUC-ROC）作为评价指标，衡量ROC曲线下面积。

3、设置 1.1：AUC 与具有固定异常率的增量污染率。

异常比率设置为0.1。再将污染率从 0 逐渐增加到 0.1，间隔设置为 0.02，以测试 ADAS 与其他算法相比是否对未标记数据中的污染更加稳健。
DADS在平均AUC-ROC和平均排名上均排名第一，这证明了DADS在抵抗未标记数据的高污染方面的优势。

	Known AD Scenario								
	DADS	DPLAN	DeepSAD	DevNet	SSAD	STOC+GOAD	STOC+IForest	Supervised	Unsupervised
annthyroid	0.881 ±0.011	0.824 ±0.042	0.894 ±0.025	0.832 ±0.025	0.739 ±0.059	0.610 ±0.028	0.833 ±0.023	**0.984±0.010**	0.804 ±0.037
cardio	**0.973 ±0.017**	0.769 ±0.098	0.879 ±0.054	0.967 ±0.021	0.898 ±0.047	0.945 ±0.015	0.919 ±0.018	0.912 ±0.076	0.920 ±0.014
satellite	0.822 ±0.041	0.852 ±0.032	**0.903 ±0.033**	0.845 ±0.015	0.859 ±0.015	0.775 ±0.014	0.786 ±0.011	0.838 ±0.044	0.769 ±0.014
satimage2	0.990 ±0.006	0.813 ±0.102	0.980 ±0.018	0.974 ±0.047	0.973 ±0.015	0.953 ±0.023	**0.994 ±0.003**	0.969 ±0.033	**0.994 ±0.003**
thyroid	0.993 ±0.007	0.940 ±0.068	0.953 ±0.031	**0.995 ±0.006**	0.960 ±0.028	0.901 ±0.031	0.978 ±0.007	0.966 ±0.043	0.974 ±0.005
Average	0.932 ±0.016	0.839 ±0.068	0.922 ±0.032	0.922 ±0.023	0.886 ±0.033	0.837 ±0.022	0.902 ±0.012	**0.934 ±0.041**	0.892 ±0.015
AverageRank	**3.0**	7.0	4.4	3.4	5.8	7.4	4.2	4.8	5.0

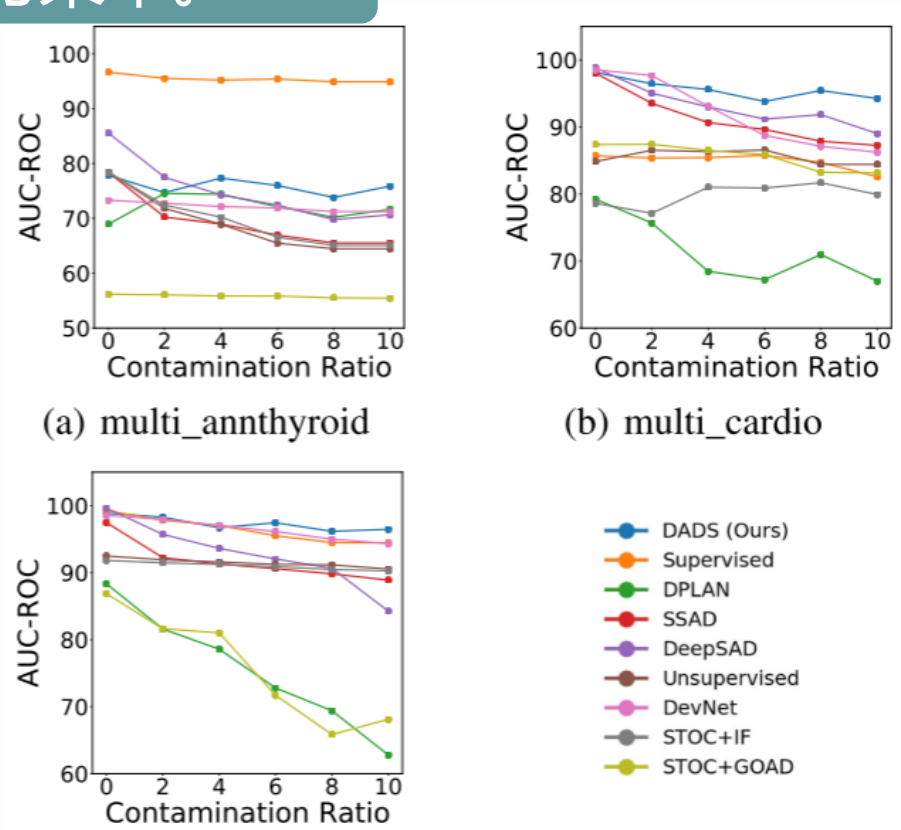
4、设置 1.2：AUC 与固定污染率下的增量异常率。

将污染率设置为0.04，并在[0.01,0.5]范围内调整异常率，看看搜索可能的异常是否对DADS的性能有贡献。从结果中我们可以看出，当只有少量标记异常可用时，DADS 表现良好。一些方法（如 IForest 和 STOC+IF）的性能随着异常率的增加而下降，而 DADS 随着异常率的增长而单调增加。

	Known AD Scenario								
	DADS	DPLAN	DeepSAD	DevNet	SSAD	STOC+GOAD	STOC+IForest	Supervised	Unsupervised
annthyroid	0.857 ±0.013	0.671 ±0.069	0.868 ±0.041	0.809 ±0.035	0.722 ±0.051	0.613 ±0.024	0.866 ±0.015	**0.932 ±0.075**	0.843 ±0.026
cardio	**0.956 ±0.037**	0.574 ±0.185	0.846 ±0.064	0.929 ±0.067	0.897 ±0.044	0.949 ±0.012	0.935 ±0.017	0.887 ±0.088	0.937 ±0.010
satellite	0.845 ±0.032	0.630 ±0.079	**0.903 ±0.018**	0.849 ±0.018	0.858 ±0.011	0.769 ±0.012	0.795 ±0.019	0.803 ±0.014	0.786 ±0.016
satimage2	0.990 ±0.006	0.886 ±0.105	0.970 ±0.034	0.989 ±0.022	0.973 ±0.012	0.957 ±0.023	**0.994 ±0.003**	0.960 ±0.029	**0.994 ±0.003**
thyroid	0.976 ±0.053	0.639 ±0.191	0.911 ±0.038	**0.986 ±0.019**	0.946 ±0.024	0.900 ±0.030	0.976 ±0.006	0.896 ±0.091	0.974 ±0.005
Average	**0.925 ±0.028**	0.680 ±0.126	0.900 ±0.039	0.912 ±0.032	0.879 ±0.029	0.838 ±0.020	0.913 ±0.012	0.895 ±0.060	0.907 ±0.012
AverageRank	**3.0**	8.8	4.6	3.8	5.0	6.8	3.2	5.6	4.2

5、设置 2.1：AUC 与具有固定异常率的增量污染率。

将异常率设置为 0.1，并将污染率从 0 调整到 0.1。
随着污染率的上升，除DADS之外的所有方法都出现了不同程度的下降，这再次证明了DADS对污染的鲁棒性。
DADS虽然平均 AUC 略逊于 XGBoost，但 DADS 的平均排名仍然排名第一。



	Unknown AD Scenario								
multi_annthyroid	0.759 ±0.047	0.716 ±0.055	0.706 ±0.064	0.712 ±0.047	0.655 ±0.047	0.554 ±0.037	0.651 ±0.039	**0.949 ±0.043**	0.645 ±0.040
multi_cardio	**0.943 ±0.021**	0.670 ±0.058	0.890 ±0.03	0.862 ±0.044	0.873 ±0.022	0.831 ±0.023	0.799 ±0.047	0.826 ±0.036	0.844 ±0.018
multi_har	**0.965 ±0.017**	0.628 ±0.135	0.843 ±0.139	0.943 ±0.016	0.889 ±0.019	0.681 ±0.148	0.903 ±0.011	0.944 ±0.020	0.906 ±0.012
Average	0.889 ±0.028	0.671 ±0.082	0.813 ±0.078	0.839 ±0.036	0.806 ±0.029	0.689 ±0.070	0.784 ±0.032	**0.906 ±0.033**	0.798 ±0.023
AverageRank	**1.3**	7.0	4.7	3.7	5.0	7.7	6.7	3.3	5.7

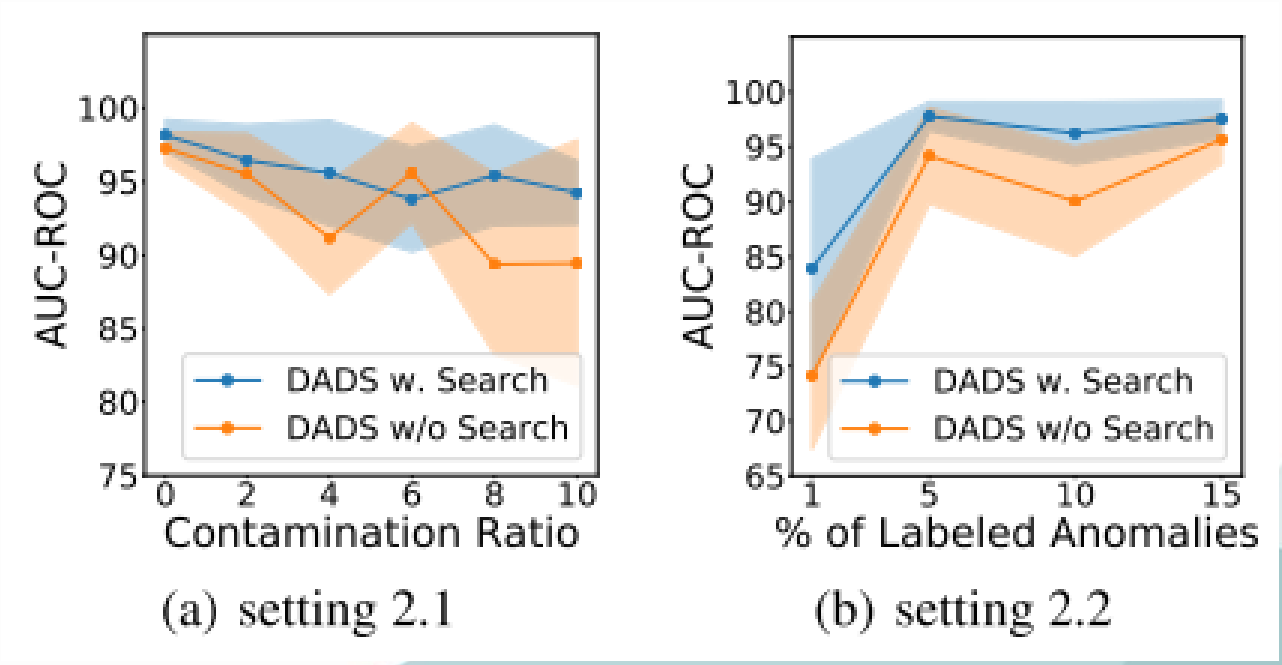
6、设置 2.2：AUC 与固定污染率下的增量异常。

将污染率固定为 0.04，并将异常率从 0.01 增加到 0.5。DADS 表现最好。归功于高效的搜索机制，该机制使得 DADS 能够在存在多个异常类别时搜索可能的异常。

	Unknown AD Scenario								
multi_annthyroid	0.703 ±0.075	0.617 ±0.128	0.693 ±0.059	0.693 ±0.069	0.664 ±0.05	0.562 ±0.035	0.690 ±0.028	**0.927 ±0.043**	0.685 ±0.046
multi_cardio	**0.978 ±0.013**	0.561 ±0.091	0.921 ±0.030	0.927 ±0.049	0.897 ±0.045	0.860 ±0.029	0.802 ±0.029	0.823 ±0.028	0.881 ±0.016
multi_har	0.977 ±0.011	0.712 ±0.13	0.898 ±0.127	0.966 ±0.015	**0.987 ±0.002**	0.766 ±0.058	0.916 ±0.014	0.882 ±0.081	0.920 ±0.013
Average	**0.886 ±0.033**	0.630 ±0.116	0.837 ±0.072	0.862 ±0.044	0.849 ±0.032	0.730 ±0.041	0.803 ±0.024	0.877 ±0.051	0.829 ±0.025
AverageRank	**1.7**	8.7	4.3	2.7	4.0	7.7	6.0	5.0	5.0

7、消融实验

删除了 DADS 的搜索机制，并保持其他组件不变。
设置 2.1 和 2.2 的实验在原始 DADS 和无需搜索的 DADS。



五、结论

1、结论

使用集成学习和设计环境的思想以及分层组织的数据集来实现对受污染的未标记数据更好的鲁棒性。
通过将无监督AD方法结合到采样函数中，我们进一步提高了搜索机制的有效性。

2、算法分析

对污染的鲁棒性强：在未标记的数据集中搜索可能的异常时，会对其进行清理。奖励函数设计简单，不包含任何未标记数据集的先验假设。

搜索效率和准确率高：借用集成学习的思想来提高搜索的准确率。DADS在其采样功能中结合了不同的无监督方法来提高搜索效率，而SAC的熵项也起到了提高的作用。

探索和利用的良好平衡：在强化学习的帮助下，DADS 的整个训练过程以自然的方式集成了探索未标记数据集和利用标记异常。