Perspective-based Multi-task Learning for Outlier Interpretation

Zhuoling Li[1], Lili Guan[1], Xinye Wang[1], Zhengyong Pan[2], and Lei Duan[1(✉)]

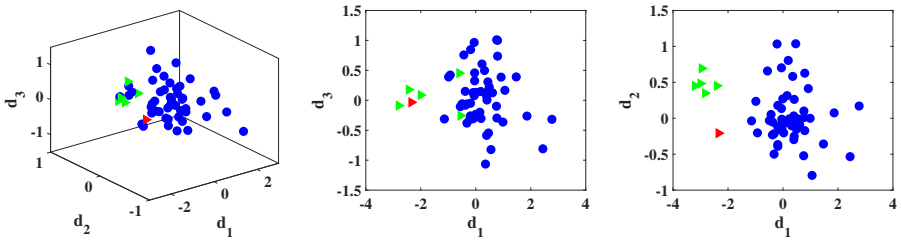
[1]School of Computer Science, Sichuan University, Chengdu, China
{lizhuoling, guanlili, wangxinye}@stu.scu.edu.cn, leiduan@scu.edu.cn
[2]JD.com, Chengdu, Sichuan, China
cdpanzhengyong@jd.com

Abstract. Anomaly detection has been intensively studied in recent years, yet answering the reason why an instance is detected to be an outlier, *i.e.*, outlier interpretation, remains challenging. To this end, existing model-agnostic methods mainly explain outliers by mining the feature subspace. However, most of them solely use normal instances to explain the query outlier. This single-perspective consideration often leads to a lack of specificity, as it fails to capture the nuanced distinctions needed to explain why the query outlier is different from other outliers. Furthermore, outliers similar to or different from the query outlier may share the same subspace for interpretation, resulting in explanations that lack targeted insights. To address these limitations, we propose a novel model-agnostic outlier interpretation method named PML (Perspective-based Multi-task Learning for outlier interpretation). Specifically, PML samples and transforms datasets into quintuplets and learns representations of quintuplets. Furthermore, PML introduces a multi-task learning strategy to learn the feature subspace from both normal and abnormal perspectives. Finally, PML uses the optimal feature subspace to explain why the query outlier is anomalous. Through extensive experiments, we demonstrate the effectiveness of PML. The source code is available at https://github.com/scu-kdde/OAM-PML-2025.

Keywords: Outlier interpretation · Feature subspace · Multi-task learning · Quintuplet.

1 Introduction

Anomaly detection is intensively studied in the past decades [15]. For the detected outliers, how to explain the reason for their anomalies is equally important, which provides valuable insights for analysts to better understand, solve, and prevent these outliers [11]. Existing model-agnostic outlier interpretation methods offer advanced techniques such as attention mechanisms [16], sum-product networks [9] and association rules [1] to improve the performance of interpretation.



(a) Original space (b) Suboptimal subspace (c) Optimal subspace

Fig. 1: Distribution of "Fraud" and "Non-Fraud" transactions from real credit card dataset in different space. ("▶" is a query "Fraud" transaction, "▶" is the other 'Fraud" transaction, and "●" is a "Non-Fraud" transaction.)

Though various techniques have been proposed, however, most of them solely consider the deviation of the query outlier from normal instances, ignoring that the query outlier is likely to either be different from or similar to the other outliers. It is worth noting that the single perspective weakens the ability of the model to further analyze the abnormality of the query outlier relative to other outliers. Take the transaction statistics of the six "Fraud" and fifty "Non-Fraud" transactions[1] as an example.

Example 1. In the original space as shown in Figure 1(a), there is no obvious distance between the query "Fraud" and "Non-Fraud" transactions. The query "Fraud" is difficult to be interpreted, since we can not tell which features lead to the "Fraud" transaction. Figure 1(b) shows the distribution in the suboptimal feature subspace. The query "Fraud" ("▶") is revealed from all the "Non-Fraud" transactions ("●"). However, the other "Fraud" transactions ("▶") either gathering near the query "Fraud" or mixing in the "Non-Fraud" transactions are ignored, which makes it hard for end-users to distinguish the query "Fraud" from the other "Fraud" transactions while separating other "Fraud" from the "Non-Fraud" transactions. Figure 1(c) shows the distribution result in an optimal feature subspace after further analyzing the abnormality of the detected "Fraud" ("▶") relative to other "Fraud" ("▶"). The interpretation from both two perspectives helps end-users to know the reason why the query "Fraud" is distinct from both "Non-Fraud" and other "Fraud" transactions, and separates "Fraud" and "Non-Fraud" transactions apart.

This example stresses the importance of interpreting the query outlier from two perspectives simultaneously. To effectively complete the interpretation task, we need to consider the following challenges:

(CH1) How to obtain the representations of quintuplets to capture rich information? Most methods directly feed raw data into neural networks

[1] https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets

to capture complex patterns, ignoring the need to learn representations of quintuplets from both perspectives. This obscures essential distinctions needed to generate a targeted interpretation.

(CH2) How to learn the feature subspace where the query outlier is revealed from both normal instances and other outliers? The performance of interpretation results heavily depends on the way how to optimize the task from both two perspectives. We look for a strategy that shows advantages in scenarios where multiple objectives need to be satisfied simultaneously.

To address these challenges, we propose a novel model-agnostic outlier interpretation method, named PML (Perspective-based Multi-task Learning for outlier interpretation). To tackle **CH1**, PML goes through sampling and concatenation procedures before the representation learning process to ensure that the model can effectively capture the desired representations for the generated quintuplets. To tackle **CH2**, PML leverages a multi-task learning strategy to enable the model to adaptively assign appropriate weights to optimize the interpretation task from two perspectives. Thus, PML outputs a feature subspace where clear separations between normal instances and the query outlier exit, while the query outlier shows its distinct abnormality relative to other outliers.

The main contributions of this paper are as follows:

- We design a model-agnostic outlier interpretation method PML to find the feature subspace, where the query outlier can be revealed from normal instances and other outliers. PML is model-agnostic and can be applied to explain any detected outlier.
- We propose a new perspective for outlier interpretation and introduce a multi-task learning strategy to further distinguish the query outlier from other outliers.
- Extensive experiments are conducted to compare PML with state-of-the-art outlier interpretation methods on 7 real-world datasets, validating the effectiveness of PML.

2 Problem Statement

Let $\mathcal{D}$ be a d-dimensional original space. Given a set of outliers $\mathbf{X}_+$ and a set of normal instances $\mathbf{X}_-$ in the full space $\mathcal{D}$, each instance $\boldsymbol{x} \in \mathbf{X}_+ \cup \mathbf{X}_-$ is represented as a vector in the form $[\boldsymbol{x}(1), \cdots, \boldsymbol{x}(d)]^\top$. Typically, we have $|\mathbf{X}_+| \ll |\mathbf{X}_-|$. To obtain scalable high-level information and rich semantic representations within the original space $\mathcal{D}$, we map all instances into an embedding space $\mathcal{E}$. For the query outlier $\boldsymbol{x}_q \in \mathbf{X}_+$, we use the outlying degree $OutD(\boldsymbol{x}_q|\mathbf{X}_+ \cup \mathbf{X}_-)$ to measure its outlyingness. Without loss of generality, we assume that the lower the value of $OutD(\boldsymbol{x}_q|\mathbf{X}_+ \cup \mathbf{X}_-)$, the more outlying $\boldsymbol{x}_q$. We formalize the outlier interpretation task as follows:

Definition 1 (Outlier Interpretation). *Given $\mathbf{X}_+$ and $\mathbf{X}_-$ in a d-dimensional space $\mathcal{D}$ and a query outlier $\boldsymbol{x}_q \in \mathbf{X}_+$, outlier interpretation for the query outlier $\boldsymbol{x}_q$ is to obtain an optimal feature subspace $\mathcal{E}^*$ where the query outlier $\boldsymbol{x}_q$ is*

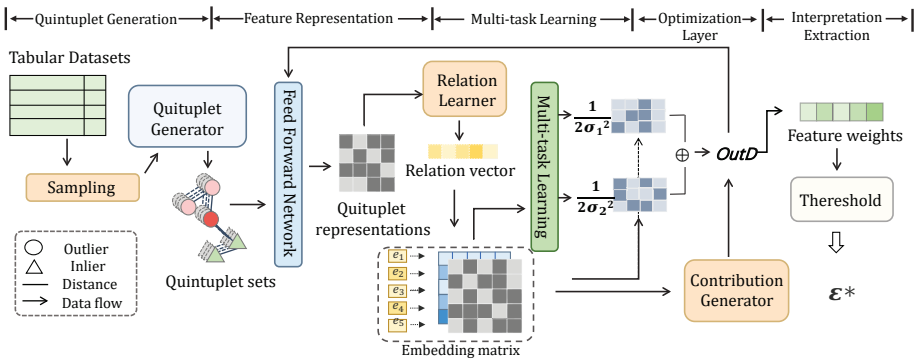


Fig. 2: The architecture of PML.

revealed from $\mathbf{X}_-$ *and shows differences from other outliers* $\mathbf{X}_+$.

$$\mathcal{E}^* = \arg\min_{\mathcal{E}} OutD(\boldsymbol{x}_q|\mathbf{X}_+ \cup \mathbf{X}_-)$$

We obtain an optimal feature subspace $\mathcal{E}^*$*, answering the reason why the query outlier* $\boldsymbol{x}_q$ *is distinct from normal instances and the other outliers.*

3 Outlier Interpretation Framework

3.1 The Design of PML

In this section, we discuss five components of PML. The overall framework of PML is shown in Figure 2.

Quintuplet Generation Given a query outlier $\boldsymbol{x}_q$, the objective is to obtain a feature subspace for interpretation from both normal and abnormal perspectives. Since outliers account for a small proportion of the imbalanced datasets [4], we randomly sampled $\mathbf{X}_-$ and $\mathbf{X}_+$ to generate candidate sets C_- and C_+, respectively. Then we utilize the Euclidean distance as the metric to select the nearest neighbors of $\boldsymbol{x}_q$, denoted as C'_- and C'_+, respectively. The set of quintuplets is:

$$\mathbf{Q} = \{\langle \boldsymbol{x}_q, \boldsymbol{\beta}, \boldsymbol{\beta'}, \boldsymbol{\gamma}, \boldsymbol{\gamma'} \rangle | \boldsymbol{\beta} \in C_+, \boldsymbol{\beta'} \in C'_+, \boldsymbol{\gamma} \in C_-, \boldsymbol{\gamma'} \in C'_- \} \tag{1}$$

The sampling size of the candidate sets C_+ and C'_+ is m, and the sampling size of the candidate sets C_- and C'_- is n. To some extent, the larger size of the training samples tends to yield better results. However, excessively large size leads to reduced efficiency.

Feature Representation PML employs a feed-forward neural network $\zeta(\cdot)$ to map quintuplets from the original space into a new embedding space. Specifically, the feed-forward neural network is carried out using two learnable linear layers, parameterized by the embedding matrix W and the relation matrix W', respectively. The first layer is utilized to transfer an instance $\boldsymbol{v} \in \mathcal{R}^d$ with the

original feature dimension of d to an embedding space with a new dimension d'. Then the vector concatenation function $cat(\cdot)$ is used to generate the union of the elements in the quintuplet, denoted as $U \in \mathbb{R}^{5d'}$.

$$U = cat(W\boldsymbol{x}_q, W\boldsymbol{\beta}, W\boldsymbol{\beta}', W\boldsymbol{\gamma}, W\boldsymbol{\gamma}') \tag{2}$$

Then U is fed into the relation layer parameterized by the relation weight matrix $W' \in \mathbb{R}^{d' \times 5d'}$ to learn the inner relationship of the elements from both normal and abnormal perspectives. We utilize the normalization function θ to ensures the outputs of the relation layer are normalized and the relation learner $AvgRow$ to obtain the relation vector $\boldsymbol{e} \in \mathbb{R}^{d'}$.

$$\boldsymbol{e} = AvgRow(\theta(W'U))$$

where $AvgRow$ is to compute the average of all rows in each column of U. Fully expanded out, PML yields quintuplets representations $\zeta(\mathbf{Q})$ as follows:

$$\zeta(\mathbf{Q}) = \{\langle \zeta(\boldsymbol{x}_q), \zeta(\boldsymbol{\beta}), \zeta(\boldsymbol{\beta}'), \zeta(\boldsymbol{\gamma}), \zeta(\boldsymbol{\gamma}') \rangle\}$$

Multi-task Learning Performance of PML crucially hinges on assigning the appropriate weights to each task. We draw inspiration from a principled multi-task learning method [5]. Let $h_{W''}(\boldsymbol{q})$ be the output of a nerual network with weight matrix W'' for each quintuplet element $\boldsymbol{q} \in \zeta(\mathbf{Q})$. The likelihood is modeled as a Gaussian distribution with mean $h_{W''}(\boldsymbol{q})$:

$$p(y \mid h_{W''}(\boldsymbol{q})) = \mathcal{N}(h_{W''}(\boldsymbol{q}), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - h_{W''}(\boldsymbol{q}))^2}{2\sigma^2}\right)$$

where y is the predicted label and σ denotes the observation noise. For multiple tasks, the likelihood factorizes across task outputs:

$$p(y_1, \ldots, y_c \mid h_{W''}(\boldsymbol{q})) = p(y_1 \mid h_{W''}(\boldsymbol{q})) \times \cdots \times p(y_c \mid h_{W''}(\boldsymbol{q}))$$

where $y_1, y_2, \ldots, y_c$ are the outputs of c tasks. Supposing our objective is decomposed into two tasks, y_1 and y_2 with the observation noise σ_1 and σ_2, the joint likelihood becomes:

$$p(y_1, y_2 \mid h_{W''}(\boldsymbol{q})) == -\frac{1}{2}\sum_{i=1}^{2} \frac{(y_i - h_{W''}(\boldsymbol{q}))^2}{\sigma_i^2} - \log(\sigma_1\sigma_2) \tag{3}$$

Optimization Layer To dynamically adjust the importance of the two tasks, we design the following outlying degree function:

$$OutD(\boldsymbol{x}_q | \mathbf{X}_+ \cup \mathbf{X}_-) = \sum_{\boldsymbol{q} \in \zeta(\mathbf{Q})} \frac{1}{2\sigma_1^2}\mathcal{T}_1(\boldsymbol{q}) + \frac{1}{2\sigma_2^2}\mathcal{T}_2(\boldsymbol{q}) + \log\sigma_1\sigma_2 \tag{4}$$

$$\mathcal{T}_1(\boldsymbol{q}) = \max(0, \alpha' + dis(\zeta(\boldsymbol{\beta}), \zeta(\boldsymbol{\beta}')) - dis(\zeta(\boldsymbol{\beta}), \zeta(\boldsymbol{x}_q)))$$
$$\mathcal{T}_2(\boldsymbol{q}) = \max(0, \alpha + dis(\zeta(\boldsymbol{\gamma}), \zeta(\boldsymbol{\gamma}')) - dis(b(\boldsymbol{\gamma}), \zeta(\boldsymbol{x}_q)))$$

where α and α' are the margin and $dis(\cdot)$ represents the Euclidean distance in the embedding space. Two triplet losses are used to achieve two tasks, denoted as $\mathcal{T}_1(\boldsymbol{q}), \mathcal{T}_2(\boldsymbol{q})$, respectively. Take $\mathcal{T}_1(\cdot)$ for example. Given an anchor point $\boldsymbol{\beta}$ (a random outlier) , the projection of a positive point $\boldsymbol{\beta}'$ (a neighbor outlier) is closer to the anchor's projection than that of a negative point $\zeta(\boldsymbol{x}_q)$ (the query outlier), by at least a margin α'. We utilize the coefficients in Equation (3) to control weights of the two tasks and use the Softplus function and a regularization function to ensure positivity, promote stability and prevent overfitting. The final term $\log \sigma_1 \sigma_2$ acts as a regularizer for curbing excessive noise increments.

Interpretation Extraction We design a contribution generator to generate the contribution vector $\boldsymbol{\eta}$. Each dimension of $\boldsymbol{\eta}$ corresponds to the contribution of each feature in the original space. η is defined as follows:

$$\boldsymbol{\eta} = \left|W^{*\top}\right| \boldsymbol{e}^*$$

where $W^{*\top}$ is the transpose of the final optimal embedding matrix and $\boldsymbol{e}^*$ is the relation vector learned in Section 3.1. We use $|\cdot|$ to compute the absolute values of the elements and measure the impact of every original feature on its corresponding dimension of the embedding space. The objective can be further expressed as $\mathcal{E}^* = \{f_j \in \mathcal{F} | \sum_{j=1}^{|\mathcal{E}|} \eta_j > \delta\}$, where the threshold $\delta = \sqrt{\frac{2}{D}} \sum_{f \in \mathcal{F}} \eta(f)$ is defined in [16] to extract the feature subspace for interpretation. Consequently, we not only obtain an optimal feature subspace $\mathcal{E}^*$, but also the contribution degree of each dimension in the feature subspace.

3.2 Convergence Analysis

Since $\mathcal{T}_1(\boldsymbol{q})$ and $\mathcal{T}_2(\boldsymbol{q})$ are both max functions and convex, their sum $OutD(\boldsymbol{x}_q | \mathbf{X}_+ \cup \mathbf{X}_-)$ is also convex. We prove the convergence of PML by analyzing this convex function.

Theorem 1. *Let step $\alpha_l = \frac{1}{L_g}$. l and L_g denote the l-th epoch and the smoothing factor, respectively. s denotes the maximum number of epochs. If $OutD(\cdot)$ is convex, $\forall r(\boldsymbol{q}), r'(\boldsymbol{q}) \in$ dom $OutD(\cdot)$*

$$OutD(r^s(\boldsymbol{q})) - OutD^*(r(\boldsymbol{q})) \leq \frac{L_g \|r^0(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2}{2s}$$

Proof. Given $r(\boldsymbol{q}), r'(\boldsymbol{q}) \in$ dom $OutD(\cdot)$, and θ $(0 \leq \theta \leq 1)$,

$$\begin{aligned} &OutD(r^{l+l}(\boldsymbol{q})) \\ &\leq OutD(r^l(\boldsymbol{q})) + \langle \nabla OutD(r^l(\boldsymbol{q})), r^{l+1}(\boldsymbol{q}) - r^l(\boldsymbol{q}) \rangle + \frac{L_g}{2} \|r^{l+1}(\boldsymbol{q}) - r^l(\boldsymbol{q})\|^2 \\ &\leq OutD(r^l(\boldsymbol{q})) - \frac{1}{2L_g} \|\nabla OutD(r^l(\boldsymbol{q}))\|^2 \end{aligned}$$

Let $\Delta^l = OutD(r^l(\boldsymbol{q})) - OutD^*(r(\boldsymbol{q}))$. We have:

$$\Delta^{l+1} - \Delta^l \leq -\frac{1}{2L_g} \|\nabla OutD(r^l(\boldsymbol{q}))\|^2 \tag{5}$$

Table 1: Description of the 6 benchmark datasets.

Dataset	#Instances	#Features	#Outliers	%Rate
Ionosphere	351	32	126	35.90
vowels	1,456	12	50	3.43
letter	1,600	32	100	6.25
thyroid	3,772	6	93	2.47
Wilt	4,819	5	257	5.33
pendigits	6,870	16	156	2.27

Since function $OutD(r(\boldsymbol{q}))$ is convex, we have:

$$\begin{aligned}\Delta^l &\leq -\langle \nabla OutD(r^l(\boldsymbol{q})), r^*(\boldsymbol{q}) - r^l(\boldsymbol{q})\rangle \\ &\leq \frac{L_g}{2}(\|r^l(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2 - \|r^{l+1}(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2) + \frac{\|\nabla OutD(r^l(\boldsymbol{q}))\|^2}{2L_g} \end{aligned} \quad (6)$$

Combined Equations (5) and (6), we have:

$$\Delta^{l+1} \leq \frac{L_g}{2}(\|r^l(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2 - \|r^{l+1}(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2)$$

Then, we have:

$$\sum_{l=0}^{s-1} \Delta^{l+1} \leq \frac{L_g}{2}\sum_{l=0}^{s-1}(\|r^l(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2 - \|r^{l+1}(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2) \leq \frac{L_g}{2}\|r^0(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2$$

Since $\Delta^{l+1} \leq \Delta^l$, we have:

$$OutD(r^s(\boldsymbol{q})) - OutD^*(r(\boldsymbol{q})) \leq \frac{L_g\|r^0(\boldsymbol{q}) - r^*(\boldsymbol{q})\|^2}{2s}$$

Therefore, PML is convergent and able to obtain the optimal subspace.

4 Experiments

4.1 Experimental Setups

Datasets We conduct experiments using 6 real-world datasets available in ADBench[2]. All the above mentioned datasets are summarized in Table 1. The outliers in the above mentioned datasets are with real semantics based on corresponding domain knowledge.

Baselines According to the form of the interpretation, outlier interpretation methods can be divided into two categories, feature weights and feature subspaces. PML falls into the former category. However, feature weights can be transformed into feature subspaces by setting the size [16], which makes PML

Table 2: The performance of methods output **Feature Weights** as interpretation. Each row represents the average value of AUPRC and AUROC using the ground-truth annotations generated by outlier detectors (HBOS, MCD, SOD, and ABOD), respectively.

	AUPRC				**AUROC**			
	SHAP	LIME	ATON	**PML**	SHAP	LIME	ATON	**PML**
thyroid	0.358	**0.450**	0.295	0.363	0.360	**0.515**	0.323	0.445
letter	0.586	0.475	**0.640**	0.630	0.663	0.575	0.710	**0.715**
Ionosphere	0.453	0.430	0.618	**0.640**	0.543	0.595	0.750	**0.765**
pendigits	**0.705**	0.605	0.550	0.610	**0.788**	0.738	0.713	0.742
vowels	0.625	0.550	0.670	**0.708**	0.658	0.658	0.720	**0.725**
Wilt	0.320	0.358	0.445	**0.625**	0.325	0.460	0.548	**0.729**
Average	0.503	0.475	0.533	**0.593**	0.552	0.585	0.625	**0.682**
Improvement	15.18%	19.90%	10.12%	-	19.06%	14.2%	8.4%	-

comparable to both categories. We select six baseline methods, including ATON [16], SHAP [10], LIME [12], SiNNE [14], Anchors [13] and PARs [1].

Ground-truth Annotations We utilize an evaluation method to obtain the ground-truth of interpretation. Evaluating the performance of outlier interpretation results requires benchmark datasets with ground-truth annotations (real feature subspace). However, to the best of our knowledge, there is no publicly available real-world dataset with such annotations. We use the labeling strategy to obtain the ground-truth annotations [16]. To eliminate the impact of distinct characteristics of different detection models, we use five different anomaly detection models (HBOS [2], MCD [3], SOD [6], ABOD [7], iForest [8].) to search for exceptional subspaces and record the most outlying subspace of each outlier as the ground-truth annotations.

Evaluation Metrics For methods that output feature weights, we use the AUPRC and AUROC as the evaluation metrics. The precision is equal to the recall after the size of the feature subspace is set to be the same as that of the ground-truth subspace during the transformation process mentioned in Section 4.1. For methods that output feature subspaces, we use F1 computed by precision and recall to evaluate the effectiveness. Suppose $\mathcal{G}$ represents the ground-truth subspace and $\mathcal{E}$ represents the generated subspace, then $p = \frac{|\mathcal{G} \cap \mathcal{E}|}{|\mathcal{E}|}$, $r = \frac{|\mathcal{G} \cap \mathcal{E}|}{|\mathcal{G}|}$, and $\text{F1} = 2\left(\frac{p \times r}{p+r}\right)$, respectively.

Parameter Settings and Implementations We utilize the Adam optimizer and apply an early stopping mechanism to train PML. The sampling numbers $m = 10$, $n = 10$, respectively. The dimension of the embedding space d' is 64. PML is trained by 25 epochs and 512 quintuplets per batch. The parameters of ATON and Anchors are set by default. In SiNNE, the width of the beam search, the sampling number, and the ensemble number are set to 10, 8 and 100,

[2] https://github.com/Minqi824/ADBench

Table 3: The performance of methods output **Feature Subspaces** as interpretation. Each row represents the average value of F1 using the ground-truth annotations generated by outlier detector (iForest).

	F1			
	PARs	SiNNE	Anchor	**PML**
thyroid	0.563	0.571	<u>0.572</u>	**0.587**
letter	0.532	**0.661**	0.460	<u>0.654</u>
Ionosphere	0.353	<u>0.496</u>	0.332	**0.584**
pendigits	0.318	<u>0.619</u>	0.598	**0.635**
vowels	0.486	<u>0.583</u>	0.486	**0.622**
Wilt	0.583	**0.677**	<u>0.603</u>	0.574
Average	0.468	0.597	0.503	**0.605**
Improvement	22.64%	1.32%	16.86%	-

respectively. The SHAP and LIME classifiers use SVM with the RBF kernel. We use the subspaces extracted from the learned rules in PARs as the interpretation.

4.2 Experiment Evaluation

Effectiveness Analysis Table 2 shows the performance of PML compared to methods output feature weights [10,12,16]. Table 3 shows the performance of our method compared to methods output feature subspaces [1, 13, 14]. All methods are independently executed 5 times on datasets shown in Table 1. PML demonstrates substantial improvements over baseline models in terms of AUPRC, AUROC, and F1. ATON fails to yield satisfactory results in datasets with high outlier rates, as it only considers the difference between the query outlier and the normal instances, ignoring a number of outliers can also help in interpretation. The interpretation results from SHAP and LIME could be subject to biases. Intriguingly, SHAP exhibits good performance in pendigits. This is because of the analogous nature between classifier explanations and outlier interpretations. LIME also shows its advantages in thyroid with its ability in approximating the outlier locally with an interpretable model. SiNNE reachs suboptimal outcomes in vowels due to the limitation of setting the maximum subspace length and utilizing pruning strategies. Additionally, Anchors is unsuitable for certain datasets such as Inonsphere since the performance of Anchors deteriorates due to variations in data distribution. However, different outlier detectors have different detection methods and evaluation criteria. As a result, the ground-truth annotations generated from the five outlier detectors can not substitute the real interpretation labels.

5 Conclusion

In this paper, we propose a novel model-agnostic outlier interpretation method named PML, which is model-agnostic and does not require domain knowledge. We introduce a multi-task learning strategy which optimizes the task in different perspectives to provide the interpretation for the query outlier. Extensive experiments show that PML achieves significant performance over the state-of-art outlier interpretation methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (62472294).

References

1. Feng, C.: PARs: Predicate-based association rules for efficient and accurate anomaly explanation. In: CIKM. p. 612–621 (2024)
2. Goldstein, M., Dengel, A.: Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm pp. 59–63 (2012)
3. Hardin, J., Rocke, D.M.: Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator **44**(4), 625–638 (2004)
4. Huang, C., Li, Y., Loy, C.C., Tang, X.: Learning deep representation for imbalanced classification. In: CVPR. pp. 5375–5384 (2016)
5. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: CVPR. pp. 7482–7491 (2018)
6. Kriegel, H., Kröger, P., Schubert, E., Zimek, A.: Outlier detection in axis-parallel subspaces of high dimensional data. In: PAKDD. pp. 831–838 (2009)
7. Kriegel, H.P., Schubert, M., Zimek, A.: Angle-based outlier detection in high-dimensional data. In: SIGKDD. pp. 444–452 (2008)
8. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data **6**(1), 1–39 (2012)
9. Lüdtke, S., Bartelt, C., Stuckenschmidt, H.: Outlying aspect mining via sum-product networks. In: PAKDD. pp. 27–38 (2023)
10. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: NeuraIPS. pp. 4765–4774 (2017)
11. Panjei, E., Gruenwald, L., Leal, E., Nguyen, C., Silvia, S.: A survey on outlier explanations. The VLDB Journal **31**(5), 977–1008 (2022)
12. Ribeiro, M.T., Singh, S., Guestrin, C.: " why should I trust you?" explaining the predictions of any classifier. In: SIGKDD. pp. 1135–1144 (2016)
13. Ribeiro, M.T., Singh, S., Guestrin, C.: Anchors: High-precision model-agnostic explanations. In: AAAI. pp. 1527–1535 (2018)
14. Samariya, D., Ma, J.: A new dimensionality-unbiased score for efficient and effective outlying aspect mining. Data Science and Engineering **7**(2), 120–135 (2022)
15. Wang, Y., Wang, X., He, C., Chen, X., Luo, Z., Duan, L., Zuo, J.: Community-guided contrastive learning with anomaly-aware reconstruction for anomaly detection on attributed networks. In: DASFAA. pp. 199–209 (2024)
16. Xu, H., Wang, Y., Jian, S., Huang, Z., Wang, Y., Liu, N., Li, F.: Beyond outlier detection: Outlier interpretation by attention-guided triplet deviation network. In: WWW. pp. 1328–1339 (2021)