گزارش سمینار

نام درس: داده کاوی پیشرفته

استاد درس: دکتر بهروز مینایی

نام دستیار: حسین فنایی

نام: محمد حقیقت

شماره دانشجویی: 403722042

گرایش: هوش مصنوعی

دانشکده: مهندسی کامپیوتر

نیم سال دوم 1404-1403

DTOR: Decision Tree Outlier Regressor to explain anomalies

چکیده (Abstract):

این بخش خلاصه‌ای از مقاله را ارائه می‌دهد. اهمیت توضیح داده‌های پرت را برجسته می‌کند، DTOR را به عنوان یک راه‌حل معرفی می‌نماید، به‌طور خلاصه روش‌شناسی آن (رگرسیون‌ساز درخت تصمیم که امتیازات ناهنجاری را تخمین زده و مسیرها را استخراج می‌کند) را تشریح کرده و به نتایج کلیدی مانند استحکام، اعتبار قاعده و عملکرد قابل مقایسه با Anchors با زمان اجرای سریع‌تر اشاره می‌کند.

مقدمه (Introduction):

تشخیص ناهنجاری در فعالیت حسابرسی داخلی بخش بانکداری:

این بخش به اهمیت حسابرسی داخلی در بانکداری برای حفظ یکپارچگی عملیاتی، مدیریت ریسک‌ها و تشخیص تقلب می‌پردازد. توضیح می‌دهد که چگونه تکنیک‌های تشخیص ناهنجاری برای شناسایی رکوردهای غیرمعمول جهت بررسی ارزشمند هستند و بر نیاز به امتیازات ناهنجاری برای رتبه‌بندی به جای طبقه‌بندی صرفاً باینری تأکید می‌کند. به طور حیاتی، بر ضرورت قابلیت توضیح تأکید می‌کند تا حسابرسان داخلی، که ممکن است متخصص داده نباشند، بتوانند یافته‌ها را درک کرده و بر اساس آن عمل کنند. همچنین به‌طور خلاصه الگوریتم‌های رایج تشخیص ناهنجاری مانند جنگل ایزوله‌سازی، ماشین بردار پشتیبان تک‌کلاسه (One-Class SVM) و مدل‌های ترکیبی گوسی (GMM) را معرفی می‌کند.

هوش مصنوعی قابل توضیح برای تشخیص ناهنجاری:

این زیربخش به نیاز گسترده‌تر به هوش مصنوعی قابل توضیح (XAI) در بانکداری برای شفافیت می‌پردازد. تکنیک‌های موجود XAI مانند SHAP (مستقل از مدل، اهمیت ویژگی) و DIFFI (مختص جنگل ایزوله‌سازی) را بررسی می‌کند. به محدودیت‌های روش‌های مبتنی بر اهمیت ویژگی برای مدل‌های پیچیده اشاره کرده و XAI مبتنی بر قاعده مانند Anchors را به عنوان جایگزین معرفی می‌کند. سپس انگیزه برای DTOR را بیان می‌کند: ایجاد یک چارچوب XAI مستقل از مدل که به‌طور خاص برای تشخیص ناهنجاری طراحی شده و تفاسیر مبتنی بر قاعده ارائه دهد، و به محدودیت‌های Anchors برای وظایف رگرسیون بپردازد. همچنین به کارهای مرتبط مانند LORE و RuleXAI اشاره می‌کند.

روش (Method):

این بخش جزئیات روش‌شناسی پیشنهادی DTOR را شرح می‌دهد. توضیح می‌دهد که DTOR از یک رگرسیون‌ساز درخت تصمیم برای یادگیری امتیازات ناهنجاری از هر نوع مدل تشخیص ناهنجاری استفاده می‌کند. یک مرحله حیاتی، تخصیص وزن بالاتر در تابع زیان به نقطه داده خاصی است که توضیح داده می‌شود تا اطمینان حاصل شود که توضیح محلی دقیق است. سپس قاعده برای یک امتیاز ناهنجاری به عنوان مسیری که نقطه داده در درخت آموزش‌دیده طی می‌کند، استخراج می‌شود (الگوریتم ۱).

Algorithm 1: DTOR method to generate explanations for a given instance.

def *explain_instance*:
input : $(x_{\text{expl}}, \hat{y}_{\text{expl}})$: an instance and its anomaly score to be explained;
$(X_{\text{train}}, \hat{y}_{\text{train}})$: a training dataset and its anomaly scores;
β: relative weight given to x_{expl};
h-args: set of hyperparameters for the Decision Tree Regressor;
output: a set of rules explaining $(x_{\text{expl}}, \hat{y}_{\text{expl}})$
$N \leftarrow \text{len}(X_{\text{train}})$;
$\text{DT} \leftarrow \text{DecisionTreeRegressor}(h\text{-}args)$;
`/* extend the training set with the instance to be explained */`
$X \leftarrow$ concatenate X_{train} with x_{expl};
$y \leftarrow$ concatenate $\hat{y}_{\text{train}}$ with $\hat{y}_{\text{expl}}$;
`/* define Decision Tree loss weights assigning more importance to the instance to be explained */`
$w \leftarrow$ concatenate $\mathbf{1}_N$ with β;
`/* fit the decision tree to the 'weighted' training set */`
$\text{DT.fit}((X, y), \text{sample\_wights}{=}w)$;
`/* extract the path of the fitted DT followed by` x_{expl} `*/`
$\text{rules} \leftarrow \text{extract\_path}(\text{DT}, x_{\text{expl}})$;
return rules

این بخش معیارهای کلیدی برای ارزیابی قواعد را تعریف می‌کند: دقت (precision) (قاعده چقدر خروجی مدل اصلی را برای نمونه‌های مشابه پیش‌بینی می‌کند)، پوشش (coverage) (چه نسبتی از داده‌ها قاعده را برآورده می‌کنند)، و اعتبار (validity) (آیا نمونه مورد توضیح، قاعده را برآورده می‌کند).

Algorithm 2: DTOR method to generate precision sampling from a generated synthetic dataset. $A[k]$ denotes the k-th predicate of rule A.

def *compute_precision*:

input : $(x_{\text{expl}}, \hat{y}_{\text{expl}})$: an instance and its anomaly score to be explained;
$(X_{\text{train}}, \hat{y}_{\text{train}})$: a training dataset and its anomaly scores;
N_gen: number of synthetic samples to generate;
AD: the anomaly detector model, outputs a score;
t: threshold for the anomaly score;
A: DTOR rule explaining $(x_{\text{expl}}, \hat{y}_{\text{expl}})$ as by algorithm 1;

output: precision score for the rule A in explaining $(x_{\text{expl}}, \hat{y}_{\text{expl}})$

$X \leftarrow$ concatenate X_{train} with x_{expl};

$y \leftarrow$ concatenate $\hat{y}_{\text{train}}$ with $\hat{y}_{\text{expl}}$;

value-grid $\leftarrow \{\}$; `/* initialize the dictionary containing the values per each feature of the rule A */`

ids-cond $= [\,]$; `/* initialize the list containing the indices that satisfy at least one sub-rule A[k] */`

`/* for loop per each feature in the rule. An example of the rule A could be: "feature_1 > 5 AND feature_4 < 0" */`

for *k in features in rule A*:

`/* select the samples that satisfy the partial rule for the specific feature, e.g., "feature_1 > 5". Note that there can be multiple condition on the same feature can appear, in this case all should go in AND condition */`

$X_{\text{cond}} \leftarrow$ elements of X satisfying $A[k]$;

ids-cond $\leftarrow$ concatenate ids-cond with the ids of X_{cond};

`/* store the values compatible with sub-rule A[k] */`

if *k is categorical*:

value-grid$[k] \leftarrow$ unique value of feautre k in X_{cond};

else:

value-grid$[k] \leftarrow$
$[\min(X_{\text{cond}}[k]), \text{quantile}_{1/4}(X_{\text{cond}}[k]), \text{quantile}_{1/2}(X_{\text{cond}}[k]), \text{ave}(X_{\text{cond}}[k]),$
$\text{quantile}_{3/4}(X_{\text{cond}}[k]), \max(X_{\text{cond}}[k])]$;

`/* Give precedence to ids that satisfy multiple sub-rules in A */`

ids-cond $\leftarrow$ sort ids-cond by frequency;

if *length of ids-cond* $< N_{gen}$:

ids-cond $\leftarrow$ concatenate ids-cond with random ids of X to fill the N_{gen} samples;

$X_{\text{synth}} \leftarrow$ first N_{gen} rows of X[ids-cond];

`/* fix all values of predicates in A of` X_{synth} `so that all samples in` X_{synth} `satisfy rule A */`

for *k in features in rule A*:

$X_{\text{synth}}[k] \leftarrow$ sample N_{gen} times from value-grid$[k]$;

if $\hat{y}_{expl} < t$: `/* If the sample to explain is an outlier */`

precision $= \text{ave}\,(\text{AD}(X_{\text{synth}}) < t)$;

else:

precision $= \text{ave}\,(\text{AD}(X_{\text{synth}}) >= t)$;

return precision

همچنین یک رویه نوین (الگوریتم ۲) برای تولید یک مجموعه داده مصـــنوعی جهت محاســـبه معیار دقت توصـیف شـده اسـت که هدف آن حفظ بهتر همبسـتگی‌ها در مقایسـه با روش Anchors اسـت. شکل ۱ این تولید داده مصنوعی را نشان می‌دهد.

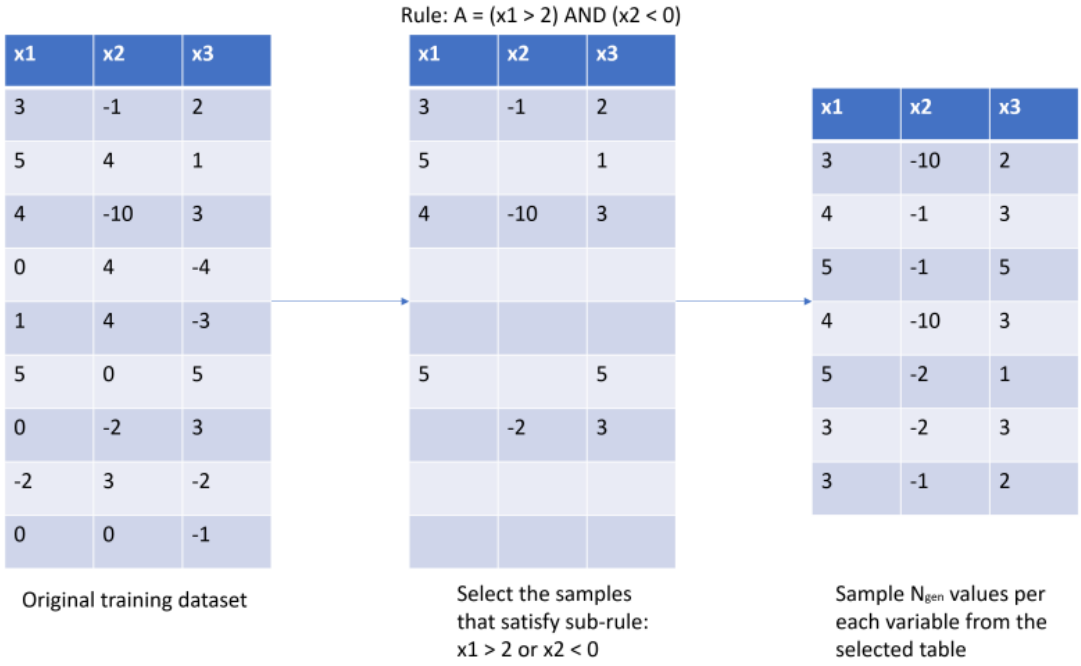

Fig. 1: A simplified illustration of synthetic data generation is presented. Initially, samples from the original dataset are selected based on sub-rules (e.g., $x_1 > 2$ or $x_2 < 0$ in the given example). Subsequently, N_{gen} samples are drawn for each variable to satisfy the overarching rule A. Notably, the image does not depict the discretization of continuous variables or the preservation of inter-variable correlations. However, for illustrative purposes, it is evident that negative values of x_3 do not occur under rule A, as observed in the synthetic dataset.

آزمایش‌ها (Experiments)

این بخش تنظیمات آزمایشگاهی برای ارزیابی DTOR را توصیف می‌کند.

مجموعه داده‌ها و مدل‌های تشخیص ناهنجاری: مجموعه داده‌های مورد استفاده را لیست می‌کند: شش مجموعه داده عمومی (Ionosphere، Glass Identification، lymphography، musk v2، breast cancer wisconsin diagnostic، arrhythmia) و یک مجموعه داده خصوصی بانکی از اینت3a سانپائولو (جدول ۱).

Table 1: Dataset Details: Each row provides information about a specific dataset, including its identifier (Dataset ID), the number of samples (Num samples), the number of features (Num features), and a brief description (Description). The datasets are sourced from the UCI Machine Learning Repository [1].

Dataset ID	# samples	# features	Description
banking	100,000	26	Banking dataset from Intesa Sanpaolo for identifying anomalies and better analyze the client for possible fraud or criminal behavior.
Ionosphere	351	34	Classification of radar returns from the ionosphere.
Glass Identification	214	9	From the USA Forensic Science Service, this dataset comprises six types of glass, each defined in terms of their oxide content, including Na, Fe, K, and others.
lymphography	148	19	This lymphography domain was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia.
musk v2	6,598	168	The goal is to learn to predict whether new molecules will be musks or non-musks.
breast cancer wisconsin diagnostic	198	33	Prognostic Wisconsin Breast Cancer Database.
arrhythmia	452	279	Distinguish between the presence and absence of cardiac arrhythmia and classify it into one of the 16 groups.

مدل‌های تشخیص ناهنجاری (AD) انتخاب شده را مشخص می‌کند: جنگل ایزوله‌سازی (IF)، ماشین بردار پشتیبان تک‌کلاسه (One-Class SVM) و مدل‌های ترکیبی گوسی (GMM)، عموماً با پارامترهای پیش‌فرض. جزئیات مربوط به ایجاد مجموعه آزمون، تعریف ناهنجاری و پیش‌پردازش داده‌ها (استانداردسازی برای GMM) ارائه شده است. همچنین در مورد انتخاب اعمال توضیحات بر روی فضای ورودی اصلی برای تفسیرپذیری بهتر بحث می‌کند (شکل ۲).

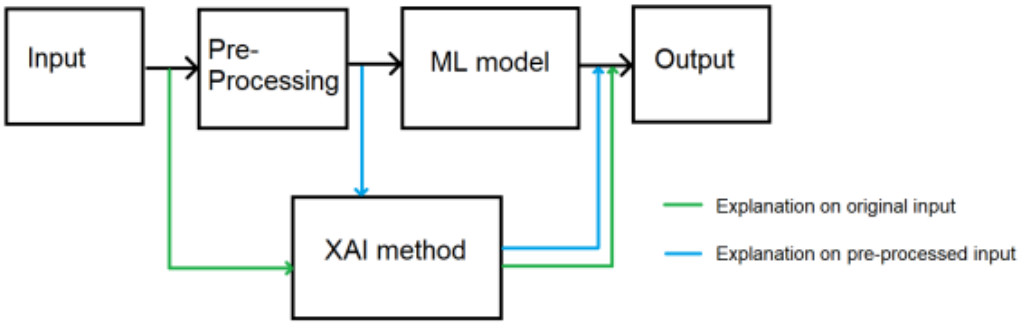

Fig. 2: Illustration of a machine learning application where the XAI method can provide explanations either in the original input space or the pre-processed one. If the latter option is chosen, the explanation must be converted back into the original feature space, particularly when a rule-based explanation is expected.

هوش مصنوعی قابل توضیح مبتنی بر قاعده: این زیربخش بر مقایسه با Anchors تمرکز دارد. دلایلی که چرا سـایر روش‌های مبتنی بر قاعده مانند LORE و RuleXAI اسـتفاده نشـده‌اند (قابلیت اسـتفاده، نگهداری) ذکر شده است. معیارهای مقایسه را تعریف می‌کند: زمان اجرا، دقت، پوشش، اعتبار و طول قاعده. انتخاب‌های فراپارامترها برای DTOR و Anchors برای اطمینان از مقایسه منصفانه مورد بحث قرار گرفته‌اند. جدول ۲ نمونه‌هایی گویا از قواعد تولید شده توسط Anchors و DTOR را ارائه می‌دهد.

Table 2: Illustrative examples of Isolation Forest explanation on the Lymphography Dataset. The table showcases three dataset samples along with the rules derived from Anchors and DTOR. Each sample is denoted by its corresponding dataset row index, indicating its position within the dataset. While some explanations exhibit similarities, others differ; notably, Anchors often fail to provide any explanation, as observed in example ID 0.

Index row sample	Anchors Rule	DTOR Rule
36	feature_9 > 1 AND feature_7 > 1	feature_9 > 2 AND feature_7 > 1.5
139	feature_8 <= 1 AND feature_7 > 1	feature_0 > 3.5 AND feature_12 <= 3
0	-	feature_4 <= 1.5 AND feature_15 <= 1.5 AND feature_10 > 1.5 AND feature_13 > 3.5

بحث و نتیجه‌گیری (Discussion and conclusion):

این بخش نتایج آزمایشگاهی (عمدتا از جدول ۳) را تفسیر کرده و مقاله را به پایان می‌رساند.برجسته می‌کند که DTOR به طور کلی عملکردی مشـابه یا بهتر از Anchors دارد، با کشـف قواعد به طور قابل توجهی سـریع‌تر و امتیازات اعتبار بالاتر، به‌ویژه برای مدل‌های IF و GMM. به تمایل Anchors به عدم ارائه توضـیح برای برخی نمونه‌ها یا داشـتن مشـکل با مجموعه داده‌های خاص (مانند arrhythmia) اشاره می‌کند. یک موازنه برای مدل‌های SVM ذکر شده است. بحث خاطرنشان می‌کند که Anchors بیشـــتر تمایل به توضـــیح نقاط ناهنجار در مجموعه داده‌های نامتعادل دارند، در حالی که DTOR همچنین نقاط "نرمال" را به‌طور مؤثر توضـــیح می‌دهد. این بخش مجدداً تأکید می‌کند که DTOR به عنوان یک رگرسـیون‌سـاز جایگزین عمل می‌کند و امکان توضـیحات دقیق‌تری مانند میانگین امتیازات ناهنجاری برای یک قاعده را فراهم می‌آورد. مقاله همچنین بر رویکرد نوین خود در محاســـبه دقت قاعده با تولید همسـایگی‌ای که سـعی در حفظ همبسـتگی‌های داده دارد، تأکید می‌کند. نتیجه‌گیری، DTOR را به عنوان یک تکنیک XAI ساده، مؤثر و همه‌کاره برای تشخیص ناهنجاری تأیید می‌کند.

		Isolation Forest		GMM		SVM	
		Anchors	**DTOR**	**Anchors**	**DTOR**	**Anchors**	**DTOR**
banking	Exec. time	**13.17 (19.20)**	24.88 (32.24)	17.39 (36.86)	**3.79 (5.98)**	15.44 (23.75)	**14.18 (19.58)**
	Precision	0.54 ± 0.03	**0.90 ± 0.18**	**0.64 ± 0.09**	0.31 ± 0.14	0.55 ± 0.04	**0.68 ± 0.26**
	Coverage	0.00 ± 0.00	**0.15 ± 0.17**	0.12 ± 0.09	**0.92 ± 0.23**	**0.80 ± 0.21**	0.30 ± 0.25
	Validity %	6	**100**	6	**100**	**100**	**100**
	Rule length	0.40	6.14	0.16	8.00	1.00	8.00
glass	Exec. time	14.75 (20.94)	**3.64 (5.13)**	**0.13 (0.15)**	2.21 (3.51)	**0.76 (1.19)**	2.49 (3.92)
	Precision	0.16 ± 0.08	**0.89 ± 0.20**	**0.74 ± 0.03**	0.17 ± 0.19	0.61 ± 0.10	**0.74 ± 0.20**
	Coverage	0.01 ± 0.01	**0.10 ± 0.12**	0.26 ± 0.00	**0.33 ± 0.28**	**0.53 ± 0.23**	0.14 ± 0.14
	Validity %	6	**100**	4	**100**	**100**	**100**
	Rule length	0.26	5.42	0.04	7.70	1.00	6.96
ionosphere	Exec. time	169.40 (275.22)	**11.25 (13.61)**	15.53 (28.24)	**3.19 (5.03)**	5.76 (8.03)	**3.85 (6.62)**
	Precision	0.73 ± 0.11	**0.91 ± 0.21**	0.64 ± 0.04	**0.86 ± 0.27**	0.54 ± 0.04	**0.61 ± 0.23**
	Coverage	0.00 ± 0.00	**0.08 ± 0.07**	0.02 ± 0.01	**0.18 ± 0.17**	**0.62 ± 0.31**	0.06 ± 0.07
	Validity %	8	**100**	12	**100**	**100**	**100**
	Rule length	0.90	6.20	0.40	7.56	1.00	7.74
lymphography	Exec. time	16.94 (40.04)	**3.90 (5.72)**	4.20 (7.98)	**2.36 (3.72)**	**1.07 (1.35)**	2.70 (4.40)
	Precision	0.51 ± 0.37	**0.89 ± 0.20**	0.77 ± 0.26	**0.83 ± 0.15**	0.56 ± 0.03	**0.70 ± 0.22**
	Coverage	0.01 ± 0.01	**0.08 ± 0.07**	0.05 ± 0.03	**0.07 ± 0.08**	**0.68 ± 0.23**	0.03 ± 0.03
	Validity %	8	**100**	8	**100**	**100**	**100**
	Rule length	0.54	6.50	0.28	7.06	1.00	7.10
musk	Exec. time	5273.15 (5273.15)	**11.92 (16.45)**	**4.97 (5.19)**	11.05 (18.60)	472.87 (1374.55)	**15.54 (27.80)**
	Precision	**0.99 ± 0.00**	0.92 ± 0.20	**0.97 ± 0.02**	0.21 ± 0.26	**0.94 ± 0.03**	0.87 ± 0.15
	Coverage	0.01 ± 0.00	**0.18 ± 0.11**	0.03 ± 0.02	**0.10 ± 0.12**	**0.27 ± 0.16**	0.05 ± 0.08
	Validity %	2	**100**	8	**100**	**100**	**100**
	Rule length	0.42	3.78	0.16	7.82	1.72	7.96
breast cancer	Exec. time	108.08 (108.08)	**4.49 (6.78)**	**0.50 (0.50)**	2.85 (4.65)	4.80 (9.64)	**3.67 (6.16)**
	Precision	0.54 ± 0.00	**0.96 ± 0.13**	**0.73 ± 0.00**	0.30 ± 0.34	0.59 ± 0.07	**0.86 ± 0.18**
	Coverage	0.02 ± 0.00	**0.12 ± 0.10**	**0.25 ± 0.00**	**0.25 ± 0.21**	**0.63 ± 0.15**	0.06 ± 0.10
	Validity %	2	**100**	2	**100**	**100**	**100**
	Rule length	0.20	6.62	0.02	7.98	1.00	7.90
arrhythmia	Exec. time	-	15.98 (26.03)	-	9.70 (17.18)	-	13.19 (24.46)
	Precision	-	0.91 ± 0.26	-	0.65 ± 0.44	-	0.54 ± 0.34
	Coverage	-	0.23 ± 0.22	-	0.02 ± 0.06	-	0.49 ± 0.32
	Validity %	-	100	-	100	-	100
	Rule length	-	6.32	-	3.70	-	7.80

Table 3: Summary of the experiments over seven datasets, three anomaly detectors, and the two compared XAI techniques: Anchors and DTOR. The best performing value between the two methods is highlighted in bold. Execution time refers to the average time (maximum time over the test set in parenthesis). Precision and coverage are reported with standard deviation over the test set. Validation is reported as a percentage. Rule length is the average length over the explanation rules found for the test set.