A Data Distillation Enhanced Autoencoder for Detecting Anomalous Gas Consumption

Yujue Zhou, Jie Jiang, Shuang-Hua Yang, *Senior Member, IEEE,* Ligang He, *Member, IEEE,* Yulong Ding, Kai Liu, Guozhong Zhu, Yali Qing

***Abstract*—The number of natural gas users has been growing rapidly in China due to the promotion of clean energy and the economic benefits of natural gas, especially in businesses and industries. Though the infrastructures for gas supplies have been highly improved, gas providers are still suffering from various problems such as malfunctioning gas meters, gas leakage, gas theft etc. With the development of the Internet of Things, smart gas meters have been widely adopted by gas providers to collect real-time gas consumption data for billing purposes which can also serve as a basis for anomaly detection. One challenge of using such data for anomaly detection is that it is difficult to obtain sufficient labelled data for model training. To address this challenge, we propose D^{2}AE, a data distillation enhanced autoencoder for detecting anomalous gas consumption, which consists of three modules. The first module preprocesses the raw meter readings and carries out a rule-based anomaly detection. The second module extracts the normal gas usage patterns via an integration of correlation and clustering based consistency evaluation methods. The extracted normal usage patterns are then used in the third module to train an autoencoder for anomaly detection. D^{2}AE intends to provide a method to detect anomalous gas consumption induced by various causes such that manual inspection can be largely reduced. Moreover, D^{2}AE does not require user-specific information and can be applied to different types of gas users. Based on a real-world gas consumption dataset, we carry out a set of experiments and show that D^{2}AE outperforms the existing and improves the F_1 score by an average of 7.4% for restaurant users and 5.7% for canteen users.**

***Index Terms*—Anomaly Detection, Unsupervised Learning, Gas Consumption, Time Series, Urban Computing**

Yujue Zhou is with National Pilot School of Software, Yunnan University, Kunming 650091, China. E-mail: yujue_zhou@163.com

Jie Jiang is with the Department of Artificial Intelligence, China University of Petroleum-Beijing, 18 Fuxue Road, Beijing 102249, P.R. China. E-mail: jiangjie@cup.edu.cn

Shuang-Hua Yang is with Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, Southern University of Science and Technology, Shenzhen, China, and also with the Department of Computer Science, University of Reading, UK. E-mail: yangsh@sustech.edu.cn

Ligang He is with the Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK. E-mail: ligang.he@warwick.ac.uk

Yulong Ding and Kai Liu are with Shenzhen Key Laboratory of Safety and Security for Next Generation of Industrial Internet, and the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China. E-mail: dingyl@sustech.edu.cn, 12032502@mail.sustech.edu.cn.

Guozhong Zhu and Yali Qing are with Shenzhen Gas, Shenran Building, No. 268, Mei'ao 1st Road, Shenzhen 518000, P.R. China. E-mail: zhuguoz,qingyl@szgas.com.cn

Corresponding author: Jie Jiang (jiangjie@cup.edu.cn) and Shuang-Hua Yang (yangsh@sustech.edu.cn).

This research is supported in part by the National Natural Science Foundation of China (Grant No. 92067109, 61873119), in part by Key Laboratory of Shenzhen(Grant No. ZDSYS20210623092007023), in part by the Science and Technology Planning Project of Guangdong Province(Grant No. 2021A0505030001), in part by the Educational Commission of Guangdong Province (Grant No. 2019KZDZX1018), and in part by the Department of Education of Guangdong Province (Grant No. 2019KQNCX132).

I. Introduction

The number of natural gas users, especially in businesses and industries, has been growing rapidly in China due to the promotion of clean energy and the economic benefits of natural gas. According to the National Bureau of Statistics of China, the natural gas consumption in China in 2020 exceeded 325.9 billion cubic meters [1]. The rapid expansion of gas distribution networks also brings an increase of problems such as malfunctioning gas meters, gas leakage, gas theft etc., which not only lead to economic losses but also endanger the public safety.

For detecting anomalous gas consumption, regular on-site inspections and annual gas meter check-ups are carried out, which are not only labour-intensive but also lack real-time feedback. Thanks to the development of the Internet of Things, smart gas meters have been widely adopted by gas providers to collect real-time gas consumption data for billing purposes which can also serve as a basis for anomaly detection. Recently several data-driven methods have been proposed for detecting anomalous gas consumption based on such digitised systems in the literature [2], [3].

However, these methods are limited to specific types of gas users such as restaurants [2] and boiler rooms [3] and focus on gas theft detection. Furthermore, the method proposed in [3] requires detailed user-specific information to be collected such as the number of gas equipment. The method in [2] adopts a supervised learning approach. Therefore, it requires labelled anomalies for model training. Such data can be difficult to be collected as they are extremely sparse in real-world situations.

To these ends, this paper proposes D^{2}AE, a data distillation enhanced autoencoder which can detect anomalous gas consumption induced by various causes such as ageing gas meters, ageing gas pipes, data transmission failures, gas theft etc. D^{2}AE is composed of three modules. The first module preprocesses the raw meter readings and carries out a rule-based anomaly detection to identify reading-level anomalies. The second module extracts the normal gas usage patterns to distil the training data via an integration of correlation and clustering based consistency evaluation methods from an intra-user as well as an inter-user perspectives. Thereafter, with the extracted normal gas usage patterns, the third module trains an autoencoder for pattern-level anomaly detection. In particular, D^2AE does not require user-specific information and can be applied to different types of gas users. Based on a real-world gas consumption dataset, we compare D^2AE with a set of anomaly methods proposed in the literature. Experiment results show that D^2AE achieves the best performance in terms of precision, recall and F_1 score of the top N ranked anomalous samples returned by the anomaly detection methods.

In summary, the contributions of this paper are as follows:

- We propose a data distillation enhanced autoencoder D^2AE that can be used for detecting anomalous gas consumption induced by various causes such as (un)intentional damages of the gas meters, gas leakage, gas theft and etc. D^2AE distils the training samples via an integration of intra-user and inter-user consistency evaluations, such that the purity of the normal samples can be improved for the training of the autoencoder based anomaly detection model and the problem of unlabeled data is alleviated.
- Different from the existing data-driven approaches which require detailed user-specific information to be collected such as the number of gas-powered devices, the application of D^2AE only relies on the availability of gas consumption data from the users of the same category which are widely available due to the employment of smart gas meters.
- We evaluate the effectiveness of D^2AE based on a real-world gas consumption dataset collected from two categories of gas users, i.e., restaurant and canteen users. We also release the dataset (anonymized) and the source code[1] to facilitate the reproducibility of our work. To the best of our knowledge, there is no public dataset available for the research of gas consumption anomaly detection in the literature.
- We develop an anomaly detection system based on D^2AE, providing the anomaly list weekly for a gas company situated in southern China. The anomaly list can support the staff of the gas company to carry out efficient on-site inspections in contrast to fixed schedules of on-site inspections.

The rest of the paper is organised as follows: Section II discusses the related work. Section III illustrates the anomaly detection problem investigated in this study. Section IV presents the proposed anomalous gas consumption detection method D^2AE. Section V describes the dataset used in the experiments. Sections VI illustrates the evaluation metrics, the baseline methods, variants of D^2AE for the ablation study, parameter setting, and analyses of the experiment results. Section VII introduces the anomaly detection system which has been developed based on D^2AE and presents a few case studies from the detection results. Finally, in Section VIII, we conclude the paper with the possibilities of future work.

[1]https://github.com/yujue-zhou/Gas_Anomaly_Detection

II. Related Work

A. Data-Driven Gas Anomaly Detection

In the literature, early research mainly focuses on gas consumption prediction and anomaly detection at a large scale such as a building, a village, or a city. For example, with gas flow and outdoor temperature data, Baldacci et al. [4] proposed a gas consumption forecasting method for anomaly detection based on a nearest neighbour approach and local regression analysis. Akouemo et al. [5] proposed a two-stage method for detecting anomalies in gas time series data. The first stage determines the probability that a data point is anomalous based on a linear regression model and a geometric probability distribution of the residuals. Based on the anomaly types identified in the previous stage, the second stage trains a Bayesian maximum likelihood classifier to identify the anomalies. Recently, researchers investigated data-driven methods for gas theft detection [2], [3], [6], [7], which are closely related to this study. In [2], Yang et al. proposed a method based on normal user modeling and RankNet for detecting gas theft suspects among restaurant users. Though this method provides a mechanism to differentiate unstable users, it still relies on the availability of labelled anomalies for model training and thus cannot apply to situations where labelled data are difficult to obtain. In [3], Yi et al. proposed a method based on One-Class SVM (OCSVM) for detecting gas theft suspects among boiler room users. Though this method adopts an unsupervised approach, it can only be applied to boiler room users as it is based on the strong correlation between gas consumption and outdoor temperature. Moreover, this method relies on the availability of user-specific information such as heating area, number of boilers etc.

In [6], Pan et al. proposed a neural clustering and ranking approach to detect gas-theft suspects of boiler room users. Similarly, this method relies on labelled abnormal data for model training, and is limited to the detection of gas-theft suspects among boiler room users. In [7], Xu et al. proposed an anomaly detection model based on LSTM-WGAN for gas load. Although this method can be used for different types of gas users, it requires that the training data does not contain anomalies. Moreover, this method relies on the availability of user specific information such as building type, and industry type, temperature, weather conditions, wind force, and average air pressure.

B. Time Series Anomaly Detection

Another line of relevant research is on time series anomaly detection, which has received a lot of attention in the literature [8]. According to Lai et al. [9], anomaly detection methods for time series data can be divided into three categories: prediction deviation, discords analysis, and majority modeling. Prediction deviation identifies anomalies by the difference between the predicted value and the true value. Autoregressive (AR) [10] assumes that the data at each moment is linearly related to the data at several past moments. The methods based on recurrent neural networks (RNNs) [11], [12] can model the nonlinear time correlation between data points in order to better simulate the variability that exists in complex variable systems. The assumption of this type of methods is that if the predicted value at a certain moment deviates largely from the original value, it may be an anomaly. Discords analysis identifies discordance as an anomaly through the similarity between sub-sequences. For example, Matrix profile (MP) [13], [14] constructs distance profiles by calculating the minimum distance between each sub-sequence and the rest of the sub-sequences to identify abnormal sub-sequences. Majority modeling is to identify the decision boundary between anomalous and normal values by modeling the distribution of normal data. It assumes that normal data is compact in hyperspace. For example, OCSVM [15], [16] maximises the boundary between the origin and normality, and defines the decision boundary as the hyperplane that determines the boundary. Recently, autoencoders have been widely adopted for time series anomaly detection [17], [18]. They map the data to the low-dimensional latent space and reconstruct the data through the representation of the latent space [19]–[21]. The decision criteria is defined based on the assumption that the reconstruction error of anomalies is significantly larger than the normal samples. Generative Adversarial Neural Networks (GAN) [22] uses generators and discriminators to perform min-max optimization, and the decision criterion is also determined by the reconstruction error. Based on graph convolution, Chen et al. [23] proposed a framework for multivariate time-series anomaly detection that can learn the topology between sensors.

In this study, we adopt autoencoders as a basis for building unsupervised models for detecting anomalous gas consumption. In particular, we enhance the training of autoencoders with an intra-user and inter-user consistency evaluation methods, which improves the anomaly detection accuracy.

III. Problem Statement

Given a time series of gas consumption indicated as $x = (x_1, x_2, \ldots, x_T)$, this study aims to detect whether the time series contains anomalies without labelled training data. Table I summarizes the notation used in the paper.

The causes of gas anomalies vary, which include but are not limited to: (1) ageing gas meters may malfunction and generate anomalous readings, (2) users may accidentally or intentionally damage the gas meters in some cases, resulting in anomalous readings, (3) problems may also occur in the data transmission process and lead to anomalous readings being transferred, (4) gas leakage is another cause of anomalies, which is often related to ageing gas pipes, (5) the cases of gas theft are also a cause of anomalies that have recently been investigated in the literature [2], [3].

Although there are various causes for gas anomalies, this study is based on the assumption that the anomalies are rare compared to the normal cases and the patterns of anomalous gas consumption deviate from the patterns of normal gas consumption. This assumption theoretically guarantees to separate the anomalous patterns from the normal patterns. In reality, however, gas consumption patterns, both normal and anomalous, are highly variable. Some anomalies can be detected by examining a few data points while the detection of some other anomalies require sophisticated pattern matching techniques. This poses challenges to establish a comprehensive anomalous gas consumption detection model that can cover various anomalous scenarios.

TABLE I
NOTATIONS TABLE

Notation	Explanation
$x = (x_1, x_2, \ldots, x_T)$	a time series of gas consumption
$X = \{x^1, x^2, \ldots, x^M\}$	the set of weekly gas consumption samples
M	the total number of weekly gas consumption samples
$x^i = (x_1^i, x_2^i, \ldots, x_T^i)$ $(1 \leq i \leq M)$	the ith week data sample
T	the length of one weekly sample
x_t^i $(1 \leq t \leq T)$	the difference between the meter reading at time step t and $t-1$
$\mathbb{1}(\cdot)$	indicator function
D	the number of gas consumption data collected daily
$max(\cdot)$	maximum function
ψ	the average of the daily maximum gas consumption over a week
c_1, c_2	hyper-parameters in detecting bursty consumption anomalies
$x^i(s)$	gas consumption of the ith week of user s
$r^i(s)$	self correlation coeffient of the ith week sample of user s
$\rho(\cdot, \cdot)$	pearson correlation coeffient
$cov(\cdot, \cdot)$	covariance
σ	standard deviation
$\bar{x}^i$	the normalization of x^i
$e(\cdot)$	permuation entropy
α_o $(1 \leq o \leq J)$	the intra-cluster distance of cluster o
J	the number of clusters
$\mu_{\text{-}o}$	the mean of the intra-cluster distance of all the clusters excluding o
$\sigma_{\text{-}o}$	the standard deviation of the intra-cluster distance across all the clusters excluding o
β_i	the distance between an instance i and its cluster centroid
μ_d	the mean of the distances between the set of instances contained in the same cluster as i and the corresponding cluster centroid (i.e. intra-cluster distance of the cluster)
P	the total number of instances in the cluster
σ_d	the standard deviation of the distances between the instances and the corresponding cluster centroid
$\varepsilon(\cdot, \cdot)$	cosine similarity
$\mathcal{L}$	reconstruction error

IV. Methodology

In this section, we describe the proposed method D^{2}AE for detecting anomalous gas consumption. D^{2}AE provides a generic and unsupervised approach for anomaly detection across all types of non-residential gas users. As shown in Figure 1, D^{2}AE consists of three modules. The first module is *preprocessing & rule-based anomaly detection*, which converts the raw meter readings into weekly gas consumption data and further applies a set of rules to detect reading-level anomalies. Through the first module, both the training set and the test set are divided into a set of reading-level normal weekly sequences and a set of abnormal sequences. The set of reading-level normal weekly sequences from the training set are then sent to the second module, namely *normal usage pattern extraction*. This module distils the training data and removes the samples deviating from the normal usage patterns based on an integration of correlation and clustering based consistency evaluation methods. Through the second module, the set of reading-level normal weekly sequences from the training set

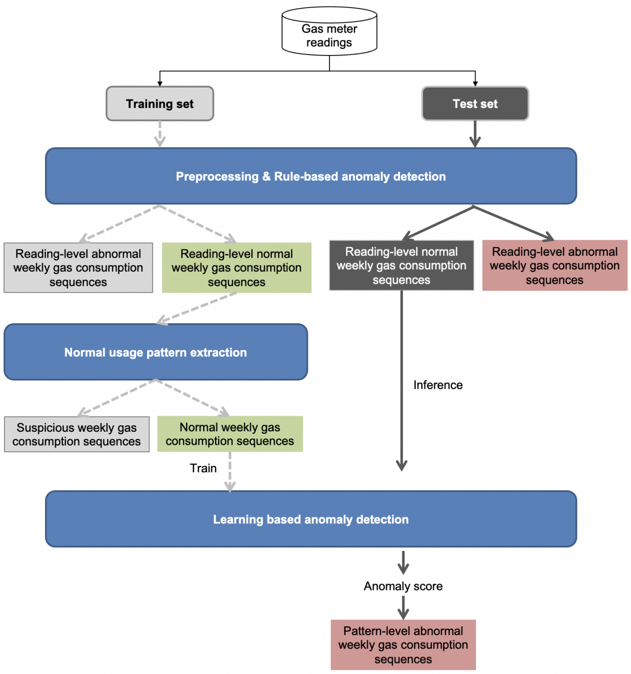

Fig. 1. The framework of D^{2}AE

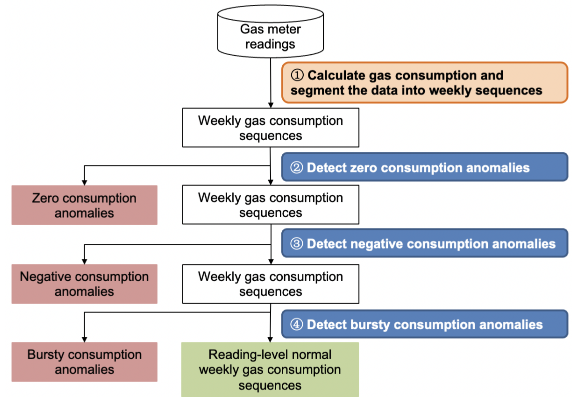

Fig. 2. Workflow of the preprocessing & rule-based anomaly detection module

is further divided into a set of suspicious sequences and a set of normal sequences. The third module, namely *learning based anomaly detection*, trains an autoencoder with the set of normal weekly sequences extracted from the second module to detect pattern-level anomalous gas consumption. During the training phase of D^{2}AE, the training data passes through each of the three modules, resulting in a trained autoencoder. During the testing phase, only the first and third modules are used for anomaly detection. The test data go through the first module for data preprocessing and rule-based anomaly detection. Then, pattern-level anomaly detection is performed through the trained autoencoder in the third module. The anomalies detected from the test set are indicated by the two rectangular boxes highlighted in red in Figure 1.

A. *Preprocessing & Rule-based Anomaly Detection*

Preprocessing & rule-based anomaly detection first convets the raw meter readings into weekly gas consumption data. Furthermore, based on expert knowledge, a set of rules is formulated to detect reading-level anomalies that may be caused by data collection and transmission failures or gas theft. The workflow of this module is shown in Figure 2. It consists of four steps. Step 1 is for data preprocessing, and the rest three steps are for rule-based anomaly detection.

The original data samples are meter readings collected from the gas meters with a fixed sampling rate and the measurement is in cubic meters m^3. As gas consumption has a strong weekly cycle [24], we follow the previous research on anomaly detection of gas consumption [2] and segment the meter readings on a weekly basis to obtain the weekly gas consumption sequences, indicated as $X = \{x^1, x^2, \ldots, x^M\}$ where $x^i = (x^i_1, x^i_2, \ldots, x^i_T)$, $1 \leq i \leq M$ represents one weekly sample. M represents the total number of weekly gas consumption data samples and T represents the length of the data samples. In specific, x^i records the amount of gas consumed by a gas user from Monday to Sunday during a week, and x^i_t, $1 \leq t \leq T$ is derived by the difference between the meter readings at time step t and $t-1$. In this study, as the gas meter readings are sampled hourly, thus $T = 168$.

Based on the weekly gas consumption sequences, a set of rules is formulated with the help of domain experts. For example, it is rather rare that the gas consumption of an active gas user over a whole week is zero, i.e., there is no gas consumption. Such a phenomenon is likely to occur when the gas meter has a mechanical failure due to ageing or the gas meter has been damaged intentionally. Therefore, the following rule is formulated to detect zero consumption anomalies:

$$\sum_{t=1}^{T} |x^i_t| = 0 \tag{1}$$

In real applications, the rule for detecting zero consumption anomalies can be adjusted to avoid false alarms. For example, during holidays such as the lunar new year, it is possible that some gas users shut down their businesses and the gas consumption stays at zero for more than a week.

Another type of anomalies that can be easily recognised from the gas consumption data is the existence of negative consumption, i.e., the meter readings from a gas meter decrease. Such anomalies are considered as an important indicator for gas theft. Therefore, the following rule is formulated to detect negative consumption anomalies:

$$\sum_{t=1}^{T} \mathbb{1}(x^i_t < 0) > 0 \tag{2}$$

where $\mathbb{1}(\cdot)$ is an indicator function which returns 1 if the condition holds, and 0 otherwise.

The daily gas consumption of a gas user usually stays within a certain range, i.e., the maximum consumption of each day during a week should not diverge too much. In contrast to negative consumption anomalies, there are cases where the gas consumption of a gas user increases dramatically within a week. This may be caused by the malfunctions of the gas meter or the data transmission module. Therefore, the following rule is formulated to detect bursty consumption anomalies:

$$\sum_{t=1}^{T} \mathbb{1}(x_t^i > max(c_1\psi, c_2)) > 0 \tag{3}$$

$$\psi = \frac{\sum_{k=1}^{T/D} \max_{t=1}^{D}(x_{t+D*(k-1)})}{T/D} \tag{4}$$

where D indicates the number of data samples collected daily, ψ represents the average of the daily maximum gas consumption over a week, c_1 and c_2 are two hyper-parameters. In this study, as the sampling rate of the meter readings is 1 hour, thus $D = 24$ and $T/D = 7$.

This rule indicates that when the gas consumption at a sampling point during a week is greater than c_1 times the average daily maximum consumption of that week, and the consumption at that sampling point is greater than c_2 m^3, it is considered as a bursty anomaly. In this study, c_1 is set to 5 and c_2 is set to 10 based on the knowledge of domain experts.

Through the above four steps of preprocessing and rule-based anomaly detection, D^{2}AE converts the raw meter readings into weekly gas consumption sequences and identifies three types of reading-level anomalies by encoding the knowledge of the domain experts. The anomalies detected from this module are only a part of the final results of D^{2}AE. Moreover, this module also helps prepare the input data for the subsequent modules which target more complex pattern-level anomalies.

B. Normal Usage Patterns Extraction

The heuristic rules introduced in Section IV-A are devised by the domain experts. They only cover a small fraction of the anomaly cases observed in the dataset. To detect more types of anomalies, a promising solution is to resort to machine learning approaches which can automatically extract knowledge from the gas consumption data to assist anomaly detection. However, as the data samples of anomalous gas consumption are rather rare and it is not possible to obtain sufficient labelled samples for training an anomaly detection model in a supervised manner. In the literature, unsupervised machine learning methods such as OCSVM and autoencoder have been widely applied to anomaly detection of time series [3], [16]. These methods often rely on the availability of normal data samples and a mixture of abnormal data samples can heavily influence the detection accuracy [18], [25]. Thus distilling the training data to obtain normal data samples for model training will improve the model performance.

To this end, the second module *normal usage patterns extraction* adopts an integration of correlation and clustering methods to obtain the weekly gas consumption sequences that are considered to be normal from both intra-user and inter-user perspectives, as shown in Figure 3. The training data through this module is divided into a set of suspicious samples and a set of normal samples, and the extracted normal data samples are then used for training the subsequent anomaly detection model. This process of extracting normal data samples is described as **data distillation** in this paper. The notion of data distillation refers to the process of extracting normal samples from a set of mixed data samples (normal and abnormal) such that the training data can achieve a tighter bound to the normal patterns. In particular, this module distills data from both an intra-user perspective and an inter-user perspective. The intra-user perspective is implemented by means of a correlation based method, which is used to evaluate the gas usage consistency of individual gas users in adjacent weeks. The inter-user perspective, on the other hand, is implemented by means of a clustering method, which is used to evaluate the gas usage consistency between users from the same category.

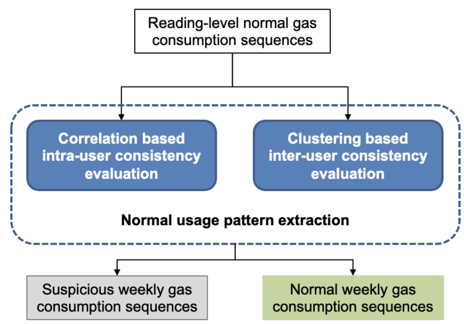

Fig. 3. Workflow of the normal pattern extraction module

1) Intra-User Consistency Evaluation: As gas consumption is closely related to human activities, sequences of weekly gas consumption usually exhibit similar patterns for the same user. For example, for a restaurant user, gas consumption often peaks around lunch and dinner time across the week while the amount of gas consumed during the weekdays and the weekends may differ slightly. Therefore, it is expected that the gas consumption of a normal user exhibits a certain level of consistency across adjacent weeks. To this end, we evaluate whether the weekly gas consumption of a gas user is self-consistent based on Pearson correlation [26]. Suppose the gas consumption of user s during the ith week is denoted as $x^i(s)$, its self-consistency coefficient is defined as follows:

$$r^i(s) = \frac{\rho(x^i(s), x^{i-1}(s)) + \rho(x^i(s), x^{i+1}(s))}{2} \tag{5}$$

where $\rho(x^i(s), x^{i+1}(s))$ is the Pearson correlation coefficient between two adjacent weekly gas consumption sequences as shown in Equation 6, and its value ranges from -1 to 1. The larger values indicate a stronger correlation.

$$\rho(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y} \tag{6}$$

where $cov(x, y)$ is the covariance of x and y, and σ_x and σ_y are the standard deviations of x and y.

According to the literature [27], when correlation $\rho(x, y) < 0.3$, x and y are considered to have weak association. Therefore, given a weekly gas consumption sequence $x^i(s)$, when its self-consistency coefficient $r^i(s)$ is less than 0.3, $x^i(s)$ is considered to be suspicious data samples and will not be used for subsequent model training.

2) Inter-User Consistency Evaluation: The self-consistency coefficient only evaluates the normality of the weekly gas consumption sequences from a single user. It is possible that the consumption sequences of a gas user over adjacent weeks have a strong correlation but they exhibit abnormal characteristics compared to the other gas users of the same type. Therefore, we introduce an inter-user consistency coefficient based on clustering. The intuition is that the gas consumption patterns of the users with the same type (e.g., Cantonese cuisine restaurants) usually follow similar patterns.

For clustering, the choice of distance metric is important, especially for time series data. In the literature, various distance metrics have been explored for time series clustering, including Euclidean distance, Dynamic Time Warping (DTW) [28], k-shape [29], and etc. However, these distance metrics mostly focus on pointwise comparison, which cannot well reflect the fact that even with the same type (e.g. Cantonese cuisine restaurants) different gas users may still differ in the amount of gas used each day and the time of usage to some extent. Therefore, the consistency evaluation across the gas users of the same type should focus more on the overall dynamics rather than the detailed changes. To this end, we adopt permutation entropy to extract new features from the weekly gas consumption sequences that serve as a basis for clustering. Permutation entropy [30] is a metric widely used to measure the dynamics of a time series by capturing the order relations of the consisting values.

In specific, given a weekly gas consumption sequence x^i, we first normalise x^i by scaling each data point by the maximum value from x^i, and the resulting sequence is denoted as $\bar{x}^i$. Then we split $\bar{x}^i$ into a set of daily gas consumption sequences, each of which consists of D data points depending on the data sampling rate. For each of these sequences, we calculate its permutation entropy following the procedure illustrated in [30]. As a result, the weekly gas consumption sequence x^i is transformed into a sequence of permutation entropy, defined as follows:

$$e(x^i) = (e(\bar{x}^i_{(k-1)*D+1}, \dots \bar{x}^i_{(k-1)*D+D}))_{k=1,\dots,T/D} \quad (7)$$

where $e(\bar{x}^i_{(k-1)*D+1}, \dots \bar{x}^i_{(k-1)*D+D})$ indicates the permutation entropy of the kth day of the week.

With the new features based on permutation entropy, we then apply Kmeans [31] for clustering such that the temporal dynamics of the weekly gas consumption sequences are taken into account for evaluating the distances between the data samples. It is expected that the normal data samples can be well grouped into different clusters with a small intra-cluster distance and a normal data sample should not be far away from the cluster centroid. To ensure a high level of normality of the data samples, we exclude the instances from the chaotic clusters with a large intra-cluster distance as well as those instances that are far away from the cluster centroids. Suppose the data samples are clustered into J clusters, we first evaluate the cluster-level consistency such that the data samples belonging to one of the clusters meeting the following criterion are considered as suspicious samples.

$$\alpha_o > \mu_{\text{-}o} + 2\sigma_{\text{-}o} \quad (8)$$

$$\mu_{\text{-}o} = \frac{\sum_{j \neq o} \alpha_j}{J} \quad (9)$$

$$\sigma_{\text{-}o} = \sqrt{\frac{\sum_{j \neq o}(\alpha_j - \mu_{\text{-}o})^2}{J}} \quad (10)$$

where $\alpha_o, 1 \leq o \leq J$ indicates the intra-cluster distance of cluster o, $\mu_{\text{-}o}$ indicates the mean of the intra-cluster distance of all the clusters excluding o, and $\sigma_{\text{-}o}$ indicates the standard deviation of the intra-cluster distance across all the clusters excluding o.

Thereafter, we evaluate the instance-level consistency within a cluster such that the data samples meeting the following criterion are considered as suspicious samples.

$$\beta_i > \mu_d + 2\sigma_d \quad (11)$$

$$\mu_d = \frac{\sum_{i=1}^{P} \beta_i}{P} \quad (12)$$

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^{P}(\beta_i - \mu_d)^2}{P}} \quad (13)$$

where β_i indicates the distance between an instance i and its cluster centroid, μ_d indicates the mean of the distances between the set of instances contained in the same cluster as i and the corresponding cluster centroid (i.e. intra-cluster distance of the cluster), P indicates the total number of instances in the cluster, and σ_d indicates the standard deviation of the distances between the instances and the corresponding cluster centroid.

For a given weekly gas consumption sequence, both the intra-user consistency evaluation and the inter-user consistency evaluation are applied to distil the data samples from the perspective of a single user as well as the perspective of a group of users with the same type. As a result, the normal gas usage patterns are extracted and will be fed to the subsequent model for training.

C. Learning Based Anomaly Detection

Usually with a sufficient number of both normal and abnormal data samples, a model can be trained to learn the decision boundary between the normal class and the abnormal class such that the trained model can be used for anomaly detection. However, in the case of detecting anomalous gas consumption, it is difficult to obtain enough anomalous data samples in practice, and thus unsupervised learning would be more applicable. That is, training the model with only the normal data samples. In the literature, unsupervised learning methods autoencoders have been shown to work well for the problem of anomaly detection of time series [32]–[34]. Following the literature in time series anomaly detection [35]–[38], in this study we adopt an autoencoder based on CNN (convolutional neural network) [39] for anomalous gas consumption detection and train the autoencoder with the normal data samples extracted by the module illustrated in Section IV-B.

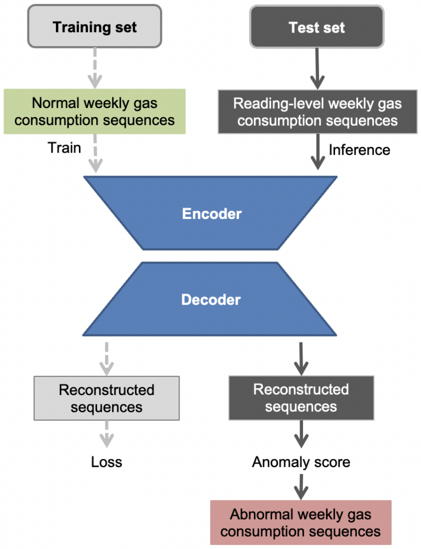

Fig. 4. Workflow of the learning based anomaly detection module

Figure 4 shows the structure the CNN-based autoencoder. We use Conv1D (a, b, c, d, f) to denote a 1D convolution layer activated by function f, where a, b, c and d represent the number of channels, kernel size, stride size and dilation size respectively. Similarly, we use Dconv1D (a, b, c, d, f) to denote a 1D deconvolution layer. FC (i, o, f) is used to represent a fully-connected layer with i input neurons and o output neurons activated by function f. The structures of the encoder and decoder are as follows: Conv1D (32, 4, 2, 2, ReLU) - Conv1D (32, 4, 2, 2, ReLU) - Conv1D (64, 8, 3, 2, ReLU) - FC (512, 64, None) and Dconv1D (64, 4, 2, 2, ReLU) - Dconv1D (32, 8, 2, 2, ReLU) - Dconv1D (16, 8, 2, 3, ReLU) - Dconv1D (1, 8, 2, 3, None).

To train the CNN-based autoencoder, we use cosine similarity for reconstruction error between the input sequence and the reconstructed sequence since it measures the angle between the vectors rather than the magnitude. This aligns with the fact that gas users may differ in the amount of gas they consume due to their capacity differences but they still have similar usage patterns. The cosine similarity between a weekly gas consumption sequence and the reconstructed sequence by the autoencoder is defined as follows:

$$\varepsilon(\bar{x}^i, \hat{x}^i) = \frac{\bar{x}^i \cdot \hat{x}^i}{\|\bar{x}^i\| \|\hat{x}^i\|} \tag{14}$$

where $\bar{x}^i$ is a normalised weekly gas consumption sequence, $\hat{x}^i$ is the reconstructed sequence, and $\varepsilon(\bar{x}^i, \hat{x}^i)$ ranges from -1 to 1. A larger value of cosine similarity indicates that the two sequences are more similar.

Based on the definition of cosine similarity, the reconstruction error (i.e. the loss function) used for model training is as follows:

$$\mathcal{L} = 1 - \varepsilon(\bar{x}^i, \hat{x}^i) \tag{15}$$

When a sequence and its reconstruction have a small cosine similarity, they are considered to be dissimilar. Accordingly, the reconstruction error is large. Therefore, the reconstruction error can be used to indicate the anomaly score of a weekly gas consumption sequence.

V. Dataset

The dataset used in this study is collected from the real-world gas meters deployed in a big city in southern China. It is composed of hourly meter readings from two categories of gas users, i.e., 481 **restaurant** users and 501 **canteen** users. For training, we used one year of meter readings between June 1, 2020 and June 1, 2021. Note that due to data collection and storage problems the meter readings of some users do not cover the complete one year period. After applying the preprocessing procedure as illustrated in Section IV-A, we obtain 21,285 weekly gas consumption sequences for the restaurant users and 24,436 sequences for the canteen users.

For testing, we use one week of meter readings from the same set of users, covering a period from November 1, 2021 to November 8, 2021. Through on-site visits and domain experts analysis, the data samples in the test set are manually labelled. Table II gives a summary of the dataset used in this study.

TABLE II
Dataset Overview

	Training Set		Test Set	
	Restaurant	Canteen	Restaurant	Canteen
Number of Gas Users	481	501	481	501
Number of Samples	21285	24436	481	501
Number of Anomalies	\	\	34	16

VI. Experiments

A. Evaluation Metric

To provide a unified comparison between the proposed method and the baselines, we use *precision@N%*, *recall@N%*, and F_1@N% as the evaluation metrics, namely the precision, recall and F_1 score of the top N ranked anomalous samples returned by the anomaly detection methods. In this study, we report the results of the top 3%, top 5%, top 10% and top 15%.

B. Baselines

In this study, we compare D^{2}AE with five baselines which are unsupervised methods widely used for anomaly detection of time series data. Among these five baselines, OCSVM has been investigated by [3] to detect gas thefts among boiler room users. The five baselines include:

- **LOF [40]**: Local Outlier Factor (LOF) is an unsupervised anomaly detection method that computes the local density deviation of a given data sample relative to its neighbours. It considers the data samples with a significantly lower density than their neighbours as outliers. The number of neighbours is chosen experimentally and set to 20.
- **iForest [41]**: Isolated Forest (iForest) uses decision trees to partition the data samples, and the depth of the data samples in the trees reflects their likelihood of being outliers. The number of decision trees is chosen experimentally and set to 100.

TABLE III
THE EXPERIMENT RESULTS OF D^{2}AE AND THE BASELINES ON THE RESTAURANT DATA

	Top3%			Top5%			Top10%			Top15%		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
rule & LOF	54.1%	35.3%	42.7%	41.1%	38.2%	39.6%	27.1%	44.1%	33.6%	19.0%	44.1%	26.6%
rule & iForest	49.6%	32.4%	39.2%	37.9%	35.3%	36.6%	25.3%	41.2%	31.4%	19.0%	44.1%	26.6%
rule & OCSVM	54.1%	35.3%	42.7%	37.9%	35.3%	36.6%	36.2%	58.8%	44.8%	26.6%	61.8%	37.2%
rule & DONUT	58.6%	38.2%	46.3%	47.4%	44.1%	45.7%	38.0%	61.8%	47.0%	29.1%	67.6%	40.7%
rule & DAGMM	67.6%	44.1%	53.4%	60.0%	55.9%	57.9%	41.6%	67.6%	51.5%	30.4%	70.6%	42.5%
D^{2}AE	**81.1%**	**52.9%**	**64.1%**	**69.5%**	**64.7%**	**67.0%**	**45.2%**	**73.5%**	**56.0%**	**34.2%**	**79.4%**	**47.8%**

TABLE IV
THE EXPERIMENT RESULTS OF D^{2}AE AND THE BASELINES ON THE CANTEEN DATA

	Top3%			Top5%			Top10%			Top15%		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
rule & LOF	11.8%	12.5%	12.1%	11.1%	18.8%	14.0%	5.8%	18.8%	8.8%	5.2%	25.0%	8.6%
rule & iForest	11.8%	12.5%	12.1%	11.1%	18.8%	14.0%	7.7%	25.0%	11.8%	6.5%	31.3%	10.8%
rule & OCSVM	11.8%	12.5%	12.1%	11.1%	18.8%	14.0%	7.7%	25.0%	11.8%	6.5%	31.3%	10.8%
rule & DONUT	11.8%	12.5%	12.1%	14.8%	25.0%	18.6%	11.6%	37.5%	17.7%	7.8%	37.5%	12.9%
rule & DAGMM	35.4%	37.5%	36.4%	22.3%	37.5%	27.9%	17.3%	56.3%	26.5%	11.7%	56.3%	19.4%
D^{2}AE	**41.2%**	**43.8%**	**42.5%**	**29.7%**	**50.0%**	**37.3%**	**19.3%**	**62.5%**	**29.5%**	**14.3%**	**68.8%**	**23.7%**

TABLE V
THE EXPERIMENT RESULTS FOR THE ABLATION STUDY ON THE RESTAURANT DATA

	Top3%			Top5%			Top10%			Top15%		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
D^{2}AE	**81.1%**	**52.9%**	**64.1%**	**69.5%**	**64.7%**	**67.0%**	**45.2%**	**73.5%**	**56.0%**	**34.2%**	**79.4%**	**47.8%**
variant V1	72.1%	47.1%	56.9%	60.0%	55.9%	57.9%	39.8%	64.7%	49.3%	32.9%	76.5%	46.0%
variant V2	76.6%	50.0%	60.5%	66.4%	61.8%	64.0%	41.6%	67.6%	51.5%	**34.2%**	**79.4%**	**47.8%**
variant V3	54.1%	35.3%	42.7%	37.9%	35.3%	36.6%	34.4%	55.9%	42.6%	26.6%	61.8%	37.2%
variant V4	58.6%	38.2%	46.3%	44.2%	41.2%	42.7%	27.1%	44.1%	33.6%	20.3%	47.1%	28.3%

TABLE VI
THE EXPERIMENT RESULTS FOR THE ABLATION STUDY ON THE CANTEEN DATA

	Top3%			Top5%			Top10%			Top15%		
	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1	Precision	Recall	F_1
D^{2}AE	**41.2%**	**43.8%**	**42.5%**	**29.7%**	**50.0%**	**37.3%**	**19.3%**	**62.5%**	**29.5%**	**14.3%**	**68.8%**	**23.7%**
variant V1	**41.2%**	**43.8%**	**42.5%**	26.0%	43.8%	32.6%	17.3%	56.3%	26.5%	11.7%	56.3%	19.4%
variant V2	**41.2%**	**43.8%**	**42.5%**	**29.7%**	**50.0%**	**37.3%**	17.3%	56.3%	26.5%	13.0%	62.5%	21.5%
variant V3	17.7%	18.8%	18.2%	14.8%	25.0%	18.6%	9.6%	31.3%	14.7%	9.1%	43.8%	15.1%
variant V4	17.7%	18.8%	18.2%	11.1%	18.8%	14.0%	7.7%	25.0%	11.8%	6.5%	31.3%	10.8%

- **OCSVM [3]**: One-Class Support Vector Machine (OCSVM) fits a hyperplane to separate the given data samples from the origin by mapping the data samples to a high-dimensional feature space through a kernel function. The RBF kernel function is used. The upper bound on the fraction of outliers is set to 0.1 following [3].
- **DONUT [17]**: DONUT is an unsupervised anomaly detection method based on VAE (Variational Auto-Encoder). It was proposed to detect the anomalies in various KPIs of large Internet companies' web applications. We follow [17] for the parameter setting.
- **DAGMM [18]**: Deep Autoencoding Gaussian Mixture Model (DAGMM) combines a deep autoencoder and a Gaussian Mixture Model (GMM) for unsupervised anomaly detection. We follow [18] for the parameter setting.

C. Ablation Study

As illustrated in Section IV, D^{2}AE consists of the rule-based anomaly detection module (*rule*), the normal usage pattern extraction module with two unsupervised methods, i.e. the correlation based intra-user consistency evaluation (*intra*) and the clustering based inter-user consistency evaluation (*inter*), and the learning based anomaly detection model based on autoencoder (*AE*) with cosine similarity (*cos*). To evaluate the effectiveness of the individual components of D^{2}AE, we also include several variants by removing or replacing the components of D^{2}AE. With D^{2}AE represented in the form of *rule* & *intra+inter* & *AE+cos*, the variants are expressed as follows:

- **V1 - *rule* & *AE+cos***: This variant removes the normal usage pattern extraction module of D^{2}AE.
- **V2 - *rule* & *intra* & *AE+cos***: This variant removes the inter-user consistency evaluation component in the normal usage pattern extraction module.
- **V3 - *rule* & *intra+inter* & *OCSVM***: This variant

replaces the autoencoder model used in the learning based anomaly detection module with OCSVM.

- **V4 - *rule & intra+inter & AE+L1***: This variant replaces the cosine similarity with L1 as the loss function used for training the autoencoder model in the learning based anomaly detection module.

D. Parameter Setting of D^{2}AE

All the parameters of the rule-based anomaly detection module are set based on the statistical analysis of the training data and the help of the domain experts. In particular, the hyper-parameters c_1 and c_2 in Equation (3) are set to 5 and 10 respectively. As a result, 1441 out of 19874 samples are considered as anomalies for the restaurants users and 2355 out of 22081 samples are considered as anomalies for the canteen users. They will be removed from the training set for subsequent model training.

Thereafter, the normal usage pattern extraction module is applied. In specific, the threshold for intra-user consistency evaluation is set to 0.3 following the literature [27] as illustrated in Section IV-B1. When Pearson correlation $\rho(x, y) < 0.3$, it is considered that x and y have a weak association [27]. As for the K-means clustering method used in the inter-user consistency evaluation, the number of clusters J is set to 10 for restaurant users and set to 14 for canteen users based on the silhouette coefficients [42]. As a result, 2175 samples are removed for restaurant users and 3526 samples are removed for canteen users.

With the normal data samples obtained from the previous two modules, the autoencoder model in the learning based anomaly detection module is trained. We use a learning rate of 0.0001, a batch size of 64, and the Adam optimizer. These hyper-parameters are chosen experimentally.

E. Results Analysis

When applying D^{2}AE after training, only the rule-based anomaly detection module and the learning-based anomaly detection module will be used. Firstly, with the data samples of the test dataset feeding into the rule-based anomaly detection module, 8 samples are considered as abnormal samples for the restaurant users and they are all labelled as anomalies in the test set. Similarly, 2 abnormal samples are considered as abnormal samples for the canteen users and they are all labelled as anomalies in the test set.

Based on the results of the rule-based anomaly detection module, Tables III and IV show the performance of D^{2}AE and the other baseline methods with respect to the datasets of the restaurant users and the canteen users respectively. It can be seen that in general D^{2}AE achieves the best performance on both datasets. Compared with OCSVM, which is used in the previous work for detecting gas theft among boiler room users [3], D^{2}AE improves the F_1 score by an average of 18.4% for restaurant users and 21.1% for canteen users.

In general, the autoencoder based methods (D^{2}AE, DAGMM and DONUT) achieve better performance than the other methods, which shows the superiority of autoencoders in the task of anomalous gas consumption detection. Compared with the unsupervised anomaly detection model DAGMM, D^{2}AE improves the F_1 score by an average of 7.4% for restaurant users and 5.7% for canteen users. All the methods achieve better performance on the restaurant data. One explanation is that the gas usage patterns of the canteen users are more diverse, which can be reflected from the fact that the number of clusters for the canteen users is larger than that for the restaurant users.

As for the ablation study, Tables V and VI show the performance of D^{2}AE and the variants as described in Section VI-C. In general, D^{2}AE outperforms all the variants. The superiority of D^{2}AE over variant V1 indicates the importance of the normal usage pattern extraction module. Without this module, more abnormal samples will be left in the training set for the autoencoder based anomaly detection model, and degrade its performance in detecting the anomalies. This can also be confirmed by the superiority of variant V3 over the baseline OCSVM. D^{2}AE outperforms variant V2 showing the effectiveness of extracting normal usage patterns from both the intra-user and inter-user perspectives. We also found that cosine similarity is a better loss function in the task of anomalous gas consumption detection, as D^{2}AE outperforms variant V4. An explanation is that with L1 the reconstruction errors of the data samples with more intensive gas consumption are likely to be higher while cosine similarity focuses more on the angle rather than the magnitude difference.

VII. Real Applications

Based on D^{2}AE, we developed an anomalous gas consumption detection system, which has been deployed in a gas company of a big city in China. This system automatically extracts the latest week's meter readings of the non-resident users every week, and applies D^{2}AE for anomaly detection. The detection results serve as a basis for the gas company to plan on-site inspection, which helps identify abnormal gas consumption more efficiently and reduce the labor cost.

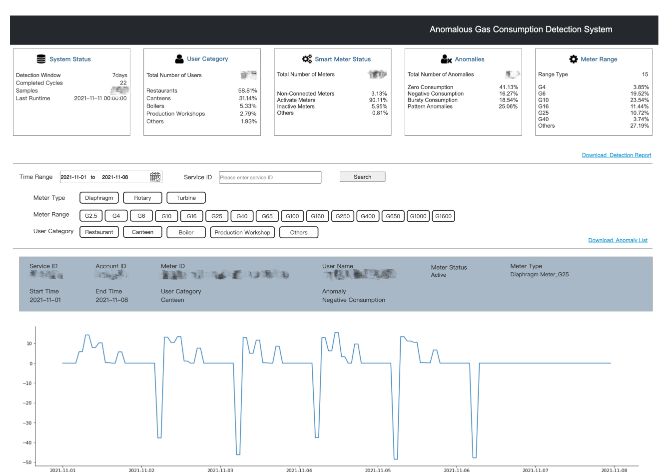

Fig. 5. User interface of the anomalous gas consumption detection system

Figure 5 shows the user interface of the system, which mainly consists of two parts. The upper part displays the time range of the current round of anomaly detection, user type

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2023.3296538

distribution, user status distribution, ratio of different types of anomalies, and etc. The second part presents the anomalies that have been detected with visualisations of gas consumption data of the current week. The system also enables the users to investigate particular anomalies by user ID, meter type, meter range, the purpose of gas usage. The anomaly list can support the staff of the gas company to carry out efficient inspections.

Next we present a few cases of the detection results from the anomalous gas consumption detection system to provide an intuitive understanding.

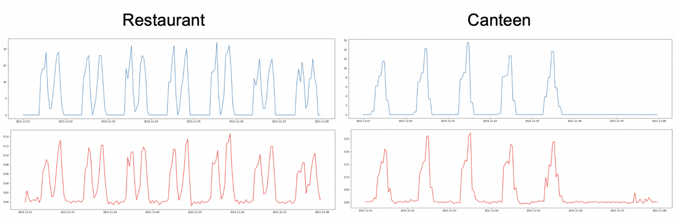

Fig. 6. Case 1: Normal gas consumption data

Figure 6 shows two cases of normal weekly gas consumption sequences. The left side of the figure is from a restaurant user, and the right side is from a canteen user. The blue plot is based on the real gas consumption data of the user, and the red plot is based on the reconstructed data returned by AE from D^{2}AE. It can be seen that the restaurant user has a regular gas consumption throughout the week and the canteen user only has a regular gas consumption during the weekdays while its gas consumption during the weekends is low. The reconstructed sequences for both users match well with the original data samples. This indicates that both samples are predicated as normal gas consumption since their patterns are well captured by D^{2}AE.

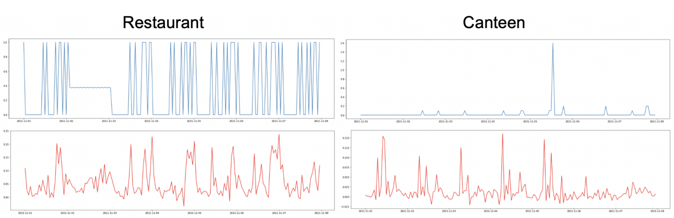

Fig. 7. Case 2: Detected abnormal gas consumption data

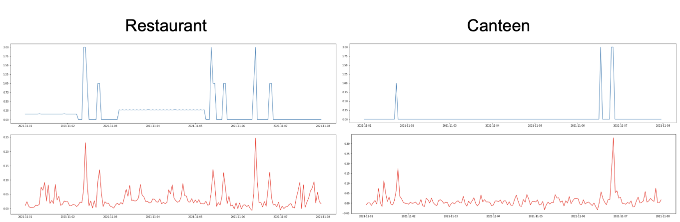

Fig. 8. Case 3: Detected abnormal gas consumption data

Figure 7 and Figure 8 respectively show four cases of abnormal weekly gas consumption sequences that have been successfully detected by D^{2}AE. From the left side of Figure 7, we can see that the gas consumption of the restaurant user maintained at a relatively high constant for a long time on Tuesday even through midnight. This anomaly is caused by the failure of the data transmission module. From the right side of Figure 7, it can be seen that the gas consumption of the canteen user seldomly consumes any gas and during the whole week and there is only one moment when the consumption is larger than $1m^3$. The cause of this anomaly is due to the worn-out gear system of the aging gas meter. From the left side of Figure 8, it can be seen that the gas consumption of the restaurant user is maintained at a relatively small constant for an extended period of time. This anomaly is caused by the transmission failure of the gas meter. From the right side of Figure 8, we can see that the canteen user only has gas consumption data on Monday and Saturday. This anomaly is due to that the canteen was temporarily out of service. For the above four cases, it can be seen that the reconstructed sequences have much difference from the original sequences, which can be considered as evidence of abnormal gas consumption.

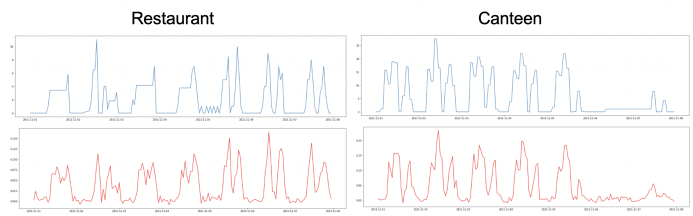

Fig. 9. Case 4: Undetected abnormal gas consumption data

Figure 9 shows two cases of abnormal weekly gas consumption sequences that have not been detected by D^{2}AE. From the left side of Figure 9, it can be seen that the usage patterns of this restaurant user are inconsistent during the week and the gas consumption was constantly on and off during midnight on November 5th, 2021. From the right side of Figure 9, we can see that the canteen user kept a relatively low gas consumption for a long time during the weekend. In these two cases, the reconstructed sequences are similar to the original sequences, which causes D^{2}AE failing to recognise the anomalies. One possible reason is due to the insignificant abnormality that exist in these two scenarios.

VIII. Conclusion and Future Work

In this paper, we propose D^{2}AE, a data distillation enhanced autoencoder for detecting anomalous gas consumption of non-residential gas users induced by various causes such as malfunctioning gas meters, gas leakage, gas theft etc. By encoding the knowledge of the domain experts, D^{2}AE first applies a rule-based anomaly detection module such that reading-level anomalies are detected. Thereafter, D^{2}AE introduces an integration of correlation based intra-user evaluation and clustering based inter-user consistency evaluation to distil the reading-level normal data samples. Finally, an autoencoder based anomaly detection model is trained with the distilled samples. With a real-world gas consumption dataset, we carry out extensive experiments and show the effectiveness of D^{2}AE compared to the baselines. For future work, we intend to improve the interpretability of D^{2}AE by learning to annotate the anomalous data points of the weekly gas consumption sequences based on the reconstructed data.

REFERENCES

[1] "National Bureau of Statistics," http://www.stats.gov.cn/tjsj/, 2021.

[2] X. Yang, X. Yi, S. Chen, S. Ruan, J. Zhang, Y. Zheng, and T. Li, "You are how you use: Catching gas theft suspects among diverse restaurant users," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 2885–2892.

[3] X. Yi, X. Yang, Y. Huang, S. Ke, J. Zhang, T. Li, and Y. Zheng, "Gas-theft suspect detection among boiler room users: A data-driven approach," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[4] L. Baldacci, M. Golfarelli, D. Lombardi, and F. Sami, "Natural gas consumption forecasting for anomaly detection," *Expert systems with applications*, vol. 62, pp. 190–201, 2016.

[5] H. N. Akouemo and R. J. Povinelli, "Probabilistic anomaly detection in natural gas time series data," *International Journal of Forecasting*, vol. 32, no. 3, pp. 948–956, 2016.

[6] L. Pan, X. Yi, S. Chen, Y. Huang, and Y. Zheng, "Neural clustering and ranking approach for gas-theft suspect detection," *Human-Centric Intelligent Systems*, pp. 1–12, 2023.

[7] X. Xu, X. Ai, and Z. Meng, "Research on abnormal detection of gas load based on lstm-wgan," in *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, vol. 12645. SPIE, 2023, pp. 819–824.

[8] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for iot time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2020.

[9] K.-H. Lai, D. Zha, J. Xu, Y. Zhao, G. Wang, and X. Hu, "Revisiting time series outlier detection: Definitions and benchmarks," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*, 2021. [Online]. Available: https://openreview.net/forum?id=r8IvOsnHchr

[10] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2005, vol. 589.

[11] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*, vol. 89, 2015, pp. 89–94.

[12] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective anomaly detection based on long short-term memory recurrent neural networks," in *International conference on future data and security engineering*. Springer, 2016, pp. 141–152.

[13] C.-C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. Keogh, "Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets," in *2016 IEEE 16th international conference on data mining (ICDM)*. Ieee, 2016, pp. 1317–1322.

[14] Y. Zhu, C.-C. M. Yeh, Z. Zimmerman, K. Kamgar, and E. Keogh, "Matrix profile xi: Scrimp++: time series motif discovery at interactive speeds," in *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 837–846.

[15] J. Ma and S. Perkins, "Time-series novelty detection using one-class support vector machines," in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, vol. 3. IEEE, 2003, pp. 1741–1745.

[16] L. Shen, Z. Li, and J. Kwok, "Timeseries anomaly detection using temporal hierarchical one-class network," *Advances in Neural Information Processing Systems*, vol. 33, pp. 13 016–13 026, 2020.

[17] H. Xu, W. Chen, N. Zhao, Z. Li, J. Bu, Z. Li, Y. Liu, Y. Zhao, D. Pei, Y. Feng *et al.*, "Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 187–196.

[18] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.

[19] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.

[20] Y.-H. Yoo, U.-H. Kim, and J.-H. Kim, "Recurrent reconstructive network for sequential anomaly detection," *IEEE transactions on cybernetics*, 2019.

[21] L. Shen, Z. Yu, Q. Ma, and J. T. Kwok, "Time series anomaly detection with multiresolution ensemble decoding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 11, 2021, pp. 9567–9575.

[22] B. Zhou, S. Liu, B. Hooi, X. Cheng, and J. Ye, "Beatgan: Anomalous rhythm detection using adversarially generated time series." in *IJCAI*, 2019, pp. 4433–4439.

[23] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning graph structures with transformer for multivariate time-series anomaly detection in iot," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9179–9189, 2022.

[24] Y. Zhou, J. Jiang, S.-H. Yang, L. He, and Y. Ding, "Musdri: Multi-seasonal decomposition based recurrent imputation for time series," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23 213–23 223, 2021.

[25] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 665–674.

[26] K. Pearson, "Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia," *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, no. 187, pp. 253–318, 1896.

[27] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.

[28] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[29] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, 2015, pp. 1855–1870.

[30] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.

[31] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[32] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Outlier detection for time series with recurrent autoencoder ensembles." in *IJCAI*, 2019, pp. 2725–2732.

[33] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly detection based on convolutional recurrent autoencoder for iot time series," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 112–122, 2020.

[34] M. Thill, W. Konen, H. Wang, and T. Bäck, "Temporal convolutional autoencoder for unsupervised anomaly detection in time series," *Applied Soft Computing*, vol. 112, p. 107751, 2021.

[35] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, "Autoencoder-based network anomaly detection," in *2018 Wireless telecommunications symposium (WTS)*. IEEE, 2018, pp. 1–5.

[36] S. Russo, A. Disch, F. Blumensaat, and K. Villez, "Anomaly detection using deep autoencoders for in-situ wastewater systems monitoring data," in *Proceedings of the 10th IWA Symposium on Systems Analysis and Integrated Assessment (Watermatex2019)*, 2019, pp. 1–4.

[37] F. Morawski, M. Bejger, E. Cuoco, and L. Petre, "Anomaly detection in gravitational waves data using convolutional autoencoders," *Machine Learning: Science and Technology*, vol. 2, no. 4, p. 045014, 2021.

[38] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan, "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises," *Robotics and Computer-Integrated Manufacturing*, vol. 79, p. 102441, 2023.

[39] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International conference on artificial neural networks*. Springer, 2011, pp. 52–59.

[40] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[41] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.

[42] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.