Outlier Detection for Water Supply Data Based on Joint Auto-Encoder

Shu Fang[1], Lei Huang[1], Yi Wan[2], Weize Sun[1,*] and Jingxin Xu[3]

Abstract: With the development of science and technology, the status of the water environment has received more and more attention. In this paper, we propose a deep learning model, named a Joint Auto-Encoder network, to solve the problem of outlier detection in water supply data. The Joint Auto-Encoder network first expands the size of training data and extracts the useful features from the input data, and then reconstructs the input data effectively into an output. The outliers are detected based on the network's reconstruction errors, with a larger reconstruction error indicating a higher rate to be an outlier. For water supply data, there are mainly two types of outliers: outliers with large values and those with values closed to zero. We set two separate thresholds, τ_1 and τ_2, for the reconstruction errors to detect the two types of outliers respectively. The data samples with reconstruction errors exceeding the thresholds are voted to be outliers. The two thresholds can be calculated by the classification confusion matrix and the receiver operating characteristic (ROC) curve. We have also performed comparisons between the Joint Auto-Encoder and the vanilla Auto-Encoder in this paper on both the synthesis data set and the MNIST data set. As a result, our model has proved to outperform the vanilla Auto-Encoder and some other outlier detection approaches with the recall rate of 98.94 percent in water supply data.

Keywords: Water supply data, outlier detection, auto-encoder, deep learning.

1 Introduction

The rapid development and widespread application of science has contributed to the outbreak of the era of big data. The massive and highly complex nature of big data presents many challenges and new opportunities for existing machine processing and computing power. The process of water supply data or the water resource consumption data is becoming more and more important, because these data reflect the water

[1] Guangdong Laboratory of Artificial-Intelligence and Cyber-Economics (SZ), College of Electronics and Information Engineering, Shenzhen University, Shenzhen, 518061, China.

[2] Water Resources Management Center of Ministry of Water Resources, Beijing, China.

[3] Departmet of Housing and Public Works, Queensland, Australia.

[*] Corresponding Author: Weize Sun. Email: proton198601@hotmail.com.

Received: 10 February2020; Accepted: 08 March 2020.

consumption status and trends of different companies and/or regions of a country, and the data mining result of such data might seriously affect the policy decision. Therefore, achieving high-precision prediction of water consumption becomes very important for the insurance of science and rationality of water resource comprehensive planning, water resource management and other policy-making. In order to achieve this goal, a complete and high-quality historical time series data is required.

Water resources management system [Raciti, Cucurull and Nadjm-Tehrani (2012)] and water resource data processing [Haddad, Afshar and Marino (2009)] have attracted lots of attention, and the collection and processing technology of water supply data has improved greatly with the development of remote sensing, telemetry, network, database and other technologies. However, due to the constraints of existing monitoring devices and methods, there are inevitably data abnormalities in the collection of historical data. Usually, outliers have a very large impact on data analysis [Alameddine, Kenney, Gosnell et al. (2010)] and information processing, and thus the methods for detecting these abnormalities of water supply data [Bakker, Lapikas, Tangena et al. (2012); Milln-Roures, Epifanio and Martinez (2018)] becomes very important. Generally speaking, data anomalies can be divided into two categories: actual mutation anomalies and anomalies to be corrected. The former is the actual change of indicator data due to actual consumption which needs to be retained, while the latter is the “abnormality” of data due to influence of human operation, equipment use, statistical caliber difference and/or other factors. In this article, we will focus on the outlier detection, or authenticity problem of water supply data. Note that for actual mutation anomalies, they are also required to be detected and further confirmed by human beings, therefore, the classification of the two kinds of anomalies is not important and thus not tackled here.

The intuitive definition of an outlier would be an observation that deviates so much from other observations [Hawkins (1980)]. Usually, the outliers in water supply data refer to some unusual large or small data that deviates from the most of the data pieces. The water supply data is collected under different perspectives, which is, from different sites, with different parameters relating to monitoring, collected by different individuals from multiple agencies, and in different time frames. Therefore, the form of anomalies might vary. To overcome this problem and obtain the real and effective information from the outliers in water resource data, many outlier detection methods have been proposed on water resources data and water resources management [Cho, Oh, Kim et al. (2013); Ayadi, Ghorbel, BenSaleh et al. (2017); Wright and Booth (2001)]. Among all these methods, the deep neural network methods are most widely used. Deep learning or deep neural network is very popular in various fields and applications [Deng and Yu (2014); Xu, Zhang, Xin et al. (2019); Huang, Sun and Huang (2020)], such as speech recognition [Sak, Senior, Rao et al. (2015); Taherian (2016); Qian, Bi, Tan et al. (2016)], information retrieval [Hang and Lu (2016)], medical data processing [Mamoshina, Vieira, Putin et al. (2016)] and computer vision [Vincent, Larochelle, Lajoie et al. (2010); Shrivastava, Pfister, Tuzel et al. (2017); Huang, Liu, Maaten et al. (2017)], and fraud detection [Webb, Pazzani and Billsus (2001)]. In this paper, we will propose a method named as Joint Auto-Encoder, to detect the outliers in water supply data. It is a variant of Auto-Encoder and is consists of two weight-sharing Auto-Encoders. Auto-Encoder extracts useful features from training data and then go on for the reconstruction [Hinton and Salakhutdinov (2006)], and the outliers can then be found by sorting the reconstruction errors.

The reminder structure of this article is organized as follows. Section II reviews the state-of-the-art water supply outlier detection and some other applications of outlier detection. In Section III, the Joint Auto-Encoder network structure are introduced in detail. The outlier detection approach that employing two thresholds τ_1 and τ_2 is also explained. Section IV presents experiments by comparing the Joint Auto-Encoder with different types of Auto-Encoder, and evaluate the proposed model on three different datasets. In the end, conclusion is given in Section V.

2 Related work

In order to detect outliers of water supply data, various works have been proposed. Mounce et al. [Mounce, Boxall and Machell (2009)] presented a method that trains a mixture density network by a continually updated historic database for detection of leaks or bursts at district meter area (DMA). Christodoulou et al. [Christodoulou, Kourti and Agathokleous (2017)], on the other hand, addressed the automatic detection of water losses in water distribution networks (WDN) by using a wavelet change-point detection classifier to identify anomalies. To void the critical infrastructure, Zohrevand et al. [Zohrevand, Glasser, Shahir et al. (2016)] proposed a method based on supervisory control and data acquisition (SCADA) for water supply system anomaly detection.

The problem of outlier detection is also widely seem in many other fields such as fraud detection [Dal, Boracchi, Caelen et al. (2018)], video surveillance [Kiran, Thomas and Parakkal (2018)], intrusion detection [Jabez and Muthukumar (2015)], face detection [Cheng, Ratha and Pankanti (2016)] and online social media analysis [Liu and Chawla (2017); Yu, He and Liu (2015)]. The outlier detection models can be roughly divided into three categories, namely, supervised models, semi-supervised models and unsupervised models. The supervised models are proved to be effective on drug name recognition [Chalapathy, Borzeshi and Piccardi (2016)] and health-care transaction applications [Chalapathy, Borzeshi and Piccardi (2016)]. However, they require a labelling of all the data manually, which is of high cost and sometimes might not be applicable. In addition, in order to achieve a good classification performance, it is assumed that the size of normal data has a same order of magnitude as the number of outliers, which is unusual because outliers are usually hard to be observed. Moreover, label costs too much. Therefore, they are not commonly used in real world data sets. To overcome these problems, the unsupervised outlier detection approaches are proposed, however, this kind of model might fail to give satisfying performance. Combining the advantages of these two, the semi-supervised outlier detection methods are proposed and they are proved to be suitable for detecting abnormal clinical electroencephalography [Wulsin, Blanco and Mani (2010)]. These methods employ the assumption that there are some non-trivial relationship between the labels and the unlabeled distributions [Lu (2009)], and thus is more efficient than the unsupervised ones. The model of Auto-Encoder [Hinton and Salakhutdinov (2006)] is one of the most widely used semi-supervised learning model on outlier detection. These kind of methods are of high computationally efficient in the testing stage and do not require too much label information of the data, thus are suitable for many applications.

Siamese network, which measures similarity, is widely used in object tracking [Tao, Gavves and Smeulders (2016)] and image matching [Zagoruyko and Komodakis (2015)]. Siamese networks are typically used in data sets that contain a large number of categories, with only a few training samples per category [Appalaraju and Chaoji (2017)]. It maps the input data to a target space through a function or a network, and uses a distance (e.g., European distance, etc.) in the target space for the comparison of the similarities. This kind of network has a good scalability and can be trained under a small dataset. In the water supply data, outliers can be regarded as an unknown category, and the number of this category is particularly small. In this work, we will employ the Siamese network together with the Auto-Encoder model to propose a new outlier detection model and solve the authenticity problem of water supply data.

3 Outlier detection method

3.1 Basic network structure

Auto-Encoder is a feed-forward neural network widely used in data mining, data compression and dimension reduction, and proved to be effective in outlier detection [Williams, Baxter, He et al. (2002)]. It mainly includes two subsections, namely, encoder and decoder. The encoder part conducts a dimension reduction to capture meaningful features, while the decoder part reconstructs the data from the reduced feature vectors. The structure of the network is shown in Fig. 1.

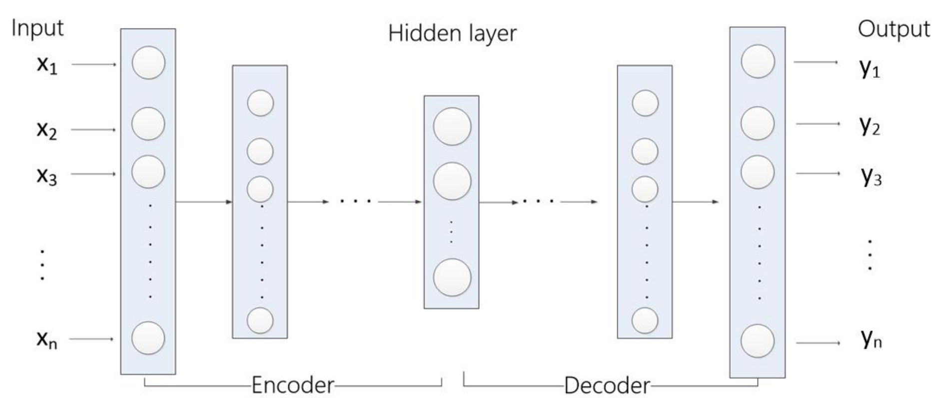

Figure 1: Vanilla auto-encoder network architecture

The purpose of this network is to generate an output that is as similar as possible to the input. The residual error between input and output, which is commonly used to detect outliers, can be expressed as follows:

$$\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{y} \tag{1}$$

where $\boldsymbol{x}$ is the raw input data and $\boldsymbol{y}$ is the prediction value. We detect the outliers by sorting the $\boldsymbol{r}$ for the input data in descending order, as the sample with a bigger $\boldsymbol{r}$ is more likely to be an outlier. However, there are still some problems to be addressed: although part of the outliers in the training data can be first removed or deleted before the training process, in real world applications, there will be many other unknown outliers that are not easy to be removed or observed by human beings, which may cause unexpected errors in the encoder. In addition, when the training data set is small, the Auto-Encoder network model will become overfitting quickly. Furthermore, the positive and negative residual error of water supply data usually play different roles and thus some necessary changes are to be made for the Auto-Encoder model.

3.2 Improved network structure and outlier detection method

In this part, we will introduce a joint network, named as Joint Auto-Encoder (JAE) network, for the problem of outlier detection. This proposed network employs the technique of the Siamese network, and is constructed by two weight sharing Auto-Encoders as shown in Fig. 2. During training, the weights of the JAE network are adjusted to minimize the joint loss function:

$$l = \|x_c - y_c\|_2 + \alpha \cdot \|x_c - y_o\|_1 + \beta \cdot \|z_c - z_o\|_1 \tag{2}$$

where x_c and y_c are the input and output of the clean data, respectively. We take the ℓ_2 norm of their difference as one part of loss function. Note that although there might be some unknown outliers in the original training data, we still define them as clean data because the outliers are unknown and cannot be removed. The y_o, on the other hand, is the output of the clean data with outliers (named as outlier data). These outlier data can be generated by adding some synthetic outliers, such as some random noise with large values or 0, to the clean data. We take the ℓ_1 norm of x_c and y_o as one part of loss function. Furthermore, the z_c and z_o are the feature vector of the network that are extracted from the encoder. Here we aim to minimize this loss function to ensure that the output of the network is as close as possible to the input data. The α and β are the hyper-parameters.

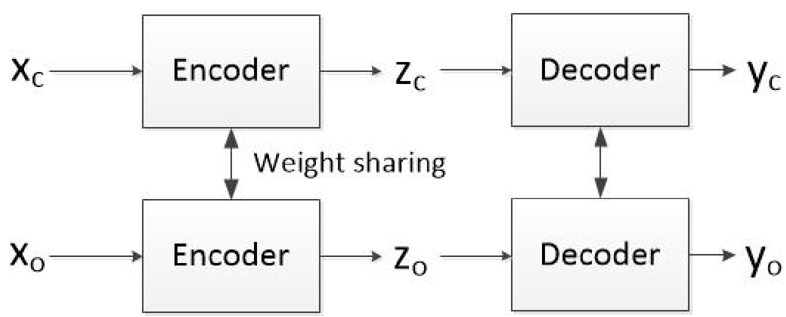

Figure 2: Proposed joint auto-encoder architecture

Generally speaking, the outlier detection problem can be considered as a two-class classification problem, where the outliers belong to one class and the normal values belong to the other. The two-class problem employs precision and recall as performance metrics, and uses the Receiver Operating Characteristic (ROC) curve to evaluate the outlier detection capability of the network model. Here we follow the idea in An et al. [An and Cho (2015)] and use the residual probability as the threshold for anomaly detection. Furthermore, the classification confusion matrix is employed for calculating the ROC curve, from which the threshold can be determined. The ROC curve is defined by plotting the False Positive Rates (FPR) against the True Positive Rates (TPR) or recall rates, where the TPR and FPR are defined as follows:

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$FPR = \frac{FP}{TN + FP} \tag{4}$$

Here the outlier samples are considered as the positive data and the normal data samples are considered as the negative data. True Positive (TP) number is the number of the true outliers detected as outliers, and False Positive (FP) number is the number of the normal item detected as outliers. The True Negative (TN) number, on the other hand, refers to the number of normal items detected as normal items, while the False Negative (FN) number is the number of outliers detected as normal items.

It is worth noting that there are at lease two types of outliers, which are, outliers with large values and those with values closed to zero. Therefore, two thresholds are employed in the proposed model. We conduct research by setting different values for positive and negative thresholds in experiments. All residual errors can be divided into two parts, the positive residual, which is used for detecting the large values, and the negative residual, which is used to detect small outliers, such as inappropriate 0 s. It is worth noting that some 0s are not outliers and thus this negative threshold is necessary. By drawing two ROC curves, the best thresholds can be found at locations where the true positive rate (TPR) is the highest and the false positive rate (FPR) is the lowest, and then the outliers in the water supply data can then be detected using the two thresholds as follows:

$$label = \begin{cases} 1\ ,\ if\ \ pr > \tau_1 \text{ or } nr < \tau_2 \\ 0\ ,\ else \end{cases} \tag{5}$$

where $label = 1$ means x being detected as an outlier, pr denotes the positive residual, and nr denotes the negative residual. By setting the best positive threshold τ_1 for big outliers and the best negative threshold τ_2 for outliers referring to inappropriate small value, the authenticity problem of water supply data can be solved.

4 Experiment results

In this section, experiments are carried out to verify the performance of the proposed method on synthesis data, the MNIST data set and the water supply data set.

4.1 Synthesis data

In this sub-section, synthesis data is generated to verify the effectiveness of JAE network for outlier detection. The synthetic data is generated as:

$$x_c(t) = max\{x(t), 0\} \tag{6}$$

where

$$x(t) = s(t) + n(t) = \sum_{n=1}^{3} A_n \sin(\omega_n t + \theta_n) + b + n(t) \tag{7}$$

for $t = 1, 2, ..., T$, and T=738 is the length of one piece of data, $A_n = 1$ is the amplitude, ω_n=$2\pi/738$, and θ_n is the random phase uniformly distributed in [0, 2π]. The $b = 6$ is the offset, and $n(t)$ is an additive Gaussian noise with mean 0 and standard deviation 1. Note that the operation in Eq. (6) is to make sure that all the $x(t)$, t=1,2,⋯ ,T are positive values for the simulation of real water supply data, whose elements are all non-negative values. A number of 1000 sample pieces are generated, and 700 of them are used for training while the remaining 300 samples are used for testing. We denote the $x_c(t)$ in Eq. (6) as clean data x_c in Fig. 2. During the training process, the x_o in Fig. 2 is generated by replacing 4% of the data points in x_c by 0 or $n_o(t)$, where the outlier data $n_o(t)$ is a large value uniformly distributed in [10, 20]. Note that the positions of outlier data points are recorded for verification purpose.

In the experiment, we build a two layers JAE with α=0.9 and β=0.5. The number of units of the hidden layer is set to 20. The network model is trained by the 700 pieces of data and then used to detect the outliers in the testing data set, and 4% of the elements of the test data pieces are replaced by outliers zeros or large values uniformly distributed in {10, 20}. The resulting ROC curves is plotted in Fig. 3, and it is shown that the propose JAE model is very efficient in outlier detection for the synthesis data.

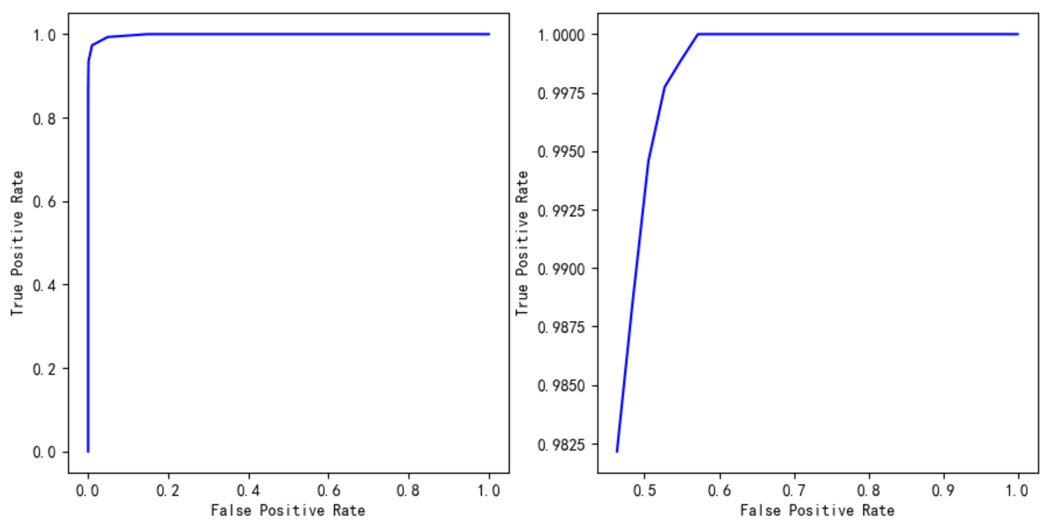

Figure 3: ROC curve of the JAE model. (Left: the ROC of positive residual, namely, τ_1; Right: the ROC of negative residual, namely, τ_2)

4.2 MNIST dataset

In this sub-section, the outlier detection methods are tested using the MNIST data set, and the proposed method is compared to One-Class Support Vector Machine (OC SVM)

[Scholkopf and Williamson (2000)] and One-Class Deep Support Vector Domain Description (OC Deep SVDD) [Ruff and Vandermeulen (2018)] approaches. The MNIST data set has 10 different classes, with 5500 and 1000 images of each class for training and testing, respectively. Therefore, the outlier detection methods will be trained and tested for 10 times, namely, one time for one class only.

In each time, one of the ten classes are selected as the normal class and the others are all outliers. During the training, the clean data x_c in Fig. 2 refers to the 5500 figures belong to the normal class, while the outlier x_o in Fig. 2 is generated by adding salt and pepper noise to 20% of the pixels in x_c. The network structure of the proposed method and the parameter settings are the same as those in the previous test. For testing, the data to be test, denoted as x', is fed into the network and the residual is now calculated as $r = \|\hat{x}' - x'\|_2 / m$ where $\hat{x}'$ is the output of the network. Note that only one ROC curve will be obtained this time, and the Area Under the ROC curve (AUC) is also calculated. As long as the test set contains 1000 normal images (the normal class) and 9000 abnormal images (the 9 other classes), we test the model for 9 times, with 1000 normal data and 1000 abnormal data in each time. The average results of AUC and its standard deviation of all the methods are shown in Tab. 1, which shows that in 7 out of 10 normal classes, the JAE approach outperforms the state-of-the-art methods. It is worth noting that the structure of outliers for training and testing data sets are different, showing the robustness of the proposed method.

Table 1: The AUCs results (mean±std) of different methods

NORMAL CLASS	OC-SVM	OC Deep SVDD	JAE
0	98.6±0.0	98.0±0.7	**99.2±0.0**
1	99.5±0.0	99.7±0.7	**99.8±0.0**
2	82.5±0.1	91.7±0.8	**91.9±0.1**
3	88.1±0.0	91.7±0.8	**94.3+0.0**
4	94.9±0.0	94.9±0.8	93.9±0.1
5	77.1±0.0	88.5±0.9	**95.4±0.0**
6	96.5±0.0	98.3±0.9	**98.5±0.0**
7	93.7±0.0	94.6±0.9	**96.8±0.1**
8	88.9±0.0	93.9±0.9	83.1±0.2
9	93.1±0.0	96.5±0.3	**96.1±0.1**

4.3 Water supply data

In this sub-section, the proposed approach is employed to solve the problem of outlier detection of the water supply data set.

4.3.1 Data and preprocessing

The raw water supply data set contains data from 204 different companies with 738 days/data points for each company. Since the water consumption of different company varies, we normalize the median of the 738 data points of each company to 1. The training, validation and testing data sets contain 153, 35 and 16 companies, respectively. It is assumed that the training and validation sets contains only clean data, or says, the obvious outliers are replaced by the median 1 in the data preprocessing stage. The testing set, on the other hand, contains only the raw data without any preprocessing, which means that the actual position of the outliers are unknown.

4.3.2 Training

For JAE network, the 153 pieces of training data are referred to as the clean data x_c and the x_o in Fig. 2 is generated by randomly replacing 2% and 2% of the elements of the data by zeros and large values uniformly distributed in [5, 15], respectively. The network structure of the proposed method and the parameter settings are the same as those in the previous test. For comparison, the situation of JAE with β=0 and the AE model [Hinton and Salakhutdinov (2006)] using ℓ_1 and ℓ_2 norm as loss functions are also included.

4.3.3 Outlier detection for validation data

Similarly, 2% and 2% of the elements of the validation data are replaced by zeros and large values uniformly distributed in [5, 15], respectively. Note that the positions of outlier data points are recorded for verification purpose. In this time, we separate the residual errors into positive parts and negative parts to discover the two separate thresholds τ_1 and τ_2 for outliers with large value and the outliers with zero value, respectively. The ROC curves are shown in Fig. 4. According to the ROC results, we set the positive threshold τ_1=2.8 and the negative threshold τ_2=-0.7, and obtain a final AUC of 99.99% and 99.10% for τ_1 and τ_2. A similar procedure is performed for the JAE model with β=0, which gives a corresponding AUC 99.03% and 98.33% and its ROC curves are shown in Fig. 5. It is shown that by a proper selected β, the JAE model can achieve a better performance, indicating the effectiveness of the loss term $\|z_c - z_o\|_1$ in Eq. (2). The confusion matrices, TPRs and FPRs of AE and JAE methods are now shown in Tab. 2. It is shown that the JAE model performs the best among all the methods with a TPR and FPR of 98.94% and 1.45%, respectively. To visualize the result, we select six typical pieces of data from the validation set and feed them into the JAE model to display the detected outliers, and the results are shown in Fig. 6. The green curves, black lines and red dots are the clean data elements, outliers and detection results, respectively. It is shown that most outliers can be correctly detected, showing the usefulness of the proposed method.

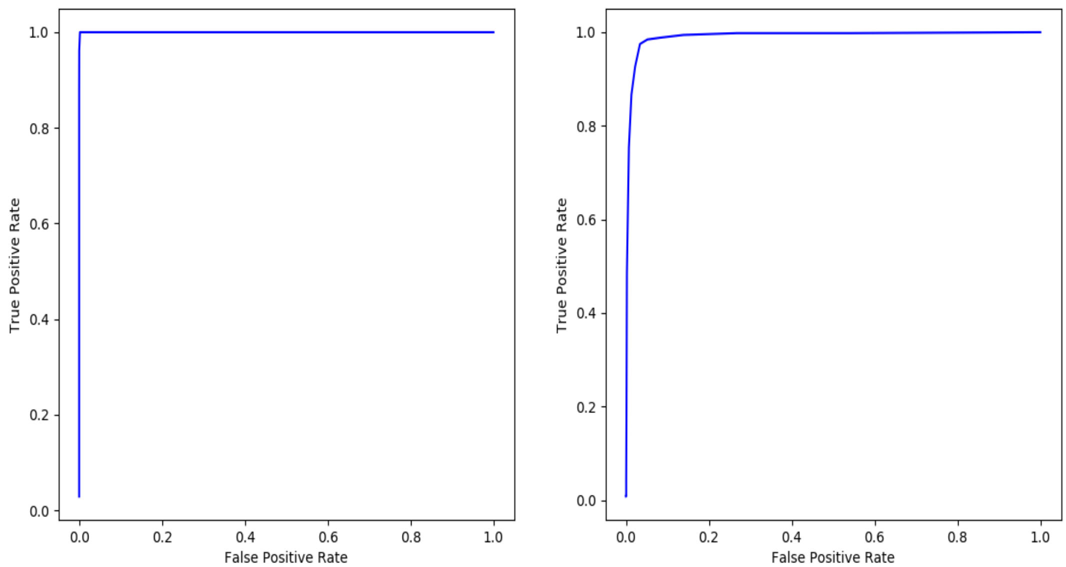

Figure 4: ROC curve of the JAE model on water supply data. (Left: the ROC of positive residual, namely, τ_1; Right: the ROC of negative residual, namely, τ_2)

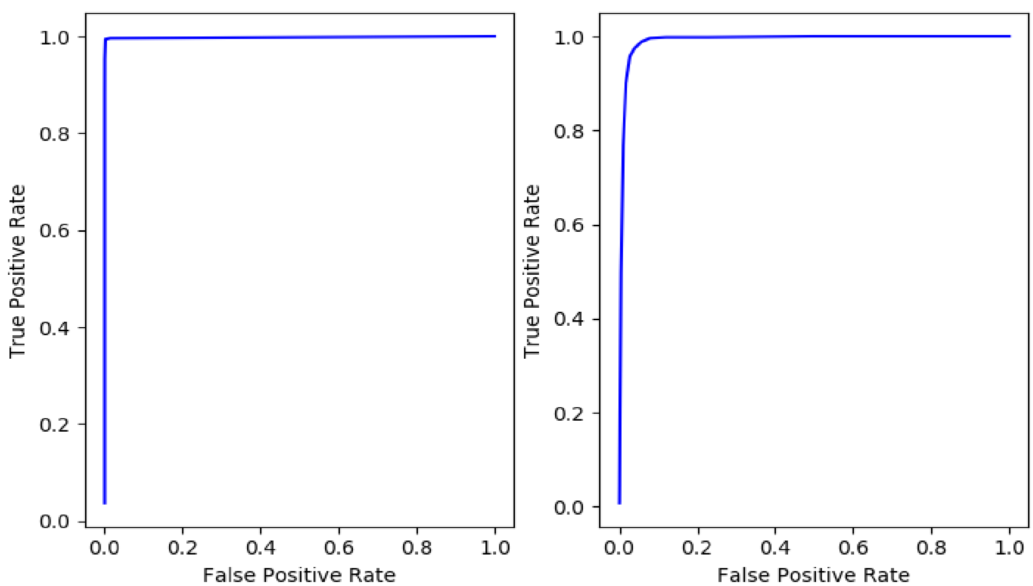

Figure 5: ROC curve of the JAE model under β=0 on water supply data. (Left: the ROC of positive residual, namely, τ_1; Right: the ROC of negative residual, namely, τ_2)

Table 2: The confusion matrices, TPRs and FPRs of JAE and AE network

Network model	TP number	FN number	FP number	TN number	TPR	FPR
JAE	1022	11	360	24437	**98.94%**	**1.45%**
JAE (β=0)	1008	25	364	24433	97.58%	1.47%
AE (ℓ_1)	868	602	602	24195	84.03%	2.43%
AE (ℓ_2)	889	11	763	24034	86.04%	3.08%

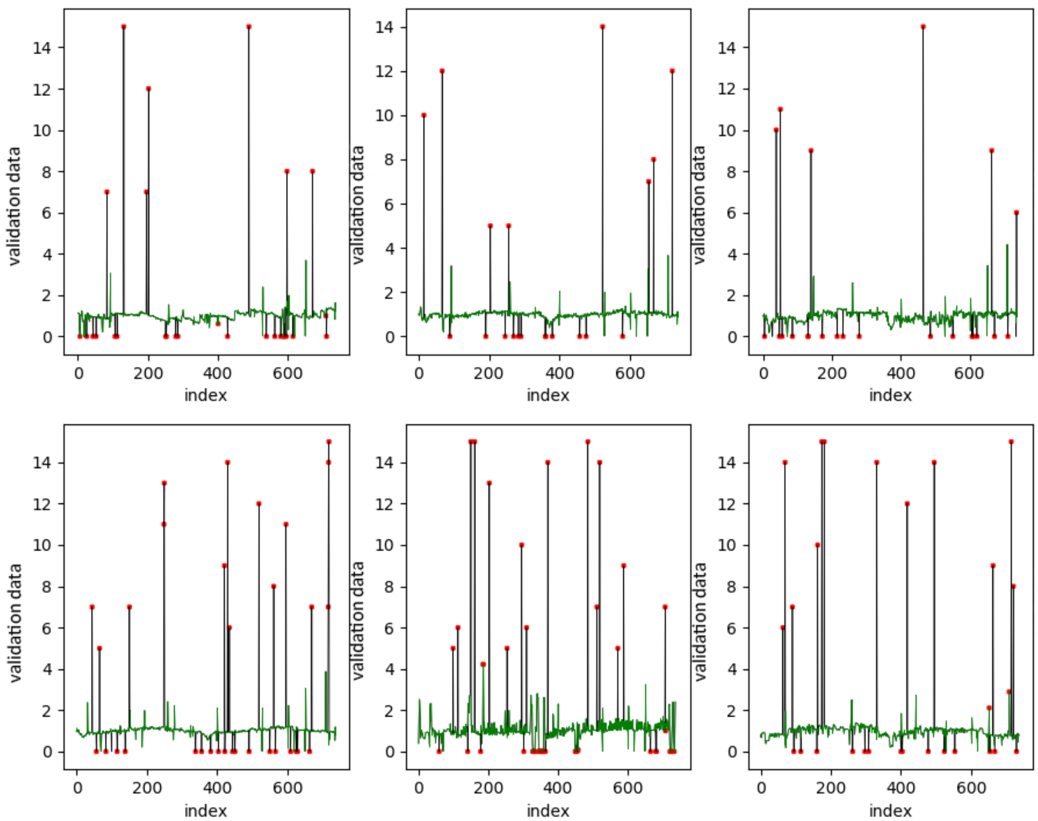

Figure 6: Outlier detection results of six typical companies from the validation set

4.3.4 Apply to testing set

Finally, we apply the proposed model to the testing set. Note that the outliers are not labeled, therefore, we can only show the detection result of six typical companies of the set. We follow the τ_1 and τ_2 values from the validation set, and the results are shown in Fig. 7. The green curves and red dots are the clean data elements and detection results, respectively. It is clearly shown that some large values and zero values are labeled as outliers by the proposed method, while some others are not, showing that the proposed method can separate the large and zero values into two classes, which are, those are detected as outliers and those are not.

According to the results from the above three experiments, the outliers detected by the JAE model should be noticed and checked carefully.

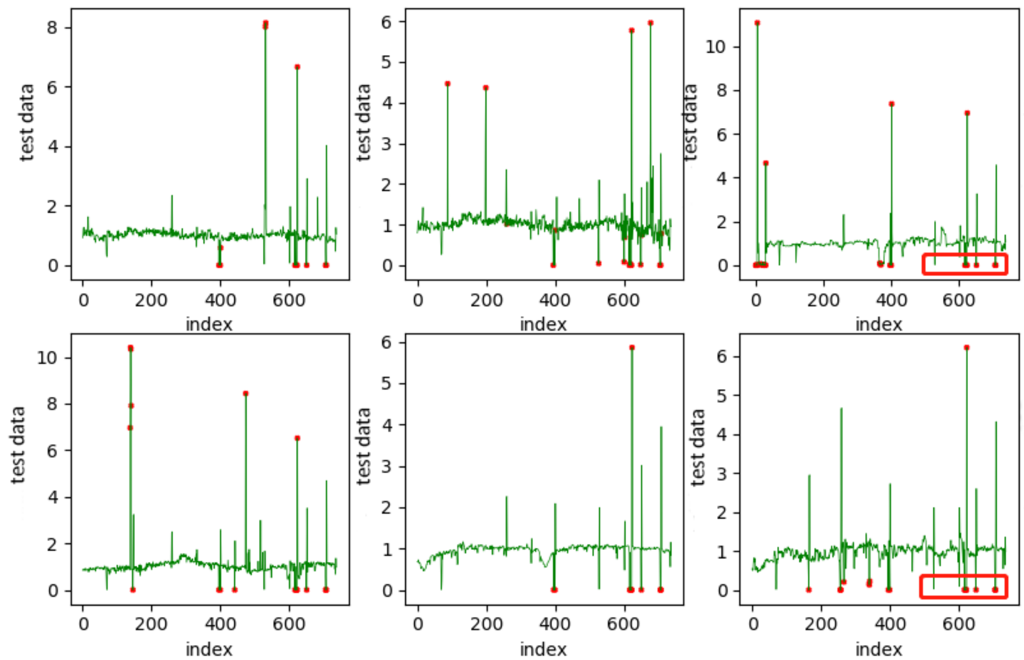

Figure 7: Outlier detection results of six typical companies from the testing set

5 Conclusion

In this paper, a Joint Auto-Encoder network that combines both the advantages of Auto-Encoder and Siamese network is proposed for outlier detection in water supply data. This network first encode the data pieces and then encode the encoded result for the data reconstruction. The outliers, can then be determined by the residual between the original data pieces and the reconstructed one. A mixture loss function is proposed, together with a dual thresholding outlier decision criteria. Experiments are carried out on three data sets, namely, synthesis data, the MNIST data set and the water supply dataset, and experimental results show the feasibility and effectiveness of the proposed model comparing to the state-of-the-art methods.

Funding Statement: The work described in this paper was supported by the National Natural Science Foundation of China (NSFC) under Grant No. U1501253 and Grant No. U1713217.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Alameddine, I.; Kenney, M. A.; Gosnell, R. J.; Reckhow, K. H. (2010): Robust multivariate outlier detection methods for environmental data. *Journal of Environmental Engineering*, vol. 136, no. 11, pp. 1299-1304.

An, J.; Cho, S. (2015): Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, vol. 2, no. 1, pp. 1-18.

Appalaraju, S.; Chaoji, V. (2017): Image similarity using deep CNN and curriculum learning. https://arxiv.org/abs/1709.08761.

Ayadi, A.; Ghorbel, O.; BenSaleh, M. S.; Obeid, A.; Abid, M. (2017): Outlier detection based on data reduction in WSNs for water pipeline. *Proceedings of 25th International Conference on Software, Telecommunications and Computer Networks*, pp. 1-6.

Bakker, M.; Lapikas, T.; Tangena, B.; Vreeburg, J. (2016): Monitoring water supply systems for anomaly detection and response. *Proceedings of IWC International Water Conference: New Developments in IT & Water*.

Bozorg Haddad, O.; Afshar, A.; Marino, M. (2009): Optimization of non-convex water resource problems by honey-bee mating optimization (HBMO) algorithm. *Engineering Computations*, vol. 26, no. 3, pp. 267-280.

Chalapathy, R.; Borzeshi, E. Z.; Piccardi, M. (2016): An investigation of recurrent neural architectures for drug name recognition. https://arxiv.org/abs/1609.07585.

Chalapathy, R.; Borzeshi, E. Z.; Piccardi, M. (2016): Bidirectional LSTM-CRF for clinical concept extraction. https://arxiv.org/abs/1610.05858.

Cheng, Y.; Ratha, N.; Pankanti, S. (2016): Outlier faces detector via efficient cohesive subgraph identification. *Proceedings of IEEE International Conference on Image Processing*, pp. 1659-1663.

Cho, H. Y.; Oh, J. H.; Kim, K. O.; Shim, J. S. (2013): Outlier detection and missing data filling methods for coastal water temperature data. *Journal of Coastal Research 65(sp2)*, pp. 1898-1904.

Christodoulou, S. E.; Kourti, E.; Agathokleous, A. (2017): Waterloss detection in water distribution networks using wavelet change-point detection. *Water Resources Management*, vol. 31, no. 3, pp. 979-994.

Dal, P. A.; Boracchi, G.; Caelen, O.; Alippi, C.; Bontempi, G. (2018): Credit card fraud detection: a realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks & Learning Systems*, vol. 29, no. 8, pp. 3784-3797.

Deng, L.; Yu, D. (2019): Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, vol. 7, no. 2-4, pp. 197-387.

Hang, L.; Lu, Z. (2016): Deep learning for information retrieval. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 1203-1206.

Hawkins, D. M. (1980): Identification of outliers. *Springer*, London.

Hinton, G. E.; Salakhutdinov, R. R. (2006): Reducing the dimensionality of data with neural networks. *Science*, vol. 313, no. 5786, pp. 504-507.

Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Q. (2017): Densely connected convolutional networks. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708.

Huang, J.; Sun, W.; Huang, L. (2020): Deep neural networks compression learning based on multiobjective evolutionary algorithms. *Neurocomputing*, vol. 378, pp. 260-269.

Jabez, J.; Muthukumar, B. (2015): Intrusion detection system (IDS): anomaly detection using outlier detection approach. *Procedia Computer Science*, vol. 48, pp. 338-346.

Kiran, B.; Thomas, D.; Parakkal, R. (2018): An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging*, vol. 4, no. 2, pp. 36.

Liu, Y.; Chawla, S. (2017): Social media anomaly detection: challenges and solutions. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 817-818.

Lu, T. T. (2009): *Fundamental Limitations of Semi-Supervised Learning* (*Master's Thesis)*. University of Waterloo.

Mamoshina, P.; Vieira, A.; Putin, E.; Zhavoronkov, A. (2016): Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, vol. 13, no. 5, pp. 1445-1454.

Millan-Roures, L.; Epifanio, I.; Martinez, V. (2018): Detection of anomalies in water networks by functional data analysis. *Mathematical Problems in Engineering*, pp. 1-3.

Mounce, S.; Boxall, J.; Machell, J. (2010): Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, vol. 136, no. 3, pp. 309-318.

Qian, Y.; Bi, M.; Tan, T.; Yu, K. (2016): Very deep convolutional neural networks for noise robust speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 12, pp. 2263-2276.

Raciti, M.; Cucurull, J.; Nadjm-Tehrani, S. (2012): Anomaly detection in water management systems. *Critical Infrastructure Protection Springer*, pp. 98-119.

Ruff, L.; Vandermeulen, R. (2018): Deep one-class classification. *Proceedings of International Conference on Machine Learning*, vol. 80, pp. 4393-4402.

Sak, H.; Senior, A.; Rao, K.; Beaufays, F. (2015): Fast and accurate recurrent neural network acoustic models for speech recognition. https://arxiv.org/abs/1507.06947.

Scholkopf, B.; Williamson, R. C.; Smola, A. J. (2000): Support vector method for novelty detection. *Advances in Neural Information Processing Systems*, pp. 582-588.

Shrivastava, A.; Pfister, T.; Tuzel, O.; Susskind, J.; Wang, W. et al. (2017): Learning from simulated and unsupervised images through adversarial training. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2107-2116.

Taherian, H. (2016): End-to-end attention-based distant speech recognition with highway LSTM. https://arxiv.org/abs/1610.05361.

Tao, R.; Gavves, E.; Smeulders, A. W. (2016): Siamese instance search for tracking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1420-1429.

Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P. A. (2010): Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, vol. 11, pp. 3371-3408.

Webb, G. I.; Pazzani, M. J.; Billsus, D. (2001): Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-2, pp. 19-29.

Williams, G.; Baxter, R.; He, H.; Hawkins, S.; Gu, L. (2002): A comparative study of RNN for outlier detection in data mining. *Proceedings of IEEE International Conference on Data Mining*, pp. 709-712.

Wright, C.; Booth, D. (2001): Water treatment control using the joint estimation outlier detection method. *Environmental Modeling & Assessment*, vol. 6, no. 1, pp. 77-82.

Wulsin, D.; Blanco, J.; Mani, R.; Litt, B. (2011): Semi-supervised anomaly detection for EEG waveforms using deep belief nets. *Proceedings of Ninth International Conference on Machine Learning & Applications*, pp. 436-441.

Xu, F.; Zhang, X. F.; Xin, Z. H.; Yang, A. (2019): Investigation on the Chinese text sentiment analysis based on convolutional neural networks in deep learning. *Computers, Materials & Continua*, vol. 58, no. 3, pp. 697-709.

Yu, R.; He, X.; Liu, Y. (2015): Glad: group anomaly detection in social media analysis. *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 2, pp. 18.

Zagoruyko, S.; Komodakis, N. (2015): Learning to compare image patches via convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4353-4361.

Zohrevand, Z.; Glasser, U.; Shahir, H. Y.; Tayebi, M. A.; Costanzo, R. (2016): Hidden Markov based anomaly detection for water supply systems. *Proceedings of IEEE International Conference on Big Data*, pp. 1551-1560.