CS550 Survey Report Unsupervised Anomaly Detection on Time Series Data

Kutay Taşcı - *22101359*, Gün Kaynar - *22101351*

Abstract—In this survey, we give an overview of the anomaly detection, time-series data and finally how anomaly detection is done on time-series data. We examine the supervised and unsupervised anomaly detection algorithms and show that unsupervised models are better for prediction on time-series data. We categorize anomaly detection models into 4 main groups, proximity-based, prediction-based, reconstruction-based and reconstruction-based GANs. For each category, we discuss the state-of-the-art methods and the advantages of using them. For the evaluation of models from each category, we refer to Garg et al. [22], and discuss that the reconstruction-based and prediction-based unsupervised models are highly robust in the context of time in data.

Index Terms—Machine Learning, Unsupervised Learning, Anomaly Detection, Time Series, Deep Learning

1 Introduction

Anomaly detection is a popular topic in all fields where data is collected. The recent proliferation of temporal observation data has led to an increasing demand for time series anomaly detection in many domains, from energy and finance to healthcare and cloud computing. A time series anomaly is defined as a time point or period where a system behaves unusually [1] . In this survey our aim is to make a literature survey on unsupervised anomaly detection on time series data.

Definition of anomaly according to Hawkings is "An anomaly or outlier is a data point which is significantly different from the remaining data" [2]. Aim of the anomaly detection techniques is to detect and classify this anomalies and outliers. Analyzing and detecting anomalies is important. Because, they can give us additional insights on our data and they allow us to detect patterns or data points which differ from the overall data. These techniques are widely used in data mining to find abnormal behaviour in data, gathered from wide range of applications. The problem of detecting outliers (rare events) has been variously called in different research communities: novelty detection, chance discovery, outlier/anomaly detection, exception mining, mining rare classes [3].

Time-series data has its own implicit features unlike regular data. Time-series data is periodical, seasonal and irregular. Apart from those, it has a trend characteristic which means that the regular anomaly detection models on time-series data cannot perform and generalize well as they would not capture the trend in different time points and act accordingly [4], [5].

Anomaly detection models mainly use unsupervised algorithms. First, in both regular and time-series data, it is clearly impossible to have all samples labeled since there is a huge amount of data to be fed into models. Second, when anomalies are cyclic, the distance dependent anomaly models will not function. Finally, on different domains of time-series data, there are different trends and contexts. Unsupervised anomaly detection models would handle these differences more robustly than the supervised models.

Most important problem when detecting anomalies is because the definition of anomaly, we don't know what we are looking for before training process. Thus, usage of the supervised algorithms are not desired because it becomes an unbalanced detection problem. Second problem on detecting anomalies is the urgency of detection task.

Usually in this task we want to detect anomalies as fast as possible. Because of this we want to get rid of labeling process and make training process as fast as possible. In the other hand in detection problems also reconstruction based methods outperform prediction based methods [6]. Because we don't wait for the future data points. Also in terms of detection speed there are methods like Hierarchical Temporal Memory (HTM) [7] which performs anomaly detection directly in streaming data almost without training in real time.

2 Background Information

There are many approaches for anomaly detection tasks in the literature. But because of the reasons we mentioned in the introduction, we are going to focus on unsupervised machine learning based techniques. In this survey we are going to analyse these algorithms based on three categories. There categories are; Proximity based algorithms, Prediction Based Algorithms and Reconstruction Based algorithms [6]. Each category has certain subcategories. Of course they have their pros and cons but overall concepts are build on similar ideas.

Also in this categories there is two important concept for given state of the art solutions. First is the machine learning based algorithms. Most of the machine learning algorithms does not have a general solution to anomaly detection problem. Rather they have different approaches and their success on evaluation step is mostly dependent on the nature of given problem. Also in most cases they fail to exploit the sequence based nature of the time series data. An exception to this is Markovian techniques [8].

Other approach to these problems are deep neural network based solutions. Compared to machine learning techniques, deep learning can provide more general solutions that can be applied to different problems. Also RNN's and LSTM's can exploit the sequence based nature of the time series data [9] . Current state of art approaches in unsupervised anomaly detection problems mostly depends on deep learning techniques. An addition to this approach recent development in Generative Adverserial Networks provides a new approach in this area which outperforms other approaches [10]. With combining this approaches with RNN's and LSTM's ability to exploit sequence information is currently the most well performing approach.

3 Related Work

In this section we are going to analyze related works on Anomaly Detection on TS data. For this task we are going to analyse past works on four categories.

- Proximity-based methods
- Prediction-based methods
- Reconstruction-based methods
- Reconstruction-based methods (GANs)

We are going to follow this order because base intuition of all approaches are based on past works in this order. Algorithms we will discuss in this section are related with each other and builds on the idea of previous approaches. In addition to these in the next section we are going to additional different approaches.

The algorithms analysed in this section mostly focuses on to provide distance over a normal data. Although some methods transform this distance into probability distribution [12], most of the methods use thresholds over a distance value.

There are two major drawback of this approaches. Firstly, these methods are unable to exploit the sequence based nature of time series data, because they can't capture temporal correlations. Secondly, choosing hyper parameters significantly effects the outcome of given method. Because of this we need the more information in training process, about anomaly duration and characteristics.

3.1 Proximity-based methods

Proximity based methods uses the distance as a measure to find anomalies. In this approach algorithm measures the distance of a point to a cluster. Also there are different approaches based on this approaches which uses this intuition. These methods mostly use machine learning techniques, specifically cluster based methods. Downside of this approach is, different algorithms are used in different types of data types perform significantly different from each other. [11].

3.1.1 Distance Based

This method uses KNN algorithm to measure the distance of a data point to its nearest neighbours [13]. This type of distance based methods uses a threshold value hyper-parameter in classification. Data points that are away from nearest point gets classified as an outlier.

3.1.2 Density Based

Another approach is density-based approaches like Local Outlier Factor (LOF) [14] and Clustering-Based Local Outlier Factor [15]. In Local Outlier Factor method, a value for how anomalous each sample in the dataset is calculated. It is a local value because only a neigbouring samples are taken into account when calculating it. But different than clustering based method, this method does not use the distribution in the data. In CLOF however, data is split into clusters and each sample's outlierness is calculated only considering other samples in the same cluster.

3.1.3 Angle Based

One of the approaches uses the angles between the points in a cluster to find outliers. Such as Angle-Based Outlier detection (ABOD) [16]. This method outperformes most of the similar methods, but algorithmic complexity of given method is much higher than distance and density-based approaches. This major drawback conflicts with the needs of applications in this area. Because in most of the cases aim is to find anomalies as fast as possible.

3.1.4 PCA Based

In this method, each data point is projected with PCA using the least detained variant. Non-anomalous samples will not have any correlation with the projected patterns whereas anomalies will. This way, it is possible to draw anomalies on the regular data. But PCA-based anomaly detection does not benefit finding the anomalies on time-series data. [17]

3.2 Prediction-based Methods

In prediciton-based methods aim is to use a time series data window to predict the next window or data point. Intuition behind this approach is that given model can successfully predict future data-point, if given sequence follows same patterns with the overall data. With using this prediction error we can detect anomalies.

There are two common problems that occur most of the works for prediction-base methods. First problem is we need the future data point to make predictions about anomalies. In real-time applications this creates an overhead for prediction time. Second is localizing an anomaly is much harder ,since prediction error will occur before, during and after the anomaly point.

3.2.1 HTM Based

One of the models that is used is Hierarchical Temporal Memory model [7]. This approach provides and unsupervised online sequence memory algorithm, to detect anomalies on streaming data. This method uses current input and predicts next time frame. Process being online is the biggest advantage of this implementation. Also a second advantage is, this method is resistant to concept drift [18] in time series data. In many scenarios the statistics of the system can change over time, a problem known as concept drift. In such cases models must adapt to a new definition of "normal" in an unsupervised, automated fashion [7].

3.2.2 LSTM RNN Based

Most common neural network tpye in this task is RNN's and LASTM RNNS's. Their ability to use sequence information provided by time series data makes this networks is the best choice for this prediction task. Spesifically LSTM's are proved to be an successful algorithm on this type of applications [4] [19]. In this methods we use Recurrent Neural Networks to predict next time frame. Compared to other methods this method outperforms other prediction based models in terms of accuracy. Disadvantage of this type of networks is training and launch time compare to HTM's.

3.2.3 CNN Based

One of the CNN based anomaly detection models is DeepAnt. It uses two different modules to capture both the trend and seasonality in time series data, along with the anomalies. For this purpose, one module named time series predictor uses a convolutional neural network that guesses the next time point. The prediction then is fed into the other module named anomaly detector. Anomaly detector module contains a dense layer after the CNN structure for Euclidean distance calculation. The biggest advantage of the model is that training does not require a lot of data since time series predictor module predicts the time stamps and does not use the actual time frame for the anomaly detection [5].

3.3 Reconstruction-based Methods

In these methods, a reconstructed data point is compared to the input, by the difference in the reconstruction, the anomaly can be found. This approach is fueled by the fact that the reconstructers such as auto encoders will remember the previous non-anomalous samples and use the features learnt from them for the reconstruction. When an anomalous data point comes into play, the reconstruction will be similar to a non outlier sample, in this case, the difference will be higher for the actual input and a normal data point, the model will catch the sample as anomaly.

3.3.1 Variational Auto-Enconder Based

As we mentioned above we use reconstructional methods to predict the input itself and use the prediction error to detect anomalies. For this task we are using threshold as a hyper-parameter. In this method aim is to encode input space into a probability distribution in latent space. In, this way model can represent an input as a probability of being in a latent distribution [12]. This means instead of using a threshold over a prediction error model can directly output probability of anomalies presence.

3.3.2 Adversely Trained Auto-Encoder Based

This method consists of training and detection models. First, two auto encoders are trained to reproduce the input at a minimal cost. Then secondly, one of the auto encoders is trained to differentiate the reproduced data from the actual input. The other auto encoder at the same time tries to fool the first auto encoder and improves the reproduction. After the training, two outputs are taken from the decoding of the trained auto encoders. These outputs are then compared, and when there is an anomaly in the data the reconstruction would not catch the difference and act rapidly, so the input and reconstruction will be non similar. At that point the decoding will detect the anomaly. This method is fast training, accurate and stable. [20]

3.4 Reconstruction-based Methods (GAN's)

In previous methods we first used proximity based methods, then we used prediction and reconstruction errors to detect anomalies. Development of deep learning methods is a huge step for Anomaly Detection applications. Recent development in Generative Adverserial Networks is also proved to be superior to the past methods [6] [10]. This methods builds on the idea of reconstructional methods. In addition training process of GAN's provides an extra Discriminator neural network.

In GAN architecture there are two neural networks. One neural network generates a data from a random latent vector, this network is called "Generator". Second neural network tires to classify if a incoming output is fake or real, a fake data is one generated by "Generator". This second model is called "Discriminator". Basically one network learns to generate fake data, other one learns to discriminate real data with fake data.

Also it is possible to find corresponding latent vector z of an existing(real) data. As you can guess we can reconstruct the data sequence with using "Generator" and measure reconstruction error. This part works similar to previous reconstructional methods. In addition to that this method uses the "Discriminator" to discriminate between anomalies and normal data. Because it is also trained with normal data, so abnormal data is treated as fake data.

3.4.1 MAD-GAN

In this work author's implemented this GAN architecture with using RNN LSTM Recurrent Neural Networks as base models for Generators and Discriminator [10]. Idea of using GAN's with combined to RNN's ability to use sequence information of time series data is implemented. It is proved to be superior to other methods like PCA, KNN, Auto-Encoder based methods [10].

3.4.2 Tad-GAN

In this approach author's expanded the idea of using Generative Adverserial Networks. They included an extra encoder to encode input to latent space. For training this they included cycle consistency in the model [6]. With the help of encoder model can both encode input data to latent vector more precisely, also since encoding process is much

	NASA		Yahoo S5				NAB					
Baseline	MSL	SMAP	A1	A2	A3	A4	Art	AdEx	AWS	Traf	Tweets	Mean±SD
TadGAN	**0.623**	**0.704**	**0.8**	0.867	0.685	0.6	**0.8**	**0.8**	0.644	0.486	**0.609**	**0.700±0.123**
(P) LSTM	0.46	0.69	0.744	**0.98**	0.772	0.645	0.375	0.538	0.474	**0.634**	0.543	0.623±0.163
(P) Arima	0.492	0.42	0.726	0.836	**0.815**	**0.703**	0.353	0.583	0.518	0.571	0.567	0.599±0.148
(C) DeepAR	0.583	0.453	0.532	0.929	0.467	0.454	0.545	0.615	0.39	0.6	0.542	0.555±0.130
(R) LSTM AE	0.507	0.672	0.608	0.871	0.248	0.163	0.545	0.571	**0.764**	0.552	0.542	0.549±0.193
(P) HTM	0.412	0.557	0.588	0.662	0.325	0.287	0.455	0.519	0.571	0.474	0.526	0.489±0.108
(R) Dense AE	0.507	0.7	0.472	0.294	0.074	0.09	0.444	0.267	0.64	0.333	0.057	0.353±0.212
(R) MAD-GAN	0.111	0.128	0.37	0.439	0.589	0.464	0.324	0.297	0.273	0.412	0.444	0.35±0.137
(C) MS Azure	0.218	0.118	0.352	0.612	0.257	0.204	0.125	0.066	0.173	0.166	0.118	0.219±0.145

Fig. 1: F1-SCORES OF BASELINE MODELS USING WINDOW-BASED RULES. COLOR ENCODES THE PERFORMANCE OF THE F1 SCORE. ONE IS EVENLY DIVIDED INTO 10 BINS, WITH EACH BIN ASSOCIATED WITH ONE COLOR. FROM DARK RED TO DARK BLUE, F1 SCORE INCREASES FROM 0 TO 1. [6]

faster than previous approaches it is more feasible solution to anomaly detection tasks.

4 Evaluation

In this section we are going to review the analysis of some of the algorithms we talked about in this survey. For this task we are going to use the provided analysis in TadGAN [6] paper. In this work we see the comparison of algorithms on benchmark datasets. There are some algorithms in this analysis we did not mention. Reason that we haven't covered these algorithms, because they are statistical methods rather than machine learning approaches.

You can see the experiments in Fig. 1. TadGan clearly outperforms most of the methods that are used in Anomaly Detection. Other best performing methods uses LSTM's and auto-encoder structure. This table clearly shows that deep learning based methods outperforms the other approaches. Also using GAN's as a reconstruction-based method is the best performing approach.

Important to note that GAN-based approaches wtihout regularization on latent space embedding performs less accurate than Auto-encoders. Reason behind this is cycle consistency trains an encoder to embed generated data back into the latent space.

5 SoTA and possible future work

In this survey we go over a lot of methods and approaches on the topic of Unsupervised Anomaly Detection on Time Series data. Overall we can say that development of RNN LSTM's are played a huge part in current state of the art models. Their ability to use time series sequence information is proved to be a good approach in this problem. In addition to this GAN's is the current best approach for this task.

Currently GAN's are a recently researched topic. There are many techniques that are developed in and being developed in GAN's. For example in TAD-GAN authors included cycle consistency [6] in their model. Usage of recently developed methods together or applying new techniques on this field is an open research topic.

6 Conclusion

In this survey, we show different methods of anomaly detection and their advantages. The most problematic part in anomaly detection algorithms on time-series data is that time-series data is not regular nor balanced. In different time frames of the data, there is another trend and season. Some anomaly detection models can not capture anomalies taking into account these seasonality and trends, on the other hand some models such as reconstruction-based methods can. To overcome this problem, these methods use a time scale in data instead of the whole data.

The most precise and stable methods in anomaly detection on time-series data have been Reconstruction-based GAN's and Prediction based methods. Reconstruction is a very popular method for anomaly detection because, it is unsupervised and training time is not limiting the prediction time [22].

In our project, we aim to detect anomalies on MIT-BIH arrhythmia database [21]. This dataset contains 48 hours of heartbeats from 47 subjects. Even though the dataset has corresponding labels for anomalous heartbeats, we use an unsupervised approach and do not consider the labeled arrhytmias.

There have been various anomaly detection works on MIT-BIH arrhythmia dataset as detection of anomalous heartbeat is very useful and important in healthcare. One study shows that LSTM autoencoder method has been successful in classifying heartbeats as anomaly or healthy in the dataset [23]. Another shows that supervised anomaly detection is also successful on the dataset [24]. Lastly, there has been a work that shows the unsupervised anomaly detection is also possible on the dataset [25].

References

[1] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41.3 (2009):

1-58.
[2] Hawkins, Douglas M. Identification of outliers. Vol. 11. London: Chapman and Hall, 1980.
[3] Lazarevic, Aleksandar, and Vipin Kumar. ”Feature bagging for outlier detection.” Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. 2005.
[4] Malhotra, Pankaj et al. “Long Short Term Memory Networks for Anomaly Detection in Time Series.” ESANN (2015).
[5] M. Munir, S. A. Siddiqui, A. Dengel and S. Ahmed, ”DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series,” in IEEE Access, vol. 7, pp. 1991-2005, 2019, doi: 10.1109/ACCESS.2018.2886457.
[6] Geiger, Alexander, et al. ”TadGAN: Time series anomaly detection using generative adversarial networks.” 2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020.
[7] Ahmad, Subutai, et al. ”Unsupervised real-time anomaly detection for streaming data.” Neurocomputing 262 (2017): 134-147.
[8] Chandola, Varun, Varun Mithal, and Vipin Kumar. ”Comparative evaluation of anomaly detection techniques for sequence data.” 2008 Eighth IEEE international conference on data mining. IEEE, 2008.
[9] Malhotra, Pankaj, et al. ”Long short term memory networks for anomaly detection in time series.” Proceedings. Vol. 89. 2015.
[10] Li, Dan, et al. ”MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks.” International Conference on Artificial Neural Networks. Springer, Cham, 2019.
[11] Chandola, Varun, Varun Mithal, and Vipin Kumar. ”Comparative evaluation of anomaly detection techniques for sequence data.” 2008 Eighth IEEE international conference on data mining. IEEE, 2008.
[12] An, Jinwon, and Sungzoon Cho. ”Variational autoencoder based anomaly detection using reconstruction probability.” Special Lecture on IE 2.1 (2015): 1-18.
[13] Angiulli, Fabrizio, and Clara Pizzuti. ”Fast outlier detection in high dimensional spaces.” European conference on principles of data mining and knowledge discovery. Springer, Berlin, Heidelberg, 2002.
[14] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “Lof: identifying density-based local outliers,” in Proc. of the ACM SIGMOD, 2000, pp. 93–104.
[15] Z. He, X. Xu, and S. Deng, “Discovering cluster-based local outliers,” Pattern Recognition Letters, vol. 24, no. 9-10, pp. 1641–1650, 2003.
[16] Kriegel, Hans-Peter, Matthias Schubert, and Arthur Zimek. ”Angle-based outlier detection in high-dimensional data.” Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 2008.
[17] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15 (July 2009)
[18] J. Gama, I. Žliobaite,˙ A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM Comput. Surv. 46 (2014) 1–37
[19] Chauhan, Sucheta, and Lovekesh Vig. ”Anomaly detection in ECG time signals via deep long short-term memory networks.” 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015.
[20] Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A. Zuluaga. 2020. USAD: UnSupervised Anomaly Detection on Multivariate Time Series. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD ’20). Association for Computing Machinery, New York, NY, USA, 3395–3404.
[21] Guanglong, Ma Xiangqing, Wang Junsheng, Yu. (2019). ECG Signal Classification Algorithm Based on Fusion Features. Journal of Physics: Conference Series. 1207. 012003. 10.1088/1742-6596/1207/1/012003.
[22] Astha Garg, Wenyu Zhang, Jules Samaran, Savitha Ramasamy, Chuan-Sheng Foo. ”An Evaluation of Anomaly Detection and Diagnosis in Multivariate Time Series” Machine Learning (cs.LG); Artificial Intelligence, IEEE, 2011
[23] Pengfei Liu, Xiaoming Sun, Yang Han, Zhishuai He, Weifeng Zhang, Chenxu Wu, Arrhythmia classification of LSTM autoencoder based on time series anomaly detection, Biomedical Signal Processing and Control, Volume 71, Part B, 2022, 103228, ISSN 1746-8094,
[24] Li, Hong Zu, and Pierre Boulanger. “A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG).” Sensors (Basel, Switzerland) vol. 20,5 1461. 6 Mar. 2020, doi:10.3390/s20051461
[25] Matias, Pedro Folgado, Duarte Gamboa, Hugo Carreiro, André. (2021). Robust Anomaly Detection in Time Series through Variational AutoEncoders and a Local Similarity Score. 91-102. 10.5220/0010320500910102.