Anomaly Detection in Time Series

Aarohan Jain
Chair of Electronic Measurement and Diagnostic Technology
Department of Energy and Automation Technology
School of Electrical Engineering and Computer Science
Technische Universität Berlin
Berlin, Germany
aarohan.jain@campus.tu-berlin.de
Matr-Nr: 375232

***Abstract*—This paper introduces the reader to the concept of anomaly detection in time series by outlining its uses and approaches to tackle the problem. Then, time series and techniques specific to anomaly detection in time series are presented. An emphasis is placed on one technique - ARIMA, as it helps illustrate multiple approaches to tackling this relatively new problem. Finally, the context of predictive maintenance is introduced in order to illustrate the importance and potential of further research and development in this area.**

***Index Terms*—anomaly detection, time series, ARIMA, predictive maintenance**

I. Introduction

An anomaly is defined as an observation that deviates significantly from other observations, which intuitively should cause a statistician to suspect that the data generation mechanism behind the anomaly is different [1]. Anomalies are referred to by many other names, often used interchangeably, such as outlier, abnormality, discordant, and others [2] [3].

The occurrence of an anomaly is usually of interest in the application field of the observations. It points to an unusual behavior in the generating process behind the data, which in turn can be used to identify abnormal system characteristics. The insight provided by anomalies can be useful in a vast range of disciplines and applications, such as health, finance, defense, science, and many others. [2]

This research paper discusses the importance and applications of anomaly detection. Types of anomaly detection in data are presented, with a focus on algorithms that specifically tackle time series data. A practical example in the world of Predictive Maintenance (PDM) concerning bearing fault analysis is explored. Finally, a possible direction of future research in the subject matter is discussed. The paper also acts as a literature survey/review, as the reader is advised to refer to the cited resources for in-depth information beyond the scope of this work.

Owing to the broad nature of the field of anomaly detection, various methods are mentioned solely to compare and contrast with others that are explored in more depth. The topics to explore in depth are picked with the intention of providing a representative overview that can be used as a stepping stone for further research. These topics are clearly marked in the respective sections.

II. Related Works

Anomaly detection in time series is a relatively new field: Aggarwal [4] cites an article by Dasgupta et al. published in 1999 [5] as one of the first proposals for the process. Despite it being a new field, multiple surveys have been conducted for anomaly detection; these offer a well-organized and intuitive overview of the field, such as: Hodge et al. [6], Chandola et al. [7], and Gupta et al. [8], with the lattermost focusing on temporal data (of which time series is a part). Additionally, books on anomaly detection [2] [3] that contain information about algorithms for both multidimensional and time-series data [4], [9] - [13] are helpful. For more specific research, references are presented in relevant sections of the paper.

III. Anomaly Detection

This section explores the field of anomaly detection in more detail. Firstly, a caveat that separates this field from the field of supervised classification is discussed. Then, methods to evaluate algorithms to detect anomalies are presented. Lastly, the section attempts to streamline and classify the vast field by exploring approaches for anomaly detection along with problem classes.

A. Anomaly Detection vs. Binary Classification

Anomaly detection frequently deals with the comparison between real-world data and expected data, with the latter being an important aspect of the process of detection [3]. Aggarwal [2] defines two types of outputs an anomaly detection algorithm can generate:

- **Outlier scores**: The algorithm quantitatively evaluates the anomalousness of any given data; this information can therefore also be used to rank the data by anomalousness.
- **Binary labels**: The algorithm produces a Boolean for any given data, and therefore only provides information about whether or not the data is anomalous. Outlier scores can be used to provide binary labels by implementation of threshold values for anomalousness.

The mention of binary labels implies a classical problem of binary classification within the field of un-/supervised learning. However, this is misleading. For instance, it can be assumed that the number of data points denoting normal behavior are

far more numerous than those denoting anomalous behavior. In this case, classical classification algorithms would produce too many false negatives to be of practical use. Additionally, it is virtually impossible to know all types of anomalous behavior that may occur in a system; therefore, a binary classifier would not be capable of detecting anomalous behavior for which it has not been trained. Furthermore, anomalies can lie in the same range as normal data but be anomalous on the basis of neighboring data instead. These challenges necessitate the development and implementation of specialized algorithms in the field of anomaly detection. [3] [14]

Even with this caveat in mind, the results of an anomaly detection algorithms can be viewed as analogous to the results of a binary classifier, as mentioned earlier. As no algorithm can be 100% correct, the results can be divided into 3 familiar cases: [9]

- **Correct detection**: a detected anomaly is correctly identified as such.
- **False positives**: a detected anomaly is a true normal data point
- **False negatives**: a true anomaly is not identified as such

False positives and negatives can occur owing to noise in the system that, for example, prevents an anomaly from crossing a threshold the algorithm would detect [9].

B. Evaluation Metrics for Anomaly Detection Algorithms

The existence of falsely classified data affords the opportunity to evaluate and compare multiple anomaly detection algorithms with one another. Three relatively simple metrics that achieve this are:

- Precision
- Recall
- Rank-Power

Precision is defined as the ratio of true positives to the sum of the true and false positives. **Recall** is defined as the ratio of true positives to the sum of true positives and false negatives. Equations for both metrics are in 1 and 2 respectively. [15]

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

As an aside, it may be desirable to have a higher precision at the expense of lower recall or vice versa, depending on whether the objective is to minimize false positives or negatives respectively [16].

A metric that utilizes outlier scores in its calculation is the rank-power metric. Here, anomalies are ranked by anomalousness score, with L_i representing the rank of the i-th true outlier. The value is maximized and equal to 1 when all true outliers are identified and ranked in the correct order by the algorithm. The calculation of this metric can be seen in equation 3. [17]

$$RankPower(m) = \frac{n(n+1)}{2 \cdot \sum_{i=1}^{n} L_i} \tag{3}$$

Here, m represents the number of outliers returned by the anomaly detection algorithm, and n represents the number of true outliers in those m values.

In addition to the aforementioned methods to calculate outlier scores, it is important to consider optimizations, and evaluate algorithms on the basis of computational effort. [9] This evaluation can take place with metrics such as the Akaike Information Criterion (AIC) [18], Bayesian Information Criterion (BIC) [19], and Minimum Description Length (MDL) [20], among others. These metrics penalize model complexity either on the basis of number of parameters or bits required to represent the model and are outside the scope of the paper.

C. Approaches for Anomaly Detection and Problem Classes

Mehrotra et al. [9] - [12] define 3 approaches for anomaly detection, namely:

- **Distance-based**: anomalies are detected on the basis of distance from other data points.
- **Density-based**: all data is clustered; points in low-density areas/clusters are marked as anomalies.
- **Model-based**: assumptions are made about the type of generation process behind the data; true values are compared with predictions to identify anomalous data.

Additionally, the problem of anomaly detection may be tackled with different levels of supervision, namely: [9]

- **Supervised**: a training dataset is used to generate test data that acts as a prediction of system behavior. Aggarwal [21] explores the implementation of supervised learning methods to detect anomalies in multi-dimensional data.
- **Unsupervised**: no training dataset is used; attributes of the entire data set are used instead to make predictions. Rebbapragada et al. [22] present an unsupervised process of anomaly detection in periodic time series.
- **Semi-supervised**: Some data can be used for training, however, not all information about the status of outliers is known.

In the following, the aforementioned approaches for anomaly detection are discussed in more detail.

1) Distance-Based Measures: Distance-based measures apply the intuitive definition of outliers by detecting points that are distant from others in the input space [10]. Mehrotra et al. [10] list multiple methods of calculating distance, which include but are not limited to Euclidean and Mahalanobis [23] distances, which have different benefits ranging from ease of use to the choice of application [24].

Further, Mehrotra et al. [10] lay out and describe the following distance-based approaches:

- Distance to all points
- Distance to nearest neighbor
- Average/median distance to k nearest neighbors

Calculating the distance between one point and all others for each point in the database appears to be computationally

expensive. However, the sum of all the distances produces an intuitive metric for the anomalousness of a point. A simple example of this anomaly detection algorithm can be viewed in figure 1.

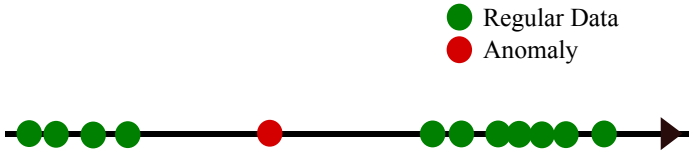

Fig. 1. Anomaly Detection by calculating distance between a point and all other points for all points in a database.

The distance to the nearest neighbor assumes that the furthest anomaly is the point that has the largest distance to its nearest neighbor. This is another intuitive approach. However, it ignores cases where anomalies may occur in groups/clusters.

The average/median distance to k nearest neighbors is a more sophisticated approach that inherits its logic from the nearest neighbor approach. All approaches listed here produce outlier scores for the investigated point by way of ranking the distance calculation. [10]

2) Density-Based Measures: Density or cluster-based approaches are similar to distance-based approaches. The clustering method assumes that similar points should belong to the same cluster, and points are assigned a degree of membership to any given cluster. [11] Jain et al. [25] discuss multiple types and approaches to clustering in detail. This paper will not delve into clustering algorithms. However, the following is a brief discussion of anomaly detection with clustering methods.

A hard clustering algorithm like k-Means [26] assigns each point to a cluster; an example clustering can be seen in figure 2. Therefore, anomaly detection can take place on the basis of the distance of a point to the centroid of its cluster, which assumes that clusters tend to be symmetric. A threshold for the size of a cluster can also be set, with points in the cluster out of bounds considered to be anomalies. Another approach is to compute the distance between a point and the boundary of a cluster, with a smaller distance implying the point being an outlier. [11]

3) Model-Based Measures: Distance-based and clustering approaches are prone to be affected by the curse of dimensionality [27]. This is especially true for clustering approaches that utilize an intrinsic distance-based measure, for example for the case of distance of cluster centroids mentioned above [28].

On the other hand, model-based approaches attempt to capture the process responsible for the generation of the available data. Examples of such approaches are linear regression, support vector machines, splines, and many others. The idea behind capturing the data generation process is to make predictions on that basis and compare actual data with predictions to identify anomalous occurrences. This is illustrated in figure 3. These approaches are especially important and useful in modeling time series data, as measurement of distance may not be meaningful when considering chronological data. [9] [12]

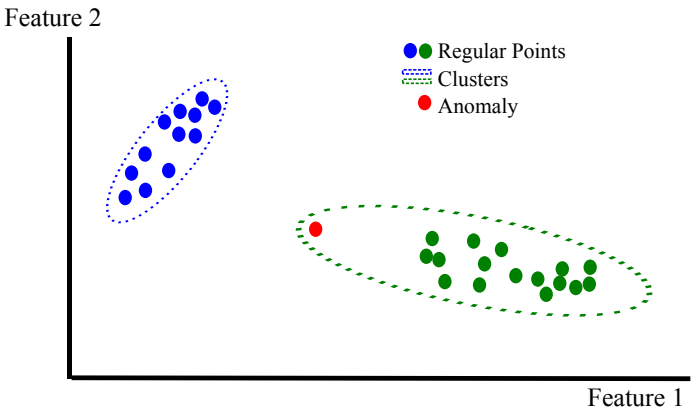

Fig. 2. Anomaly Detection by calculating distance between a point and the cluster boundary or centroid. The anomaly need not be part of a cluster; however, this depends on the type of clustering algorithm used.

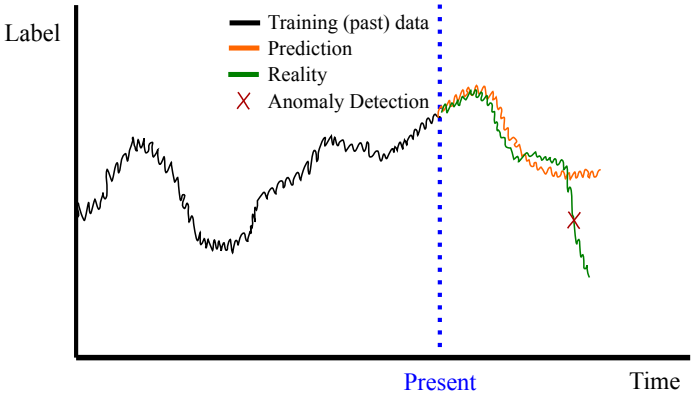

Fig. 3. A model-based approach attempts to predict the behavior of data using a training set. In this figure, the data looks like a time series, the true values of which begin to deviate from the predictions after a point in the future. After a threshold between prediction and truth is crossed, the anomaly is detected.

In the next section, time series are discussed along with their applications. The uniqueness of time series data as well as that of the problem of anomaly detection with such data is also discussed.

IV. Time Series Data

A time series is a series that indexes its data in chronological order. Data in time series have 3 attributes: the time, which is used to order the series, the label, which indicates the type of signal captured, and the values, which are the values of a label at each given time. [12] Shumway et al. [29] outline several applications of time series, for which the utility of anomaly detection can be intuitively extracted. Two examples are discussed below:

Financial data: The DOW Jones industrial average is a time series dataset, where a label may be the returns, the time vector with an entry for each day, and the values containing the size of returns for each day. According to [29], the 2008 financial crisis is easy to discern from the time series data. Needless to say, the detection of an anomalous sequence in the time series data would have been of massive use.

Science: Critical paradigm shifts such as climate change can be observed with time series. Observing trends in the available data can have ramifications for the lives of billions of people.

Box et al. [30] list five areas of research concerning time series. These are:

- **Forecasting**: prediction of future values on the basis of past values. This is also an important aspect of anomaly detection in time series.
- **Determining the transfer function**: for a given set of inputs, the effect on the output can be demonstrated.
- **Usage of indicator input values**: This is used to evaluate the effects of intervention events on the data in time series.
- **Relationship between multiple time series datasets**: For example, do two time series have trends in lockstep with one another?
- **Control design**: The compensation of input values if outputs deviate from the desired range.

Of these areas, forecasting and the relationship between multiple time series are relevant to this paper. For the latter, Gupta et al. [8] and Mehrotra et al. [9] argue that anomaly detection in time series can be classed in two problem sets:

- **Single time series**: all data points have the same label; anomalies are either singular points or subsequences of the time series.
- **Multiple time series**: there exist many sets of time series data; the anomaly is a time series dataset that does not appear to behave similar to the others.

V. Anomaly Detection in Time Series: Approaches

Various approaches to detect anomalies in time series exist, depending on whether the data at hand is a singular time series or a collection of time series data sets. Some of these approaches are introduced below.

A. Singular Anomalous Point in Single Time Series

Mehrotra et al. [12] outline an intuitive method to detect a singular anomalous point by a geometric measure. Here, the perpendicular distance between the investigated point and the line segment connecting its predecessor (data point for the previous time index) and its successor (next time index) is measured. This process is illustrated in figure 4 and is an algorithm that can produce outlier scores. [12] also states that one disadvantage of this approach is that the first and last points in the time series cannot be investigated as they have no predecessor and successor respectively. Additionally, the approach is only capable of detecting singular anomalous points (as opposed to anomalous subsequences or time series), making its application domain largely restrictive.

B. Anomalous Subsequences or Time Series Within Multiple Time Series

Detection of an anomalous subsequence or time series presents a more challenging problem that has been tackled by, among others, the approaches that are discussed in this section. A comparative evaluation of time series data by Chandola et al. [31] groups a majority of anomaly detection algorithms in three classes:

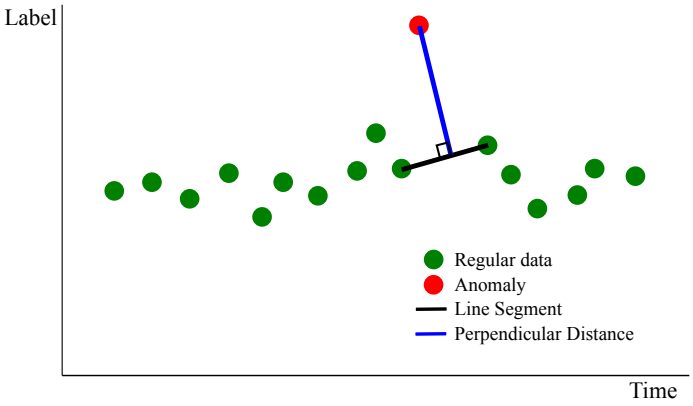

Fig. 4. A method to calculate the anomalousness of a given single point in a time series. The perpendicular distance between the point and the line segment connecting its predecessor and successor determine the anomalousness. [12]

- **Kernel-based**: Distance- and density-based approaches discussed in III-C1 and III-C2 belong in this group. Two distance-based methods specifically for time series will be discussed in V-D.
- **Window-based**: Slices of the time series data are taken and predictions are made on the test data. Though not explicitly mentioned in [31], it can be hypothesized that models such as ARIMA, which will be discussed in detail, belong to this group.
- **Markovian**: Subsequences are assigned a probability given preceding data/observations [12]. An example of a Markovian technique used for anomaly detection in time series is a paper by Gao et al. [32]. Such techniques are outside the scope of this paper.

C. ARIMA - Autoregressive Integrated Moving Average

In this section, one method of anomaly detection in time series is chosen and explored. ARIMA is a relevant choice as it combines multiple statistical models and ideas, from which it inherits the capability to capture many features of time series. [12]

The model attempts to predict the behavior of time series data using past (training data). The output of this algorithm can therefore be considered analogous to that expressed in figure 3. Anomaly detection therefore is a matter of comparing predictions with reality, and flagging reality as anomalous if it deviates too far from predicted behavior. [12]

1) AR - Autoregression: The first part of the term is 'AR', short for autoregressive. The core idea of an autoregressive model is that the value at time t, x_t, is a function of p past values of the time series. The parameter p is the order of the autoregressive model, whose general form can be expressed as follows in equation 4. [33]

$$x_t = \alpha + \phi_1 \cdot x_{t-1} + \phi_2 \cdot x_{t-2} + ... + \phi_p \cdot x_{t-p} + w_t \quad (4)$$

Where:

- α is the mean of the time series (which is stationary, i.e. the mean value is stable [34]).
- $[\phi_1 ... \phi_p]$ are the parameters of the regressor.
- $\phi_p \neq 0$.
- w_t is noise.

Calculating the parameter p involves the use of the partial (Pearson) autocorrelation function (PACF) [33].

2) MA - Moving Average: Moving average is an technique that utilizes the noise term w_t over q past values to evaluate the model. It can be summarized generally in an equation that looks similar to that of the AR model (equation 5). [33]

$$x_t = \alpha + \theta_1 \cdot w_{t-1} + ... + \theta_q \cdot w_{t-q} \tag{5}$$

Where:

- α is the mean; the time series is stationary.
- w_{t-q} is the noise q values ago.
- $[\theta_1 ... \theta_q]$ are the parameters of the model.
- $\theta_q \neq 0$.

In short, the moving average model can be thought of as an iterative process that, for a current point, weights the deviations of previous points to judge what the value at the current point should be.

3) I - Integrated: Equations for $AR(p)$ (4) and $MA(q)$ (5) can be added together to produce an $ARMA(p, q)$ model. This has the following benefits: [12]

- Value is dependent on its history.
- There is a mechanism for assigning greater importance to recent values preceding the investigated value (by way of choosing p and q).
- The MA component smooths the time series signal and eliminates some noise.
- The equation includes the effects of random noise by way of $w_t, ...$

More about ARMA models can be read in [35] and [36].

However, both models require the time series to be stationary; i.e. the time series cannot have any trend or seasonality. This means that such a model cannot be used for a variety of applications, such as quarterly earnings, a rising stock, census data, etc. [34].

Box et al. [37] outline the $ARIMA(p, d, q)$ model, asserting that the use case is for nonstationary time series. The 'I' component has a parameter d, which signifies the number of times the difference is taken between the points in the time series. [37] further assumes that d is chosen such that the d-th difference of the time series results in a time series that can be represented and modeled by a (stationary) ARMA model.

The ARIMA model has benefits in addition to those of ARMA [12]:

- Non-stationarity.
- Calculation of a term that signifies the drift of the time series over time.

To summarize, the $ARIMA(p, d, q)$ model can be tuned by choosing appropriate values for parameters p, d, and q. The result is a model that uses available time series data to make a prediction about the future. When that data is available, it is compared with the predictions in order to detect anomalies. If a significant deviation such as the one visible in figure 3 is visible, then the data in the time series is flagged as anomalous.

4) Extensions of the ARIMA Model: There are further extensions of the ARIMA model, such as: [12]

- **VARIMA**: The values observed at a time point are **v**ectors instead of scalars. For further research, see [38].
- **SARIMA**: The **s**easonality of the time series data (yearly, monthly, etc.) is modeled explicitly. See [39].
- **ARFIMA**: **F**ractional values of d are allowed. See [40].

D. Transformational Anomaly Detection Algorithms and Dimensionality Reduction

In addition to predictive models such as ARIMA, there are transformational algorithms that attempt to reduce the time series data into a multidimensional vector, upon which 'traditional' distance-based algorithms can be applied [4]. Two examples of these, along with resources for further research, are: [4] [12]

- Discrete Fourier Transformation (DFT). See [41].
- Discrete Wavelet Transformation (DWT). See [42].

Keogh et al. [43] discuss the need for dimensionality reduction, as time series data size can be prohibitively high concerning computational expense. The idea behind dimensionality reduction according to [43] is to create a compact approximation of the time series data that contains its important features.

DFT and DWT are examples of methods for dimensionality reduction [44].

Some additional methods of dimensionality reduction (when comparing multiple time series) are presented here:

1) STREND - SameTrend: Huang et al. [45] define a measure to calculate anomalousness between time series in lockstep with one another using an intuitive discrete function. Firstly, for each time series, the difference between a chosen point and its predecessor, i.e. $\Delta x(t) = x(t+1) - x(t)$ is taken for all but the last sample of the time series (to stay within bounds). The comparison between two time series, $x(t)$ and $y(t)$ takes place with the function $S(t)$ defined in equation 6. [13]

$$S(t) = \begin{cases} 1 & \text{if } \Delta x(t) \cdot \Delta y(t) > 0 \\ -1 & \text{if } \Delta x(t) \cdot \Delta y(t) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Therefore, the sign of the STREND function indicates whether two time series move in the same direction at various times t. These values can be aggregated over all times between two time series $\mathcal{X}$ and $\mathcal{Y}$; this is shown in equation 7 and can be used as a scalar metric to evaluate anomalousness of time series. The lower the value, the more closely the two time series move in lockstep. [13]

$$d_S(\mathcal{X}, \mathcal{Y}) = 1 - \frac{1}{n-1} \sum_{t=1}^{n-1} S(t) \tag{7}$$

A disadvantage of the STREND measure is its limited flexibility. Although the metric is easy to implement, it cannot be directly used in time series with lagging. [13]

2) DIFFSTD - Standard Deviation of Differences: A similar measure to STREND is DIFFSTD, where the standard deviation of the differences between two time series $\mathcal{X}$ and $\mathcal{Y}$ are calculated; this can be seen in equation 8. [13]

$$dist(\mathcal{X}, \mathcal{Y}) = \sqrt{\sum_t \frac{(\delta(t) - \mu)^2}{n}} \tag{8}$$

where:

- $\delta(t) = |x(t) - y(t)|$
- $\mu = \sum_t \frac{\delta(t)}{n}$

E. Other Algorithms and Best Practice

Multiple dimensionality reduction algorithms can be used depending on the nature of the time series being observed. Examples are: [13]

- **EUC** (Cross-Euclidean Distance): Easy to implement, used only for time series in lockstep. See [46].
- **SAXBAG** (Symbolic Aggregate Approximation, Bag-of-Features): Can be used with time series with lag, and can tolerate noise given a standard deviation threshold. However, it cannot detect anomalies of different shapes that look similar when transformed into the frequency domain. See [47].

Mehrotra et al. [13] assert that three dimensionality reduction measures should be used in the process of anomaly detection in time series. These are chosen with the a-priori knowledge of the capabilities of each measure, in order to avoid overlap and maximize functionality. Each measure is capable of detecting anomalies in a domain that the others may not.

VI. Applications in Predictive Maintenance

Predictive Maintenance (PDM) is a strategy involving the prediction of failure or errors and the remaining useful lifetime of a technical system. Using these predictions, maintenance measures are to be carried out. This proactive strategy is a more efficient use of possibly expensive resources, and so is of great interest in the industrial world. [48]

An example of a technical system that can benefit from PDM is any mechanical system consisting of (rolling) bearings. These components are used in a wide variety of applications, such as gearboxes and power trains in the automotive industry. Determining whether or not bearings will continue to be functional is therefore an important aspect in determining the reliability of the applications. The following possible types of failures are possible with bearings: fatigue, abrasion, adhesive wear, electrical corrosion, plastic deformation, and fracture [49].

Various studies have been conducted that demonstrate the importance of proper bearing maintenance. It is argued that over 40% of all machine failures are caused by bearing faults [50].

Thus, the problem of detecting anomalies in bearing behavior cannot be thought of as a binary classification issue. Rather, it is a one-class classification issue, which is for example explored in [51]. This is so because normal behavior is known, but *all* various sources of abnormal behavior are not. Such a problem falls into the domain of anomaly detection, as has been explained earlier in III-A. [52].

In addition, another reason why anomaly detection approaches are a good fit for the field of predictive maintenance is the disparity of sample sizes of 'normal' and anomalous behavior. As normal behavior would be present in far higher proportions to anomalous behavior in most cases, a binary classifier would suffer the aforementioned problem of too many false negatives. It would thus offer poor information about the status of a rolling bearing. [3] [14]

The application of anomaly detection in time series in the field of predictive maintenance, specifically here concerning rolling bearings, can be conjectured as follows: sensor data in the form of a time series is generated by the mechanical system, with a label that offers information about the status of the bearings. This data is used to train an ARIMA model to predict sensor data for the next day. The next day, true data is compared with predictions. According to thresholds set by the user, an anomaly can be detected if the real data deviates from predictions. The user may consider additional anomaly detection approaches, such as DIFFSTD, STREND, etc. using other available sensor data in order to make a more informed decision about whether or not the detected anomaly can be trusted. If yes, then maintenance can be performed on the machine that produces the sensor data. If no, then a possibly unnecessary and expensive equipment replacement that would have been mandated by preventive maintenance practices would be prevented. The industry also saves valuable resources by having implemented a relatively more automated procedure of detecting machine failure (as opposed to, for example, a person hearing machine parts progressively making more noise). The use of sophisticated anomaly detection techniques would also offer insight that superficial human observation could not. [53]

VII. Conclusion

This literature review introduced the reader to the field of anomaly detection, discussing its utility, evaluation metrics, and various classes of methods that can be applied. Then, time series data and its uniqueness to the problem of anomaly detection was introduced, together with specific algorithms (such as ARIMA) used to detect anomalies in time series. Additionally, various approaches to make this process computationally viable were introduced (transformational algorithms and dimensionality reductions). Then, the problem was

discussed in the increasingly important context of Predictive Maintenance, using bearing fault analysis as an example to help illustrate the usefulness of the field.

It is hoped that this review offers an informative introduction into the field, and that the sources provided are good for deeper research into particular topics.

As this area of research is relatively novel, there are multiple areas where further research can be conducted to tackle the problem better. One such example is outlined by Blasquez et al. [54]. In this work, it is suggested that the calculation of thresholds to determine the anomalousness of a point, subsequence, or time series is usually done manually. Multiple thresholds can be chosen by people working on the same data set, and this would produce different results. Therefore, the paper suggests research into the area of dynamic and adaptive selection of thresholds, where no prior specification is required by the data scientist. The aim of such an approach would be to standardize anomaly detection and reduce the arbitrary nature of the choice of threshold.

Lastly, for a more hands-on look into anomaly detection, Google Trends [55] can be used to obtain time series data about searching activity.

REFERENCES

[1] D. M. Hawkins, 'Introduction', in Identification of Outliers, D. M. Hawkins, Ed. Dordrecht: Springer Netherlands, 1980, pp. 1–12. doi: 10.1007/978-94-015-3994-4_1.

[2] C. C. Aggarwal, 'An Introduction to Outlier Analysis', in Outlier Analysis, C. C. Aggarwal, Ed. Cham: Springer International Publishing, 2017, pp. 1–34. doi: 10.1007/978-3-319-47578-3_1.

[3] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Introduction', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 3–19. doi: 10.1007/978-3-319-67526-8_1.

[4] C. C. Aggarwal, 'Time Series and Multidimensional Streaming Outlier Detection', in Outlier Analysis, C. C. Aggarwal, Ed. Cham: Springer International Publishing, 2017, pp. 273–310. doi: 10.1007/978-3-319-47578-3_9.

[5] D. Dasgupta and S. Forrest, "Novelty Detection in Time Series Data using Ideas from Immunology", 1999. Available: https://forrest.biodesign.asu.edu/data/publications/1999-dasgupta-novelty-detection.pdf. [Accessed 30 May 2021].

[6] V. J. Hodge and J. Austin, 'A Survey of Outlier Detection Methodologies', Artif Intell Rev, vol. 22, no. 2, pp. 85–126, Oct. 2004, doi: 10.1007/s10462-004-4304-y.

[7] V. Chandola, A. Banerjee, and V. Kumar, 'Anomaly Detection: A Survey', ACM Comput. Surv., vol. 41, Jul. 2009, doi: 10.1145/1541880.1541882.

[8] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, 'Outlier Detection for Temporal Data: A Survey', IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 9, pp. 2250–2267, Sep. 2014, doi: 10.1109/TKDE.2013.184.

[9] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Anomaly Detection', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 21–32. doi: 10.1007/978-3-319-67526-8_2.

[10] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Distance-Based Anomaly Detection Approaches', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 33–39. doi: 10.1007/978-3-319-67526-8_3.

[11] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Clustering-Based Anomaly Detection Approaches', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 41–55. doi: 10.1007/978-3-319-67526-8_4.

[12] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Model-Based Anomaly Detection Approaches', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 57–94. doi: 10.1007/978-3-319-67526-8_5.

[13] K. G. Mehrotra, C. K. Mohan, and H. Huang, 'Algorithms for Time Series Data', in Anomaly Detection Principles and Algorithms, K. G. Mehrotra, C. K. Mohan, and H. Huang, Eds. Cham: Springer International Publishing, 2017, pp. 153–189. doi: 10.1007/978-3-319-67526-8_9.

[14] T. Olsson, E. Kallstrom, D. Gillblad, and P. Funk, 'Fault Diagnosis of Heavy Duty Machines: Automatic Transmission Clutches', p. 10, Sep. 2014.

[15] K. M. Ting, 'Precision and Recall', in Encyclopedia of Machine Learning, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 781–781. doi: 10.1007/978-0-387-30164-8_652.

[16] M. Santos, "Explaining Precision vs. Recall to Everyone," Medium, 06-Jun-2020. [Online]. Available: https://towardsdatascience.com/explaining-precision-vs-recall-to-everyone-295d4848edaf. [Accessed: 30-May-2021].

[17] J. Tang, Z. Chen, A. Fu, and D. Cheung, 'Capabilities of outlier detection schemes in large datasets, framework and methodologies', Knowledge and Information Systems, vol. 11, pp. 45–84, Jan. 2007, doi: 10.1007/s10115-005-0233-6.

[18] H. Akaike, 'A new look at the statistical model identification', IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716–723, Dec. 1974, doi: 10.1109/TAC.1974.1100705.

[19] G. Schwarz, 'Estimating the Dimension of a Model', The Annals of Statistics, vol. 6, no. 2, pp. 461–464, Mar. 1978, doi: 10.1214/aos/1176344136.

[20] J. Rissanen, "Modeling by shortest data description." Automatica 14(5), 465-471 (1978)

[21] C. C. Aggarwal, 'On Abnormality Detection in Spuriously Populated Data Streams', in Proceedings of the 2005 SIAM International Conference on Data Mining (SDM), 0 vols, Society for Industrial and Applied Mathematics, 2005, pp. 80–91. doi: 10.1137/1.9781611972757.8.

[22] U. Rebbapragada, P. Protopapas, C. E. Brodley, and C. Alcock, 'Finding Anomalous Periodic Time Series: An Application to Catalogs of Periodic Variable Stars', Mach Learn, vol. 74, no. 3, pp. 281–313, Mar. 2009, doi: 10.1007/s10994-008-5093-3.

[23] P. C. Mahalanobis, "On the Generalized Distance in Statistics," vol. II, no. 1, Apr. 1936.

[24] J. Walters-Williams and Y. Li, 'Comparative Study of Distance Functions for Nearest Neighbors', in Advanced Techniques in Computing Sciences and Software Engineering, Dordrecht, 2010, pp. 79–84. doi: 10.1007/978-90-481-3660-5_14.

[25] A. K. Jain and R. C. Dubes, Algorithms for clustering data. USA: Prentice-Hall, Inc., 1988.

[26] S. Lloyd, 'Least squares quantization in PCM', IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.

[27] Keogh E., Mueen A. (2017) Curse of Dimensionality. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_192

[28] 'k-Means Advantages and Disadvantages — Clustering in Machine Learning', Google Developers. https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages (accessed May 31, 2021).

[29] R. H. Shumway and D. S. Stoffer, 'Characteristics of Time Series', in Time Series Analysis and Its Applications: With R Examples, R. H. Shumway and D. S. Stoffer, Eds. Cham: Springer International Publishing, 2017, pp. 1–44. doi: 10.1007/978-3-319-52452-8_1.

[30] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Introduction," in Time Series Analysis: Forecasting and Control, Hoboken, NJ: Wiley, 2016.

[31] V. Chandola, V. Mithal, and V. Kumar, "Comparative Evaluation of Anomaly Detection Techniques for Sequence Data," in 2008 Eighth IEEE International Conference on Data Mining, Dec. 2008, pp. 743–748. doi: 10.1109/ICDM.2008.151.

[32] B. Gao, H.-Y. Ma, and Y.-H. Yang, "HMMs (Hidden Markov models) based on anomaly intrusion detection method," in Proceedings. International Conference on Machine Learning and Cybernetics, Nov. 2002, vol. 1, pp. 381–385 vol.1. doi: 10.1109/ICMLC.2002.1176779.

[33] R. H. Shumway and D. S. Stoffer, “ARIMA Models,” in Time Series Analysis and Its Applications: With R Examples, R. H. Shumway and D. S. Stoffer, Eds. Cham: Springer International Publishing, 2017, pp. 75–163. doi: 10.1007/978-3-319-52452-8_3.
[34] 8.1 Stationarity and differencing — Forecasting: Principles and Practice (2nd ed). Accessed: Jun. 02, 2021. [Online]. Available: https://Otexts.com/fpp2/
[35] “Related Statistical Techniques,” in Robust Regression and Outlier Detection, John Wiley & Sons, Ltd, 1987, pp. 248–291. doi: 10.1002/0471725382.ch7.
[36] J. D. Hamilton, “Stationary ARMA Processes,” in Time Series Analysis, vol. I, Princeton, NJ: Princeton University Press, 1994, pp. 43–71.
[37] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, “Linear Nonstationary Models,” in Time Series Analysis: Forecasting and Control, Hoboken, NJ: Wiley, 2016.
[38] W. W. S. Wei, “Vector Time Series Models,” in Time Series Analysis Univariate and Multivariate Methods, 2nd ed., Pearson.
[39] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, “Analysis of Seasonal Time Series,” in Time Series Analysis: Forecasting and Control, Hoboken, NJ: Wiley, 2016.
[40] J.W. Galbraith, V.Zinde-Walsh, “Autoregression-based estimators for ARFIMA models”, CIRANO Working Papers, No: 2011s-11, Feb 2001.
[41] A. Collins Jackson and S. Lacey, “Seasonality and Anomaly Detection in Rare Data Using the Discrete Fourier Transformation,” in 2019 First International Conference on Digital Data Processing (DDP), Nov. 2019, pp. 13–17. doi: 10.1109/DDP.2019.00013.
[42] M. Thill, W. Konen, and T. Bäck, “Time Series Anomaly Detection with Discrete Wavelet Transforms and Maximum Likelihood Estimation,” Jan. 2019.
[43] E. Keogh and S. Kasetty, “On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration,” Data Mining and Knowledge Discovery, vol. 7, no. 4, pp. 349–371, Oct. 2003, doi: 10.1023/A:1024988512476.
[44] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, “Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures,” p. 11, Aug. 2008.
[45] H. Huang, K. Mehrotra, C. Mohan, “Detection of anomalous time series based on multiple distance measures,” in 28th International Conference on Computers and Their Applications (CATA-2013), Honolulu, Hawaii, USA, 2013
[46] C. Faloutsos, M. Ranganathan, Y. Manolopoulos., “Fast subsequence matching in time- series databases,” in Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, New York, NY, USA, pp. 419–429, 1994
[47] J. Lin, R. Khade, and Y. Li, ‘Rotation-invariant similarity in time series using bag-of-patterns representation’, J Intell Inf Syst, vol. 39, no. 2, pp. 287–315, Oct. 2012, doi: 10.1007/s10844-012-0196-5.
[48] R. Bink and P. Zschech, ‘Predictive Maintenance in der industriellen Praxis’, HMD, vol. 55, no. 3, pp. 552–565, Jun. 2018, doi: 10.1365/s40702-017-0378-2.
[49] G. Jacobs and M. Plogmann, ‘Rolling Bearings: Overview’, in Encyclopedia of Lubricants and Lubrication, T. Mang, Ed. Berlin, Heidelberg: Springer, 2014, pp. 1655–1663. doi: 10.1007/978-3-642-22647-2_2.
[50] I. Y. Önel, K. B. Dalci, and İ. Senol, ‘Detection of bearing defects in three-phase induction motors using Park’s transform and radial basis function neural networks’, Sadhana, vol. 31, no. 3, pp. 235–244, Jun. 2006, doi: 10.1007/BF02703379.
[51] D. Tax, ‘One-Class Classification; Concept-Learning In The Absence Of Counter-Examples’, Jan. 2001.
[52] G. Georgoulas and G. Nikolakopoulos, ‘Bearing fault detection and diagnosis by fusing vibration data’, in IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Oct. 2016, pp. 6955–6960. doi: 10.1109/IECON.2016.7794118.
[53] Vibration Analysis - Bearing Failure Analysis by Mobius Institute. 2016. [video] Directed by J. Tranter. Australia: Mobius Institute. Accessed: Jun. 14, 2021. [Online]. Available: https://youtu.be/dEn2Qvh_qjc
[54] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, ‘A review on outlier/anomaly detection in time series data’, arXiv:2002.04236 [cs, stat], Feb. 2020, Accessed: Jun. 14, 2021. [Online]. Available: http://arxiv.org/abs/2002.04236
[55] ‘Explore what the World is Searching’, Google Trends. https://trends.google.com/trends/ (accessed Jun. 14, 2021).