The Dialectics of Unsupervised Learning: A Synthesis of Anomaly Detection Methodologies

By ***Dr. Yu Han***
Associate Professor of Computer Science, Shanghai Jiao Tong University, China

1. Introduction to Unsupervised Learning

The most frequently used category of learning is by far supervised learning. Supervised learning has become a field of applied research everywhere where computation facilities are available. So why do we care about unsupervised learning? Can unsupervised learning do what supervised learning can do (classification) and does every learned classifier have a supervised learning's unsupervised counterpart? The answer to this last question concerns the family of the product of base classifiers, namely the ensemble classifier. Unlike classifiers learned by other supervised learning methodologies, ensemble classifiers do not always have a ready-made unsupervised counterpart. This is so because they differ from models learned by other supervised methodologies in that their learning objective is indirect training error minimization. Unsupervised learning, while often overshadowed by the prominence of supervised learning, offers unique advantages and capabilities that cannot be ignored. It enables us to venture into uncharted territories and uncover hidden patterns and structures within data without the need for explicit labels or guidance. By autonomously exploring and analyzing the intrinsic properties of the data, unsupervised learning opens up a whole new dimension of possibilities. In the realm of classification tasks, supervised learning has long held the crown as the go-to approach. However, this does not negate the potential of unsupervised learning to tackle classification challenges. While unsupervised learning may not directly mirror the classification capabilities of supervised learning, it possesses the ability to discover underlying patterns and groupings that can indirectly contribute to classification tasks. Think of unsupervised learning as a powerful detective, tirelessly working to uncover the subtle nuances and relationships within the data, ultimately aiding in the classification process. Now, let's address the query regarding the existence of unsupervised counterparts for every learned classifier in supervised learning. While it would be ideal for every supervised learning classifier to have a neatly packaged unsupervised counterpart, this is not always the case. Enter the ensemble classifier, a family of base classifiers that combine their collective wisdom to make highly accurate predictions. Unlike classifiers learned through traditional supervised methodologies, ensemble classifiers have a distinct learning objective: the minimization of training error through an indirect approach. This distinction in learning objectives is what sets ensemble

This work is licensed under CC BY-NC-SA 4.0.

classifiers apart and makes the creation of a ready-made unsupervised counterpart more challenging. The intricate nature of ensemble learning necessitates a specialized unsupervised counterpart that can effectively capture the collaborative wisdom of the base classifiers. Despite this hurdle, researchers continue to push the boundaries of unsupervised methods in the quest for adaptable and robust ensemble classifiers that can empower diverse applications. In conclusion, while supervised learning reigns supreme in popularity, the importance of unsupervised learning should not be underestimated. From unlocking hidden insights to indirectly aiding in classification tasks, unsupervised learning offers a rich landscape of possibilities. Although every supervised learning classifier may not have a readily available unsupervised counterpart, the ensemble classifier stands as a testament to the evolving nature of learning methodologies. As the field progresses, the synergy between supervised and unsupervised learning continues to pave the way for exciting advancements in the realm of machine learning.

Unsupervised learning, on the contrary, is not solely about labels - it is primarily focused on discovering relevant and valuable information within vast amounts of raw data. This process of extracting meaningful insights from unlabeled data may initially seem abstract, but its concrete applications become more apparent when one understands that the primary objectives of unsupervised learning are category induction and automatic component discovery. These crucial tasks often involve working with data sets for which labels or predetermined classifications are simply not available, emphasizing the fundamental nature and significance of unsupervised learning methods. Unsupervised learning algorithms play a vital role in various fields, including data mining, pattern recognition, and computational biology. In the realm of data mining, these algorithms help identify hidden patterns, relationships, and structures that exist within unexplored datasets. By analyzing the inherent properties of the data, unsupervised learning techniques enable researchers to gain valuable insights and make informed decisions. For example, in the field of pattern recognition, unsupervised learning algorithms can be used to group similar data points together, allowing for the creation of meaningful clusters. Furthermore, unsupervised learning methods are extensively utilized in computational biology, where they contribute to the understanding of biological systems and processes. By analyzing large-scale genomic and proteomic datasets, researchers can identify underlying structures and relationships that guide the functioning of living organisms. This knowledge facilitates advancements in personalized medicine, disease prevention, and genetic engineering, among other critical areas. In the realm of artificial intelligence, unsupervised learning serves as a foundation for many advanced techniques, such as generative modeling and anomaly detection. Generative models, such as Variational Autoencoders and Generative Adversarial Networks, learn the underlying distribution of the data and can generate new, realistic samples. This capability has far-reaching applications, including image synthesis, text generation, and even music composition. Anomaly detection, on the other hand, involves identifying unusual patterns or outliers within a dataset. Unsupervised learning algorithms

This work is licensed under CC BY-NC-SA 4.0.

enable the detection and classification of anomalous instances, which can be useful for fraud detection, network security, and equipment maintenance. By leveraging unsupervised learning, organizations can proactively identify and mitigate potential risks or issues before they escalate. In conclusion, unsupervised learning holds immense potential in various domains, offering a means to extract valuable knowledge from unlabeled data. As researchers and practitioners continue to develop and improve unsupervised learning algorithms, the impact and significance of these methods will only grow. By uncovering hidden patterns and structures, unsupervised learning enables us to make sense of complex datasets, driving innovation and progress in numerous fields.

Different types of learning are often very different. Unsupervised, supervised, semi-supervised, and reinforcement learning are all quite different in research objectives and learning methodologies employed. Both supervised and unsupervised learning have become rather mature fields that have undergone extensive research and development in recent years. The former, supervised learning, focuses on learning to map input data to human-assigned labels. This involves training a model using labeled data examples, with the goal of enabling the model to accurately classify new, unseen data based on its learned patterns and relationships. Input data refers to the kind of input features in which the label prediction is interested. Input features can take various forms, such as instances represented in a feature space, contexts in which a sequence is interested, or states that are associated with scheduled operations in a dynamic system. These input features provide valuable information for the model to make accurate predictions and improve its overall performance. On the other hand, unsupervised learning approaches aim to discover patterns, structures, and relationships within the data without any provided labels. By exploring the inherent structure and organization of the data, unsupervised learning algorithms can uncover hidden patterns and gain insights into the data distribution. This type of learning is particularly useful in scenarios where labeled data is scarce or difficult to obtain. Additionally, semi-supervised learning techniques combine elements of both supervised and unsupervised learning. They leverage a small amount of labeled data along with a larger set of unlabeled data to improve the model's performance. By using the labeled data to guide the learning process and the unlabeled data to discover additional patterns and structure, semi-supervised learning strikes a balance between the two approaches. Lastly, reinforcement learning operates differently from both supervised and unsupervised learning. It involves an agent interacting with an environment to learn from trial and error. Through a series of actions and observations of the resulting rewards or punishments, the agent learns to make decisions that maximize the cumulative reward over time. Reinforcement learning has gained significant attention due to its applications in fields such as robotics, game playing, and autonomous systems. With its unique approach to learning, reinforcement learning empowers agents to adapt and improve their decision-making capabilities in dynamic and uncertain environments. Overall, the various types of learning each offer their own set of advantages

This work is licensed under CC BY-NC-SA 4.0.

and techniques for tackling different problems and domains. By understanding the characteristics and methodologies of each approach, researchers and practitioners can leverage the right learning paradigm to address their specific objectives and achieve desired outcomes.

1.1. Definition and Characteristics

By unsupervised learning, we mean a class of problems where some sort of structure or pattern within the input data is being discovered. This definition encompasses a great number of methodologies and their applications, perhaps disproportionately many to pattern recognition, clustering, and other associated problems of automatic processing of readily available data. Remarkably, this field also underlies an equivalent number of intelligent data acquisition systems that autonomously search for new data items whose similarity qualifies them as potential novel members. We will argue and illustrate in this paper that the unsupervised learning problem has a good deal of dialectic interaction with the challenging problem of detecting data anomalies. Unsupervised learning refers to the fascinating realm of problems where an inherent structure or pattern in the provided data is unveiled through ingenious algorithms. The vast landscape of this discipline encompasses a plethora of methodologies, greatly outnumbering those devoted to pattern recognition, clustering, and other related aspects of automated processing of readily available information. Astonishingly, this field also underpins the creation of an equivalent number of sophisticated data acquisition systems capable of autonomously hunting for novel data elements that exhibit a substantive degree of similarity. In the following discourse, we will expound upon and vividly demonstrate the intricate relationship between the unsupervised learning problem and the formidable challenge of effectively detecting data anomalies.

Unsupervised learning plays a crucial role in most contemporary AI applications. We examine several important unsupervised learning strategies in terms of their connection and application to the task of anomaly detection. Besides their technical characteristics, we draw attention to the dialectic nature of the interaction between unsupervised learning systems and the data anomaly detection task. This two-way knowledge transfer appears useful and helps uncover the limitations and particularities of both methods.

1.2. Applications in Various Fields

Anomaly or outlier detection, as is often pointed out, is the problem of finding patterns in data which do not conform to the expected behavior. The definitions of "interest" and "out of the ordinary" depend on the application and context but the problem remains constant: very few samples exist to learn from

This work is licensed under CC BY-NC-SA 4.0.

explicitly labeled "good" (normal behavior) or "bad" (outliers or anomalies) patterns, since an overwhelming majority of examples are unlabelled and are often difficult to interpret for training predictive models. In some sense, the "novelty detection algorithms" are like one-class classification models, meaning one is interested in determining whether a new observation belongs to the category "new normal data" as opposed to other established canonical categories. They generally refer to finding new objects that differ significantly from the objects seen during training.

Abstract: Anomaly detection is crucial for the proper functioning of a vast amount of systems in fields as diverse as e-health, finance, and network security. A diverse range of theories, algorithms, and implementations have addressed this oversight. These studies have taken place in diverse application fields and with various algorithm types. In fact, unsupervised learning provides the tools used in the creation of both anomaly-detection-specific and domain-specific algorithms. In this work, after a profound literature review of the topic, different studies are analyzed with the aim of providing a global view of the field. In this way, mass unsupervised learning is carried out.

Isabel Valle, Department of Computer Architecture and Technology, University of the Basque Country, UPV/EHU, 20009 Donostia–San Sebastián, Spain. Iago Calvo, Department of Computer Science and Artificial Intelligence, University of the Basque Country, UPV/EHU, 20009 Donostia–San Sebastián, Spain. Department of Health of the Basque Government, 01009 Vitoria-Gasteiz, Spain. IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain.

2. Anomaly Detection in Unsupervised Learning

A perfect fraud detection model performs with extremely high accuracy and extremely low latency. Essentially, it would immediately identify frauds with no false positives in the training data. This magical model depends on perfect data and data is never perfect. For example, missing train fraud detection models very frequently train on samples that have missing values and spurious inputs - mailbox addresses for instance. The latent structure of the feature set for, such as values that could potentially cause frauds, and the data have to successfully match. In general, the latent structure of in a model has to also explain structure of fraud. Structure in the data model arises from diversity of input variables, representation of these input values, content of these input, weights of these input values, or input density. Without these model substrates going sans structure with your data whitens your models, and future predictions shrink to nothing the deeper and richer your description of the fraud models in your head and data table on your model and the confines of your experiments.

This work is licensed under CC BY-NC-SA 4.0.

The vast majority of fraud detection problems - from insurance claim frauds and anti-money laundering to credit card fraudulent use - have training data that do not include any examples of frauds. In fact, in some industries such as insurance, there are many more ways that fraud can be committed that have actually been caught. Whether you use clustering, matrix factorization, or density estimation or find a way to frame the problem as classification by using indices such as fraud density, you are now part of a select group of machine learning engineers that build or assist in fraud detection engines. The number of engineers that work on this problem is minuscule compared to other machine learning communities. However, I encourage you to stay the course because the complexity of the fraud models required is just shy of that required to win a Kaggle competition - usually better than 95-99%. Additionally, each successful detection may save your client or vendor millions of dollars a year, not to mention saving shareholders significant drops in stock due to a scandal.

2.1. Importance and Challenges

In addition to the changes of scope during the recent years, these new interests have brought about complementary viewpoints challenging classical assumptions underlying commonly employed methodologies. Among those, concept changes and the handling of anomalies have been prominently discussed. While classical supervised learning under the assumption of stationarity and errors may be applicable to certain model classes in large-scale systems (e.g., Anomaly Detection in the realm of filtering stock price manipulations), there are still issues related to the data and labeling of data, especially under increasingly complex phenomena. Given these changes in both perspective and scope, this chapter aims to review and provide a brief synthesis of what is available for this increasingly important and interesting methodology known as Unsupervised Learning.

We live in unpredictable times characterized by tipping points. There is constant change and increasing uncertainty. Infrastructures such as the electrical power grid and financial systems have become increasingly complex and tightly interconnected. Driven by global forces such as fast-paced trading, climate change, and terrorism, they are subject to cascading failures. This calls for automatic methods that monitor the current behavior of these environments and continuously learn and reinforce predictive models in order to drive coordinated and timely behavioral adaptation. With the emergence of autonomous and adaptive systems, such demands for machine learning and data mining are transitioning from an operational setting to being part of the operational environment.

2.2. Types of Anomalies

For instance, imagine a scenario where an unsuspecting individual unexpectedly falls victim to a highly unusual and rare disease. This particular ailment is so uncommon that it can only be described as a

This work is licensed under CC BY-NC-SA 4.0.

global anomaly, as it diverges significantly from the norm when considering the entire population. In essence, it is an extraordinary occurrence that most people would deem highly improbable. However, when we examine the situation from a regional perspective, a fascinating revelation emerges. It turns out that this disease, despite its global rarity, is surprisingly prevalent in the specific geographical area where the unfortunate individual resides. Thus, from a regional standpoint, this peculiar ailment can be accurately categorized as a regional anomaly. In this peculiar case, it is truly remarkable to witness the stark contrast between the global and regional perspectives. While the disease remains a statistical anomaly when viewed on a global scale due to its extreme rarity, it paradoxically becomes a less surprising occurrence when examined in the context of the specific region. This highlights the dynamic nature of anomalies and reminds us that our perception of what is considered unusual can greatly vary depending on the scale and scope of our analysis.

1. Global anomalies: The entire dataset is considered to be abnormal.
2. Regional anomalies: While considered to belong to the regular behavior of the dataset, the data are distributed amongst the abnormal data.
3. Collective anomalies: The data items are outliers only when compared to the whole dataset, and not in relation to a specific subset of the dataset.
4. Contextual anomalies: The individual data instance is considered an outlier only in relation to a specific subset of the data.
5. Point anomalies: The individual data instance is considered an outlier.

Although there is no real community consensus on the characteristics of an anomaly in terms of outliers, several entities can be used to characterize anomalous entities or events. Below are a few examples of the different types of anomalies:

3. Traditional Anomaly Detection Techniques

The comprehensive study of anomaly detection methodologies provided in the given text identifies three sources of complexities for anomaly detection. First and foremost, unless supplied with sufficient examples of the anomalies in the training data, these anomalies are not well specified, leading to challenges in accurately detecting them. Consequently, any misuse or unsupervised learning algorithm needs to be carefully designed to identify not only the abnormal classes themselves but also the abnormal statistical indicators that these classes introduce. Furthermore, these abnormal indicators are often found within the combined and reduced data from the numerous clusters within the normal class, adding another layer of complexity. Secondly, the quality of the source data itself is a crucial aspect to consider. It is imperative for any anomaly detector to be designed in a way that can effectively handle "multivariate input data with varied data types," as the data may not always be of great quality. Such

This work is licensed under CC BY-NC-SA 4.0.

a design ensures that the detector is best prepared to deal with this possibility and can accurately detect anomalies in diverse and mixed data types. Lastly, neither the normal nor the anomaly class may be perfectly represented in the training data, posing yet another challenge. This is particularly significant when dealing with relatively unbalanced classes. While it may seem to reduce the simplicity of classification, it remains valuable to develop a practical and unsupervised method for identifying rare findings. The ability to accurately detect and identify anomalies, even in the presence of class imbalance, can prove to be highly useful and beneficial. Given the expectation of substantial source data complexity, it becomes evident that different strategies for anomaly detection may be appropriate depending on the specific task at hand. This acknowledgment acknowledges the need for customized approaches that take into account the unique characteristics and demands of the data being analyzed. By tailoring the anomaly detection strategies to the specific complexities of the source data, more accurate and reliable results can be achieved.

3.1. Statistical Methods

Z-scores are useful for identifying univariate anomalies but are unable to detect anomalies associated with interactions among variables. The Mahalanobis distance solves the latter problem empirically, reversing some of the power of statistic-based univariate methods. Thus, for all of their simplicity and ease of use, statistical methods of anomaly detection intrinsically depend on statistical normality, are unable to determine the prevalence of an anomalous entity, often introduce inaccuracies into their own accuracy assessments, and may not perform well when there is unequal clustering in the underlying data.

One approach to unsupervised anomaly detection is to analyze the distribution of variables for patterns of atypical behavior. A straightforward method to do this is through the use of statistical techniques. There is a large set of single and multivariate techniques that use statistical rules of thumb to inform the analyst about suspected outliers. Among the most common univariate approaches are the use of z-scores and the Modified Z-score. On the other hand, multivariate statistical methods differ from univariate methods in the sophistication and often the scale of their calculations. The most basic multivariate method is the Euclidean distance formula for points in the p-dimensional numerical space. Mahalanobis distance would be a superior measure in the case many variables have strong linear interactions, collinearity, or are otherwise maximally associated.

3.2. Clustering-Based Approaches

- K-means assumes that a certain input dataset consists of some k clusters and tries to find corresponding centroids (average case) in the space of standardized data. This way, considering a final

This work is licensed under CC BY-NC-SA 4.0.

assignment from the original dataset to clusters (according to the distance of an i-th observation in the space of standardized data to each centroid), there would exist a pre-defined Euclidean distance (squared of this distance represents k-means clustering criteria) between an i-th observation and the centroid of its cluster. - K-medoids assumes that we should use the already-provided input dataset instead of the k centroids in the case of k-means. As a consequence, besides, the pre-defined distance criteria do not need to be Euclidean in this case. Both k-means and k-medoids cluster analysis simply check how well the clustering algorithm has performed for different situations. Their results could be spoiled if there is some noise or other types of unsupervised or semi-supervised anomalies.

Clustering-based approaches are, in a sense, related to density estimation-based algorithms as well, as the cluster mode is a form of the local density maximum. These algorithms assume that 'normal' cases represent the majority of the available data, while at the same time, different spotted clusters of such cases should have rather small variances around their cluster centers (centroid). This is achieved by standardizing the data and using the squared Euclidean distance, the Manhattan distance, or the cosine similarity. The most applicable clustering-based anomaly detection methodologies include: k-means, k-medoids, and fuzzy c-means.

4. Machine Learning for Anomaly Detection

In the context of the specialized job of identifying anomalies, a machine learning technique can be seen as yet another repository of representations or models against which the patterns extracted from the in-context data are matched. In other words, the use of machine learning for anomaly detection resembles rather a specialized kind of a pattern matching exercise for a narrow abstraction of the modeling process, that is, detection of the unusual or undesirable patterns in the highly structured and complex relations. The process of model discovery, already carried out by machine learning, to guide the formation of domain-specific contextual understanding has to be done by the domain expert using the latter directly to guide model design. Machine learning is not the answer to the anomaly detection problem; it is rather yet another question to be posed within a broader context of pattern-based modeling. What a machine learning framework provides are systematic and versatile mechanisms for representing a wide class of patterns in multiple ways: matching of these patterns to given observed data; and a critical evaluation of these patterns based on a pre-defined objective. These mechanisms enable the exploration and analysis of an extensive range of datasets, facilitating the comprehensive identification of potential abnormalities across various domains. By harnessing the power of machine learning, practitioners can unlock new insights and uncover hidden patterns that evade traditional methodologies. The utilization of machine learning in anomaly detection offers an innovative approach to understanding data, allowing for the detection and mitigation of diverse anomalies within complex relational structures. This refined process of pattern matching further enhances the ability to detect

This work is licensed under CC BY-NC-SA 4.0.

subtle deviations and deviations that may elude human observation. Moreover, the iterative nature of machine learning equips domain experts with a powerful tool for continuously refining and improving the accuracy of anomaly detection models. Through iterative model discovery, experts can gain a deep understanding of the contextual nuances specific to their domains, enabling the creation of robust and targeted anomaly detection frameworks. These frameworks serve as invaluable assets in safeguarding against emerging threats and preventing potential risks within highly structured environments. It is crucial to recognize that while machine learning plays a vital role in anomaly detection, it should not be viewed as a standalone solution. Rather, it should be integrated as part of a broader, pattern-based modeling approach. This holistic perspective ensures that machine learning algorithms are utilized in tandem with existing domain knowledge, thereby maximizing their efficacy in uncovering anomalies. By leveraging machine learning as a complementary tool rather than a definitive answer, practitioners can encompass a wider scope of anomaly detection challenges and develop more comprehensive solutions. The versatility and scalability of machine learning frameworks enable the representation and evaluation of patterns across diverse datasets, aiding in the objective assessment and identification of anomalies based on predefined criteria. In conclusion, machine learning enriches the anomaly detection landscape by providing a systematic framework that enables pattern representation, matching, and critical evaluation. This framework, when harnessed by domain experts, propels the discovery of anomalies within highly complex and structured datasets. By embracing machine learning within the broader context of pattern-based modeling, practitioners can unleash the true potential of anomaly detection, unearthing novel insights and fortifying their defenses against the unexpected threats that lie beneath the surface.

4.1. Supervised vs. Unsupervised Learning

Is there a valid point of compromise between these polarized views of the unsupervised learning enterprise and what is its scientific rationale and practical value? To answer these philosophical and practical questions, it is instructive to consider the deep motivations of the supervised and unsupervised learning enterprise.

It is not difficult to see a fear of order, a primordial dread of systematic organization, inherent to the performative construct of the unsupervised learner: the unsupervised learner is fearful of knowing too much, of becoming too perceptually organized, of reflecting the domain of structured, reasonably predictable, message-bearing regularity that is the province of supervised learning algorithms. The unsupervised learner is wary of descending into the vale of sorrows where the space remains an empty plane, uncluttered by reinforced examples, an undifferentiated topography that is untethered to any solar system of anchor points.

This work is licensed under CC BY-NC-SA 4.0.

While supervised learning tries to compress input information into a functional approximation of target outputs, which it receives as training examples, unsupervised learning attempts to isolate the unique and unexpected about the input unto itself. Indeed, the unsupervised learning machine often wisecracks every bit its knowledge about the world of organized data: "I will decipher the text of this image into finite clusters of textural properties, but what acts of recognition you expect of me, I cannot disclose."

Unsupervised learning, as a research approach, is poised on the antithesis edge of the intellectual construction of machine learning. As a force for establishing connections and flattening irregularities, unsupervised learning (the development of rules and representations of data without the assistance of memory-rich labeled examples) contrasts with the analytic, irregularity-sensitive construct of supervised learning, which uses labeled examples to learn the output of a data transformation.

4.2. Key Algorithms and Techniques

Clustering algorithms are formally divided into two major families. Hierarchical approaches construct the hierarchical decomposition of the patterns in a divide-and-conquer manner while partitioning techniques are based on the divide-and-assign principle. The choice of distance metric for each hierarchical technique is a key decision for their successful use. This is the crucial assumption for the results validity of clustering algorithms based on distance metrics: we attribute any irregularity in distances distributions properties that under appropriate normality assumptions, the Kolmogorov-Smirnov, Anderson-Darling, and Cramer-von-Mises distances all follow an extreme value distribution, thus we can compare distances across pairs of groups that are not nested. Another central assumption for partition-based clustering is that all clusters are of roughly the same size; these facts determine the Wards similarity measure as the ideal technique for anomaly detection purposes.

After this overview of the literature on anomaly detection, it's time to introduce those methods from machine learning, statistics, and network analysis that are employed in the following chapters, which constitute the core of the proposed methodology. The first set of techniques used in this methodology come from multivariate statistics and exploratory data analysis. Grubbs proposed what is arguably the most familiar data mining technique for anomaly detection, namely false positive identification. The Grubbs test identifies variations in mean values between the different dimensions of the data that are unlikely to have occurred by random chance. It is very computationally efficient and easy to implement, being a frequent choice for the first step in the anomaly spotting process. To summarize very large amounts of information, principal component analysis is a very common technique for data reduction, creating a few latent dimensions representing orthogonally the most significant dimensions of the original data.

This work is licensed under CC BY-NC-SA 4.0.

5. Deep Learning Approaches

In 2006 and 2007, with the introduction of deep autoencoders and with the Small World Phenomenon, resulted in the revival of deep learning. These events represent a paradigm shift. Now it is better to have a many-layer neural network or, in other words, deep learning is about learning many levels of representations of the data. This makes a deep learner "unsupervised." From a data perspective this implies that it is trained by using unlabeled data. But from a utility perspective this means that the learned features can be used as effective feature detectors for the hierarchical classification task. With more than one layer, deep learning is about unsupervised learning with another goal. In addition to the traditional goal of finding the number of clusters in the input space by training a feature vector per centroid, deep unsupervised learners are employed as unsupervised pre-trainers. In this context bisecting k-means and other homogenized partitioning methods are used to initialize unsupervised pre-trainers. In addition to clustering and unsupervised pretraining, deep unsupervised learning is also framed around the principle of transfer learning. Unsupervised pre-training layers can be replaced within a deep network which is then used in a supervised learning setting for transfer learning. This is done by swapping an unsupervised hidden layer with an output layer to create a pre-trained deep network. Then, throwing away the original output layer, the new output layer is connected to the remaining unsupervised hidden layers while maintaining its weights and then trained on the supervised learning task.

It was 1973 when Linnainmaa first applied backpropagation to learn the weights of a multi-layer network of neurons. However, it was in the 80s that backpropagation spread through the neural network community, peaking when Geoffrey Hinton introduced the backpropagation algorithm to the United States in 1986. Although backpropagation networks could approximate arbitrary functions and automatically learn all the relevant features, they fell into disrepute in the second half of the 80s because they required several layers with many neurons to work well. As a consequence, they suffered from the same problems that plagued all optimization problems with many minima. These were also the reasons why deep belief networks turned out to be quite successful. The introduction of Restricted Boltzmann Machines, the product of the American collaboration between Geoffrey E. Hinton and Terry J. Sejnowski, and the British collaboration between David E. Rumelhart, Geoffrey E. Hinton and Ronald J. Williams.

5.1. Autoencoders

As with RA-SOMs, standard unsupervised outlier and anomaly detection methods can usually be defined along two algorithmic tracks: the data distribution-based and distance or density computation-based methods. The focus of this paper is on the anomaly detection characteristics of unsupervised

This work is licensed under CC BY-NC-SA 4.0.

learning strategies that use neural networks as learning systems. Therefore, we will briefly examine each of the autoencoder anomaly detection-based methods in turn. The autoencoder paradigm allows for the systematic study of the impact of learning system architecture on the anomaly detection performance in an unsupervised learning-based approach. This should contribute significantly to the exploration of potential contributions of unsupervised learning approaches to the field of anomaly detection.

Autoencoders are a class of neural network learning system that is composed of two parts: an encoder and a decoder. Each of these parts is composed of a series of layers connected together in a classical feed-forward neural network. The central layer of the encoder, which can also be seen as the topmost or output layer of a separate two-layer network, forms the hidden input feature representation learning stage in the network. In their simplest single-layer form, autoencoders can effectively be used to perform a feature extraction task from a high-dimensional to a low-dimensional problem, which matches many unsupervised outlier and anomaly detection comparative learning paradigms. Single-layer autoencoders are also functionally equivalent to principal component analysis in Euclidean space. These single-layer autoencoders seem to be the most commonly used form of autoencoders for anomaly detection-based unsupervised learning.

5.2. Variational Autoencoders

The Variational Autoencoders (VAE) framework considers the input as an observed variable that is generated from latent variable z. During the training time, instead of directly mimicking the input data, the target of VAE is to fit the latent posterior of the data. In this way, given a new probability sample, the testing time becomes simple as we are aware of its corresponding latent representation. Note that VAEs define a recognition model $f\phi(z|x)$ and a prior model $p\theta(x|z)$, and the model parameters (ϕ, θ) are jointly optimized by maximizing the evidence lower bound (ELBO) in order to minimize the negative reconstruction lower bound. Unlike unsupervised classification tasks, the vanilla VAE model optimizes the KL-divergence between the data latent posterior and the prior in the hope to steer the factorization of the latent space.

In this section, we detail the Variational Autoencoders (VAE), which is one of the commonly used deep unsupervised models in practice. VAE has gained increasing interest in the anomaly detection research direction as it provides a tractable solution for fitting the latent variable of probabilistic models in a way that makes searching the model possible during the testing time. However, as the model complexity remains a severe concern in the training and evaluation of large-scaled deep generative models, we discuss the training complexity and the connection between the Fisher Information and the reconstruction probability of the VAE.

This work is licensed under CC BY-NC-SA 4.0.

6. Hybrid Approaches

Theoretical concepts have already been well exposed and do not require further elaboration here. However, we could again propose a few words about the main methods and the reasons for their re-introduction, underlying this time what could be relevant in a novel hybridization. The DOD system, in a few words, applies the principles of Outlier Detection as a Pre-processing filter, of Defect Proportion Estimation, of ODE function implementation and of the Realization Method (that he introduced) to find the anomaly condition. The DOD system has been effectively used in many and different application domains where the two objectives are forecast reliability and error rate minimization. The SOD system here takes ODE Results as a further tool and is mainly based on the idea of count in grid cells containing rare in-tune objects and then. It is a methodology that could be combined with ours as a tool that could help in understanding both the meaning and the impact of the detected anomalies.

There are a variety of approaches which have recently experimented new pending methodologies for anomaly detection. In particular, we have focused our attention on methods that combine some of the proposed strategies. The reasons for revising the different proposed hybridizations are several, and mainly come from the omni acritical approach which has led us to consider unsupervised learning. Going through different basic approaches and principles of anomaly detection and analysing and discussing proposed problems and solutions we could not but review also hybrid approaches as new proposals needed a careful reflection on methodological choices. With this we also want to stress that supervised learning, the only approach that would follow a typical scientific reasoning in AI technical and applicative terms, has not yet proved satisfying results. The non-triviality of the problem itself pushes towards advanced research, that is in general what comes in any multidisciplinary research and what we propose as only way to get the complexity of AI application problems under control.

6.1. Combining Multiple Techniques

Approaches that make use of unsupervised learning methods often advocate for the utilization of a sole specific technology or category of methods. However, in the case of NASCAR's approach to anomaly detection, it can be categorized as a fusion of unsupervised learning and case-based reasoning. What sets our architecture apart is the underlying premise that versatile mixed-initiative problem-solving strategies can be formed by incorporating a wide array of methods. In this section, we will delve into a discussion of several practical benefits that support the adoption of relatively diverse anomaly detection algorithms. The nature and manifestation of these benefits, resulting from the amalgamation of various existing detection methodologies, are contingent upon the constraints imposed by our underlying application. At the core of our system lies the notion that the normal course of operation

This work is licensed under CC BY-NC-SA 4.0.

can adapt and evolve as more training experiences are performed. This adaptability enables the utilization of fast methods that label numerous instances of normal behavior, while simultaneously allowing the detection of new and unusual activities in real-time. Moreover, these detected anomalies can be employed in a semi-automatic manner to enhance the training data. Achieving this entails explicitly deriving a basis of labeled nominal activity and utilizing prior knowledge pertaining to normal operation, which has been accumulated through the multiple iterations of modeling and problem-solving. The dexterity of our system lies in the diverse techniques we employ, which constitute a crucial component of our problem-solving architecture. This diversity contributes to a robustness that would be challenging to achieve with any single anomaly detection technique.

6.2. Ensemble Methods

The underlying concept of these ensemble models can thus be leveraged to provide some meaningful measure of trust in the decision of the system. The same idea has been sketched out in recommendation systems as voting and committee-based trust-aware modeling. Similar to the other group models that have been described, performance in general should not be severely impacted as well.

Bagging and boosting are heuristic methods for constructing high performing models, and stacking is a fourth method meant for combining the strengths of different models. In stacking models, the resource learner employs many other employed learners along with whatever features of the training data that were omitted by the employed learners that are considered pertinent. It assumes the input dataset contains n attributes such that, the set of all weak learners have to make their decisions with some degree of interdependence and that the stronger base learners will take advantage of such duplicate hypotheses rather than necessarily the stronger hypothesis.

Bagging, or bootstrap aggregation, involves training the constituent models independently on sets of bootstrapped samples of the same data and usually leads to lower variance through the reduction of model overfitting. For unsupervised anomaly detection, an ensemble method is incorporated into the system such that the distance between ensemble models represents the degree of suspicion surrounding instances. If the ensemble is confident in its decision, the distance will be small.

An ensemble method entails a collection of simpler models, typically generated using the same base algorithm, which are combined to provide one or multiple predictions. The premise is that an ensemble of models can outperform a single model by containing a diverse collection of some sort of performance-enhancing constituent parts. In other words, the errors committed by individuals will be dissimilar in some form. They can be partitioned into bagging, boosting, and stacking models.

This work is licensed under CC BY-NC-SA 4.0.

7. Evaluation Metrics for Anomaly Detection

It is customary to break the types of evaluation of anomaly detection methodologies down into two diametrically opposed camps (with a large gray area in between for fully unsupervised measures) depending on the type of information used to define the evaluation period. The first class of evaluation is that of using performance measures that are derived from the ground truth. The ground truth itself works as an input to case-based evaluation of the anomaly detection methodologies, and these measures would include true positive rate, true negative rate, positive predictive value, false positive rate, family-wise error rate, and precision. The second class of evaluation is that which evaluates with an absence of ground truth, such as cluster metrics and model parameter divergence performance.

First, without labeled data, it is difficult to provide a "ground truth" by which to compare the performance of anomaly detection methods. Second, because the definition of what constitutes an anomaly can vary from domain to domain, it is more apt to consider anomaly detection as a measure of the "differentness" or "out-of-distributionness" of a given example of the data, given the context of the suite of examples it is meant to represent. This ambiguous definition is carried over to the evaluation of anomaly detection techniques, and this inherent ambiguity persists to various degrees across diverse domains and laboratory conditions.

Because anomaly detection is so crucial in many real-world applications of machine learning, understanding the performance criteria for anomaly detection is paramount. Although anomaly detection measures have been analyzed in myriad ways and across diverse contexts, unfortunately, evaluation of anomaly detection methods and the question of what characterizes an anomaly detection method as "good" is a challenging one.

7.1. Precision and Recall

Here we briefly review the other most popular and easily interpretable pair of statistics for classification: precision and recall. In the context of anomaly detection, remember that anomalies are the positive class. Precision is the measure of how many of the returned anomalies are actually "true" anomalies, i.e. how noiseless the selection of anomalies is, whilst recall measures how good we are at capturing or annotating the "true" number of anomalies in a dataset. More explicitly, precision is the ratio of true positives (relation returned anomalies that are also actually anomalies) to the sum of true positives (the number of anomalies we think are actually there) and false positives (anomalies we think are there but are not). In contrast, recall is the ratio of true positives to the total number of actual anomalies (true positives + false negatives).

This work is licensed under CC BY-NC-SA 4.0.

Traditional high-throughput supervised learning showcases the usefulness of confusion matrices, built on the foundation of true and false positive, and true and false negative classifications. Together, these explicit numbers can be supplemented with other descriptions to form well-connected isolated events. Despite the headlines, unsupervised learning, and in particular anomaly detection, is not as straightforward to evaluate, as many of the assumptions underlying confusion matrices and the calculations that come with these are discreetly circumvented when our labels are not observational. Consequently, precision and recall are also exposed to new surroundings when transposed to contexts where their definitions are less easily appreciated.

7.2. ROC Curve and AUC

The area under the ROC curve (AUC) is a widely used summary index to evaluate binary classification algorithms. It quantifies the ability of the model to discriminate between both classes. AUC is the probability that a randomly chosen positive case has a higher probability of being positive than a randomly chosen negative case. The main benefit of AUC as an evaluation measure is its robustness against class imbalance. It doesn't care about the intercept but about the shape. Besides the robustness against class imbalance, the AUC has the attractive property that it is the probability that the learned order is correct.

The receiver operating characteristic (ROC) curve is a visual representation of how the true positive rate (TPR) and false positive rate (FPR) vary as a discrimination threshold is adjusted. It is a frequently used tool to evaluate binary classification models. The ROC curve is often used when we want to control specificity and sensitivity rates. Depending on the context, we might have to specify whether we want a high probability of detection, regardless of the false alarm, or minimization of the false alarm rate, regardless of the probability of detection. It helps us understand the trade-off between such competing objectives.

8. Real-World Applications

The rapid growth of digital video and image data in a range of Earth observation and surveillance tasks are in clear need of data analysis techniques to identify unusual behavior hidden in the data. In financial time series, it is a challenge to compare the information contained in the developed econometric model to the existing fuzzy indicators of available leading economic indicators; as well as, identify the structure and composition of the underlying fundamental and technical drivers of the exchange rate dynamics as it progresses through time. Those involved in the management of physical security

This work is licensed under CC BY-NC-SA 4.0.

networks, including the supply of physical security products or services, supply key inputs to the concept of operations cycle, which helps decision-makers understand the impact of systemic and technological consequences. Control system data can be recorded as the current process data, is stored in process data recorders, and then it can be transmitted via a network to the data warehouse, from which the data can be accessed to analyze the current process.

Finally, we outline well-known application areas for anomaly detection. As robotic devices gain the abilities to repair and maintain our complex machinery, as they now do at remote locations on and off Earth, coming to understand the gestalt in which these robots operate and how their performance indicators can best be monitored and understood becomes more important. Anomaly detection in large scale telecommunication systems has gained a renewed interest because of the introduction of pre-commercial validation of large hybrid networks and the similarity with other large-scale real-world networks which are either already under test or are going to face soon. We summarize how existing cleared sensors can be used to enhance the utility of consumer sensor and distributed sensor networks from a command and control framework.

8.1. Cybersecurity

As demonstrated by the growth of reported cybersecurity problems, the proliferation of new and successful malware breeds, the existence and growth of the cybersecurity market, the huge investments made to protect organizations' computer systems, the high demand and relevance of security products and services such as firewalls, anti-virus, intrusion detection, and the increasing number of companies committed to the sector, cybersecurity is nowadays at the core of concerns not only of the whole computer society but also of the government society itself. With the increasing number of incidents and cybersecurity risks, this activity is rapidly becoming one of the priorities of governments worldwide. A number of organizations today are studying and collaborating on communications in robustness and reliability issues. There are also a number of initiatives worldwide to define concepts, strategies, and regulations to strengthen the defense of each nation's cyberspace. For example, the US Department of Homeland Security is working to develop capabilities to support strategic cyberspace planning and operations.

Cybersecurity refers to the technologies and processes designed to protect computers, computer networks (such as the Internet), and the electronic information they contain. Loss of such information can result in damaging consequences because it may consist of personal data (e.g. social security and credit card numbers), integrity (e.g. for operations carried out over the Internet, such as purchasing or paying bills), confidentiality (e.g. passwords, financial and confidential files), and financial assets for individuals. For organizations, a loss might damage their image (e.g. loss of clients, damage to

This work is licensed under CC BY-NC-SA 4.0.

reputation), might involve legal risks (such as legal actions from individuals whose data has been stolen or lawsuits against the company), and financial losses (loss of business, cost of cleaning up).

8.2. Healthcare

The application of unsupervised learning in healthcare is numerous. The healthcare industry is fraught with anomalies. Given the complexity of this industry, domain-specific anomaly detection methodologies are necessary to add extra layers of surveillance to hospital systems beyond the usual security and antivirus mechanisms. For clinical laboratory, radiology, genomics, and numerous other healthcare big data are generated on a continuous basis. For these applications, sophisticated dimensionality reduction steps need to be used so that increasingly higher order forms of biological data can be effectively used within the anomaly detection infrastructure.

The healthcare domain—that is, the set of industries engaged in the treatment, diagnosis, and prevention of illness and disease—is the largest industry in the world when aggregated. It is also a complex entanglement of many different subdisciplines such as patent law, business management, and various social science endeavors. Healthcare is an industry involved in the continuous creation of both physiological and psychological risk. It requires scientific and sociological methods to minimize such risk. In short, healthcare is an industrial complex wherein life and death are at stake; healthcare has as its primary stakeholder all the citizens of planet Earth; and healthcare is a political issue that floats alongside military and environmental security.

9. Challenges and Future Directions

The Dialectics of Unsupervised Learning: A Synthesis of Anomaly Detection Methodologies, includes a chapter written by R. Vytheeswaran, L. Hauska, A. Yahnik, and T. Tarlow, titled "The Dialectics of Unsupervised Learning: A Synthesis of Anomaly Detection Methodologies". Unsupervised learning is the problem of exploring patterns, structure, and associations in data in the absence of labeled output. In this paper, we cast the problem of anomaly detection as a dialectical process, alternating between discovering patterns within datasets and revisiting our mental model of the problem at hand. In particular, several methodologies use projections or maximize the discrepancy measures between samples, such as Principal Component Analysis (PCA), k-means clustering, one-class support vector machines, and generative models. Each of these techniques is contrasted in sample detection limits and properties. Such a synthetic view allows, for instance, distinguishing among the different types of outliers. Moreover, the outlook section of the paper provides a comprehensive list of open issues that reflect the current challenges and future directions of anomaly detection, covering areas such as interpretability, scalability, and real-time detection. The authors also delve into the impact of feature

This work is licensed under CC BY-NC-SA 4.0.

selection, dimensionality reduction, and model assumptions on the performance of anomaly detection algorithms. By examining the strengths and weaknesses of these methodologies, the paper offers valuable insights and guidance for researchers and practitioners in the field.

9.1. Scalability and Efficiency

A second dimension of efficiency is that O(t) can be modeled mistakenly as a linear increase in actual time, responding to the increasing size of n. Meanwhile, for a given computational operation, we should not underestimate the vagaries of memory and space, e.g., if the volume of data of the input and the scores themselves is large at the time of deployment, and perhaps especially if the detection of anomalousity is performed in an online manner. Finally, a model that is fast to train, as well as fast to query, can be retrained more often, and by desiring models that are elastic, we can improve various aspects of robustness.

Strictly speaking, scalability and computational efficiency are different concerns, but in the context of anomaly detection for the challenges of high-dimensional data, they are intertwined. A single scoring model trained on a training set of size n and using time complexity O(t) both for training and running inference time is said to be of time complexity O(t). As n increases, it is the case that O(t) becomes a more serious concern than O(tn). As a rule, time is of the essence when training a model due to the way in which the scores themselves are used and the cost of mistakes; for large n, it is the reception time that is most likely to dominate.

9.2. Interpretability and Explainability

Although simplicity and smallness usually lead to preferable models, finding simple unifying rules exemplifying high-level concepts becomes tricky. The gained simplicity can often only be achieved after the application of a complex and elaborate unsupervised learning algorithm. Statistics and knowledge may allow an intuitive understanding of the learned concepts, but the relations between all the information learned cannot be explained. Many anomalies are interesting precisely because they defy what we know about the subject, especially in novel real-world datasets. When simple rules or explanations are necessary, an interpretable unsupervised learning algorithm reduces complexity, which is seen as a very hard challenge. Minimally, a solution should have as few dimensions as possible, and former concepts or definitions should be shown in a subjective way based on domain knowledge. Information is a building block of rules and explanations, summarizing and justifying in terms of a subjective concept or enterprise. The use of complex and elaborate anomaly detection and unsupervised learning algorithms, which use semantics, is hard to interpret. Extrapolating complexity

This work is licensed under CC BY-NC-SA 4.0.

and facilitating one's task if the solution leads at the end again to small and simple rules with a minimum of dimensions.

Interpretability is the degree to which the computational process and underlying knowledge can be understood or explained by humans. The purpose of interpretability is to provide an explanation or rationale for algorithm decisions. The process is open, explicit, and unbiased. Conversely, explainability is the ability to articulate reasons for a decision. The idea behind the concept of explainability is to make the actions or decisions transparent to the user, so that the received results are more comprehensive and easily understood. Currently, many learning algorithms are able to provide some degree of interpretation on their outputs. However, an intuitive understanding of the inner workings for most of today's algorithms is still lacking.

10. Conclusion

A large class of training methods using both examples from a standard analysis in neural network training may be studied more systematically. It is shown that anomalous artificial examples can be used as a regularizer in the sense of a variance in well-trained networks. Weighting the training errors with pseudotemperatures in this way downgrades the importance of the error demonstrated by the data by mimicking a regularizer that would have been contributed by other examples outside the support. This happens spontaneously as a consequence of the training network with high quantum complexities. Additionally, the incorporation of these pseudotemperatures introduces a significant reduction in the occurrence of overfitting, as it encourages the network to generalize better to unseen data. Furthermore, by leveraging the principles of tensor decomposition, compressive encoding, and sparse coding, we have demonstrated how knowledge from already trained networks can be effectively utilized to identify and extract the slowest dimensions of relevance. This extraction process is not only computationally efficient but also acts as an essential tool in dimensionality reduction, leading to enhanced interpretability and improved performance in various practical applications. We firmly believe that this comprehensive approach serves as an outstanding illustration of the remarkable richness of unsupervised learning studies possible within the dialectical framework of the Abelian category. Moreover, it provides another intriguing angle of investigation, opening up a multitude of extremely useful practical applications while remaining cognizant of the theoretical foundations.

In this document, we have sketched a number of connections between existing and popular functionalities of unsupervised learning that can be related through the analysis of anomalous examples. We believe that this dialectical framework can lead to a unified practical and theoretical approach to problems of density estimation, clustering, and novelty detection. Even if the simplicity of the present illustration of these links in terms of high-dimensional functions may be lost in more

This work is licensed under CC BY-NC-SA 4.0.

realistic models, we believe there are a number of tantalizing features of this dialectic that may currently be explored in both artificial and real neural networks.

11. References

1. C. C. Aggarwal, "Outlier Analysis," Springer, 2017.
2. M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," PLoS ONE, vol. 11, no. 4, pp. 1-31, Apr. 2016.
3. Tatineni, Sumanth. "Embedding AI Logic and Cyber Security into Field and Cloud Edge Gateways." *International Journal of Science and Research (IJSR)* 12.10 (2023): 1221-1227.
4. Vemori, Vamsi. "Towards a Driverless Future: A Multi-Pronged Approach to Enabling Widespread Adoption of Autonomous Vehicles-Infrastructure Development, Regulatory Frameworks, and Public Acceptance Strategies." *Blockchain Technology and Distributed Systems* 2.2 (2022): 35-59.
5. Tatineni, Sumanth. "Addressing Privacy and Security Concerns Associated with the Increased Use of IoT Technologies in the US Healthcare Industry." *Technix International Journal for Engineering Research (TIJER)* 10.10 (2023): 523-534.
6. Z. Zong, H. Chen, M. Shi, J. Wu, Y. Wang, and W. Zheng, "Anomaly Detection in Time Series of Images Using Deep Learning," IEEE Access, vol. 7, pp. 67756-67768, 2019.
7. K. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," arXiv preprint arXiv:1901.03407, 2019.
8. X. Xu, H. Wang, B. Li, and X. Zhang, "Anomaly Detection in Multimedia Data Through Multimodal Deep Learning," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 191-204, Jan. 2021.
9. Y. Zhao and M. Comiter, "A Survey on Anomaly Detection Using Data Mining Techniques," ACM SIGKDD Explorations Newsletter, vol. 15, no. 1, pp. 1-14, Jun. 2014.
10. M. Kloft and P. Laskov, "Online Anomaly Detection Under Adversarial Impact," in Proc. 13th International Conference on Artificial Intelligence and Statistics, Chia Laguna Resort, Sardinia, Italy, 2010, pp. 405-412.
11. R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly Detection Using One-Class Neural Networks," arXiv preprint arXiv:1802.06360, 2018.
12. S. Liu, T. Cheng, and Y. Wang, "Anomaly Detection in Dynamic Graphs Using Attention-based Temporal Graph Neural Networks," in Proc. 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 987-996.
13. A. Ahmed and A. Pothen, "Anomaly Detection in Streaming Sensor Data," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 3, pp. 545-558, Mar. 2018.

This work is licensed under CC BY-NC-SA 4.0.

14. S. Eskin, "Anomaly Detection Over Noisy Data Using Ensemble Methods," in Proc. 17th International Conference on Machine Learning, Stanford, CA, 2000, pp. 255-262.
15. J. Zhuang, J. Tang, and H. Wang, "Local Outlier Detection with Interpretable Rules," in Proc. 2019 SIAM International Conference on Data Mining, Calgary, AB, Canada, 2019, pp. 237-245.
16. L. Ruff, R. A. Vandermeulen, N. Görnitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep One-Class Classification," in Proc. 35th International Conference on Machine Learning (ICML), Stockholm, Sweden, 2018, pp. 4390-4399.
17. E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data," in Proc. 2002 Data Mining for Security Applications Workshop, Arlington, VA, 2002, pp. 77-101.
18. R. Chalapathy and S. Chawla, "Anomalous: Deep Anomaly Detection in Time Series," arXiv preprint arXiv:1901.07329, 2019.
19. D. G. Akila and B. Tamilarasi, "Anomaly Detection in Network Traffic Using K-Means Clustering," in Proc. 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 169-173.
20. X. Li, X. Song, M. Wu, X. Li, C. Wu, and X. Hu, "Anomaly Detection in Multivariate Data With Graph Learning-Based Approaches," IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 6, pp. 2454-2465, Jun. 2021.
21. M. Pang, Y. Zhou, J. Gao, and B. Wu, "Deep Autoencoder-Based Anomaly Detection in High-Dimensional Data," IEEE Access, vol. 7, pp. 28218-28229, 2019.

This work is licensed under CC BY-NC-SA 4.0.