Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework

Ander Carreño, Iñaki Inza, Jose A. Lozano

Abstract

In recent years, a variety of research areas have contributed to a set of related problems with rare event, anomaly, novelty and outlier detection terms as the main actors. These multiple research areas have created a mix-up between terminology and problems. In some research, similar problems have been named differently; while in some other works, the same term has been used to describe different problems. This confusion between terms and problems causes the repetition of research and hinders the advance of the field. Therefore, a standardization is imperative.

The goal of this paper is to underline the differences between each term, and organize the area by looking at all these terms under the umbrella of supervised classification. Therefore, a one-to-one assignment of terms to learning scenarios is proposed. In fact, each learning scenario is associated with the term most frequently used in the literature. In order to validate this proposal, a set of experiments retrieving papers from Google Scholar, ACM Digital Library and IEEE Xplore has been carried out.

1   Introduction

Numerous applications require filtering or detecting *abnormal* observations in data. For instance, in security, intruders are abnormalities (Ribeiro et al., 2016; Pimentel et al., 2014; Luca et al., 2016; Phua et al., 2010; Yeung and Ding, 2001); in traffic data, road accidents (Theofilatos et al., 2016); in geology, the eruption of volcanoes (Dzierma and Wehrmann, 2010); in food control, foreign objects inside food wrappers (Einarsdóttir et al., 2016); in economics, bankruptcy of a company (Fan et al., 2017); or in neuroscience, an unexperienced stimulus is considered an abnormality (Kafkas and Montaldi, 2018). In some situations, the abnormalities are called rare events, anomalies, novelties, outliers, exceptions, aberrations, surprises, peculiarities, noise or contaminants among others. Of these, the most common terms in the literature are rare event, anomaly, novelty and outlier.

Considering the importance of abnormalities in different areas, a lot of research has been done, mainly in the last 10 years. However, the fact that these

contributions have been carried out in different knowledge areas, a mix-up between names and problems has occurred in the literature. Particularly, when the same term is used in distinct disciplines but with other meaning and vice versa. Moreover, the terminology has changed over time and even in the same discipline; a similar problem has been named differently in different time periods. On the one hand, different names have been used for similar problems. For instance, Van Den Eeckhaut et al. (2006) deal with a problem of predicting, in a fixed period of time, the risk factor of a landslide in an area. The authors create a landslide susceptibility map in which each area is scored based on the risk of a landslide. This is done using historical data of either normal and ground which has suffered a landslide (abnormal). In this study, the authors refer to landslides as *rare events* because landslides seldom occur. In Ribeiro et al. (2016) a similar problem is addressed, but with a different term. Here, a study in the railway industry is carried out. Train passenger doors have several subsystems in order to keep them open or closed according to a variety of safety and comfort rules. In some situations these doors fail due to the deterioration of the system. Therefore, the authors predict whether the door is going to fail in a fixed period of time or not. In order to do that, both normal and failure historical data is used to learn a model. In this case, the door failures are referred to as *anomalies*. As can be seen, both problems are very similar and different terms have been used to refer to the abnormalities. In both problems, temporal data of normal and abnormal classes is available to build the prediction models.

On the other hand, the same terms have been used to describe widely different problems. In the following two problems, the authors use the term *novelty* to describe the abnormalities. In Luca et al. (2016) a variety of patients are constantly monitored with a 3D accelerometer. Those patients eventually suffer an epileptic seizure. Due to abrupt movement during a seizure, the patient could became injured. Therefore, detecting this behavior as soon as possible is relevant in order to avoid this harmful situation. In order to predict if a patient is suffering an epileptic seizure, a model is built based on the recorded movement data of several patients. The data consists of 3D accelerometer data divided in fixed time windows in which whether or not an epileptic seizure has occurred is annotated. However, notably less abnormal (seizure) data is available due to the eventuality of these attacks. In the prediction phase, given new information about a currently monitored patient, the classifier detects if the patient is suffering an attack at that moment. Einarsdóttir et al. (2016) detect foreign objects inside food envelopes. A classifier is learned only from food-images without abnormal objects. In other words, the model is learned using information of only one class. However, in the detection phase, the model classifies new instances in two classes, normal (without foreign objects) and abnormal (with foreign objects). While both examples are named with the same term, the problems are widely different. For instance, the former has both normal and abnormal data available to train the model, whereas the latter only learns from a dataset with observations of only one class.

As we have seen in the previous paragraphs, there is an important mix-up between terms and problems. Possibly motivated by the same mix-up detected

by us, some papers that present specific learning methods have made an effort in their introduction section to discuss the differences between one or two terms, or to clearly define their learning scenario. However, to the best of our knowledge, no paper in the literature has treated the four rare event, anomaly, novelty and outlier terms under the supervised classification point of view. For instance, in Luca et al. (2016); Dufrenois and Noyer (2016) a brief discussion about the novelty term and one-class classification framework is made. In Weiss and Hirsh (1998), the authors clearly define their rare event learning scenario. In Campos et al. (2016), an effort is made to distinguish between one class classification and outlier detection. Finally, in Ribeiro et al. (2016), three methods related with outlier, anomaly and novelty detection learning scenarios are used to solve the same problem. Also, some insights are given about all these three learning scenarios. However, none of these papers frame the corresponding terms into the supervised classification framework.

This confusion calls for the repetition of research and hinders the advance of the field. Therefore, the aim of this paper is to contribute with a first step in the organization of the area. In order to do that, this work underlines the differences between each term, and organizes the area by looking at all these terms under the umbrella of supervised classification. Particularly, for each term, the most frequently used learning scenario is associated.

Paper reference	Brief description of the main characteristics of the paper	Term used
Van Den Eeckhaut et al. (2006)	The risk factor of a landslide is predicted in a fixed period of time. The data consist of historical landslided and normal land features.	Rare event
Ribeiro et al. (2016)	Failure of the train passenger doors is predicted in a fixed period of time. The data consist of both normal and failure temporal instances.	Anomaly detection
Luca et al. (2016)	Whether the patient is suffering an epileptic seizure is predicted. The data consists of normal and abnormal patient records. The patient records are timely monitored.	Novelty detection
Einarsdóttir et al. (2016)	The presence of foreign objects inside food envelopes is predicted. The data consist only of images of food envelopes without foreign objects.	Novelty detection

Table 1: An illustrative example of the mix-up between terms and problems in the literature.

This paper is organized as follows. Each section describes a supervised learning scenario: Section 2 describes rare event detection, Section 3, anomaly detection and Section 4, novelty detection. In each section, the objective of the classification task, the characteristics of the input data and the most popular techniques for the described learning scenario are reviewed. In Section 5, the related outlier term is treated. In Section 6, the one-to-one assignment of terms to learning scenarios is described coupled with a brief discussion about the main

3

evaluation techniques of each learning scenario. In Section 7, the experimental validation is described. Finally, in Section 8, the conclusions of this work are exposed.

2 Rare Event Detection

Almost all the papers that use the term *rare event* to describe the abnormalities of the problem to be solved share the time dimension as a common characteristic (Theofilatos et al., 2016; Heard et al., 2010; Dzierma and Wehrmann, 2010). For instance in Theofilatos et al. (2016), a road accident study in the Attica Tollway (Greece) is performed. The authors divide the tollway into different sections and they detect the occurrence of an accident in a certain section of the highway. A model is built based on recorded data from ground-sensors and traffic-cameras. More specifically, the data is sliced into one-hour time intervals and manually labeled by experts. Therefore, given a new one-hour time interval, the model detects an accident occurrence. In Dzierma and Wehrmann (2010), a geomorphological study is performed. The authors predict if a new volcano eruption is going to happen in a fixed period of time. A Poisson Process is learned with the historical Volcanoes Explosivity Index (VEI) of two volcanoes. Next, given a new VEI of one of the two volcanoes, the occurrence of the eruption in a fixed time interval is predicted.

In the previously described problems, the goal consists on the prediction of occurrence of a rare event in a bound period of time. A genuine characteristic of the rare event learning scenario, from a supervised classification point of view, is that the instances are time series (Hamilton, 1994). From this perspective, the objective is to classify new incoming time series as rare (when the rare event has occurred) or normal (no event has occurred) using a previously learned model. This approach is known in machine learning as supervised time series classification (Esling and Agon, 2012). However, due to the temporal nature of the problem, two different classification approaches can be found in the literature. Firstly, the *full length supervised time series classification* is dealt with. For example, in Murray et al. (2005), the SMART[1] dataset is used to detect if a hard-drive is faulty in a fixed period of time. The authors learn a model using recorded hard-drive sensor measurements at different times. Then, given new hard-drive sensor data, failure is detected. In Zhang et al. (2017), a termo-technology dataset which contains information gathered over time about heating systems is used. The objective is to detect if the heating system has failed in a fixed period of time. Secondly, another type of classification of rare events can be found in the literature, in which the objective is to classify new observations (time-series) as early as possible, preferably before the full time series is available. This approach is known as *early supervised time-series classification* in machine learning literature (Mori, 2015). For example, in Ogbechie et al. (2017) a prediction of faulty metal bars is studied. During the bar melting process,

---

[1] The SMART dataset is hard-drive self-monitoring recovered data in which both normal and failure behaviors are collected.
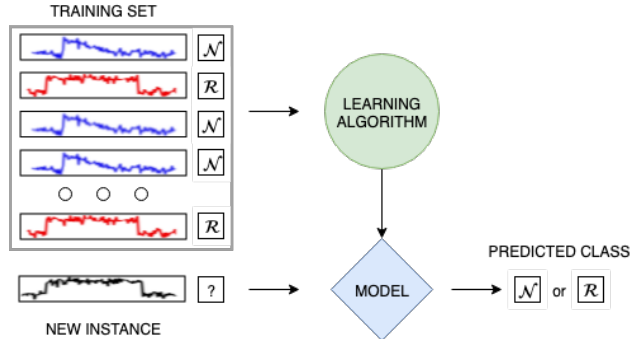
Figure 1: A flowchart of the supervised time series classification data mining task (Mori (2015)).

several sensors monitor the characteristics of each bar. These measurements, recovered from both normal and faulty bars, are used to learn a model. Next, given information about a new bar, the classifier predicts if the bar is going to be faulty. The early detection of a faulty bar is crucial because, depending on when it is detected, it can be fixed during the rest of the process.

According to the characteristics of the data, in most of the problems referred to with the *rare event* term, instances are time series and are labeled in two categories: normal ($\mathcal{N}$) and rare ($\mathcal{R}$). Furthermore, in many papers, the data shows an unbalanced distribution of classes. Formally, assuming that the data is generated by a generative mechanism $P(\mathbf{x}, c)$ (Mitchell, 1997), $P(C = \mathcal{R}) \ll P(C = \mathcal{N})$. Considering the instances during the training stage, both normal and abnormal instances are available to learn the classifier. Therefore, rare event classification can be formalized as a (highly) unbalanced supervised time series classification problem (Köknar-Tezel and Latecki, 2011; Cao et al., 2011). Formally, this scenario can be described as follows:

A time series (TS) is an ordered pair (timestamp, value) of fixed length m:

$$TS = \{(t_1, x_1), \ldots, (t_i, x_i), \ldots, (t_m, x_m)\}$$
$$\text{with } t_i \in \mathbb{N}, \text{ for } i = 1, \ldots, m \tag{1}$$

Time series classification is a supervised data mining task in which giving a training set of time series, $\mathbf{TR} = \{(\mathbf{TS}_1, y_1), \ldots, (\mathbf{TS}_n, y_n)\}$, in which y represents the label of the corresponding time series, the objective is to build a classifier that is able to predict the class label of any new time series as accurately as possible (Mori, 2015). In the particular case of a rare event, it is common to have a scenario where $P(C = \mathcal{R}) \ll P(C = \mathcal{N})$. A common classification process can be seen in Figure 1.

Besides, there are some problems in the literature in which the prediction must be output as soon as possible. This learning scenario is known as early time series classification (Mori et al., 2018).

However, even though the problem itself has the time dimension as a key

component, in some rare event detection applications, instances are transformed without considering this genuine characteristic. Therefore, the approach treats the problem as an unbalanced non-temporal classification task, similar to those found in the anomaly detection learning scenario (further described in Section 3). For instance in Murray et al. (2005), the data is composed of several hard drive sensor measurements at different time intervals. Therefore, for the same drive, many readings of the same sensors are available. However, the authors do not consider the order in which the measures have been recorded, and, given new hard-drive unordered measurements, the model classifies the drive as faulty or normal. Hence, the temporal nature of the data is not leveraged.

Regarding the rare event literature, the objective of most of the related papers is focused on classifying the rare class. Therefore, in order to evaluate the performance of the classification task, popular metrics such as AUC (Zhang et al., 2017; Xu et al., 2016; Ren et al., 2016) and the recall of the rare class (Zhang et al., 2017; Ren et al., 2016) have been commonly used.

Among the most frequently used techniques in time series classification, rare event logistic regression, an adaptation of the logistic regression for this learning scenario, is a popular choice (King et al., 2001; Theofilatos et al., 2016; Ren et al., 2016; Van Den Eeckhaut et al., 2006). However, techniques such as Kullback-Leibler divergence to discriminate between rare and normal events (Xu et al., 2016), long-short term neural networks (Zhang et al., 2017), rule-based classification learned with genetic algorithms (Weiss and Hirsh, 1998), multiple-instance naïve Bayes (Murray et al., 2005), Poisson Processes (Dzierma and Wehrmann, 2010), support vector data regression with surrogate functions (Bourinet, 2016), Bayesian networks (Cheon et al., 2009) or support vector machines (Khreich et al., 2017) have been successfully adapted for this learning scenario.

Taking into account the unbalanced distribution of classes, most of the previous methods are coupled with techniques specifically designed to deal with unbalanced time-series classification. Some of these techniques include: the Structure Preserving Over Sampling (SPO) technique (Cao et al., 2011), or an adaptation of the classical Synthetic Minority Over-sampling TEchnique (SMOTE) (Köknar-Tezel and Latecki, 2011).

Finally, another widely different rare event related learning scenario can be found in the literature. The estimation of the probability of occurrence of a rare event (Wu et al., 2003; Cadini et al., 2017; Dessai and Hulme, 2004; Dueñas-Osorio and Vemuru, 2009; Bedford and Cooke, 2001). This approach is mainly used in engineering and physics and some illustrative examples of rare event probability estimation include: the estimation of the probability of infrastructure failure in a fixed period of time (Dueñas-Osorio and Vemuru, 2009), the estimation of the probability of failure of technical systems in a fixed period of time (Bedford and Cooke, 2001), or the estimation of the probability of extreme climate developments in a specific time window (Dessai and Hulme, 2004). Since this learning scenario is beyond the supervised classification framework, it is not considered in this paper. Among the most frequently used techniques in order to estimate the rare event probability, importance sampling, Monte Carlo sim-
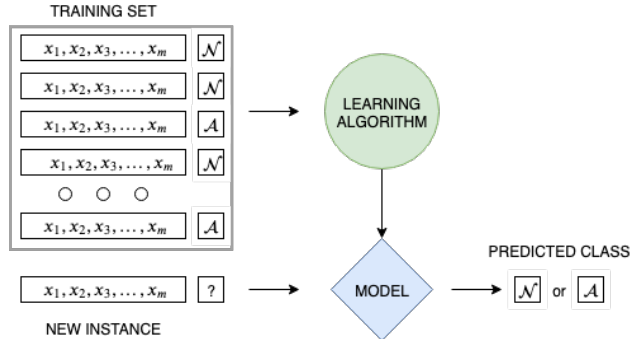
Figure 2: A flowchart of a supervised classification task. This learning scenario is assigned to the anomaly detection term.

ulations (Balesdent et al., 2016; Auffray et al., 2014), kriging (Auffray et al., 2014) or first order reliability method (FORM) (Straub et al., 2016) are found in the literature.

3   Anomaly detection

Most of the problems which describe the abnormalities with the anomaly term are non-temporal. The data is labeled in two categories: normal ($\mathcal{N}$) and anomaly ($\mathcal{A}$). For instance, in Miri Rostami and Ahmadzadeh (2018), the authors detect breast cancer using the Surveillance Epidemiology and End Results (SEER) [2] dataset. This dataset consists of patients which have been examined for cancer diseases. The patients which suffer from cancer are described with anomaly term. Hence, the data consists of cases of both normal and anomalous instances. A model is then learned which classifies new unseen cases as anomalous or normal. Fiore et al. (2017) detects credit-card transactions using a public dataset with legal and notably less fraudulent transactions. A neural network is learned to classify new incoming transactions as legal or fraudulent.

In anomaly detection learning scenario, anomalous instances are scarce due to the unbalanced distribution between normal and anomaly classes (Chandola et al., 2009). Therefore, this scenario can be formalized as (highly) unbalanced supervised classification. Formally, an instance is defined as $\mathbf{x} = (x_1, \ldots, x_m)$. Given a training set $\mathbf{TR} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, in which y represents the label of the corresponding instance, the objective is to learn a classifier that is able to predict a new class label of any new instance as accurately as possible. Regarding the probability distribution of the class variable, $P(\mathcal{A}) \ll P(\mathcal{N})$. Where $\mathcal{A}$ represents the anomaly class label and $\mathcal{N}$ the normal class label. An illustrative example can be seen in Figure 2.

In order to evaluate the performance of the classifiers, due to the (highly)

---

[2]Available here: https://seer.cancer.gov/data/

unbalanced distribution of classes, common metrics such as accuracy are not informative enough. Therefore, authors focus on the correct classification of abnormalities. A popular evaluation measure used is the maximization of the recall of the minority class (Ribeiro et al., 2016; Miri Rostami and Ahmadzadeh, 2018).

For anomaly detection, popular supervised classifiers have been adapted obtaining competitive results. For instance, support vector machines (Zhou et al., 2017), neural networks (Noto et al., 2012) or Gaussian mixture models (Reynolds, 2015) present genuine algorithms to deal with anomaly detection domains. Note that, since anomaly detection can be formalized as a (highly) unbalanced supervised classification problem, techniques that specifically deal with unbalanced domains can be used for anomaly detection. Similar to the rare event oversampling techniques, SMOTE (Miri Rostami and Ahmadzadeh, 2018; Araujo et al., 2018), is widely used to synthetically generate instances from the minority class.

4 Novelty detection

In most of the papers that use the term novelty to describe the abnormalities, the model is learned using a dataset that contains only one class. For instance, in Khreich et al. (2017), system call traces are classified as novel or normal. A novel instance corresponds to an unsupported or unexpected system call trace. To learn the model, only normal system call traces which have been gathered in a secure environment are used. When a new system call arrives, the classifier predicts it as normal or novel. Similarly, in Einarsdóttir et al. (2016), a study in food control is carried out. Specifically, in some cases, foreign objects can be found inside food envelopes. Since this situation can result in bad customer experience and legal issues, the detection of foreign objects is crucial. The authors learn a classifier using X-ray images only from normal food (without foreign objects inside). Next, giving a new unseen X-ray image, the classifier predicts it as novel (with foreign object inside) or normal. The novelty term has also been commonly used in streaming scenarios. Masud et al. (2013) start from a labeled dataset, where an initial model is learned. This model classifies the new incoming instances either among the normal known classes or as novel (the instance is not similar to any known class). If this new instance is classified as novel, it is kept in a buffer because it is considered as a candidate for a new class. When this buffer is full, new classes are sought in this buffer. The classifier is updated with new emerging novel classes for future predictions.

Regarding the two aforementioned problems, two different learning scenarios can be considered. What we call the *static novelty detection* learning scenario is considered. Here, the problem can be cast as a binary supervised classification problem. Given a dataset composed by only one class, a model is built. This model learns a decision boundary that isolates the normal behavior. For prediction, when a new instance arrives, it is classified as novel or as normal. In this framework, the efforts are focused on correctly classifying the normal class

TRAINING SET

$x_1, x_2, x_3, \ldots, x_n$ | Normal

$x_1, x_2, x_3, \ldots, x_n$ | Normal

$x_1, x_2, x_3, \ldots, x_n$ | Normal

$x_1, x_2, x_3, \ldots, x_n$ | Normal

○　○　○

$x_1, x_2, x_3, \ldots, x_n$ | Normal

LEARNING ALGORITHM

$x_1, x_2, x_3, \ldots, x_n$ | ?

MODEL

PREDICTED CLASS
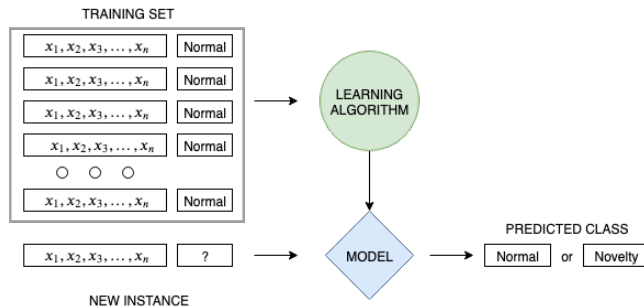
Normal | or | Novelty

NEW INSTANCE

Figure 3: A flowchart of the supervised classification framework task in which only one class is available to learn the classification model. This learning scenario is assigned to the the static novelty detection term.

(Pimentel et al., 2014; Einarsdóttir et al., 2016; Kafkas and Montaldi, 2018). Therefore, in order to evaluate the performance of the classifiers, the recall of the normal class is commonly maximized (Swarnkar and Hubballi, 2016; Luca et al., 2016). Formally, the training set is generated only from $P(\mathbf{x}|C = \mathcal{N})$. At the training stage, even though the classifier is learned using information about only one class (normal class), it is built considering that another behavior exists which is different that which is normal.

Formally, an instance is defined as $\mathbf{x} = (x_1, \ldots, x_m)$. Given a training dataset, $\mathbf{TR} = \{(\mathbf{x}_1, y_1 = \mathcal{N}), \ldots, (\mathbf{x}_n, y_n = \mathcal{N})\}$, the objective is to learn a classifier that will be able to predict between normal $\mathcal{N}$ and *novel*. Note that, in this learning scenario, only one class, the normal class $\mathcal{N}$, is available to train the model. An illustrative example can be graphically seen in Figure 3.

Besides, what we call *dynamic novelty detection* is considered. In some situations, in the literature, it is also known as *evolving classes*, *future classes* or *novel class detection* (Masud et al., 2013; Mu et al., 2018; Faria et al., 2016). This learning scenario can be formalized as a supervised classification problem in which the number of labels for the class variable is unknown. In other words, the generative probability distribution dynamically changes during the classification process. Therefore, the classifier has to adapt to these changes. When a new instance arrives, the model has to classify among the current classes or it stores it in a buffer (Masud et al., 2013; Spinosa et al., 2007; Zhu et al., 2018). Considering the life-cycle of the classes, these can drift, be born, die or reappear. Hence, the classifier must be updated for those changes, considering that the adaptation time is relevant in a streaming environment. Note that most of the existing approaches consider a dynamically (highly) unbalanced supervised classification scenario (Masud et al., 2013; Spinosa et al., 2007; Chen et al., 2008; Zhu et al., 2018) since a few instances may constitute a new emerging class (Figure 5). To evaluate the performance of the classifier in this environment, genuine metrics have been proposed. For instance, Masud et al. (2013) use the percentage of novel class instances classified as a current class; the percentage of
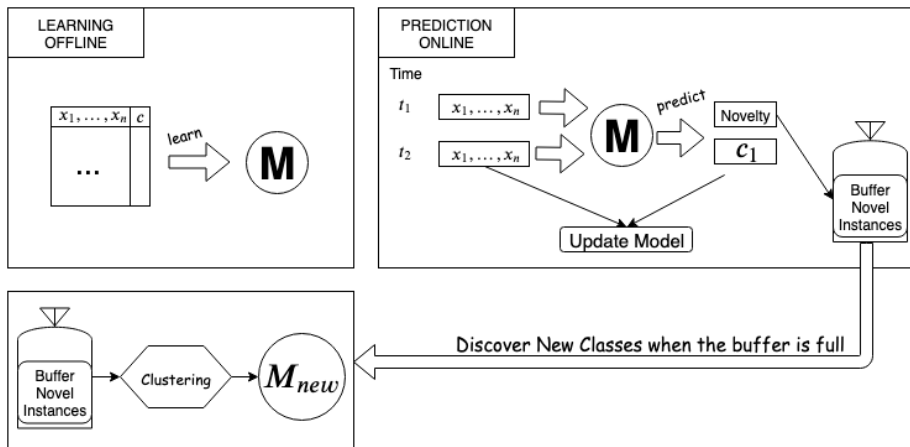
Figure 4: A flowchart of the dynamic novelty detection problem. In this problem, the number of labels for the class variable is unknown, and dynamically changes over time.

existing class instances falsely identified as novel; and, the total misclassification error. Zhu et al. (2018) use the average precision among all classes. Chen et al. (2008) output the evolution of the classification error as new events occur: the emergence of a class, disappearance or drift.
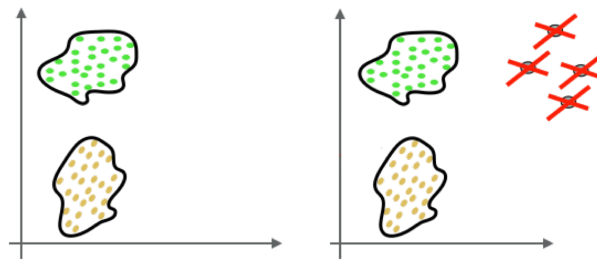
The dynamic novelty detection learning scenario can be divided in two stages. Firstly, the initial learning stage (also known as offline stage), in which given a labeled training dataset, a model is built. Secondly, the prediction stage (also known as online stage), in which new classes may emerge and disappear, and the old classes may also drift. These two phases are formalized as follows:

Initial training phase (offline): In the offline phase a classifier C_0 is learned considering a set of labels L_0.

Prediction phase (online): The online phase can be described as a prediction and adaptation stage in which a data stream (DS) is observed. A DS is a possibly infinite sequence of instances. At time t, the current classifier C_t predicts a new instance. If the instance is classified as one of the current labels, the classifier is adapted with this knowledge to create C_{t+1}. If the new instance can not be classified in the current set of labels, it is kept apart in a buffer and the model does not modify. Once the buffer is full the classifier is updated and the set of labels L_t modified.

An illustrative flowchart of this learning scenario can be seen in Figure 4.

According to the techniques used in *static novelty detection*, one class classification techniques are those which are the most representative ones in this learning scenario. For instance, one class SVM (Dufrenois and Noyer, 2016; Erfani et al., 2016; Khreich et al., 2017), K-Nearest Neighbors data description (Tax, 2001), graph embedded one class classifiers (Mygdalis et al., 2016), one class Random Forests (Désir et al., 2013) and Isolation Forest (Zhang et al.,

(a) Initial training example. (b) Classification of instances.
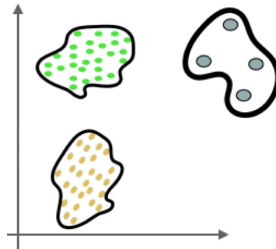
(c) Class discovery.

Figure 5: Flowchart of the dynamic novelty detection learning scenario. At the beginning, the given classes are modeled. When new instances arrive, they are classified among known classes or they are rejected as not belonging to any existing class (see the crossed instances in Figure 5b). Finally, the new emerging class is sought.

2011) have been successfully applied under the *static novelty detection* learning scenario. Besides, in *dynamic novelty detection*, techniques such as OLINDDA (Spinosa et al., 2007), a sphere-based novelty detection algorithm, in which clustering is done with the k-means algorithm; MuENLForest (Zhu et al., 2018) which discovers new labels in a multi-label classification framework by creating an ensemble of Random Forest and Isolation Forest classifiers to discover emerging new classes; or the ensemble proposed in Masud et al. (2013), which creates an ensemble of decision trees which, in each leaf node, runs a k-means algorithm to discover sphere-shaped emerging new classes, have been successfully proposed in the literature.

5   The related outlier detection scenario

The outlier term also comes up when seeking related works with rare event, anomaly and novelty terms. While the term is mainly associated with an unsupervised framework, the literature shows examples where the term is used to name other previously explained scenarios (Hodge and Austin, 2004; Zhang and Zulkernine, 2006; Billor et al., 2000). Therefore, it is briefly considered in this section.

In some papers, the term outlier has been related with *noise*, linking these observations with incorrect or inconsistent behaviors (Aggarwal, 2017). Consequently, the outlier detection task forms part of a preprocessing phase (Teng et al., 1990; Rousseeuw et al., 2011). For instance, when human errors are introduced retrieving data, these erratic observations are considered outliers (Barai and Lopamudra, 2017). In other situations, the detection of instances with high deviation are considered outliers (Radovanović et al., 2015; Xuan Hong Dang et al., 2014). In Radovanović et al. (2015), the authors detect all-star players in an unlabeled dataset composed by NBA players between 1973 and 2003. The outstanding NBA players are considered outliers. In order to detect them, clustering is pursued and those points which deviate significantly from others are considered outstanding NBA players.

Regarding the characteristics of the data in the outlier detection scenario, it can be either temporal (time-series) (Gupta et al., 2013) or non-temporal (Aggarwal, 2017; Campos et al., 2016; Radovanović et al., 2015).

An outlier detection task can be formalized as an unsupervised classification problem. Formally, given a dataset $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, the objective is to find the instance that (highly) deviates from others. An example of an outlier detection task can be seen in Figure 6.

6   The proposed assignment of terms and learning scenarios

In this paper, based in our experience and initial approach to the literature (see the list of key references at the end of the paper), we did discover two major
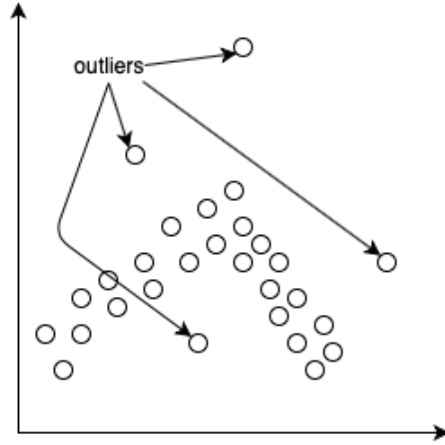
Figure 6: An example of unsupervised classification. This learning scenario is assigned to the outlier detection task. As can be seen the outliers are deviated instances without a clear pattern.

Relative to	Characteristics	Rare Events	Anomaly	Novelty	Outlier
Data	Temporal data	Yes	No	No	Possible
Data	All classes represented in training data	Yes	Yes	No	-
Problem	Unbalanced Classification	Yes	Yes	Possible	-
Problem	Supervised Classification	Yes	Yes	Yes	No

Table 2: Summary of the principal characteristics, extracted from the literature, of the reviewed terms and learning scenarios.

issues: a) the existence of a problematic mix-up between terms and learning scenarios. And b) we realize that most of these problems can be put in the same learning framework. Furthermore, we based on the assignment of terms to problems in these key papers to design our taxonomy. For each paper, we have reviewed the goal of the paper, the characteristics of the input data and the most representative techniques used in each rare event, anomaly, novelty and outlier detection works. Concretely, for each term related paper, the problem that the authors want to solve, such as, whether it is a time series classification, has an unbalanced learning characteristic, it is a classification task or a regression task, which the evaluation measures are, and, if it is a supervised or unsupervised classification problem has been reviewed.

In Figure 7, the assignment of terms to learning scenarios is graphically explained. As can be seen, each term is associated with one learning scenario. Moreover, the genuine characteristics of each learning scenario are shown in this figure. Also, an extended summary is exposed in Table 4.

In the case of the rare event term, the most relevant learning scenario under the supervised classification point of view is the (early) time series classification. In most of the papers described with the rare event term, there is a temporal nature in the problem, the classes are unbalanced and all the classes are represented in the training set.

In the problems described with the anomaly term, the most relevant learning scenario is the (highly) unbalanced supervised classification. In this learning scenario, the data is static, the distribution of classes is unbalanced, and all the classes are represented in the training set.

Regarding the problems described with the novelty term, two different learning scenarios are considered. On the one hand, the static novelty detection in which the objective is to classify an instance between novelty or normal based on a model which has been trained with only the normal class. On the other hand, the dynamic novelty detection is considered. In this learning scenario, the objective is to discover new emerging classes in an streaming environment. However, both learning scenarios share some common characteristics, such as: both of the learning scenarios are supervised, and, both of them try to discover instances from classes that were not available in the training set. Hence, both of the learning scenarios do not have all the classes represented in the training data (in the case of static novelty detection, the novel class is not available. In the case of dynamic novelty detection, the new novel classes are neither available in the training set).

Finally, the outlier detection term has been mostly associated with the unsupervised classification framework in the literature.

All these learning scenarios require specific measures in order to evaluate the performance of the classifiers that solve the related problems. Therefore, depending on the objective of the classification task, different measures are commonly computed in the literature. In Table 3, the most common evaluation measures for each term are presented. Regarding the evaluation techniques used to validate the performance of the classifier, in the majority of the papers the k-fold cross validation, stratified k-fold cross validation and the train and test
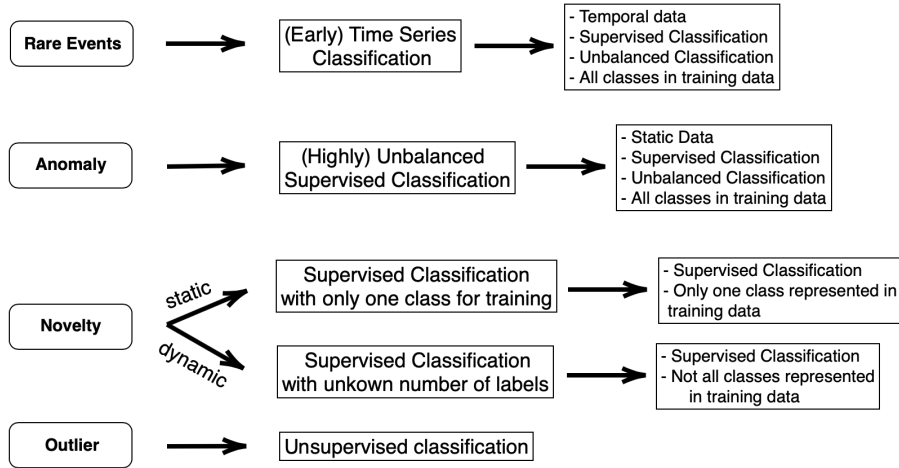
Rare Events → (Early) Time Series Classification →
- Temporal data
- Supervised Classification
- Unbalanced Classification
- All classes in training data

Anomaly → (Highly) Unbalanced Supervised Classification →
- Static Data
- Supervised Classification
- Unbalanced Classification
- All classes in training data

Novelty
→ *static* → Supervised Classification with only one class for training →
- Supervised Classification
- Only one class represented in training data
→ *dynamic* → Supervised Classification with unkown number of labels →
- Supervised Classification
- Not all classes represented in training data

Outlier → Unsupervised classification

Figure 7: Assignment of terms to learning scenarios. The main characteristics of each learning scenario have been summarized.

split are used.

7 Validation of the proposed assignment

In order to validate this proposal of assignment, an experiment has been carried out considering two different scenarios. In the first scenario, the most cited papers after the year 2000 have been gathered; while in the second scenario, the first search-results after 2014 have been considered. In both scenarios, for each paper, two terms are obtained. On the one hand, that used by the authors to describe the problem, and on the other hand, that which would have been assigned with our taxonomy. In this way, a confusion-like matrix has been formed for every scenario.

In order to retrieve these papers, Google Scholar, ACM Digital Library and IEEE Xplore search engines have been used individually. Hence, the experiment is replicated for each individual search engine. In this way, the possible differences between these three communities have been checked.

The goal of the experiment is two-fold. Firstly, we would like to validate the presented proposal of assignment of terms to learning scenarios, and check when it matches the majority of the literature papers. Secondly, we would like to identify the most frequently confused learning scenarios between pairs of terms. Finally, we have also tested if the confusion varies between different communities and, hence, different search engines have been considered.

According to the confusion matrix of the most cited papers (Tables 5, 7 and 9), the terms used to describe the different types of abnormalities mostly

Table 3: Summary of the most used evaluation measures of each term related learning scenario.

Term	Evaluation Measures	Formula	References
Rare Event	Accuracy	$\frac{\eta_{\mathcal{A}}+\eta_{\mathcal{N}}}{|D|}$	Zhang et al. (2017); Xu et al. (2016); Ren et al. (2016)
	Recall of Rare Events	$\frac{\eta_{\mathcal{A}}}{\eta_{\mathcal{A}}+\eta_{\mathcal{A}\to\mathcal{N}}}$	
	AUC	Area under ROC curve	
	Earliness	$\frac{1}{|D|}\sum_{x\in D}\frac{t_x^*}{L}\cdot 100\%$	Mori et al. (2018)
Anomaly	Accuracy	$\frac{\eta_{\mathcal{A}}+\eta_{\mathcal{N}}}{|D|}$	Swarnkar and Hubballi (2016); Luca et al. (2016)
	Recall of Anomalies	$\frac{\eta_{\mathcal{A}}}{\eta_{\mathcal{A}}+\eta_{\mathcal{A}\to\mathcal{N}}}$	
Static Novelty	Accuracy	$\frac{\eta_{\mathcal{A}}+\eta_{\mathcal{N}}}{|D|}$	Pimentel et al. (2014); Einarsdóttir et al. (2016); Kafkas and Montaldi (2018)
	Recall of Normal	$\frac{\eta_{\mathcal{N}}}{\eta_{\mathcal{N}}+\eta_{\mathcal{N}\to\mathcal{A}}}$	
Dynamic Novelty	EN_Accuracy	$\frac{A_0+A_n}{|D|}$	Masud et al. (2013, 2009); Zhu et al. (2018); Chen et al. (2008); Mu et al. (2017)
	F-measure	$\frac{2*P*R}{P+R}$	
	Miss New	$\frac{\eta_{\text{new}\to\text{old}}}{\eta_{\text{new}}+\eta_{\text{new}\to old}}$	
	False New	$\frac{\eta_{\text{old}\to\text{new}}}{\eta_{\text{old}}+\eta_{\text{old}\to\text{new}}}$	
	Global Error	$\frac{\eta_{\text{new}\to\text{old}}+\eta_{\text{old}\to\text{new}}}{|D|}$	
	Correct Between Known	Accuracy between known instances	
Outlier	Number of outliers detected		Radovanović et al. (2015); Xuan Hong Dang et al. (2014); Campos et al. (2016)

$|D|$: Number of instances.
L: Length of the Time Series.
t_x^*: Time at which the prediction is made.
$\eta_{\mathcal{A}}$: Number of instances correctly classified from the abnormal class.
$\eta_{\mathcal{N}}$: Number of instances correctly classified from the normal class.
$\eta_{\mathcal{A}\to\mathcal{N}}$: Number of instances from the abnormal class classified as normal.
$\eta_{\mathcal{N}\to\mathcal{A}}$: Number of instances from the normal class classified as abnormal.
$\eta_{\text{old}\to\text{new}}$: Number of instances from a new class classified as an old class.
$\eta_{\text{new}\to\text{old}}$: Number of instances from an old class classified as a new class.
η_{new}: Number of instances correctly classified as a new class.
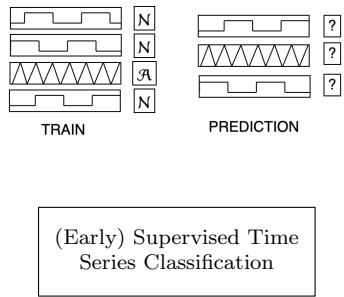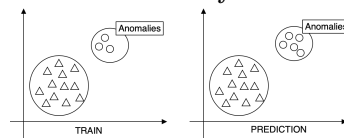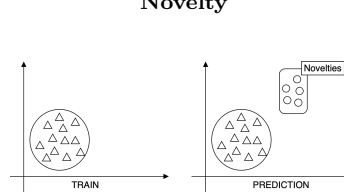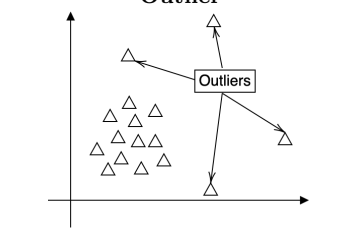η_{old}: Number of instances correctly classified as an old class.
P: Precision of the emerging class.
R: Recall of the emerging class.
A_0: Number of known instances classified as an old label.
A_n: Number of new instances classified as a new label.

Table 4: Summary of the main characteristics of each term along with the key references of the literature.

Term	Description	Key References
Rare Event (Early) Supervised Time Series Classification	• Temporal data • All classes represented in the training set • Unbalanced class distribution • Supervised Classification	Theofilatos et al. (2016); Heard et al. (2010); Dzierma and Wehrmann (2010); Hamilton (1994); Zhang et al. (2017); Ogbechie et al. (2017)
Anomaly (Highly) Unbalanced Supervised Classification	• All classes represented in the training set • Unbalanced classification • Supervised classification	Miri Rostami and Ahmadzadeh (2018); Fiore et al. (2017); Chandola et al. (2009)
Novelty Supervised Classification only one class for training	• Possible unbalanced classification • Supervised classification	Khreich et al. (2017); Einarsdóttir et al. (2016); Masud et al. (2013, 2009); Pimentel et al. (2014); Kafkas and Montaldi (2018); Swarnkar and Hubballi (2016); Luca et al. (2016); Spinosa et al. (2007); Zhu et al. (2018); Chen et al. (2008)
Outlier Unsupervised Classification	• Possible temporal data • Unsupervised classification	Hodge and Austin (2004); Zhang and Zulkernine (2006); Billor et al. (2000); Aggarwal (2017); Teng et al. (1990); Rousseeuw et al. (2011); Barai and Lopamudra (2017); Radovanović et al. (2015); Xuan Hong Dang et al. (2014); Zhang and Zulkernine (2006); Gupta et al. (2013); Campos et al. (2016)

17

match our proposal of assignment. However, in some situations, we have found discrepancies. Particularly, the highest discrepancies are found between the anomaly and rare event terms. In the case in which the authors use the anomaly term, it is frequently confused with our standardization of the rare event term. After checking the related literature, we realize that this happens when the problem has a temporal nature. Therefore, these problems would have been described with the rare event term regarding our proposal of assignment of terms. Similarly, these discrepancies are found in problems described with the rare event term but on the opposite side. When the novelty term is used by the authors to refer to the abnormalities of their problems, a minor set of papers are confused with our concept of anomaly term. In these works, instances of the novelty class are available during the training stage. Consequently, according to the presented proposal of assignment, their learning scenario is associated with the anomaly term. Finally, considering the outlier term, only a few situations are found in which the outlier detection learning scenario has been confused with the novelty detection one. In these mismatched works, a normality model is learned from labeled data. Then, instances non-conforming the normal behavior are rejected and considered outliers. Based on our proposal, this learning scenario corresponds with novelty detection.

In the second scenario with the first search-results of each term after 2014 (Table 6, 8 and 10), a similar trend can be seen. However, there is some increase in the discrepancies. The confusion of the use of the terms novelty and anomaly is noticeable. For instance, the anomaly and the novelty problem descriptors have been confused in more situations than in the previous experiment with the subset of most cited works.

Regarding the different search engines, it can be seen that the mix-up is more prominent in the ACM community. Particularly, in the first 50 search-results (Table 8), it can be seen that the mix-up between the outlier term is considerably higher than in other communities. However, this trend can not be seen in the 50 most cited papers (Table 7). Moreover, the novelty term also shows a slightly higher confusion in this community.

It can be concluded that the proposed assignment of terms to learning scenarios is supported by the literature. In addition, the confusion matrices reveal the mix-up between terms and learning scenarios. This clearly promotes the repetition of works and hinders the progress of the field. Furthermore, due to the popularity and increase of contributions in these term-related fields in recent years, this confusion is increasing. Therefore, we think that the standardization of the field is necessary and, with this review, we try to take a short step towards the solution of this mix-up.

8 Conclusions

In this paper, we have underlined those genuine characteristics of each rare event, anomaly, novelty and outlier terms that are shared by the majority of the papers in the literature and have been assigned to a learning scenario. In order

Table 5: The confusion-like matrix formed from the results obtained from Google Scholar. For each term, the 50 most cited search-results (papers) have been analyzed after the year 2000. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	37	19	4	0	60
Our	Anomaly	12	28	7	3	50
proposed	Novelty	0	1	37	1	39
term	Outlier	1	2	2	46	51
	Total	50	50	50	50	200

Table 6: The confusion-like matrix formed from the results obtained from Google Scholar. For each term, the first 50 search-results (papers) after the year 2014 have been analyzed. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	35	15	5	1	56
Our	Anomaly	13	24	16	1	54
proposed	Novelty	0	7	26	3	36
term	Outlier	2	4	3	45	54
	Total	50	50	50	50	200

Table 7: The confusion-like matrix formed from the results obtained from the ACM Digital Library. For each term, the 50 most cited search-results (papers) have been analyzed after the year 2000. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	34	10	3	2	49
Our	Anomaly	13	24	12	4	53
proposed	Novelty	2	6	24	4	36
term	Outlier	1	10	11	40	62
	Total	50	50	50	50	200

Table 8: The confusion-like matrix formed from the results obtained from the ACM Digital Library. For each term, the first 50 search-results (papers) after the year 2014 have been analyzed. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	30	15	4	6	55
Our	Anomaly	10	23	17	12	62
proposed	Novelty	1	3	20	4	28
term	Outlier	9	9	9	28	55
	Total	50	50	50	50	200

Table 9: The confusion-like matrix formed from the results obtained from the IEEE Xplore search engine. For each term, the 50 most cited search-results (papers) have been analyzed after the year 2000. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	24	15	6	4	49
Our	Anomaly	16	25	8	5	54
proposed	Novelty	5	2	25	5	54
term	Outlier	5	8	11	36	60
	Total	50	50	50	50	200

Table 10: The confusion-like matrix formed from the results obtained from the IEEE Xplore search engine. For each term, the first 50 search-results (papers) after the year 2014 have been analyzed. The terminology used by the authors is compared with respect to our proposal of assignment of terms to learning scenarios.

		Problem descriptor used in the searched paper				
		Rare Events	Anomaly	Novelty	Outlier	**Total**
	Rare Events	30	17	8	2	57
Our	Anomaly	11	24	11	4	50
proposed	Novelty	1	1	21	5	28
term	Outlier	8	8	10	39	65
	Total	50	50	50	50	200

to do that, we have reviewed the different aims of each paper, the characteristics of the input data and the most representative techniques used in each rare event, and anomaly and novelty detection works. Each term has been accompanied with a set of illustrative applications to highlight the different learning scenarios. We have argued that the learning scenarios associated to the reviewed terms can be formalized under a supervised classification framework. Finally, we hope that the discussion with the closely related outlier term can enrich the comprehension of each scenario. Finally, the main characteristics of terms and problems have been summarized in Table 2. In this table, both the features related with the available data and the characteristics of the problem have been distinguished.

With this paper, we take a short step towards the standardization of the rare event, anomaly, novelty and outlier terms. We think that our proposed assignment of terms to learning scenarios can help to resolve the muddle which hinders the progress in the term-related fields. Also, we think that the standardization of the terms and learning scenarios can strongly help to improve the progress in the field by letting the community (and especially young, newcomer researchers) to easily find what they are looking for, and by avoiding the repetition of works.

Acknowledgements

References

Aggarwal CC (2017) Outlier Analysis. Springer International Publishing

Araujo M, Bhojwani R, Srivastava J, Kazaglis L, Iber C (2018) ML Approach for Early Detection of Sleep Apnea Treatment Abandonment. Proceedings of the 2018 International Conference on Digital Health - DH '18 pp 75–79

Auffray Y, Barbillon P, Marin JM (2014) Bounding rare event probabilities in computer experiments. Computational Statistics and Data Analysis 80:153–166

Balesdent M, Morio J, Brevault L (2016) Rare Event Probability Estimation in the Presence of Epistemic Uncertainty on Input Probability Distribution Parameters. Methodology and Computing in Applied Probability 18(1):197–216

Barai A, Lopamudra D (2017) Outlier Detection and Removal Algorithm in K-Means and Hierarchical Clustering. World Journal of Computer Application and Technology 5(2):24–29

Bedford T, Cooke RM (2001) Probabilistic risk analysis : foundations and methods. Cambridge University Press

Billor N, Hadi AS, Velleman PF (2000) BACON: blocked adaptive computationally efficient outlier nominators. Computational Statistics & Data Analysis 34(3):279–298

Bourinet JM (2016) Rare-event probability estimation with adaptive support vector regression surrogates. Reliability Engineering and System Safety 150:210–221

Cadini F, Agliardi GL, Zio E (2017) Estimation of rare event probabilities in power transmission networks subject to cascading failures. Reliability Engineering and System Safety

Campos GO, Zimek A, Sander J, Campello RJ, Micenková B, Schubert E, Assent I, Houle ME (2016) On the Evaluation of Outlier Detection and One-Class Classification Methods. Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference on pp 1–10

Cao H, Li XL, Woon YK, Ng SK (2011) SPO: Structure Preserving Oversampling for Imbalanced Time Series Classification. In: 2011 IEEE 11th International Conference on Data Mining, IEEE, pp 1008–1013

Chandola V, Banerjee A, Kumar V (2009) Anomaly detection. ACM Computing Surveys 41(3):1–58

Chen S, Wang H, Zhou S, Yu PS (2008) Stop chasing trends: Discovering high order models in evolving data. In: Proceedings - International Conference on Data Engineering, pp 923–932

Cheon SP, Kim S, Lee SY, Lee CB (2009) Bayesian networks based rare event prediction with sensor data. Knowledge-Based Systems 22(5):336–343

Désir C, Bernard S, Petitjean C, Heutte L (2013) One class random forests. Pattern Recognition 46(12):3490–3506

Dessai S, Hulme M (2004) Does climate adaptation policy need probabilities? Climate Policy 4(2):107–128

Dueñas-Osorio L, Vemuru SM (2009) Cascading failures in complex infrastructure systems. Structural Safety 31(2):157–167

Dufrenois F, Noyer JC (2016) One class proximal support vector machines. Pattern Recognition 52:96–112

Dzierma Y, Wehrmann H (2010) Eruption time series statistically examined: Probabilities of future eruptions at Villarrica and Llaima Volcanoes, Southern Volcanic Zone, Chile. Journal of Volcanology and Geothermal Research 193(1-2):82–92

Einarsdóttir H, Emerson MJ, Clemmensen LH, Scherer K, Willer K, Bech M, Larsen R, Ersbøll BK, Pfeiffer F (2016) Novelty detection of foreign objects in food using multi-modal X-ray imaging. Food Control 67:39–47

Erfani SM, Rajasegarar S, Karunasekera S, Leckie C (2016) High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recognition 58:121–134

Esling P, Agon C (2012) Time-series data mining. ACM Computing Surveys 45(1):1–34

Fan S, Liu G, Chen Z (2017) Anomaly detection methods for bankruptcy prediction. In: 2017 4th International Conference on Systems and Informatics (ICSAI), 17, pp 1456–1460

Faria ER, Ponce de Leon Ferreira Carvalho AC, Gama J (2016) MINAS: multiclass learning algorithm for novelty detection in data streams. Data Mining and Knowledge Discovery 30(3):640–680

Fiore U, De Santis A, Perla F, Zanetti P, Palmieri F (2017) Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences

Gupta M, Gao J, Aggarwal CC (2013) Outlier Detection for Temporal Data : A Survey. IEEE Transactions on Knowledge and Data Engineering 25(1):1–20

Hamilton JDJD (1994) Time series analysis. Princeton University Press

Heard NA, Weston DJ, Platanioti K, Hand DJ (2010) Bayesian anomaly detection methods for social networks. Annals of Applied Statistics 4(2):645–662

Hodge V, Austin J (2004) A survey of outlier detection methodologies. Artificial intelligence review 22(2):85–126

Kafkas A, Montaldi D (2018) How do memory systems detect and respond to novelty? Neuroscience Letters

Khreich W, Khosravifar B, Hamou-Lhadj A, Talhi C (2017) An anomaly detection system based on variable N-gram features and one-class SVM. Information and Software Technology 91:186–197

King G, Langche Zeng G, Fowler J, Katz E, Tomz for research assistance M, Alt J, Freeman J, Gleditsch K, Imbens G, Manski C, McCullagh P, Mebane W, Nagler J, Russett B, Scheve K, Schrodt P, Tanner M, Tucker for helpful suggestions R, Bennett S, Huth P, Tucker R (2001) Logistic Regression in Rare Events Data. Political Analysis 9(2):137–163

Köknar-Tezel S, Latecki LJ (2011) Improving SVM classification on imbalanced time series data sets with ghost points. Knowledge and Information Systems 28(1):1–23

Luca S, Clifton DA, Vanrumste B (2016) One-class classification of point patterns of extremes. Journal of Machine Learning Research 17:1–21

Masud MM, Gao J, Khan L, Han J, Thuraisingham B (2009) Integrating Novel Class Detection with Classification for Concept-Drifting Data Streams. In: Buntine W, Grobelnik M, Mladenić D, Shawe-Taylor J (eds) Machine Learning and Knowledge Discovery in Databases, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 79–94

Masud MM, Chen Q, Khan L, Aggarwal CC, Gao J, Han J, Srivastava A, Oza NC (2013) Classification and Adaptive Novel Class Detection of Feature-Evolving Data Streams. IEEE Transactions on Knowledge and Data Engineering 25(7):1484–1497

Miri Rostami S, Ahmadzadeh M (2018) Extracting Predictor Variables to Construct Breast Cancer Survivability Model with Class Imbalance Problem. Shahrood University of Technology 6(2):263–276

Mitchell TM (1997) Machine Learning. McGraw-Hill

Mori U (2015) Contributions to time series data mining departing from the problem of road travel time modeling. PhD thesis, University of the Basque Country

Mori U, Mendiburu A, Dasgupta S, Lozano JA (2018) Early Classification of Time Series by Simultaneously Optimizing the Accuracy and Earliness. IEEE Transactions on Neural Networks and Learning Systems 29(10):4569–4578

Mu X, Ting KM, Zhou ZH (2017) Classification Under Streaming Emerging New Classes: A Solution Using Completely-Random Trees. IEEE Transactions on Knowledge and Data Engineering 29(8):1605–1618

Mu X, Zhu F, Liu Y, Lim EP, Zhou ZH (2018) Social Stream Classification with Emerging New Labels. In: PAKDD, Springer, pp 16–28

Murray JF, Hughes GF, Kreutz-Delgado K (2005) Machine Learning Methods for Predicting Failures in Hard Drives: A Multiple-Instance Application. Journal of Machine Learning Research 6:783–816

Mygdalis V, Iosifidis A, Tefas A, Pitas I (2016) Graph Embedded One-Class Classifiers for media data classification. Pattern Recognition 60:585–595

Noto K, Brodley C, Slonim D (2012) FRaC: a feature-modeling approach for semi-supervised and unsupervised anomaly detection. Data mining and knowledge discovery 25(1):109–133

Ogbechie A, Díaz-Rozo J, Larrañaga P, Bielza C (2017) Dynamic Bayesian Network-Based Anomaly Detection for In-Process Visual Inspection of Laser Surface Heat Treatment. Machine Learning for Cyber Physical Systems pp 17–24

Phua C, Lee V, Smith K, Gayler R (2010) A Comprehensive Survey of Data Mining-based Fraud Detection Research. Monash University

Pimentel MAF, Clifton DA, Clifton L, Tarassenko L (2014) A review of novelty detection. Signal Processing 99:215–249

Radovanović M, Nanopoulos A, Ivanović M (2015) Reverse nearest neighbors in unsupervised distance-based outlier detection. IEEE Transactions on Knowledge and Data Engineering

Ren Y, Wang Y, Wu X, Yu G, Ding C (2016) Influential factors of red-light running at signalized intersection and prediction using a rare events logistic regression model. Accident Analysis and Prevention 95:266–273

Reynolds D (2015) Gaussian Mixture Models. In: Encyclopedia of Biometrics, Springer US, Boston, MA, pp 827–832

Ribeiro RP, Pereira P, Gama J (2016) Sequential anomalies: a study in the Railway Industry. Machine Learning 105(1):127–153

Rousseeuw PJ, Leroy AM, John Wiley & Sons, Wiley InterScience (Online service), Hubert M (2011) Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(1):73–79

Spinosa EJ, Carvalho APDLFD, Gama J (2007) OLINDDA: a cluster-based approach for detecting novelty and concept drift in data streams. In: Proceedings of the 2007 ACM symposium on Applied computing, November 2015, pp 448 – 452

Straub D, Papaioannou I, Betz W (2016) Bayesian analysis of rare events. Journal of Computational Physics 314:538–556

Swarnkar M, Hubballi N (2016) OCPAD: One class Naive Bayes classifier for payload based anomaly detection. Expert Systems with Applications 64:330–339

Tax D (2001) One-class classification. PhD thesis, Delft University of Technology

Teng H, Chen K, Lu S (1990) Adaptive real-time anomaly detection using inductively generated sequential patterns. In: Proceedings. 1990 IEEE Computer Society Symposium on Research in Security and Privacy, IEEE, pp 278–284

Theofilatos A, Yannis G, Kopelias P, Papadimitriou F (2016) Predicting Road Accidents: A Rare-events Modeling Approach. Transportation Research Procedia 14:3399–3405

Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). Geomorphology

Weiss GM, Hirsh H (1998) Learning to Predict Rare Events in Event Sequences. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining pp 359–363

Wu J, Rehg JM, Mullin MD (2003) Learning a Rare Event Detection Cascade by Direct Feature Selection. Neural Information Processing Systems (NIPS) 16:1–17

Xu J, Denman S, Fookes C, Sridharan S (2016) Detecting rare events using Kullback-Leibler divergence: A weakly supervised approach. Expert Systems with Applications 54:13–28

Xuan Hong Dang, Assent I, Ng RT, Zimek A, Schubert E (2014) Discriminative features for identifying and interpreting outliers. In: 2014 IEEE 30th International Conference on Data Engineering, IEEE, pp 88–99

Yeung DY, Ding Y (2001) Host-Based Intrusion Detection Using Dynamic and Static Behavioral Models. Pattern Recognition 36(1):229–243

Zhang D, Li N, Zhou ZH, Chen C, Sun L, Li S (2011) iBAT: Detecting anomalous taxi trajectories from GPS traces. Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11) pp 99–108

Zhang J, Zulkernine M (2006) Anomaly based network intrusion detection with unsupervised outlier detection. In: IEEE International Conference on Communications, ieeexplore.ieee.org, pp 2388–2393

Zhang S, Bahrampour S, Ramakrishnan N, Schott L, Shah M (2017) Deep learning on symbolic representations for large-scale heterogeneous time-series event prediction. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp 5970–5974

Zhou Y, Su W, Ding L, Luo H, Love P (2017) Predicting Safety Risks in Deep Foundation Pits in Subway Infrastructure Projects: Support Vector Machine Approach. Journal of Computing in Civil Engineering 31(5):04017052

Zhu Y, Ting KM, Zhou ZH (2018) Multi-label learning with emerging new labels. IEEE Transactions on Knowledge and Data Engineering