

Chapter 8

Density-Based Anomaly Detection in the Maritime Domain

Jeroen Janssens, Eric Postma, and Jaap van den Herik

8.1 Detecting Anomalous Events in the Maritime Domain

Human operators monitoring maritime safety and security typically watch a large graphical display on which all vessel movements in the coastal region are plotted (see Chap. 2). Any unexpected deviation from normality should be detected by the operators. Such deviations from normality in the real world are generally referred to as “anomalies”. Despite visual aids, anomalies may go unnoticed by human operators due to two cognitive limitations. The first cognitive limitation is that human observers are bad at maintaining vigilance for a sustained period of time [13]. The second cognitive limitation is that humans may be blind to visual changes due to attentional limitations [9].

Computers do not suffer from these limitations. Maintaining vigilance and monitoring large volumes of data are the hallmarks of computers. Of course, in comparison with human operators, computers fall short in understanding the “gist” of maritime situations. The situation awareness of maritime patterns by experienced operators relies largely on knowledge and familiarity with vessels, sea lanes, rules and regulations, the weather, and so forth. An important lesson from the early days of artificial intelligence is that such common sense or expert knowledge is very difficult to program into computers. Simply specifying all maritime knowledge in terms of rules leads to a system that has difficulty dealing with the uncertainties of the real world. These uncertainties arise, for instance, from incomplete or wrong information, noisy sensor readings, or weather forecasts. Given these considerations, the best way to proceed is to let the computer take over the tasks requiring vigilance and cognitive processing power and to leave the interpretation of the situation largely to the operator. The existence of uncertainties in the maritime

J. Janssens (✉) • E. Postma • J. van den Herik
Tilburg center for Cognition and Communication (TiCC),
Tilburg University, Tilburg, The Netherlands
e-mail: jeroen@jeroenjanssens.com; eric.postma@gmail.com; jaapvandenherik@gmail.com

domain dictates the use of probabilistic methods, which are now commonplace in artificial intelligence [2]. We focus on a particular class of probabilistic methods for anomaly detection, called the density-based methods [10].

The outline of the rest of this Chapter is as follows. Section 8.2 describes how outlier-detection tasks are represented in density-based methods. Then, in Sect. 8.3 an overview is given of existing density-based outlier detection methods. Section 8.4 presents the SOS outlier-detection method. The outlier-detection performances achieved by the SOS method are reported in Sect. 8.5. Finally, Sect. 8.6 concludes with the statement that the SOS method provides an outlier-detection method that can be successfully applied in a wide variety of domains.

8.2 Representation Space

In so-called density-based statistical methods, maritime objects (e.g., vessels) and events (e.g., vessel turns) are generally represented as points in a (potentially high-dimensional) representation space. The dissimilarity of objects or events is represented by distance. Anomalies may manifest themselves as points that are distant from all other points, so-called “outliers”. To sketch a more concrete picture of statistical density-based methods, we consider, as an example, a straightforward two-dimensional representation space where the axes, represent the *features speed over ground* and *rate of turn* of vessels. Figure 8.1 shows such a representation space.

Let us suppose that all but one vessels form a cluster. In other words: all but one vessel have approximately the same speed over ground and rate of turn. The odd-one-out vessel is separated from the cluster by a considerable distance, indicating that the speed over ground and/or rate-of-turn of this vessel differs considerably from that of the other vessels. Figure 8.1 displays such a situation. The clustered points are the inliers (open circles). The outlier (asterisk) is separated from the cluster. In this particular example, the vessel associated with the asterisk is an

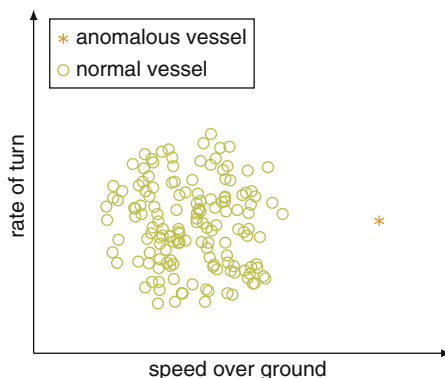


Fig. 8.1 Example of a two-dimensional representation space where the points (*open circles* and *asterisk*) represent combinations of the speed over ground (*horizontal axis*) and rate-of-turn (*vertical axis*) of vessels

outlier because it has an exceptionally large speed over ground. Its rate of turn is not anomalous, because many other vessels share approximately the same value on this feature. In realistic cases, points are anomalous on more than one feature and the detection of outliers requires taking into account multiple features, rather than just one.

The automatic detection of the odd-one-out vessel typically relies on a measure of distance, for the obvious reason that an outlier is by definition distant from all other points. The large range of outlier detection methods differ in their measurement and weighting of distance.

It is important to note that the definition of the features is crucial to the success of statistical outlier detection methods. Domain experts should be involved in the choice of the features that define the representation space. The features can be elementary, such as, the speed over ground and the rate of turn, or they can be abstractions that are known to be relevant for outlier detection, e.g., the degree to which a vessel is on a collision course with another vessel. Generally, domain experts have a good intuition about the types of information relevant to the task at hand. This intuition guides the choice of features. A useful representation for outlier detection in the maritime domain is described in Chap. 7. The number of features determines the dimensionality of the representation space and should be large enough to include the relevant information, but not too large because this hampers the ability to learn from the data [2].

8.3 Density-Based Outlier Detection Methods

This section reviews existing outlier detection methods that operate on points in representation spaces.

8.3.1 *Traditional Statistical Outlier Detection*

Traditional statistical outlier detection methods assume that points are normally distributed (i.e., the density of data points has a bell shape) and compute the average (center of the bell, μ) and standard deviation (half-width of the bell, σ) [1]. Figure 8.2 shows an example of normally distributed points on a line (i.e., a one-dimensional representation space). The bell-shaped curve represents the density of points at each position. The height of the curve is proportional to the number of points with that value of x , i.e., the feature of interest, e.g., the speed of a vessel. (x has average value μ and standard deviation σ .) The inset of Fig. 8.2 shows an enlarged view of the tail of the curve where at $x = 3\sigma$ inliers (open circles) are separated from outliers (asterisks). Statistical text books often define a point as an outlier when its distance to the average μ is more than m standard deviations. Figure 8.2 illustrates an example in which all points within a distance of 3 standard

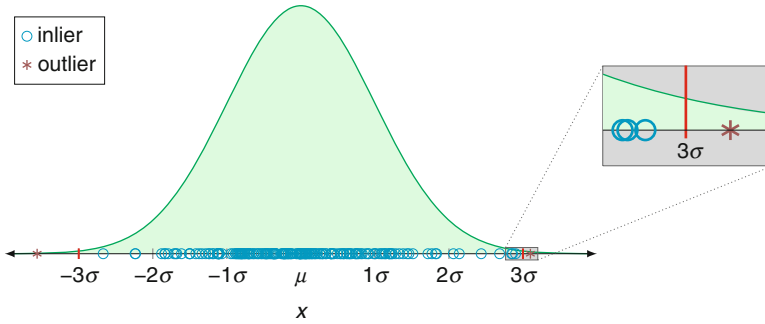


Fig. 8.2 Illustration of normally distributed points on a *line*. The bell-shaped *curve* represents the density of points at each position. The height of the *curve* is proportional to the number of points with that value of x , i.e., the feature of interest, e.g., the speed of a vessel. (x has average value μ and standard deviation σ .) The inset shows an enlarged view of the tail of the *curve* where at $x = 3\sigma$ inliers (*open circles*) are separated from outliers (*asterisks*)

deviations of the mean ($\mu \pm 3\sigma$) are considered to be inliers (represented by the open circles), those at larger distances are identified as outliers (represented by asterisks). This traditional outlier detection method is at the core of a large variety of statistical outlier detection methods. In application domains where the normality assumption holds, it offers an effective means to detect anomalies.

The main limitation of the traditional outlier detection methods is the assumption of normality, i.e., they assume that the distribution of points has a bell shape. In the maritime and many other real-world domains, data points are rarely normally distributed. Often, data points are distributed heterogeneously over space. For a single feature, the density of points does not form a single bell shape, but either multiple separated bell shapes, or totally different shapes. In two- (and higher) dimensional representation spaces, heterogeneous distributions are characterized by regions with many points (dense regions) that are interspersed with regions with few or no points (sparse regions). In terms of our example, dense regions correspond to vessels with frequently occurring speed over ground - rate of turn combinations and sparse regions correspond to the rare or no occurrence of vessels with associated speed over ground - rate of turn combinations.

8.3.2 Modern Statistical Outlier Detection: The LOF Method

Modern statistical outlier detection methods do not impose normality and deal with density variations by taking the local density into account. The most prominent density-based outlier detection method is the *Local Outlier Factor* (LOF) method [3], which originates from the domain of Knowledge Discovery and Data

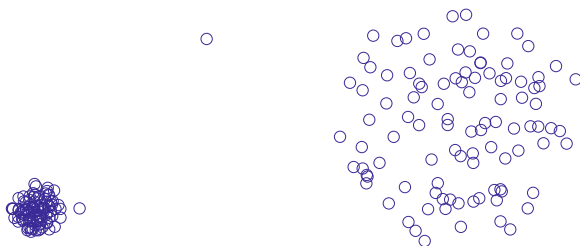


Fig. 8.3 Illustrations of two clusters of points, one with a high density (*lower left corner*) and one with a low density (*upper right corner*). Both clusters have a single outlier, but their distances from the clusters differ. LOF is capable of identifying both outliers

Mining [4]. The essence of the LOF method is that it compares the local densities of neighboring points. The local density of a point is a measure of the number of nearby points, i.e., the number of points in a predefined fixed-size spatial neighborhood. In a two-dimensional representation space, the neighborhood of a point is typically defined as a circular region around the point. A point located within a sparse region has a small local density, whereas a point located in a dense region has a high local density. The LOF method computes for each point p , an outlier value, called the Local Outlier Factor. This outlier value is obtained by dividing the averaged local densities of the points in the neighborhood (spatial vicinity) of point p by the local density of point p itself. If LOF has a value smaller or (approximately) equal to 1, the local density of point p is larger or (approximately) equal to the averaged local densities of the points in its neighborhood and the point is considered to be an inlier. Alternatively, if LOF has a value that is (much) larger than 1, the density in the neighborhood of point p is much higher than the density of the point itself, indicating that point p is an outlier.

The main advantage of the LOF method is that it can detect outliers in heterogeneous distributions of points. Returning to our two-dimensional maritime example, we consider the case of two spatially separated clusters of points representing two types of vessels, type A and type B, shown in Fig. 8.3. The speed over ground and rate of turn values of type A vessels have a small variation, whereas the speed over ground and rate of turn values of type B vessels have a large variation. As a result, the type A and B vessels give rise to clusters with high and low densities, respectively. For a type A vessel to be considered an outlier it has to be separated a certain minimal distance d_A from the type A cluster. Similarly, for a type B vessel to be considered an outlier it has to be located a certain minimal distance d_B from the type B cluster. The value of d_A is smaller than the value of d_B , because type A vessels have smaller variations in their rate of turn and speed over ground values than type B vessels. Where a purely distance-based outlier detection method would fail to take such density variations into account, the LOF method is able to identify outliers of both types.

Despite its widespread use, the LOF method suffers from two related drawbacks. The first drawback is that the LOF value is difficult to interpret. In order for a point to be an outlier, the LOF value should be much larger than 1, but how much larger depends on the problem at hand. Complex real-world domains, such as the maritime domain, are characterized by heterogeneous point distributions with unknown densities, and therefore pose a problem for the interpretation of the output of the LOF method. The second drawback is that the LOF method has no clear probabilistic foundation. As a result, LOF values cannot be interpreted in terms of probabilities. Operators assessing anomalies in maritime safety and security, would be much helped if they could assess the probability of a point being an outlier. For instance, when confronted with multiple outliers, probabilities allow them to weigh the costs of action (e.g., intercepting a vessel) against the costs of a false detection.

In recent years, a large number of density-based variants of the LOF method have been proposed. We mention three examples: the Nearest Neighbor Data Description (NNDD) method [12], the Local Correlation Integral (LOCI) method [8], and Least-Squares Outlier Detection (LSOD) method [6]. These three methods attempt to improve upon LOF in several respects, but they all suffer from the aforementioned two limitations. In the following section, we present our Stochastic Outlier Selection method, a density-based outlier detection method that does not suffer from these two limitations and we evaluate its performance by comparing it to the performances of LOF, NNDD, LOCI, and LSOD.

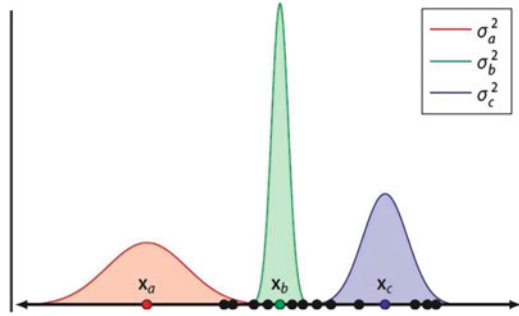
8.4 The Stochastic Outlier Selection Method

The Stochastic Outlier Selection (SOS) method [5] relies on three principles: (1) dissimilarity representation, (2) soft neighborhoods, and (3) outlier probabilities. The following three subsections describe these principles in detail.

8.4.1 *Dissimilarity Representation*

The SOS method relies on dissimilarities between points. Dissimilarities are proportional to the distances between pairs of points. The representation space is sometimes called a similarity space, because two vessels with similar speed over ground and rate of turn values are represented by nearby points, and two vessels with dissimilar values are represented by distant points. In representation space, vicinity translates into similarity, and distance into dissimilarity.

Fig. 8.4 Illustration of the bell shaped functions (soft neighborhoods) associated with three points, the circles labeled x_a , x_b , and x_c . The widths of the neighborhoods are determined by the local density, i.e., the number of neighboring points. The *solid circles* represent other points for which the soft neighborhoods are not drawn



8.4.2 Soft Neighborhoods

The SOS method does not treat all similarities equally. Inspired by insights from cognitive psychology [11], the similarity between two points separated by a distance d is given by a bell-shaped function centered at $d = 0$. In the domain of cognitive psychology, these points may represent, for instance, faces and the similarity space may be defined by two or more facial features (e.g., length of nose, size of mouth). The maximum similarity (top of the bell-shaped function) is obtained when two points are the same ($d = 0$, i.e., same lengths of nose and sizes of mouth). With growing distance between both points ($d > 0$, different lengths of nose and sizes of mouth), the similarity falls off towards zero (tail of the bell-shaped function). According to Shepard, the bell-shaped function is a universal law that relates distance to similarity [11]. In the cognitive psychology domain, the function returns the probability that two points (faces) fall in a region of representation space that are treated equally in terms of similarity judgment (“same face”, “different face”). Shepard’s similarity function is not restricted to faces, it applies to a wide variety of mental representation [11].

The bell-shaped function used in the SOS method can be interpreted as a soft version of the “hard” neighborhood used in the LOF and related methods. In a hard neighborhood, neighbor-ship changes at the circular neighborhood boundary from “neighbor” to “no neighbor”. In the soft neighborhood of the SOS method, neighboring points have a neighbor-ship value N_{val} that varies from a maximum value for $d = 0$ ($N_{SOS} = 1$, top of the bell-shaped function) towards zero values of neighbor-ship for very large values of d ($N_{SOS} \rightarrow 0$, tail of the function). In the SOS method, the widths of the soft neighborhoods centered at each point are automatically set to values to ensure that all points have the same number of neighboring points. Figure 8.4 illustrates this for a one-dimensional representation space, i.e., a line. For three points, x_a , x_b , and x_c , the associated bell-shaped soft neighborhoods are drawn. The widths of the neighborhoods depend on the local density of points. If the local density is large, the neighborhood is small, whereas if the local density is small the neighborhood is large. Through the automatic scaling of the neighborhood, the SOS method deals effectively with density variations in the data. Hence, it can deal with heterogeneous densities.

8.4.3 Outlier Probabilities

To determine the probability of a point being an outlier, the SOS method examines for each point to what extent it is part of the soft neighborhood of all other points. If a point is highly dissimilar from all other points, it is located in the tails of all the associated soft neighborhoods. Being located in a tail implies a very low neighborhood-ship value, N_{SOS} , that is near to zero. In the formal definition of the SOS method, the neighborhood-ship values are expressed by the term $(1 - N_{SOS})$, where being located in a tail translates to a value that is near to one. The outlier probability of the i -th point indexed, $P_{outlier}(i)$ is proportional to the product of all these neighborhood-ship terms and is formally defined as:

$$P_{outlier}(i) = \prod_{j=1, j \neq i}^K (1 - N_{SOS}(j)), \quad (8.1)$$

where K is the total number of points and $N_{SOS}(j)$ is the neighbor-ship value of the j -th point.

8.5 Performance of the SOS Method

We evaluated the performance of the SOS method by comparing it to the performance of state-of-the-art outlier detection methods. Two such comparative evaluations were performed: one qualitative evaluation on artificial datasets and one quantitative evaluation on realistic datasets. In all evaluations, the parameters of the outlier detection methods were optimized to yield the best performance.

8.5.1 Evaluation on Three Artificial Datasets

To get some insights into the performances of the SOS method in comparison to the other outlier-detection methods LOF, NNOD, LOCI, and LSOD, we defined three different artificial two-dimensional datasets: *Banana*, *Densities*, and *Ring*. The Banana dataset consists of a banana-shaped cluster of points. The Densities dataset consists of two separated circular clusters of points with different densities, and the Ring dataset contains points arranged in a ring-shaped form. Applying an outlier-detection method to the Banana dataset tests if distance from a cluster of points affects the outlier value appropriately. Applying it to the Densities dataset tests if the method takes the different densities into account. Finally, applying the method to the Ring dataset tests if points inside and outside the ring are evaluated similarly. For the Banana dataset, the outlier values assigned to points should vary with distance from the shape of the banana and become gradually larger with

increasing distance from the points. For the Densities dataset, the rate of change from blue to red should be slower for the low-density cluster than for the high-density cluster. Finally, for the Ring dataset, outliers within the ring should be treated similarly to outliers outside the ring.

Figure 8.5 displays the representation spaces of the three datasets (the three columns) with the color-coded outlier values superimposed. The white dots are the points forming the datasets. The top row shows for SOS the outlier values (probabilities) assigned to each representation-space location. The colors range from dark blue (inliers; smallest outlier value or probability $P_{outlier} = 0$) to dark red/brown (outliers; largest outlier value or probability $P_{outlier} = 1$). On the Banana dataset, the SOS method assigns outlier values that vary smoothly with the shape of the banana and become gradually larger with increasing distance from the points. For the Densities dataset, the rate of change from blue to red is appropriately slower for the lower left cluster (which has a low density) than for the upper right cluster (which has a high density). For the Ring dataset, outliers within the ring are treated similarly to outliers outside the ring.

The bottom four rows of the figure illustrate how other state-of-the-art methods assign outlier values to locations in similarity space. For the Banana dataset, the outlier values generated by LOCI and LSOD fail to follow the banana shape of the points. For the Densities dataset, the other methods yield quite different outlier-value assignments. Finally, for the Ring dataset, all other outlier detection methods fail to treat interior and exterior ring locations equally in terms of outlier value assignment.

These qualitative evaluations show that the different methods behave differently on different data distributions. We now turn to a quantitative assessment of their performances on realistic datasets.

8.5.2 Performance Evaluation on Realistic Datasets

Our quantitative evaluation aims to assess the outlier detection performance of the SOS method in comparison with its main alternatives. In practical outlier-detection tasks, a wide variety of heterogeneous point distributions may arise. To ensure generality of our comparative evaluation, we decided to select 18 datasets, each from a completely different realistic application domain. We evaluated the outlier-detection performance in terms of what we call the “weighted AUC”, a performance measure that takes into account the detection rate and the false positive rate and expresses the outlier-detection performance on a scale ranging from 0 (worst performance) to 1 (best performance). Figure 8.6 displays a plot of the results of the comparative evaluation. For each outlier-detection method, it shows the weighted AUC (vertical axis) achieved on each dataset (horizontal axis). The curves connect the performances of a single method. The SOS method achieves the best performance overall, because the curve associated with the SOS method (purple curve with diamond markers) is almost always on top.

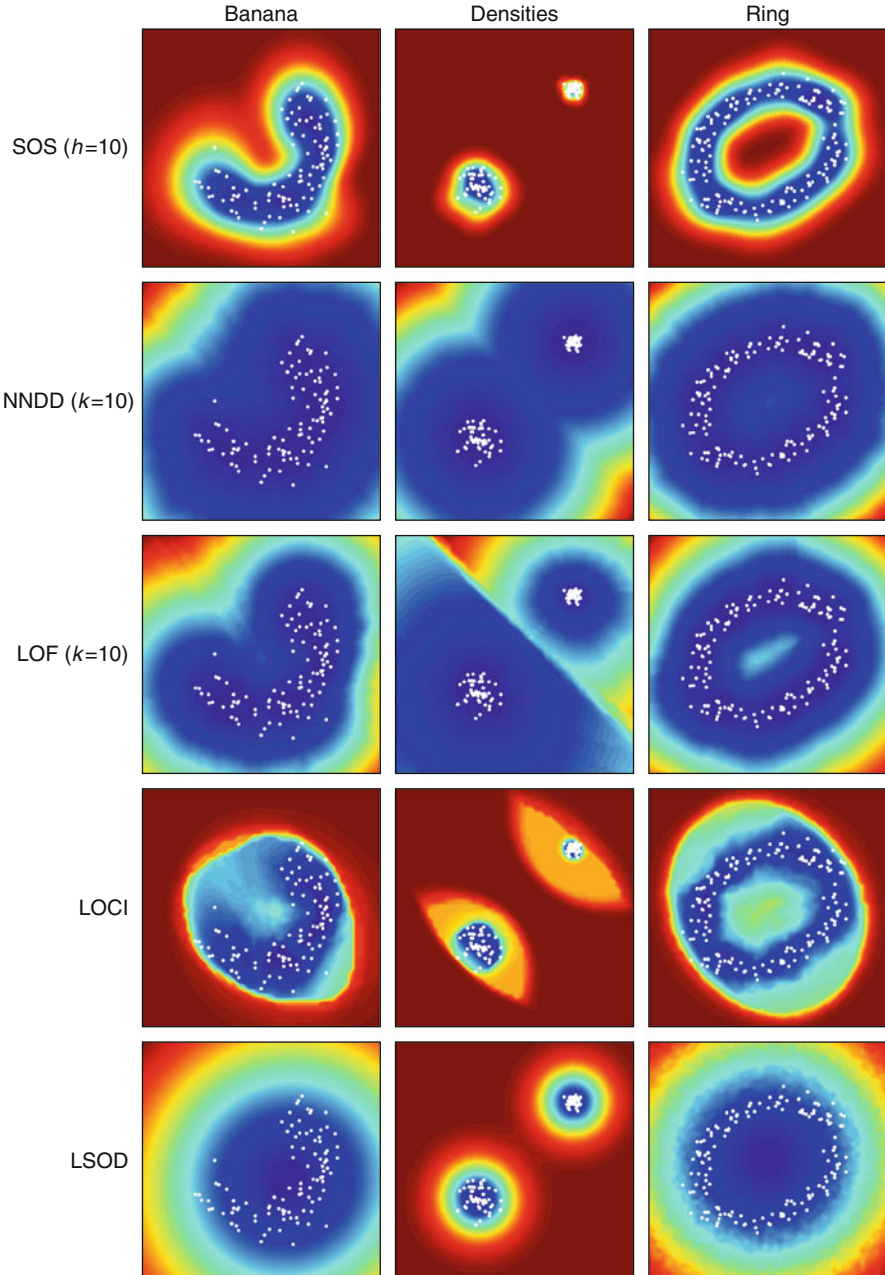


Fig. 8.5 Qualitative (visual) evaluation of outlier scores assigned to three datasets (columns) by the SOS method (*top row*) and four other state-of-the-art outlier detection methods (*bottom four rows*). Each *square* shows a two-dimensional representation space containing points (*white dots*). All other locations are colored according to the outlier value generated for that location. Outlier values are color-coded and range from *dark blue* (inliers) to *dark red/brown* (outliers)

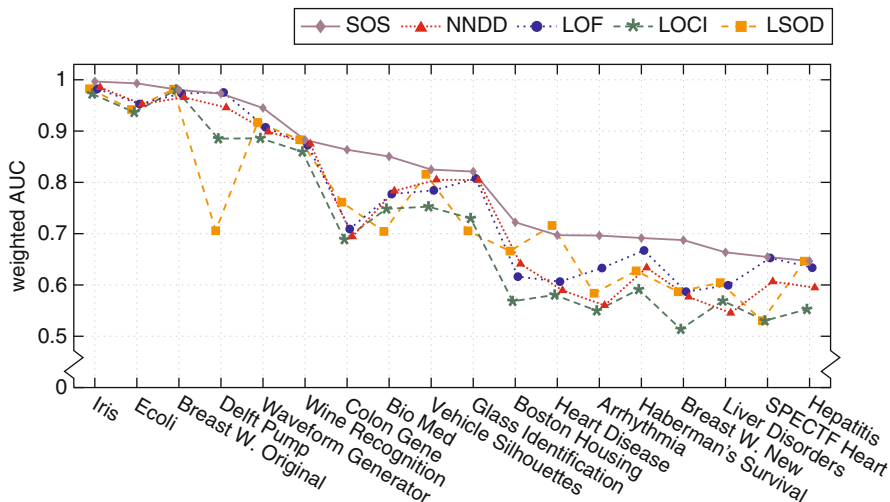


Fig. 8.6 Comparative evaluation of the SOS method and four competitive methods (NNDD, LOF, LOCI, and LSOD) on 18 realistic datasets. The outlier-detection performance is expressed in terms of the weighted AUC which ranges from 0 (worst performance) to 1 (best performance)

	Outlier-selection algorithm				
	SOS	NNDD	LOF	LOCI	LSOD
Average AUC	0.811	0.748	0.763	0.716	0.742
Average rank	1.250	3.444	2.833	4.639	2.833

Fig. 8.7 Numerical summary of the comparative evaluation of the SOS method and four competitive methods. The average AUC is the weighted AUC averaged over all 18 datasets. The average rank is obtained using a statistical method that determines the ranking of the methods on the basis of their performances. Smaller ranks correspond to better performing methods

A numerical summary of the results is presented in Fig. 8.7. The row labeled “Average AUC” lists the weighted AUC averaged over all 18 datasets. The row labeled “Average rank” specifies the ranks of the methods as obtained from a statistical method [7] that determines the ranking of the methods on the basis of their performances. The statistical method is necessary because we compare average performances of outlier-detection methods and we would like to assess the probability that differences in performance are due to chance. Smaller ranks correspond to better performing methods. The SOS method outperforms all other methods in terms of average AUC and achieves the highest rank.

8.6 Discussion and Conclusion

The SOS method has been shown to provide the best overall performance on our selection of 18 realistic datasets, as compared to state-of-the-art outlier detection methods. In addition to this result, the SOS method has an important advantage over existing density-based methods. It provides easily interpretable outlier values that correspond to probabilities. When confronted with many (potential) outliers, operators working in the maritime domain (or any other realistic domain) may prioritize the outliers using their associated probabilities, by dealing with the most probable outlier first.

It is a well-known fact in machine learning that there is no single best method for a given dataset or application domain. Similarly, we do not claim that the SOS method is the best method of choice for all domains. We have observed that the SOS method often, but not always, outperforms competitive methods.

Although we have succeeded in developing a density-based outlier-detection algorithm that performs well in comparison to state-of-the-art algorithms, a more extensive evaluation in the maritime domain has still to be performed. Provided that the maritime representation space is defined in cooperation with domain experts, we are confident that the SOS method will be successful in detecting outliers. We conclude that the SOS method provides an outlier-detection method that can be successfully applied in a wide variety of domains.

Acknowledgements This research has been carried out as a part of the POSEIDON project at Thales under the responsibilities of the Embedded Systems Institute (ESI). This project is partially supported by the Dutch Ministry of Economic Affairs under the BSIK program.

References

1. Barnett V, Lewis T (1994) Outliers in statistical data. Wiley series in probability and mathematical statistics, 3rd edn. Wiley, Chichester
2. Bishop CM (2006) Pattern recognition and machine learning. Series in information science and statistics. Springer, New York
3. Breunig MM, Kriegel HP, Ng RT, Sander J (2000) LOF: identifying density-based local outliers. ACM SIGMOD Rec. Dallas, TX 29(2):93–104
4. Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) Knowledge discovery and data mining: towards a unifying framework. Knowl Discov Data Min 82–88
5. Janssens JHM, Huszar F, Postma EO, van den Herik HJ (2012) Stochastic outlier selection. Technical report TiCC TR 2012-001, Tilburg University, Tilburg Center for Cognition and Communication, Tilburg, The Netherlands
6. Kanamori T, Hido S, Sugiyama M (2009) A least-squares approach to direct importance estimation. J Mach Learn Res 10:1391–1445
7. Nemenyi P (1963) Distribution-free multiple comparisons. Ph.D. thesis, Princeton
8. Papadimitriou S, Kitagawa H, Gibbons PB, Faloutsos C (2003) LOCI: fast outlier detection using the local correlation integral. In: Proceedings of the 19th international conference on data engineering, Bangalore, India, pp 315–326

9. Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: the need for attention to perceive changes in scenes. *Psychol Sci* 8:368–373
10. Riveiro M, Falkman G, Ziemke T (2008) Improving maritime anomaly detection and situation awareness through interactive visualization. In: 11th international conference on information fusion, 2008. IEEE, Piscataway, pp 1–8
11. Shepard RN (1987) Toward a universal law of generalization for psychological science. *Science* 237(4820):1317–1323
12. Tax DMJ (2001) One-class classification: concept-learning in the absence of counter-examples. Ph.D. thesis, Delft University of Technology, Delft, The Netherlands
13. Warm JS, Parasuraman R, Matthews G (2008) To see or not to see: the need for attention to perceive changes in scenes. *Hum Factors* 50:433–441