Directional anomaly detection

Oliver Urs Lenz[1 2] and Matthijs van Leeuwen[1]

[1] Leiden Institute of Advanced Computer Science, Leiden University
o.u.lenz@liacs.leidenuniv.nl
[2] Research Group for Computational Web Intelligence,
Department of Applied Mathematics, Computer Science and Statistics,
Ghent University

Abstract. Semi-supervised anomaly detection is based on the principle that potential anomalies are those records that look different from normal training data. However, in some cases we are specifically interested in anomalies that correspond to high attribute values (or low, but not both). We present two asymmetrical distance measures that take this directionality into account: ramp distance and signed distance. Through experiments on synthetic and real-life datasets we show that ramp distance performs as well or better than the absolute distance traditionally used in anomaly detection. While signed distance also performs well on synthetic data, it performs substantially poorer on real-life datasets. We argue that this reflects the fact that in practice, good scores on some attributes should not be allowed to compensate for bad scores on others.

Keywords: anomaly detection, average localised proximity, distance measures, nearest neighbour distance, one-class classification, risk factors

1 Introduction

A defining characteristic of semi-supervised anomaly detection (also known as *one-class classification* [22]) is that the training set only contains normal data, but that we still want to create a model that can distinguish between normal and anomalous data. Potential reasons for working in this setting include not wanting to assume that the available examples of anomalies are representative of all anomalies, and wanting to flag all data that is different from normal data as a potential anomaly. Examples of application domains of anomality detection include failure detection in industrial settings and screening patients in healthcare.

In the present paper, we are specifically concerned with quantitative tabular data, where the training set consists of a representative sample of a certain normal target class. In this context, all semi-supervised anomaly detection algorithms are necessarily based on the principle that anomality increases with distance away from the normal training data in the feature space. However, there are circumstances in which we possess additional helpful domain knowledge. In this paper, we consider how best to deal with the knowledge that only relatively high values of an attribute should be indicative of anomality, not relatively low values (or vice-versa). For instance, certain attributes may correspond to higher strain of a machine, or encode known risk factors for patients. In such cases, we may want to decide that we are only interested in detecting anomalies that express high strain, or high risk factors, i.e., that we want to detect machine failure and at-risk patients, rather than underutilisation and exceptionally healthy patients.

To this end, we propose adaptations of two effective anomaly detection algorithms, Nearest Neighbour Distance (NND) and Average Localised Proximity (ALP), that take this directionality into account.

Other attempts to take domain knowledge into account include *contextual anomaly detection*, where anomality is conditioned on one or more contextual attributes [19, 14], and *fair anomaly detection*, where the goal is to ensure that anomalies follow the same distribution over one or more sensitive attributes as normal records [16, 26, 18]. However, to the best of our knowledge, this is the first work to investigate *directional anomaly detection*.

2 Background: NND and ALP

Semi-supervised anomaly detection algorithms, also known as one-class classifiers or data descriptors, take a training set X consisting of normal data and learn a model that assigns an anomaly score to new instances. In the present paper, we will use two such algorithms: Nearest Neighbour Distance (NND) and Average Localised Proximity (ALP), both of which are directly based on distance measurements in the feature space.

2.1 NND

NND is one of the simplest anomaly detection algorithms and goes back to at least [10]. In NND, the anomaly score of a test record corresponds to the distance to its kth nearest neighbour in the training set. Despite its simplicity, NND performs surprisingly well across diverse datasets [12], making it an attractive baseline. In [11], a slight modification was proposed, taking the linearly weighted average of the first k nearest neighbour distances. Thus, the anomality score of a test record y becomes

$$\sum_{i \leq k} w_i \cdot d_i(y), \tag{1}$$

where $d_i(y)$ is the ith nearest neighbour distance of y in the training set according to some distance measure d, and w_i is the ith weight.

2.2 ALP

The idea of ALP [12] is to offset nearest neighbour distance against what is typical for normal data in that part of the feature space. It is based on the earlier form of localised nearest neighbour distance first proposed in [17] and [23].

ALP calculates a normality score in $[0, 1]$ for a test record y by taking the so-called weighted maximum of its localised proximity values:

$$w\max\left(\mathrm{lp}_i(y)\right)_{i\leq k}, \tag{2}$$

for a choice of $k \geq 1$ and weight vector w, where the weighted maximum $w\max$ is defined by

$$w\max X = \sum_{i\leq k} w_i \cdot X^{(i)}, \tag{3}$$

for any collection X, where $X^{(i)}$ is the ith largest element of X; and where $\mathrm{lp}_i(y)$ is the ith localised proximity of y to the training set, defined by

$$\mathrm{lp}_i(y) = \frac{D_i(y)}{D_i(y) + d_i(y)}, \tag{4}$$

where $d_i(y)$ is the ith nearest neighbour distance of y in the training set according to some distance measure d, and $D_i(y)$ is the average ith nearest neighbour distance in the training set local to y, defined as

$$D_i(y) = \sum_{j\leq l} w'_j \cdot d_i(\mathrm{NN}_j(y)), \tag{5}$$

for a choice of $l \geq 1$ and weights w'_j, where $\mathrm{NN}_j(y)$ is the jth nearest neighbour of y in the training set.

Note that k fulfills a similar role as k in weighted NND, whereas l controls the amount of localisation (lower values correspond to more localisation).

2.3 Hyperparameters

For both NND and ALP, the values k and l, the various weights and the distance measure are hyperparameters, which can in principle be tuned when there are anomalities available for validation purposes [13]. In general, however, we have to use sensible default values.

A typical choice for the distance d is a form of the Minkowski p-distance:

$$d(y, x) = \left(\sum_{j\leq m} |y_j - x_j|^p\right)^{\frac{1}{p}}, \tag{6}$$

for some value $p \geq 0$. The two most frequently used measures are $p = 2$, which is Euclidean distance, and $p = 1$, which is Boscovich distance, also known as city-block, Manhattan or rectilinear distance. For NND, Boscovich distance generally outperforms Euclidean distance [12, 11], while for ALP this is the standard choice anyway [12]. For both weighted NND and ALP, the default choice is to use linearly descending weights everywhere [12, 11]. For unweighted NND, $k = 1$ is optimal [12], but for weighted NND slightly higher values like $k = 8$ increase general performance [11]. For ALP, $k = 5.5 \log n$ and $l = 6 \log n$ have been established as optimal default values [12], where n is the number of training records.

While not typically seen as a hyperparameter, both NND and ALP are also dependent on the relative scale of the attributes.

3 Directional anomaly detection

As discussed in the introduction, we want to find a way to take into account *directional* attributes, for which we should only interpret extreme values in one direction as anomalous. Without loss of generality, we will assume that high values correspond with anomality — any attributes for which the opposite applies can be transformed with a sign change. We will use the following *potentially asymmetric* distance measure:[3]

$$d(y, x) = \sum_{j \leq m} d_j(y_i - x_i), \tag{7}$$

and consider three variants of the per-attribute distance $d_j(y_i - x_i)$, listed in Table 1. Absolute distance is the baseline. It corresponds to the existing Boscovich distance, i.e. Minkowski distance (6) with $p = 1$. Ramp distance is the result of applying the ramp function to the difference $y_j - x_j$. The ramp function is also known as the *rectifier* function, or *rectified linear unit* (ReLU), which is a popular activation function in neural networks [4]. Signed distance is simply the identity function applied to $y_j - x_j$.

Table 1. Variant approaches to directional attributes.

Variant	$d_j(y_i - x_i)$
Absolute	$\lvert y_j - x_j \rvert$
Ramp	$\max(0, y_j - x_j)$
Signed	$y_j - x_j$

All three variants agree for positive values of $y_j - x_j$; they interpret this as positive evidence for the anomality of y. They differ with respect to negative values of $y_j - x_j$, i.e., they differ in how to account for test values that are lower than the reference value in the training set. Absolute distance counts such

[3] Note that while this introduces ramp distance and signed distance as variants of Boscovich distance, it is also possible to substitute ramp distance in the calculation of Minkowski distance (6) with other values of p, and in particular for Euclidean distance. For signed distance, adapting Minkowski distance requires more work.

low values as positive evidence for the anomality of y just as it does high values, thereby ignoring the directionality of the attribute j. In contrast, signed distance interprets such low values as negative evidence for anomality, which can offset high values in other attributes. Finally, ramp distance stakes out the middle ground by interpreting such values as a lack of evidence either way.

Next, we consider how these variant distances affect NND and ALP. We can simply substitute the ramp distance into the overall distance for all directional attributes, and calculate the nearest neighbour distances according to the resulting measure.

For signed distance, the situation is not so straightforward. Note first that when all attributes are directional and we use signed distance, the nearest neighbour of any test record y is the training instance x that minimises $\sum_i (y_i - x_i)$. This is the training instance with the largest attribute value sum $\sum_i x_i$, and this does not depend on the test instance. Analogously, all test instances share the same subsequently nearest neighbours. Consequently, we can disregard nearest neighbours altogether, and simply compare the attribute sums of test records. Note that we still use the training set to determine the right scale of each attribute. Nevertheless, because we no longer need to perform any nearest neighbour queries we end up with a significant computational simplification.

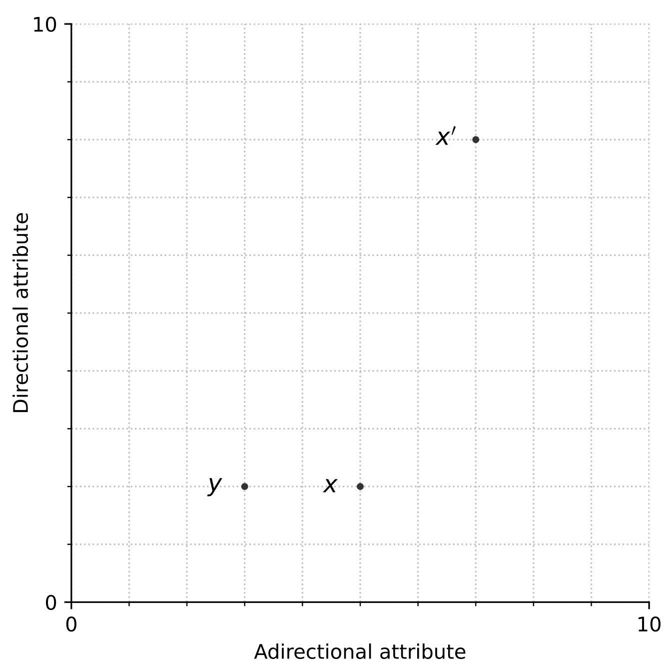

Fig. 1. A dataset with one directional and one adirectional attribute, one test record y and two training records x and x'. See text for discussion.

What should we do when we have a mixture of adirectional and directional attributes? One option is to add the signed distance from the directional attributes and the absolute distance from the adirectional attributes together to obtain a mixed distance between every test and training record and use this to identify the nearest neighbour distances of every test record. However, this would lead to the paradoxical situation illustrated by Fig. 1. The nearest neighbour of y is x', since its mixed distance $4 - 6 = -2$ is smaller than the mixed distance $2 + 0 = 2$ between y and x. However, in terms of absolute distance, y is clearly more similar to x, even if we only consider the adirectional attribute, and so it seems nonsensical to calculate the anomality of y on the basis of its distance to x'.

Therefore, we propose that for NND modified with signed distance, the right order of things is to seperately calculate the risk score for the directional attributes and the nearest neighbour distance for the adirectional attributes and add these together to obtain a single anomality score. Because of its effect on nearest neighbour calculations, we will not attempt to use signed distance with ALP.

4 Experiments on synthetic datasets

In order to evaluate our proposal, we construct a series of synthetic datasets, each containing 1000 normal training records and 200 test records — 100 normal and 100 anomalous. Each record consists of 10 attributes with randomly generated values. We will consider two types of attributes. First, we generate datasets with continuous attributes, where both normal and anomalous values follow a Gaussian distribution with standard deviation 1, centred on 0 and a respectively. We let a vary between 0 and 1 in steps of 0.1. Secondly, we generate datasets with binary attributes, where normal and anomalous values follow Bernoulli distributions with $p = 0.5 - 0.5 \cdot b$ and $p = 0.5 + 0.5 \cdot b$ respectively, for which we let b vary between 0 and 0.5 in steps of 0.05. For each value of a and each value of b we generate 100 datasets.

We evaluate anomaly detection performance using the area under the receiver operating characteristic (AUROC). This expresses the ability of an anomaly detector to separate anomalies from normal data. For this purpose, we monotonically transform the NND distance from (1) to an anomaly score in $[0, 1]$ with:

$$a \longmapsto \frac{1}{2} \cdot \frac{a}{|a| + 1} + \frac{1}{2}. \tag{8}$$

We do this for absolute distance, ramp distance and signed distance. The results for NND are displayed in Figure 2. For the Gaussian attributes, signed and ramp distance perform substantially better than absolute distance, for small and large k, in line with our directional hypothesis. Signed distance performs slightly better than ramp distance, but the difference is small. For Bernoulli attributes, the picture looks a bit different. At $k = 1$, signed distance outperforms absolute distance, but ramp distance performs worse than absolute distance. At $k = 8$,

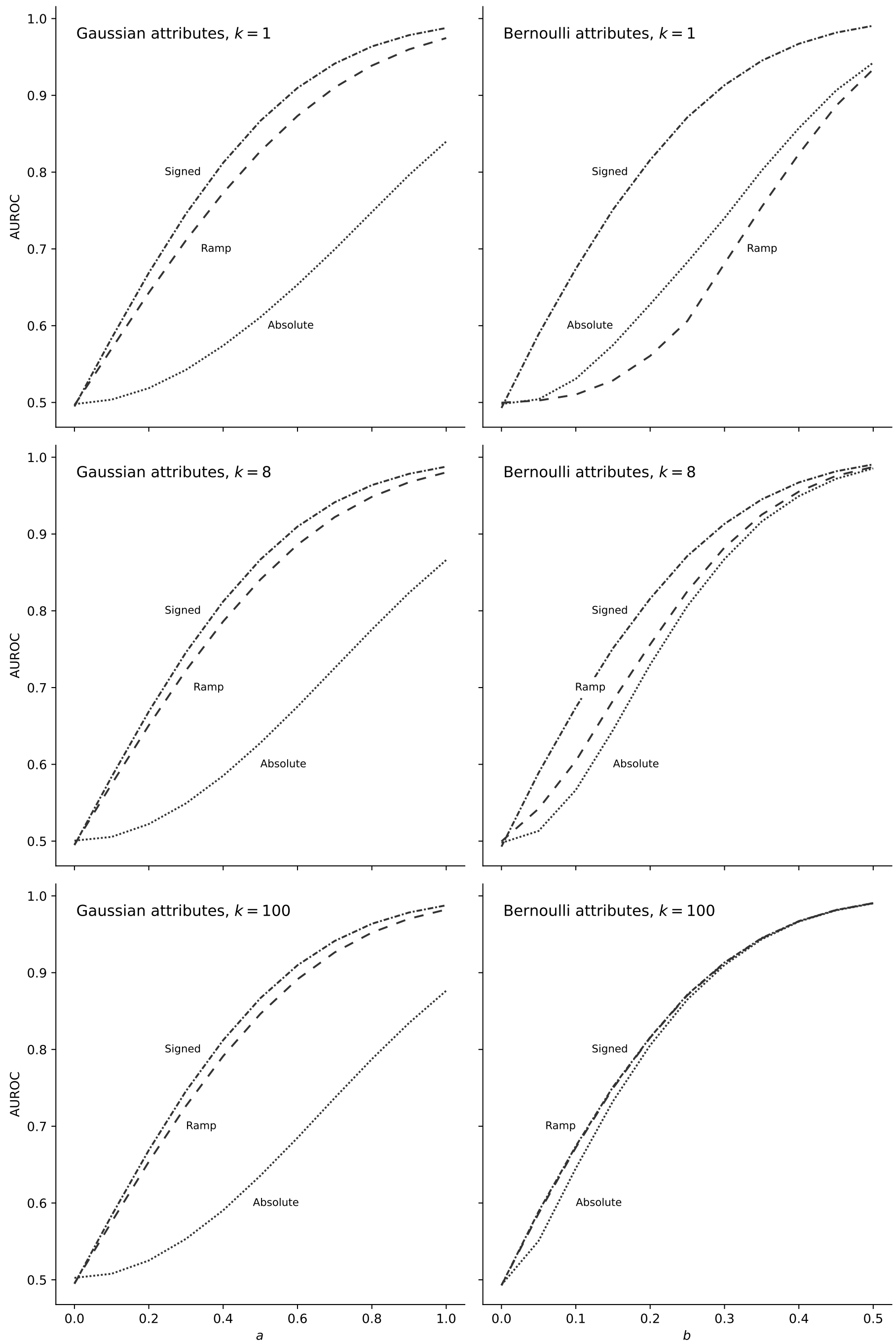

Fig. 2. Mean AUROC obtained with NND for synthetically generated datasets with Gaussian and Bernoulli attributes. a: distance between the distributions of anomalous and normal attribute values; b: difference between the probability of a positive anomalous and a positive normal attribute value.

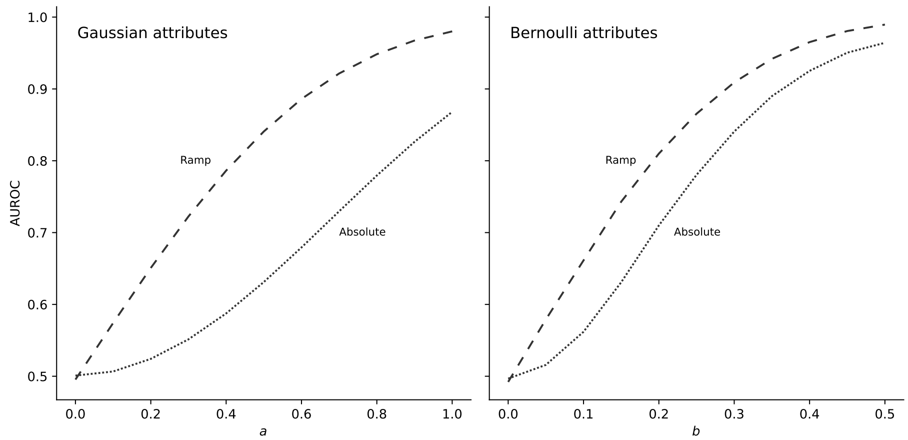

Fig. 3. Mean AUROC obtained with ALP for synthetically generated datasets with Gaussian and Bernoulli attributes. a: distance between the distributions of anomalous and normal attribute values; b: difference between the probability of a positive anomalous and a positive normal attribute value.

the gap between signed and absolute distance has tightened considerably, and ramp distance now sits in between. At $k = 100$, signed and ramp distance have become indistinguishable in terms of performance, and absolute distance is only very slightly behind.

For ALP, the results for the default hyperparameter values $k = 5.5 \log 1000 = 38$ and $l = 6 \log 1000 = 41$ are displayed in Table 3. For both Gaussian and Bernoulli attributes, ramp distance outperforms absolute distance, although more clearly so for Gaussian attributes.

Taken together, these results suggest that when we know for a given attribute that anomalies have a higher mean value than normal data, we should indeed use signed or ramp distance rather than absolute distance. Between these two variants, signed distance appears to have the advantage with NND, but we will see in the next section that the situation is different for real-life datasets.

5 Experiments on real-life datasets

We will now evaluate our proposal on a number of real-life datasets from the UCI repository [3] that can be approached as directional anomaly detection problems. The main properties of these datasets are listed in Table 2. The attributes of these datasets express machine operating conditions (*ai4i2020*), medical symptoms, comorbidities and lifestyle factors (*diabetes-risk*, *fertility*, *heart-failure*, *post-operative* and *thoraric-surgery*), tumor characteristics (*wdbc*, *wisconsin* and *wpbc*), and risk indicators provided by experts (*phishing-websites*, *qualitative-bankrupty* and *south-german-credit*). *ai4i2020* and *post-operative* contain a mix-

Table 2. Real-life datasets used in the experiment. m: number of attributes.

Dataset	Source	Records		Anomalies		m
ai4i2020	[15]	10 000	Simulated machine operation records	339	Five different failure modes (not distinguished by the label)	6
diabetes-risk	[8]	520	Patients with diabetes symptoms	320	Patients actually diagnosed with diabetes	14
fertility	[5]	100	Sperm samples	12	‘Altered’ samples	8
heart-failure	[1]	299	Heart patients	96	Patients who died in the follow-up period	11
phishing-websites	[24]	11 055	Websites	4898	Phishing websites	30
post-operative	[2]	87	Patients who have undergone surgery	63	Patients who had to stay in hospital	8
qualitative-bankruptcy	[9]	250	Companies	107	Companies that went bankrupt	6
south-german-credit	[6]	1000	Credits provided by a bank	300	Bad credits	20
thoraric-surgery	[27]	470	Primary lung cancer patients who underwent major lung resections	70	Patients who died within one year	16
wdbc	[21]	569	Images of aspirated breast tumor samples	212	Malignant tumors	30
wisconsin	[25]	683	Fine-needle aspirates of breast tumors	239	Malignant tumors	9
wpbc	[20]	138	Patients who have undergone surgery for an invasive malignant breast cancer tumor with no evidence of distant metastases	28	Patients who experienced recurrence within two years	32

Table 3. Mean cross-validation AUROC for the absolute, ramp and signed distance variants. **Bold**: highest value (before rounding) for NND and ALP, respectively.

Dataset	NND			ALP	
	Absolute	Ramp	Signed	Absolute	Ramp
ai4i2020	0.823	**0.922**	0.724	0.877	**0.924**
diabetes-risk	**0.971**	0.923	0.716	0.895	**0.926**
fertility	0.602	**0.653**	0.540	0.581	**0.636**
heart-failure	0.715	**0.769**	0.735	0.734	**0.766**
phishing-websites	0.901	**0.927**	0.804	0.927	**0.936**
post-operative	0.476	0.504	**0.557**	0.459	**0.484**
qualitative-bankruptcy	**1.000**	1.000	0.998	**1.000**	**1.000**
south-german-credit	0.648	**0.718**	0.683	0.648	**0.714**
thoraric-surgery	0.597	**0.624**	0.583	**0.634**	0.621
wdbc	0.950	**0.976**	0.969	0.957	**0.981**
wisconsin	**0.995**	0.994	0.995	0.872	**0.995**
wpbc	0.570	0.625	**0.633**	0.537	**0.654**

ture of directional and adirectional attributes, the other datasets only have directional attributes.

As in the previous section, we use AUROC to evaluate anomaly detection performance. We perform 5-fold cross-validation on the normal records from each dataset, creating at each iteration a test set by combining one fifth of the normal records with all anomalous records. For NND, we use linear weights and fix $k = 8$. We rescale all attributes by subtracting the midhinge and dividing by the semi-interquartile range of the normal training values, such that the interquartile range becomes $[-1, 1]$ in the training data.

Table 3 lists the mean cross-validation AUROC obtained by the various variants for each dataset. The first thing to note is that for NND, unlike what we saw for synthetic data in the previous section, ramp distance generally performs (much) better than signed distance. The only datasets for which signed distance performs better are *post-operative* and *wpbc*, for which the AUROC scores are very low anyway. When we perform a one-side Wilcoxon signed-rank test, we obtain $p = 0.021$. We can conclude from this that real-life anomaly detection problems are different from our synthetic datasets in one important aspect: an unexpectedly low score on one risk factor does not compensate for high scores on other risk factors.

However, these experiments do confirm the usefulness of directional anomaly detection, because ramp distance performs about as well or better than absolute distance. For NND, the signed-rank test gives $p = 0.011$, or $p = 0.023$ if we apply Holm-Bonferroni correction [7] for multiple testing, while for ALP it gives $p = 0.0029$.

There is one notable exception: *diabetes-risk* for NND. A potential explanation for this is the fact that the attributes of this dataset encode diabetes symptoms, and that moreover, the normal records do not represent average healthy people, but non-diabetes patients who nevertheless display symptoms of diabetes. Consequently, this may undermine the directionality of this dataset — diabetes patients may not necessarily have more symptoms than this particular group of other patients, but rather *different combinations* of symptoms, which would make absolute distance a better fit than ramp distance.

Indeed, when we look at the mean attribute values of normal and anomalous records in this dataset, we find that for five of the fourteen attributes, the anomalous records (diabetes patients) have lower or only slightly higher mean values than the normal records. If we rerun the experiment while treating these five attributes as non-directional, the AUROC obtained by ramp distance matches that of absolute distance.

6 Conclusion

In this paper, we have introduced directional anomaly detection, a new problem setting wherein only high (or only low) values of certain attributes are indicative of anomality, and which can therefore be viewed as risk factors. We have proposed two different ways of adapting distance-based anomaly detection algorithms to make use of this knowledge: ramp distance and signed distance. The difference between these two variants corresponds to a choice in how a practitioner wants to interpret an unexpectedly low value on a given risk factor. With ramp distance, such low values are simply discounted, whereas with signed distance, they contribute negative evidence against anomality. Both variants contrast with absolute distance, the non-directional baseline wherein such low values contribute positive evidence for anomality.

In an experiment with synthetically generated Gaussian and Bernoulli data, we found that signed distance is able to achieve slightly higher anomaly detection than ramp distance. However, in a subsequent experiment with real-life datasets, ramp distance performed substantially better. We conclude from this that in typical real-life use cases, unexpectedly low values on risk factors should not compensate for high values on other risk factors.

We also found that ramp distance generally performed as good or better than absolute distance. Therefore, we recommend the use of ramp distance when a practitioner knows that some of the attributes in their dataset are really risk factors.

Finally, we will end by briefly considering what it means if, for a given dataset, ramp distance should nevertheless result in worse anomaly detection performance than absolute distance. One possibility is that the directional hypothesis is false for one or more of the attributes, that there is a good reason after all why (some) anomalies should have lower values than normal. This is in fact what we saw with one of the real-life datasets that we evaluated, the only dataset where absolute distance clearly performed better than ramp distance for NND.

The other possibility is that there is no clear explanation, no plausible causal relation linking a lower attribute value to a higher risk of abnormality. In that case using ramp distance may still be preferable, because it better matches the domain knowledge of the practitioner, resulting in a more interpretable prediction model.

Acknowledgements

This publication is part of the project Digital Twin with project number P18-03 of the research programme TTW Perspective, which is (partly) financed by the Dutch Research Council (NWO).

References

1. Ahmad, T., Munir, A., Bhatti, S.H., Aftab, M., Raza, M.A.: Survival analysis of heart failure patients: A case study. PloS one **12**(7), e0181001 (2017)
2. Budihardjo, A., Grzymala-Busse, J., Woolery, L.: Program LERS_LB 2.5 as a tool for knowledge acquisition in nursing. In: IEA/AEI-91: Proceedings of the Fourth International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. pp. 735–740. University of Tennessee, Space Institute (1991)
3. Dua, D., Graff, C.: UCI machine learning repository (2019), `https://web.archive.org/web/20230527180226/https://archive.ics.uci.edu/ml/datasets.php`
4. Fukushima, K.: Visual feature extraction by a multilayered network of analog threshold elements. IEEE Transactions on Systems Science and Cybernetics **5**(4), 322–333 (1969)
5. Gil, D., Girela, J.L., De Juan, J., Gomez-Torres, M.J., Johnsson, M.: Predicting seminal quality with artificial intelligence methods. Expert Systems with Applications **39**(16), 12564–12573 (2012)
6. Grömping, U.: South German credit data: Correcting a widely used data set. Report 04/2019, Beuth Hochschule für Technik Berlin, Reports in Mathematics, Physics and Chemistry (2019)
7. Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics **6**(2), 65–70 (1979)
8. Islam, M.M.F., Ferdousi, R., Rahman, S., Bushra, H.Y.: Likelihood prediction of diabetes at early stage using data mining techniques. In: ISCMM 2019: Proceedings of the First International Symposium on Computer Vision and Machine Intelligence in Medical Image Analysis. pp. 113–125. Springer (2019)
9. Kim, M.J., Han, I.: The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. Expert Systems with Applications **25**(4), 637–646 (2003)
10. Knorr, E.M., Ng, R.T.: A unified notion of outliers: Properties and computation. In: KDD-97: Proceedings of the Third International Conference on Knowledge Discovery and Data Mining. pp. 219–222. AAAI (1997)
11. Lenz, O.U.: Fuzzy rough nearest neighbour classification on real-life datasets, chap. 9. Fuzzy rough one-class ensembles. Doctoral thesis, Universiteit Gent (2023)
12. Lenz, O.U., Peralta, D., Cornelis, C.: Average Localised Proximity: A new data descriptor with good default one-class classification performance. Pattern Recognition **118**, 107991 (2021)
13. Lenz, O.U., Peralta, D., Cornelis, C.: Optimised one-class classification performance. Machine Learning **111**(8), 2863–2883 (2022)

14. Li, Z., van Leeuwen, M.: Explainable contextual anomaly detection using quantile regression forests. Data Mining and Knowledge Discovery **37**(6), 2517–2563 (2023)
15. Matzka, S.: Explainable artificial intelligence for predictive maintenance applications. In: AI4I 2020: Proceedings of the Third International Conference on Artificial Intelligence for Industries. pp. 69–74. IEEE (2020)
16. P, D., Sam Abraham, S.: Fair outlier detection. In: WISE 2020: Proceedings of the 21st International Conference on Web Information Systems Engineering. pp. 447–462. Springer (2020)
17. de Ridder, D., Tax, D.M.J., Duin, R.P.W.: An experimental comparison of one-class classification methods. In: ASCI'98: Proceedings of the Fourth Annual Conference of the Advanced School for Computing and Imaging. pp. 213–218. ASCI (1998)
18. Shekhar, S., Shah, N., Akoglu, L.: FairOD: Fairness-aware outlier detection. In: AIES '21: Proceedings of the Fourth AAAI/ACM Conference on AI, Ethics, and Society. pp. 210–220 (2021)
19. Song, X., Wu, M., Jermaine, C., Ranka, S.: Conditional anomaly detection. IEEE Transactions on Knowledge and Data Engineering **19**(5), 631–645 (2007)
20. Street, W.N., Mangasarian, O.L., Wolberg, W.H.: Individual and collective prognostic prediction. Technical report 96-01, University of Wisconsin – Madison, Department of Computer Sciences, Mathematical Programming Group (1996)
21. Street, W.N., Wolberg, W.H., Mangasarian, O.L.: Nuclear feature extraction for breast tumor diagnosis. Technical report 1131, University of Wisconsin – Madison, Department of Computer Sciences (1992)
22. Tax, D.M.J.: One-class classification: Concept learning in the absence of counter-examples. Doctoral thesis, Technische Universiteit Delft (2001)
23. Tax, D.M.J., Duin, R.P.W.: Outlier detection using classifier instability. In: SSPR/SPR 1998: Proceedings of the Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition. pp. 593–601. Springer (1998)
24. Thabtah, F., Mohammad, R.M., McCluskey, L.: A dynamic self-structuring neural network model to combat phishing. In: IJCNN 2016: Proceedings of the International Joint Conference on Neural Networks. pp. 4221–4226. IEEE (2016)
25. Wolberg, W.H., Mangasarian, O.L.: Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the National Academy of Sciences of the United States of America **87**(23), 9193–9196 (1990)
26. Zhang, H., Davidson, I.: Towards fair deep anomaly detection. In: FAccT '21: Proceedings of the Fourth ACM conference on Fairness, Accountability, and Transparency. pp. 138–148 (2021)
27. Zięba, M., Tomczak, J.M., Lubicz, M., Świątek, J.: Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. Applied Soft Computing **14**, 99–108 (2014)