Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

A Probabilistic Approach to Aggregating Anomalies for Unsupervised Anomaly Detection with Industrial Applications

Tomas Olsson and **Anders Holst**
SICS Swedish ICT
Isafjordsgatan 22, Box 1263
SE-164 29 Kista, Sweden
tol@sics.se, aho@sics.se

Abstract

This paper presents a novel, unsupervised approach to detecting anomalies at the collective level. The method probabilistically aggregates the contribution of the individual anomalies in order to detect significantly anomalous groups of cases. The approach is unsupervised in that as only input, it uses a list of cases ranked according to its individual anomaly score. Thus, any anomaly detection algorithm can be used for scoring individual anomalies, both supervised and unsupervised approaches. The applicability of the proposed approach is shown by applying it to an artificial data set and to two industrial data sets – detecting anomalously moving cranes (model-based detection) and anomalous fuel consumption (neighbour-based detection).

1 Introduction

Anomaly detection is a research field that investigates various ways of identifying cases that deviate substantially from what is considered normal. Since anomaly detection does not require that all possibly anomalous classes are known in advance to be useful, it has become a quite popular approach. Thus, it has been applied in various domains over the years such as disease outbreak detection (Shmueli and Burkom 2010), intrusion detection (Garcia-Teodoro et al. 2009; Dey 2009), maritime surveillance (Holst et al. 2012a), fraud detection (Bolton and Hand 2002), and fault detection (Zaher et al. 2009; Holst et al. 2012b; Olsson et al. 2014).

The approach taken in this paper is to consider the situation when we are not interested in detecting individual cases as anomalies but in assessing the anomaly of groups of cases that can contain both normal and anomalous cases. The assumption is that cases are related to each other, for instance, by being geographically close or by being generated from the same machine, so that they can naturally be divided into groups. Then, by assessing the anomaly of groups of related cases, we assess the anomalousness as a collective.

Especially, we are interested in aggregating anomaly scores from non-statistical methods that do not in themselves provide a simple way of assessing the significance of an anomaly score or a set of anomaly scores. Many applications do not produce measurements that are easily modelled using a statistical probability distribution. Particularly for anomaly detection, the tail of the data, which contains very few cases, is often most relevant. Therefore, such data will not be fitted that easily to a global statistical model. Thus, this work does not model the individual cases using a probability distribution, but instead systematically scan through the empirical distribution to identify anomalies. To show the applicability of our approach to non-statistical approaches, we will apply the proposed approach to anomaly scores from k-nearest neighbour-based and model-based methods.

Collective or group anomalies have been investigated in some recent work, using statistical models, such as Bayesian networks (Das, Schneider, and Neill 2009), Gaussian mixture models (Vatanen et al. 2012), and an extension of the Latent Dirichlet Allocation model (Xiong, Póczos, and Schneider 2011), and ensembles of decision trees (Liu, Ting, and Zhou 2010). In contrast to our approach, which is unsupervised, most of theses are semi-supervised approaches that assumes the availability of known normal cases and in addition, do not consider aggregating the output from other anomaly detection methods. The only work, to our knowledge, on aggregating anomalies from arbitrary methods is presented in (Das, Schneider, and Neill 2008). In contrast, it assumes the existence of a training set containing mostly normal data with only a small fraction of anomalous data. It also uses a rule-based approach to find groups of cases that have a common significantly different pattern, while we assume that groups are naturally formed. In addition, the individual cases are classified as anomalous based on an individual (local) anomaly detector using a threshold learned from automatically selecting a specified false discovery rate, while our approach do not need to learn such a threshold.

In summary: the main contribution of this paper is a novel, unsupervised method for detecting anomalies at the collective level. The method probabilistically aggregates the contribution of the individual anomalies to detecting anomalous groups of cases. The only input to the method is the anomaly ranking of the individual cases. Then, the method is able to: (1) aggregate rankings from any anomaly detection method, (2) produce an aggregated anomaly score for each group, and (3) assess the significance of the estimated anomaly score. Thus, it is possible to filter out significantly anomalous groups from less anomalous groups, even if the used anomaly detection method does not provide means to

test the significance of the individual anomalies.

In the following, we first show the applicability of the approach for an artificial dataset. Then, we apply the approach in two industrial application. The first application is for detecting anomalous cranes in a container terminal by analysing the time it takes to move a crane. The second application is about detecting anomalous road segments by analysing the anomalous fuel consumption of vehicles travelling both in and outside cities.

The paper is organised as follows. The first section presents some background on collective and unsupervised anomaly detection. Next section describes the approached proposed in this paper. The section after that contains the evaluation of the proposed approach to one artificial data set and two industrial data sets. The last section ends the paper with some conclusions and future work.

2 Background

Collective Anomalies

The characteristic of detected anomalies can be classified into three types (Chandola, Banerjee, and Kumar 2009): point anomalies, context anomalies and collective anomalies. Point anomalies are cases that are deemed anomalous with respect to the whole data set. Context anomalies are anomalous when considered in a specific context, but not necessary in another context. For instance, during the winter in Sweden, it is not unusual that the temperature drops well below zero degree Celsius but not in the summer, which would be a anomaly in context of the current season. Collective anomalies are cases that are only deemed to be anomalous when considered together but not necessary as individuals. This is typically in case of systems that slowly develop faults due to material wear. Each anomalous measurement might not indicate a fault but only when several of them appear together. The latter type of anomaly is the focus of this paper, while we use both point anomalies and context anomalies as input to the system. Most of the current research has focused on point and context anomalies.

Unsupervised Anomaly Detection

Unsupervised anomaly detection assumes that no labeled data is available, but where the normal part of the data is much larger than the anomalous part (Chandola, Banerjee, and Kumar 2009). In contrast, a supervised approach would need labeled cases of both normal and abnormal instances. A semi-supervised approach typically uses the fact that it is easier to collect normal cases than anomalous cases, and thus, builds a model only using the normally labeled cases. The approach presented in this paper is a unsupervised approach but can also be applied in a supervised setting.

3 Aggregation of Anomalies

In this section, we first present and discuss a simple and intuitive but naive anomaly aggregation approach and list its shortcomings. Thereafter, we will propose a probabilistic approach that lacks the shortcomings of the naive approach. Last, we describe how to use the proposed method to identify and rank anomalous groups.

A Naive Aggregation Scheme

A straightforward approach to aggregating anomaly scores would be to only look at the top most, say 10 % (or what ever seems reasonable), of all cases with the highest point anomaly scores and use the fraction of those in each group as an aggregated anomaly score. This gives us an aggregated anomaly score from 0 to 1.0 that also can be used for ranking, but we have no idea of when to consider a group anomalous and we have no way of assessing the significance of the aggregated anomaly score. For instance, if there are 10 cases in one group and 3 belong to the to the top 10 %, and another group with 100 cases of which 30 belong to the top 10 %. Then, both have a fraction of 0.3, but presumably the latter is more significant. In the next section, we will describe a probabilistic approach improving on this simple approach for which it is possible to also assess the significance of the aggregated anomaly score.

Probabilistic Aggregation

The assumption behind the naive aggregation scheme above is an intuitively appealing idea that a more anomalous group should generate more anomalous cases. However, instead of using the fraction as a measure of anomaly, we propose using a probabilistic definition of an anomaly. In this case, a lower probability means a more anomalous group (high anomaly score), and a higher probability means a less anomalous group (low anomaly score). Thus, this approach gives a clearly defined aggregated anomaly score that can thereby also be used for assessing the significance.

Definition 1 *Let p be a predefined fraction of the top most anomalous cases among all cases, for instance, $p = 0.1$ (that is, 10%). Then, we denote all cases belonging to this fraction as "top p-outliers".*

Thus, p is the probability that a randomly drawn case is a top p-outlier. Then, the probability, by chance alone, that y number of cases belong to that fraction given n number of cases is binomially distributed with parameters n and p:

$$P(y; p, n) = \binom{n}{y} p^y (1-p)^{n-y} \tag{1}$$

Then, for assessing the normality of a group with n cases and y outliers, we compute the probability of getting y or more number of outliers:

$$\bar{A}(y; p, n) = 1 - \sum_{k=0}^{y-1} P(k) \tag{2}$$

Intuitively, this means that if the probability to get a larger value is very small, then this is a very unusual and highly unlikely number of anomalous cases. Thus, this also means that it is an unlikely group of cases that can be considered anomalous.

Identifying Anomalous Groups

The proposed method to identifying anomalous groups is shown in Algorithm 1. The approach is to scan through the

data letting the value of p vary from 0 to 1 while counting the number of times each group is identified as anomalous. Thus, a group can be identified as anomalous at several different values of p. We use a very restrictive threshold for accepting a group as anomalous: a probability of less than 10^{-6}, that is, 1 per 1.000.000, which ensures a very low false discovery rate. Finally, the groups are ranked according to the number of times they have been identified as anomalous. Concerning the values for p, in theory, it would be possible to scan through each individual anomaly score one at a time, but for large data sets that is not practical. Thus, we have selected a subset of all possible values for p such that the tail is scanned in smaller step while larger steps are taken for larger values of p. This is because, as noted in the introduction, the tail is often more interesting when doing anomaly detection.

Algorithm 1 Compute a ranked list of anomalous groups

INPUT The set of all groups of cases G
OUTPUT A ranked list of anomalous groups of cases R
$P \leftarrow \{0.0001, 0.001, 0.002, \ldots, 0.01, 0.02, \ldots, 0.1, 0.2, \ldots, 0.9, 0.99\}$ (the top fractions)
$R \leftarrow$ an empty list
for $g \in G$ **do**
  $counter \leftarrow 0$
  for $p \in P$ **do**
    $n \leftarrow$ number of cases in g
    $y \leftarrow$ number of top p-outliers in g (Definition 1)
    if $\bar{A}(y; p, n) < 10^{-6}$ **then**
      $counter \leftarrow counter + 1$
  if $counter > 0$ **then**
    append tuple $(g, counter)$ to R
Return R sorted on the second element of each pair

4 Evaluation

In this section, we evaluate the proposed group anomaly detection algorithm using an artificial data set generated from three normal distributions. Then, we apply the approach to two real world applications. First, we apply it to detecting anomalous cranes from analysing the predicted move time compared to the true value. Last, we identify road segments with anomalous fuel consumption. The experiments were implemented using the scikit-learn and the minepy libraries (Pedregosa et al. 2011; Albanese et al. 2013).

Artificial Data Set

We generate artificial cases as follows. Let us pretend that we have 1000 similar machines and that we have an algorithm that can predict the performance of the machines given their input. Each prediction is a case. Assuming that the prediction error, that is, the predicted value minus the real value, is normally distributed – $N(0, 1)$ – for the non-anomalous cases. In addition, we assume that there is a normally distributed noise: $N(4, 12)$. Further, we assume that $m = 100$ of the machines are anomalous with normally distributed anomalous cases: $N(6 + r, 4)$, where $r = 2 * i/m$ for each $i \in \{0, \ldots, m - 1\}$ is a value used for separating between different anomalous machines. The number of cases per machine is assumed to be Poisson distributed with $\lambda = 1000$. The fraction of noise in each group is 1% while the fraction of anomalous cases in an anomalous group is 5%. As point anomaly score for individual cases we use the absolute prediction error. Figure 1 shows the histogram of the point anomaly scores. As can be seen, there is a large number of anomalous cases up to three standard deviations away from zero, and just a small number of anomalous cases.

Figure 2 shows the result from applying the proposed method. The top curve shows the number machines detected as outliers with non-anomalous cases sampled from $N(0, 1)$ over different values of $p \in (0, 1)$. The largest number of detected outliers is 100 at $p = 0.02$. The other curves show what happens when the standard deviation of the non-anomalous cases increases from 1 to 6, which leads to a decrease in the number of detected outliers as the difference between anomalous and non-anomalous machines gets smaller.

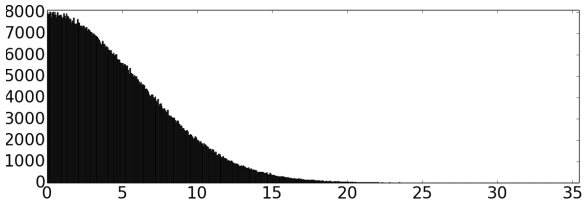

Figure 1: Histogram of anomaly scores of the artificial data.

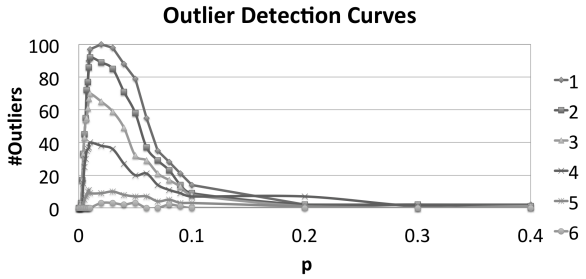

Figure 2: Outlier detection curves for different value of p and varying non-anomalous standard deviation. Curves are close to zero for $p > 0.4$.

Table 1 compares the proposed approach with a baseline algorithm in terms of the precision, false discovery rate (FDR) and recall for the different values of the standard deviation. As baseline approach, we fitted the mean prediction errors of all machines to a normal distribution using a maximum likelihood estimation in two steps. First, we removed outliers with mean prediction error three standard deviations or more away from the distribution mean. Last, we fitted a final normal distribution to the remaining mean prediction errors. Then, all machines with mean prediction errors three final standard deviations or more away from the final distribution mean are considered anomalous. Notable in Table 1 is that for the proposed algorithm, the FDR is always 0% and

the precision is always 100%, while the recall is decreasing as the distinction between anomalous and non-anomalous cases gets blurred. This is due to the very restrictive threshold we use for accepting a machine as an outlier: 10^{-6}. In contrast, the baseline shows worse recall up to a standard deviation of 5, but with a cost of lower precision and higher FDR after that. We could improve on the number of detected outlier machines for the proposed approach by increasing the threshold but that would also lead to an increase in the number of false positives, which is not desired. Assuming that anomalies are more prevalent in the tail of the empirical distribution, an increase of the threshold would lead to an even larger increase in the false discovery rate.

In order to test the robustness of the proposed approach, we also let the anomalous cases be normally distributed with zero mean and with varying standard deviation: $N(0, 4+r)$, where r is defined as above. The result can be seen in the last column of Table 1 where for the proposed approach, the recall is still high 88% with 100% precision and 0% FDR, while the baseline has a very poor performance. Thus, the proposed algorithm is more robust to variations of the anomalies than the baseline that is best suited to detect anomalies with respect to the mean value only.

Table 1: Performance (in %) for different values of the non-anomalous standard deviation: Proposed Algorithm (above) and baseline (below). The last column (*) shows the performance of the robustness test.

Std.	1	2	3	4	5	6	1*
Prec.	100	100	100	100	100	100	100
FDR	0	0	0	0	0	0	0
Recall	100	98	73	53	16	5	88
Std.	1	2	3	4	5	6	1*
Prec.	100	100	100	100	96	95	50
FDR	0	0	0	0	4	5	50
Recall	83	67	51	38	26	19	3

Crane Anomaly Detection

This section applies the proposed approach to detecting anomalous cranes in a container terminal. Each crane is analysed in terms of the time it takes to move the containers. As anomaly score, we use the prediction error (difference) between the true and the predicted time duration for each move. Then, the prediction error provides a ranked list of moves that are thereafter aggregated in order to collectively assess whether a crane is significantly anomalous.

A container terminal consists of one or more container stacks, where two cranes are working in parallel to move containers in and out of the stack. One of the cranes is moving containers from a ship into the container stack and one crane is moving containers from the stack to waiting trucks or vice versa. The expected or predicted duration of a move is computed using a physical model that takes into account moves in all directions.

Figure 3 shows the plot of the move anomaly scores as a histogram. There seems to be two types of anomalies: one type with prediction differences larger than 75 seconds and another with differences larger than 200 seconds. In order to

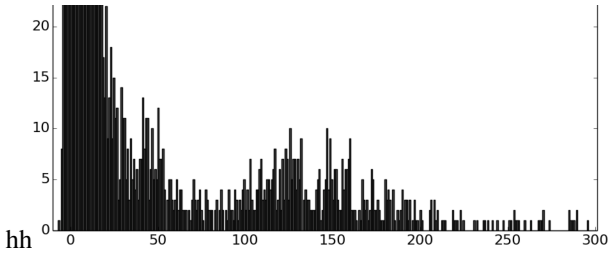

Figure 3: The zoomed tail of the histogram of anomaly scores of container move durations. The y-axis denotes number of moves and it is cut at about 20.

visualise the difference between cranes, we plotted the histograms for the individual cranes as shown in Fig. 4 and Fig. 5. Figure 4 shows the histogram for prediction error for the

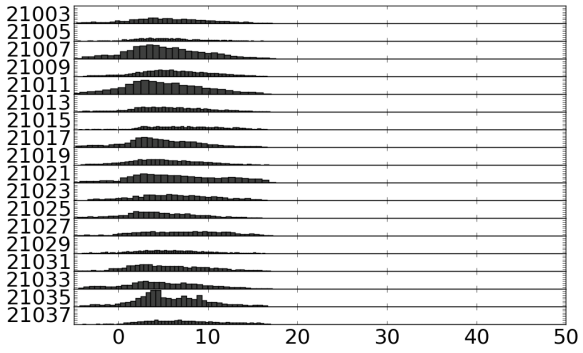

Figure 4: Histogram of anomaly scores per crane. The x-axis is cut at 50 sec and each crane has different y-axis scale.

main part of the moves with errors less than 50 seconds. As can be seen, they are quite similar, except crane 21035 that looks quite different from the rest for some unknown reason and crane 21021 that has a heavier tail at the right. However, since we are also interested in the most anomalous moves, we take a look at Fig. 5 where the histograms of the anomaly scores from 20 to 1000 seconds are shown. Now, we can see that there are some variations between the cranes, but it is not easy to spot the real differences. The cranes 21003, 21011, 21017, 21021, and 21035 seem to be different from the other in the 75 to 200 seconds interval, which corresponds to the first type of anomaly defined above. However, we cannot assess the significance from just looking at the plots.

In Fig. 6, we can see the outlier detection curve for the cranes. In total 8 cranes were identified as anomalous, where at most 6 cranes were identified at any single p. Figure 7 shows the number of times each anomalous crane has been identified as anomalous. In contrast to the previous visual identification of anomalous cranes, it is now clear that 21035 indeed is anomalous but not the most anomalous, but the least. In addition, of the five identified as anomalous in the tail, only 21021, 21017, 21035 are anomalous while 21003 and 21011 are not anomalous.

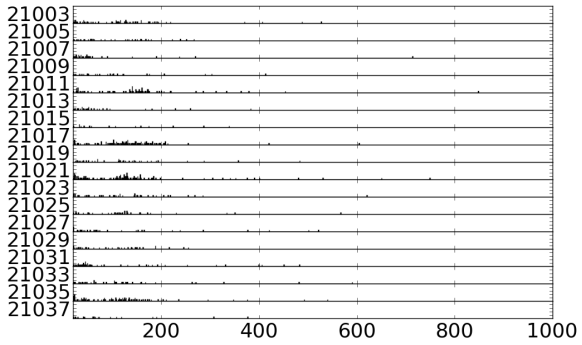

Figure 5: Histogram of anomalies scores for each crane. The x-axis shows from 20 up to 1000 sec. The y-axis has a uniform scale from 0 to 10.

If we only look at the 50 seconds threshold (corresponding to $p \approx 0.01$), then only crane 21021 and 21017 qualify as truly anomalous cranes. If we do the same for anomalies with duration differences larger than 200 seconds (corresponding to $p \approx 0.002$) we do not find any cranes that deviate significantly from the others, although crane 21021 has the lowest probability.

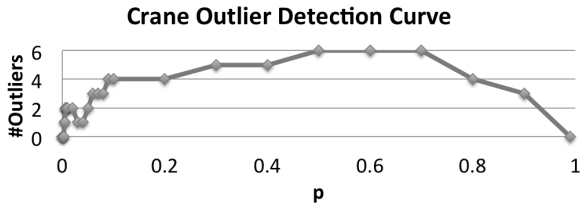

Figure 6: The outlier detection curve for cranes. The y-axis is number of identified anomalous cranes. The x-axis is the different values of p.

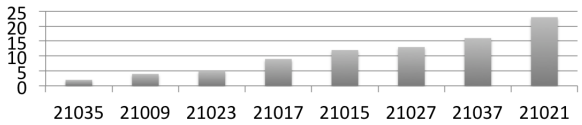

Figure 7: The sorted number of times each crane was identified as an outlier.

Road Segment Anomaly Detection

This section applies the proposed approach to detecting road segments with anomalous fuel consumption. The data consists of measurement points from five vehicles travelling between different destinations in northern Germany. As the data did not appear to adhere to any known probability distributions, it was decided to use a k-nearest neighbour approach to predict the instantaneous fuel consumption. In addition, since, we are foremost interested in detecting anomalous fuel consumption, the absolute prediction error is used as basic individual anomaly score. The travelled geographical area was then segmented into 1265 segments of 2×2 km^2

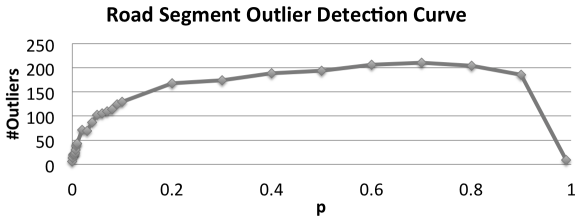

Figure 8: Outlier detection curve for anomalous road segments.

corresponding to all measurement points with GPS coordinates in that area. Then, anomaly scores were aggregated for each geographical segment. We only look at road segments with more than 100 measurement points, which resulted in 802 road segments.

A k-nearest neighbour algorithm (kNN) was trained to predict the fuel consumption given a subset of the available attributes. In order to ensure independence between measurement points, the kNN was trained to predict the fuel consumption of a vehicle using only the data points of the other vehicles as training data. As similarity metric, we have used the weighted Euclidean distance, where the weights were computed using the maximum information coefficient (MIC) between the instantaneous fuel consumption and the other attributes (Murphy 2012). The number of neighbours was selected using 5-fold cross validation. In addition, by removing one attribute at a time, only those attributes that improved the prediction performance were kept.

Figure 8 plots the result from running the proposed algorithm with varying values of p. For each p, we get a number of anomalous road segments, that is, segments with a probability less than 1 per 1.000.000 of being normal. This resulted in a total of 243 segments being identified as anomalous for at least one p. This is a quite large percentage given that we are only looking at 802 segments in total, so a means for prioritising between outliers is needed. Of the outlier segments, 63 have been identified as anomalous for all $p \leq 0.01$, that is, for the top 1% of the most anomalous cases. Figure 9 shows the number of times each anomalous road segment has been identified as an outlier for a p. So, by ranking the outliers according to number of detections, we can prioritise between anomalous road segments.

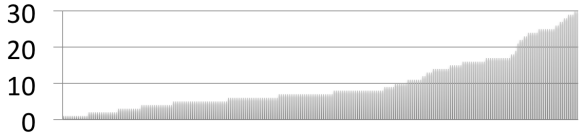

Figure 9: The sorted number of times (y-axis) that each road segment (x-axis) was identified as an outlier.

5 Conclusions and Future Work

This paper proposes a novel, unsupervised approach to probabilistically aggregating anomaly scores from non-statistical anomaly detection algorithms for groups of cases. The only

required input are: (1) a ranked list of cases ordered according to the individual anomaly scores and (2) a means to form natural groups from the cases. By scanning through the top fraction of the ranked cases, it can then *detect* groups of cases that are *significantly* anomalous at different individual rank levels. Then, groups can be *ranked* according to the number of times they are identified as anomalous when scanning through the ranking.

Using artificial data, we have shown that the proposed approach can indeed detect and identify significantly deviating groups of cases and we showed that the performance was better than a baseline algorithm that assumes normal distributed mean prediction errors. In addition, we have shown how to apply the proposed approach to two industrial applications where one used a model-based anomaly detection algorithm and the other used an instance-based (kNN) anomaly detection algorithm. In both applications, we could identify and prioritise between significantly anomalous groups of cases. Unfortunately, the ground truth of the industrial data sets was not available, so the performance could not be evaluated.

Nevertheless, comparing the curves in Fig. 2 with the curves in Fig. 6 and Fig. 8, the artificial data curve looks quite different, while the industrial data curves are quite similar. This indicate that the real data is quite much more complex than the artificial data. Thus, future work would need to investigate more realistic data where the ground truth is available. However, notice that the problem for the real data is not that too few outliers are detected but rather the opposite, especially in the second application. So, the problem of using a too restrictive threshold is not necessary a problem for real world data. Another research direction would be to investigate how the detected anomalous groups could be compared to previously identified anomalous groups using, for instance, case-based reasoning in order to support diagnosis of anomalies.

6 Acknowledgments

This research is funded by VINNOVA, grant number 2012-01277, and ABB Crane Systems, Västerås. Authors would like to thank the personal from ABB Crane Systems for their help with the information and data on the container cranes during the research.

References

Albanese, D.; Filosi, M.; Visintainer, R.; Riccadonna, S.; Jurman, G.; and Furlanello, C. 2013. minerva and minepy: a c engine for the mine suite and its r, python and matlab wrappers. *Bioinformatics* 29(3):407–408.

Bolton, R. J., and Hand, D. J. 2002. Statistical fraud detection: A review. *Statistical Science* 17(3):235–255.

Chandola, V.; Banerjee, A.; and Kumar, V. 2009. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)* 41(3):15.

Das, K.; Schneider, J.; and Neill, D. B. 2008. Anomaly pattern detection in categorical datasets. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 169–176. ACM.

Das, K.; Schneider, J.; and Neill, D. B. 2009. Detecting anomalous groups in categorical datasets. Technical Report CMU-ML-09-104, Carnegie Mellon University, School of Computer Science, Machine Learning Department.

Dey, C. 2009. Reducing ids false positives using incremental stream clustering (isc) algorithm. Master's thesis, Department of Computer and Systems Sciences, Royal Institute of Technology, Sweden.

Garcia-Teodoro, P.; Diaz-Verdejo, J.; Maciá-Fernández, G.; and Vázquez, E. 2009. Anomaly-based network intrusion detection: Techniques, systems and challenges. *computers & security* 28(1):18–28.

Holst, A.; Bjurling, B.; Ekman, J.; Rudström, A.; Wallenius, K.; Björkman, M.; Fooladvandi, F.; Laxhammar, R.; and Trönninger, J. 2012a. A joint statistical and symbolic anomaly detection system: Increasing performance in maritime surveillance. In *15th International Conference on Information Fusion (FUSION)*, 1919–1926. IEEE.

Holst, A.; Bohlin, M.; Ekman, J.; Sellin, O.; Lindström, B.; and Larsen, S. 2012b. Statistical anomaly detection for train fleets. *AI Magazine* 34(1):33.

Liu, F.; Ting, K.; and Zhou, Z.-H. 2010. On detecting clustered anomalies using sciforest. In *Machine Learning and Knowledge Discovery in Databases*, volume 6322 of *Lecture Notes in Computer Science*. Springer. 274–290.

Murphy, K. P. 2012. *Machine learning: a probabilistic perspective*. MIT Press.

Olsson, T.; Källström, E.; Gillblad, D.; Funk, P.; Lindström, J.; Håkansson, L.; Lundin, J.; Svensson, M.; and Larsson, J. 2014. Fault diagnosis of heavy duty machines: Automatic transmission clutches. In *Workshop on Synergies between CBR and Data Mining at 22nd International Conference on Case-Based Reasoning*.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Shmueli, G., and Burkom, H. 2010. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics* 52(1):39–51.

Vatanen, T.; Kuusela, M.; Malmi, E.; Raiko, T.; Aaltonen, T.; and Nagai, Y. 2012. Semi-supervised detection of collective anomalies with an application in high energy particle physics. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Xiong, L.; Póczos, B.; and Schneider, J. G. 2011. Group anomaly detection using flexible genre models. In *Advances in Neural Information Processing Systems*, 1071–1079.

Zaher, A.; McArthur, S.; Infield, D.; and Patel, Y. 2009. Online wind turbine fault detection through automated scada data analysis. *Wind Energy* 12(6):574–593.