**ISBN** 978-981-14-1684-2
**Proceedings of 2019 the 9th International Workshop on Computer Science and Engineering**
(WCSE 2019 SUMMER)
*Hong Kong, 15-17 June, 2019, pp. 124-129*
doi: 10.18178/wcse.2019.06.018

# Superimposed Rule-Based Classification Algorithm (SRBCA) for One-Class Multivariate Conditional Anomaly Detection

Ivy Kim D. Machica [1+], Bobby D. Gerardo [1,2] and Ruji P. Medina [1]

[1] Technological Institute of the Philippines, Philippines
[2] West Visayas State University, Philippines

**Abstract.** Traditional anomaly detection causes a problem of detecting too numerous false positives in many problem domains. In this work, a Superimpose Rule-Based Classification algorithm (SRBCA) is proposed for conditional anomaly detection. The algorithm is an enhancement of the traditional OneR algorithm. The traditional OneR can generate a set of rules from its attributes with multiple classes, compute the error rate and apply the rule to the attribute with the smallest error. However, OneR has a disadvantage for one-class datasets which contains values belonging to the normal class. The enhanced algorithm, SRBCA, does not embody very complex rules similar to its predecessor. Furthermore, SRBCA includes the generation and application of rules from the one-class dataset in an n-dimensional space using classification. Holdout method was used to evaluate the performance of the classifiers' accuracy which involved training multiple subsets' behavioral and indicator attributes, superimposing rules and testing by using balanced and unbalanced class data to detect and label conditional anomaly data points. This paper shows the comparison between SRBCA, One-Class Support Vector Machine (OCSVM) and other anomaly detection classification algorithms for conditional anomaly detection. It proves that the new method can handle one-class multivariate for conditional anomaly detection with better accuracy.

**Keywords:** one-class classification, conditional anomaly detection, classification algorithm.

## 1. Introduction

Traditional anomaly detection causes a problem of detecting too numerous false positives and false negatives in many problem domains. Unfortunately, the current method of one-class modeling has lower accuracy in detecting conditional anomalous instances. If anomalies are not accurately identified, critical knowledge can get lost resulting to attacks, fraud, damages and loss of life in case the case of aircraft safety or medical prescription and many more.

In data mining, classification is a method of generating a model or classifier from a training subset and then, classify the remaining data set using the learned model. Many real application scenarios have datasets which consist of normal data points or events without anomalies. Learning from normal class is the case when anomalous data is difficult to obtain. The training dataset generates a model that is learned from a normal class [1]. This kind of learning is semi-supervised or one-class classification. Therefore, anomalous data points are identified if they deviate from the model.

A straightforward and simple algorithm that can minimize the false positive rates in conditional anomaly detection is proposed in this study. The proposed algorithm is called Superimposed Rule-Based Classification Algorithm (SRBCA). This is a rule-based classification algorithm that is based on a simple but powerful classification algorithm called OneR [2]. However, unlike OneR, the SRBCA can handle training using one-class dataset. The SRBCA rules are learned from the training of multivariate datasets. The rules generated during training with domain knowledge from experts, which is used as the threshold, are

[+] Corresponding author. Tel.: +63 9171580164;
*Email address:* ikmachica@usep.edu.ph

superimposed to form a general rule. The general rule will serve as a model and evaluated by applying to the test set.

The rest of the paper is organized as follows. Section 2 reviews the related literature and works. Section 3 describes the new conditional anomaly detection algorithm. Section 4 presents the experimental results with balanced and unbalanced datasets. Section 5 concludes the paper.

# 2. Related Studies and Work

## 2.1. Anomaly detection

Anomaly detection is a process of discovering data points or events that significantly deviate from the normal data point or events [3]. This definition specifies that an anomaly detection algorithm must learn normal data points, and it labels these points as anomalous if it deviates significantly from the rest of the data. The three (3) types of anomalies are point, contextual and collective [1]. A point anomaly occurs if an individual data instance deviates to the normal data points or events. The contextual or conditional anomaly occurs if the data instance is anomalous within a context or condition. This type of anomaly requires a domain knowledge or requires a notion of context [4]. The last type of anomaly is called a collective anomaly. It contains the collection of data points or events that deviate from another group. The applications of anomaly detection involve network intrusion, credit fraud, healthcare informatics, image processing and many more [4].

## 2.2. OneR

OneR is a simple algorithm developed by Robert Holte of the University of Ottawa [5]. Holte appeals to the researchers to use “simplicity first” methodology in machine learning. OneR builds one rule for each attribute in the training data set. The algorithm will compute the error rate of each attribute. Then, the attribute which has the smallest error rate will be used to classify the remaining dataset [6]. This algorithm will process dataset which contains labeled instances that belong to multiple classes. Also, this algorithm will arbitrarily choose an attribute if in case of a tie in the smallest error rate. However, this idea will hide interesting patterns of other attributes which can be used for analysis.

## 2.3. SVM

Support Vector Machine (SVM) is a learning machine for two-group classification [7]. One Class Support Vector Machine (OCSVM) is an improvement of the conventional two-class classification algorithm. OCSVM is a well-known algorithm for one-class classification [1]. It generates a classification model during training of dataset and uses the learned model to classify new data points. OCSVM uses a hyperplane to separate a normal data point from an anomalous data point. Moreover, OCSVM uses various types of kernel and parameters to generate the model. In the taxonomy of anomaly detection algorithms, One-class Support Vector Machine (OCSVM) is identified as a commonly used classifier-based anomaly detection [1][8]. In spite of its popularity, OCSVM suffers from the drawbacks of its predecessor. One disadvantage is its performance which is affected by the type of kernel and its parameters used to generate the model [9]. OCSVM has mercer kernels and parameters that the user must set during modeling.

## 2.4. Anomaly detection classifiers

The classification algorithms provide classifiers that identify a set of categories a new instance belongs [10] [11] based on modeling the training dataset. There are many real-world application including identification of anomalous data points or events [12] can be modeled using a classification algorithm [11]. Some of the classification algorithms are OneR [13] [10] [14], Projective Adaptive Resonance Theory (PART) [14], J48 [14], Naïve Bayes [11][14], K-Nearest Neighbor, Random Forest [14], and Logistic model tree [15].

# 3. Superimposed Rule-Based Classification Algorithm

This section demonstrates the procedure of SRBCA and its application to the climatological dataset. This anomaly detection technique assumes that the training dataset is outlier-free and then uses the algorithm to detect conditional anomaly to the test dataset.

## 3.1. Pseudocode

**3.1.1 Input**: Dataset D0 = One-class Multivariate

Step 1. Assign 60% of D0 for Historical Training Data D1 (1990-2000)

**3.1.2 Process**:

Step 2. Partition D1 according to the number of behavioral datasets.

Step 3. Generate n rules for multiple behavioral datasets from its indicator features.

Step 4. Superimpose rules and pre-program threshold to form general rule.

Step 5. Inject synthetic multiclass test data (40% of D0) to form balance class test data.

**3.1.3 Output**: Testset D2

Step 6. If instance <> rule, assign anomaly to class.

## 3.2. Dataset

The datasets consist of 324 instances for 27 years of data collection from the Philippine Atmospheric, Geophysical and Astronomical Scientific Association (PAGASA) in Region XI, Davao City [16]. The environmental attributes or features are the two (2) seasons of the Philippines: Rainy from June to November and Dry seasons from December to May [17]. The indicator attributes are the following four (4) climatological features: temperature, relative humidity, rainfall and daylight hours. Moreover, the attributes are treated independently. The training dataset is 60% of the entire dataset and 40% for the test data. The 60% is based on historical data which is one (1) decade of data collection. Also, the application of the domain knowledge is based on the key findings of PAGASA which reveals that the Philippines' temperature is increasing at an average rate of 0.1°C/decade [18]. The domain knowledge and the predicted (threshold) increase in temperature and other climatological features are used in the generation of rules. Table 1 shows the descriptive statistics and correlation of the training dataset.

TABLE 1: BASIC STATISTICS OF THE TRAINING DATASET

| ***Descriptive Statistics*** | | | | | | | | ***Correlation Matrix*** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Statistic N=132** | Min | 1st Qu. | Median | Mean | 3rd Qu. | Max | StdDev. | Temp | RH | Rainfall | Daylight |
| Temp | 26.00 | 27.50 | 27.80 | 27.84 | 28.20 | 29.70 | 0.63 | 1 | -0.49 | -0.19 | 0.62 |
| RH | 68.00 | 78.00 | 80.00 | 79.59 | 82.00 | 88.00 | 3.53 | -0.49 | 1 | 0.52 | -0.68 |
| Rainfall | 1.40 | 79.08 | 133.80 | 144.40 | 195.65 | 449.30 | 88.48 | -0.19 | 0.52 | 1 | -0.28 |
| Daylight | 6046 | 10684 | 12507 | 12368 | 14014 | 17258 | 2207.50 | 0.62 | -0.68 | -0.28 | 1 |

## 3.3. The Superimposed Rules

The SRBCA will determine the minimum and maximum value for each feature in each subset. Also, a pre-programmed threshold is considered in creating the rules. The rules and pre-program threshold are superimposed to generate a general rule for the test sets as shown in Table 2.

TABLE 2: SUPERIMPOSED RULES FROM BEHAVIORAL DATASETS

| | *Temperature* | *Relative Humidity* | *Rainfall* | *Daylight Hours (min)* | *Temperature* | *Relative Humidity* | *Rainfall* | *Daylight Hours (min)* |
|---|---|---|---|---|---|---|---|---|
| *A. Behavior 1 - Dry Season (January-May and December)* | | | | | *B. Behavior 2 - Rainy Season (June to November)* | | | |
| Minimum | 26.00 | 68.00 | 1.40 | 6046 | 27.20 | 76.00 | 31.80 | 8442 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Maximum | 29.70 | 88.00 | 337.00 | 17258 | 28.8 | 84.00 | 449.30 | 15851 |
| Threshold | 0.1°C ▲ | | | | 0.1°C ▲ | | | |

# 4. Experimental Results

The experimental results show comparisons between balanced & unbalanced datasets and SRBCA & other classification algorithms for anomaly detection. The balanced datasets contain 66 instances belonging to normal and anomaly classes. Furthermore, the unbalanced datasets include 64 normal class and 68 anomaly classes The confusion matrix as shown in Table 3 provides analysis of the accuracy of the SRBCA model.

TABLE 3: CONFUSION MATRIX

| *n* | ***True Positive*** | ***True Negative*** |
|---|---|---|
| **Predicted Positive** | <True Positive - TP> | <False Positive - FP> |
| **Predicted Negative** | <False Negative - FN> | <True Negative - TN> |

## 4.1. SRBCA

The superimposed rules will label class instances as either normal or anomaly with respect to the behavioral indicators. The confusion matrix of the model of the SRBCA, as shown in Table 4, is based on the output of the superimposed rules produced during training.

```
testpredictors ← TestData //Balanced and Unbalanced Datasets
for (i = 0; i < N; i++) {// Behavioral features
    for (j = 0; j < N; j++) {// Indicator features
        if (testpredictors$<Indicators> < minTh | testpredictors$<Indicators> > maxTh)
                print (testpredictors$Class = "Anomaly")
        else
             print (testpredictors$Class ="Normal")}}
```

TABLE IV. SRBCA CONFUSION MATRIX OF BALANCED TEST SET

| *n = 132* | ***True Positive*** | ***True Negative*** |
|---|---|---|
| **Predicted Positive** | 63 | 0 |
| **Predicted Negative** | 3 | 66 |

The classification accuracy and other measures of the SRBCA and different classification algorithms are shown in Table 5. The balanced and unbalanced datasets were tested using D1 and D2 as labels in the table. It shows that SRBCA's accuracy is 98% and an error rate of 2% which is more accurate compared to the OCSVM and other classifiers. Ninety-five percent (95%) reflects the model's ability of SRBCA to detect members of the normal class as reflected in the sensitivity and recall measurement. Furthermore, the 100% value of the specificity means that the model can identify members of the anomaly class. The Negative Predicted Value (NPV) is 96% which means that test instances with anomaly class are anomalous. The False Positive Rate (FPR) is zero which means that the frequency with which the classifier makes a mistake by classifying the normal state as anomalous. The False Discovery Rate (FDR) is zero that indicates the rate of predicting normal class that is anomalous. The False Negative Rate (FNR) is 5% which reflects the frequency with which the classifier makes a mistake by classifying the anomaly class to normal. The F1 score is 98% which represents the weighted average of the precision and recall. The F1 provides a balanced view of the precision and recall. The kappa value is 0.97 which falls in the range of 0.81-1.00 means almost perfect or perfect agreement. Cohen's Kappa is a statistical coefficient that represents the degree of accuracy and reliability of the classification [11]. Also, the OneR, PART, J48, Naïve Bayes, K-Nearest Neighbor, Random Forest, and Logistic Model Tree (LMT) classification algorithms used two (2) classes for training.

The balanced testing datasets which contain both classes were appended to the training dataset which contains all normal classes to form one training dataset. The cross-validation with ten folds was used to test the classifier.

TABLE 4: PERFORMANCE OF SRBCA AND OCSVM (BALANCED CLASS FOR TEST SET 1

| | ***SRBCA*** | | *OCVSM* | | *OneR* | | *PART* | | *J48* | | *Naïve Bayes* | | *K-Nearest Neighbor* | | *Random Forest* | | *LMT* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Measure | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 | D1 | D2 |
| Sensitivity | 0.95 | 0.98 | 0.88 | 0.87 | 0.86 | 0.92 | 0.93 | 0.96 | 0.96 | 0.98 | 0.97 | 0.98 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.95 |
| Specificity | 1.00 | 1.00 | 0.89 | 0.91 | 0.29 | 0.59 | 0.82 | 0.88 | 0.82 | 0.82 | 0.58 | 0.51 | 0.71 | 0.87 | 0.82 | 0.90 | 0.83 | 0.95 |
| Precision | 1.00 | 1.00 | 0.89 | 0.91 | 0.78 | 0.87 | 0.94 | 0.96 | 0.94 | 0.94 | 0.87 | 0.85 | 0.91 | 0.96 | 0.94 | 0.97 | 0.95 | 0.98 |
| NPV | 0.96 | 0.99 | 0.88 | 0.87 | 0.40 | 0.73 | 0.79 | 0.88 | 0.89 | 0.95 | 0.86 | 0.90 | 0.89 | 0.95 | 0.87 | 0.95 | 0.93 | 0.84 |
| FPR | 0.00 | 0.00 | 0.11 | 0.09 | 0.71 | 0.41 | 0.18 | 0.12 | 0.18 | 0.18 | 0.42 | 0.49 | 0.29 | 0.13 | 0.18 | 0.10 | 0.17 | 0.05 |
| FDR | 0.00 | 0.00 | 0.11 | 0.09 | 0.22 | 0.13 | 0.06 | 0.04 | 0.06 | 0.06 | 0.13 | 0.15 | 0.09 | 0.04 | 0.06 | 0.04 | 0.05 | 0.02 |
| FNR | 0.05 | 0.02 | 0.12 | 0.13 | 0.14 | 0.08 | 0.07 | 0.04 | 0.04 | 0.02 | 0.03 | 0.02 | 0.03 | 0.02 | 0.04 | 0.02 | 0.02 | 0.05 |
| **Accuracy** | **0.98** | **0.99** | **0.89** | **0.89** | **0.72** | **0.84** | **0.90** | **0.94** | **0.93** | **0.94** | **0.87** | **0.86** | **0.91** | **0.95** | **0.92** | **0.96** | **0.94** | **0.95** |
| Error Rate | 0.02 | 0.01 | 0.11 | 0.11 | 0.28 | 0.16 | 0.10 | 0.07 | 0.08 | 0.07 | 0.15 | 0.15 | 0.10 | 0.05 | 0.11 | 0.10 | 0.08 | 0.08 |
| F1 Score | 0.98 | 0.99 | 0.89 | 0.89 | 0.82 | 0.89 | 0.93 | 0.96 | 0.95 | 0.96 | 0.92 | 0.91 | 0.94 | 0.97 | 0.95 | 0.97 | 0.96 | 0.97 |
| Kappa Value | 0.97 | 0.99 | 0.85 | 0.85 | 0.16 | 0.55 | 0.74 | 0.84 | 0.80 | 0.84 | 0.61 | 0.57 | 0.73 | 0.88 | 0.79 | 0.90 | 0.84 | 0.86 |

# 5. Findings and Conclusion

This paper presents the improvement of OneR classification algorithm. The SRBCA has higher accuracy rate in classifying conditional anomaly using a one-class training dataset with multiple behavioral datasets. Thus, reduces the false positive and false negative data points or events in a dataset. The proposed algorithm addresses the problem of using one-class as training data for conditional anomaly detection. The experimental results show that the proposed algorithm has higher accuracy in detecting conditional anomalies for balanced and unbalanced datasets. The paper presents a comparison between the proposed algorithm and powerful classification algorithms and found that the proposed algorithm has higher accuracy in detecting conditional anomaly instances.

# 6. Acknowledgments

The dataset used in this paper was provided by the Philippine Atmospheric, Geophysical, and Astronomical Services Region XI in Davao City, Philippines.

# 7. References

[1] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLoS One*, 2016.

[2] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*. 2005.

[3] R. M. Alguliyev, R. M. Aliguliyev, Y. N. Imamverdiyev, and L. V. Sukhostat, "An Anomaly Detection Based on Optimization," *Int. J. Intell. Syst. Appl.*, 2017.

[4] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 631–645, May 2007.

[5] C. G. Nevill-Manning, G. Holmes, and I. H. Witten, "The development of Holte's 1R classifier," *Proc. 1995 Second New Zeal. Int. Two-Stream Conf. Artif. Neural Networks Expert Syst.*, 1995.

[6] G. Buddhinath and D. Derry, *A Simple Enhancement to One Rule Classification*. 2019.

[7] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[8] N. Shahid, I. H. Naqvi, and S. Bin Qaisar, “One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments,” *Artif. Intell. Rev.*, 2015.

[9] S. Abe, *Support Vector Machines for Pattern Classification - Shigeo Abe - Google Books*. Springer-Verlag London Limited 2010.

[10] C. C. Aggarwal, *Data classification : algorithms and applications*. .

[11] S. Upadhyaya and K. Singh, “Classification Based Outlier Detection Techniques,” *Int. J. Comput. Trends Technol.*, vol. 3, no. 2, pp. 294–298, 2012.

[12] N. Shahid, I. H. Naqvi, and S. Bin Qaisar, “Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey,” *Artif. Intell. Rev.*, 2012.

[13] S. C. Suh, *Practical applications of data mining*. Jones & Bartlett Learning, 2012.

[14] G. Kalyani, “Performance Assessment of Different Classification Techniques for Intrusion Detection,” *IOSR J. Comput. Eng.*, vol. 7, no. 5, pp. 25–29, 2012.

[15] N. N. Jani, V. S. Parsania, and N. H. Bhalodiya, “Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis,” *Int. J. Darshan Inst. Eng. Res. Emerg. Technol.*, vol. 3, no. 1, pp. 60–64, 2014.

[16] gov.ph, “PAGASA - Climate of the Philippines.” [Online]. Available: http://bagong.pagasa.dost.gov.ph/information/climate-philippines. [Accessed: 15-Feb-2019].

[17] C. V. Dionisio, R. A. Albacite, and N. F. B. Itoralba, “climate-philippines,” *GOVPH*, 2018. .

[18] T. A. Cinco *et al.*, “Observed Climate Trends and Projected Climate Change in the Philippines.pdf,” 2018.