FAILURE CLASSIFICATION IN ELECTRIC SWITCH MACHINES USING SYMBOLIC CONCEPTS AND COMPUTATIONAL INTELLIGENCE

Nielson Soares

Eduardo Pestana de Aguiar

Leonardo Goliatt da Fonseca

soares.nielson@engenharia.ufjf.br

eduardo.aguiar@engenharia.ufjf.br

leonardo.goliatt@ufjf.edu.br

Universidade Federal de Juiz de Fora (UFJF)

Rua José Lourenço Kelmer, S/n - Martelos, 36036-330, Juiz de Fora - MG, Brazil

Abstract. *Switch machines are electromechanical equipment that is of great importance in a rail network. With the growth of the Brazilian railway sector, these machines have been more used, which tends to increase the probability of failures. An early diagnosis of failures that may occur in a switch machines can mean a reduction of costs and an increase in the productivity. Usually, the amount of data that is available is big, making the analysis an expensive process. An alternative is the use of techniques of extraction and selection of characteristics, in order to obtain a subset of data that represents the original data. However, this practice may lead to a loss of information. Another alternative is to perform a symbolic data analysis that allows to adequately represent the raw data. This article aims to reconcile the SDA technique with supervised and unsupervised learning methods. The supervised methods used were Random Forests, K-Nearest Neighbors and Support Vector Machine. The unsupervised K-Means method combined with PCA was employed with the intention of identifying and separating the different faults that may occur in switch machines. The data set was provided by a Brazilian railway company and covers four possible switch machine states. The results presented show a high accuracy regarding the classification and identification of these faults. However, the same could not be observed regarding the clustering of the defects*

Keywords: *Computational intelligence, Symbolic data analysis, Failure classification, Railroad switch machine*

1 INTRODUCTION

With the evolution of technology, companies increasingly seek to modernize their equipment and automate processes in order to become more competitive in the market. As a consequence, and with the increase of the Brazilian rail sector, companies in this sector have increased the use of equipment such as switch machines, which are electromechanical equipment of great importance in a railway network, which tends to increase the probability of failure. Such failures can represent a great cost to the company, be it economical, due to transportation delays and accidents, and also a human cost when these accidents have victims. All this interferes with the name and credibility of the company.

In order to avoid such failures, increase the productivity of the company and reduce the TST (Time Sttoped Train) index, several maintenance strategies are applied, with emphasis on Condition Based Maintenance (CBM), which is a type of preventive maintenance where the action is made from an analysis done on data collected from sensors, on the health of the equipment. Diagnosis is a fundamental component of CBM and is defined as the failure identification and its current state (Eker *et al.*, 2012).

Due to the magnitude of the problem, the identification of these faults has attracted the attention of researchers and professionals of the area, who look for ways to apply computational intelligence with intent to solve this issue. Recently, several studies have been done on fault classification using different intelligence techniques, but only in (Aguiar *et al.*, 2014), (Aguiar *et al.*, 2016) and (Tao & Zhao, 2015) that the identification of these faults by the monitoring of the motor current of the machine came to be approached. This current monitoring can generate a massive amount of data, which can make the analysis an expensive process. Usually, techniques of extraction and selection of features are employed with the intention of obtaining a set of data that satisfactorily represents the original data, and thus apply the methods based on computational intelligence. However, this practice may result in loss of information when confronted with data in its raw form (Cury *et al.*, 2010).

In data mining, different data types such as single quantitative or categorical values, interval-valued data, multi-valued categorical data and modal multi-valued (histograms) can be applied. Typically, these types of data are called ”symbolic data” and allow the variability and uncertainty present in the raw data to be represented (Cury & Crémona, 2012). The development of methods for data analysis compatible with these data types is the main scope of the symbolic data analysis (SDA).

This paper focuses on associating SDA use with the well-known supervised classification methods such as Random Forests (RF), K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), and unsupervised clustering methods such as K-Means combined with Principal Component Analysis (PCA), with the intention of identifying and isolating the different faults that may occur in switch machines. This paper is divided as follows: Section 2 introduces the database used. Section 3 and its subsections concentrate on presenting the computational methods used to solve the problem. Section 4 discusses the results obtained in the computational analysis. Section 5 presents the conclusions drawn from the proposed methods.

2 DATABASE OVERVIEW

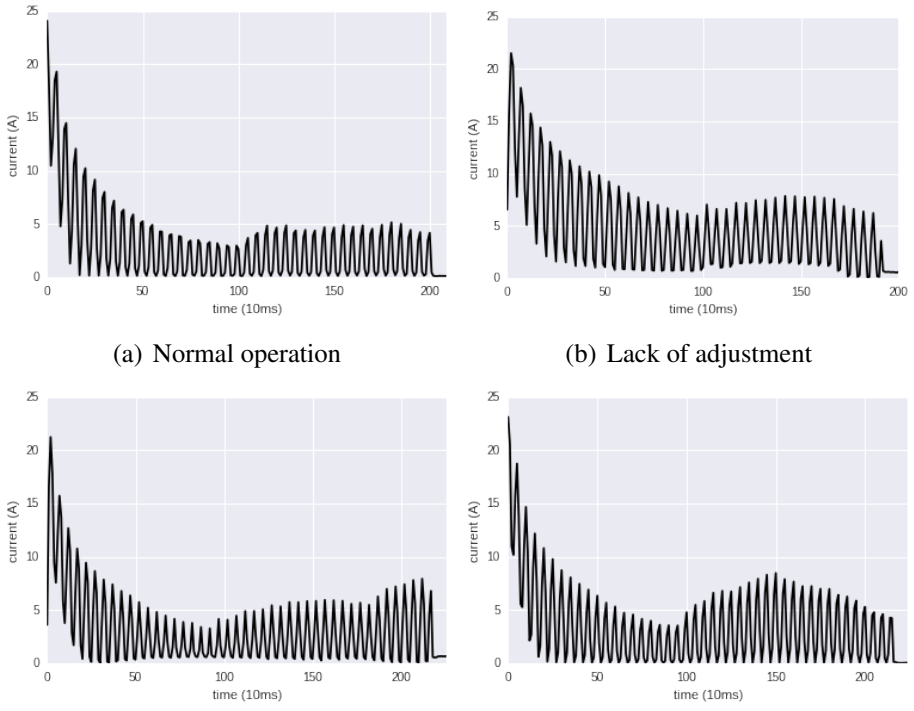
The railroad switches devices is an equipment that enables wagons to be guided from one railway to another by moving their blades before they pass, such as a rail junction. In the

past, these blades were manually moved by an operator. Nowadays, most of these devices are remotely operated by electric motor or pneumatic or hydraulic actuators Tao & Zhao (2015). This paper will focus on switch machines, electromechanical equipment, which are responsible for moving these blades from one position to the opposite position.

The data set was provided by a Brazilian railway company and consists of the current (A) signals of operation of these machines and were obtained through four channels of an industrial data acquisition board. These signals comprise four classes: normal operation and failures due to lack of lubrication, lack of adjustment and component malfunction. The data were acquired from different switch machines considering several factors, such as availability of the equipment, the complexity of the operation, favorable climatic conditions, etc Aguiar *et al.* (2017). A total of 1506 current signals were obtained, of which:

- 1389 are normal operation,
- 27 are lack of lubrication,
- 16 are lack of adjustment,
- 74 are component malfunction.

A sample of the signals of each condition is shown in Figure 1.



(a) Normal operation (b) Lack of adjustment

(c) Lack of lubrication (d) Component malfunction

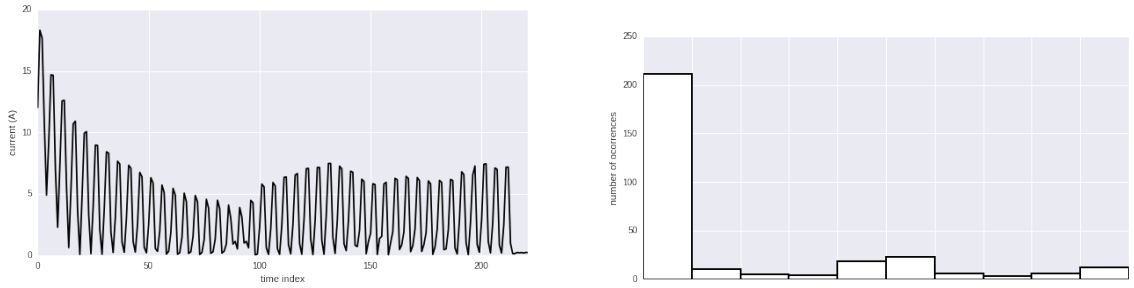
Figure 1: Typical waveform of current signals of a switch machine.

3 METHODOLOGY AND METHODS

3.1 Symbolic data analysis

As stated in Section 1, in most cases monitoring equipment for control and maintenance purposes generates a large amount of data. In this case, converting this abundant amount of data into a compact but also highly descriptive data types, such as symbolic data, become an interesting approach (Cury & Crémona, 2012).

There are several ways to convert the original data into symbolic data. The symbolic data type used in this article is the k-category histogram, which is represented by: $X = \{1(0.0025), 2(0.0721), 3(0.8546), ..., k(0.0082)\}$ (Cury & Crémona, 2012). An example can be seen in Figure 2.



(a) Raw Data (b) Symbolic Data

Figure 2: Example of a classic signal converted into a symbolic one.

The Figure 2(a) shows a signal in its classic form, as it was acquired. And the Figure 2(b) shows its representation in a 10-category histogram, where the current values were grouped, according to their frequency, in the abscissa (x) axis. In the example signal, the current ranges from 0.01 to 18.31 A, so the 10-category histogram is created according to the following ranges: 0.01-1.84 A; 1.84-3.67 A; 3.67-5.5 A; ...; 16.48-18.31 A.

3.2 K-nearest neighbors

The k-nearest neighbors (KNN) is one of the simplest classification methods to be used. The first formulation of the nearest neighbors rule was proposed in (Fix & Hodges Jr, 1951). The nearest neighbor rule consists of classifying a sample, which its classification is unknown, according to the class represented by the majority of its k-nearest neighbors. The proximity between the samples is generally calculated by the Euclidean distance. The Figure 3 shows an example of classifying a new sample using KNN.

The new sample to be classified is represented by the green circle. Depending on the value of k chosen, the class defined by the KNN to the new sample may change. For $k = 3$, the new sample would be classified as being of the class represented by the red triangle, since the nearest neighbors consist of two red triangles and one blue square. Now, for $k = 5$ the new sample would be classified as being of the class represented by the blue square, since the nearest neighbors now consist of three blue squares against two red triangles.

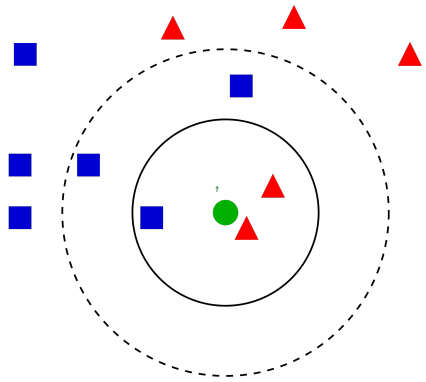


Figure 3: Classification example of a new sample using KNN.

3.3 Random forests

Random Forests (RF) are a classification technique developed by (Breiman, 2001) and consist of a combination of decision trees. Decision trees are algorithms that map the attributes of a data set, and that, from logical tests, find the attributes that best discriminate these data. Each inner node indicates a test on an attribute, each branch represents the result of this test and each leaf node has the class label (Han *et al.*, 2011). The Figure 4 shows an example of a decision tree that ranks if a customer is likely to buy a computer.

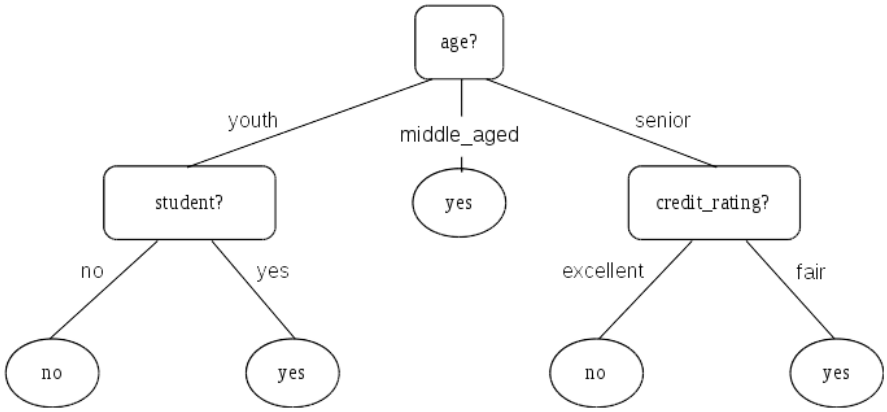


Figure 4: Example of a decision tree. Source: Han *et al.* (2011).

The ensemble of these classifiers forms the random forest, where the data set is divided into subsets that are used, at random, in the method training. Each tree is responsible for classifying the data through a vote and the final result is given by the majority of these votes.

3.4 Support vector machine

The support vector machines (SVM) were introduced by (Vapnik, 1998) and demonstrated to be very efficient in data classification. Its robust performance over sparse and noisy data makes the method extensively used. When the SVM is used for classification, it separates

classes through a hyperplane that has the maximum distance between them. This distance between the hyperplane and the first point of each class is customarily called margin and the points that are near the margin are called support vectors. The Figure 5 shows an idea of the use of a hyperplane separator in a set of data.

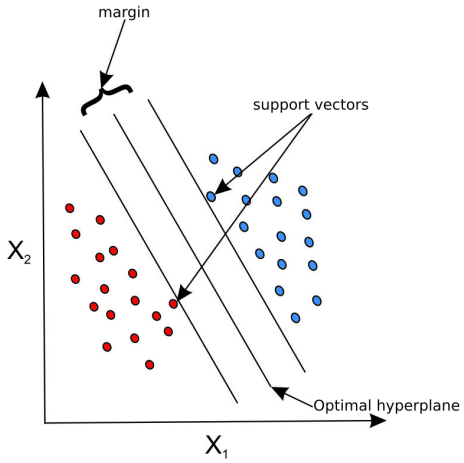
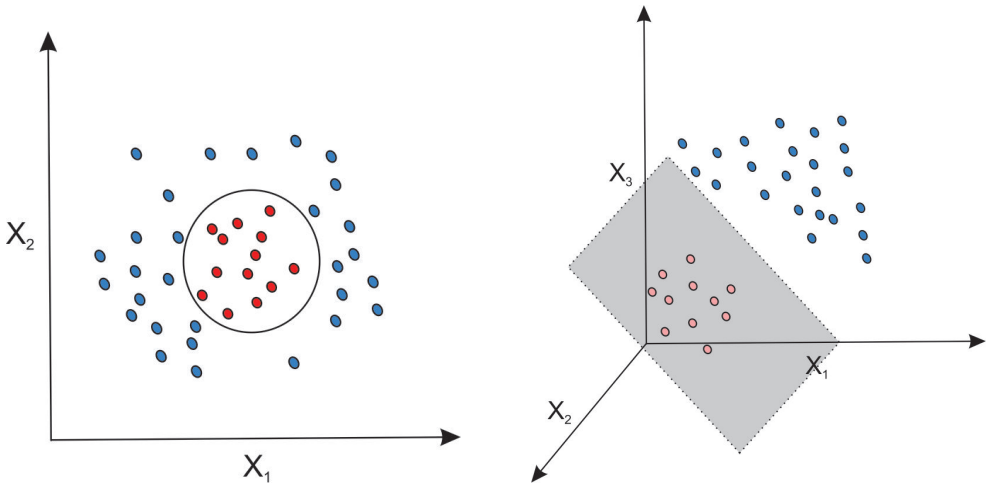


Figure 5: Example of a hyperplane separator with maximum margin.

However, there are many cases in which it is not possible to satisfactorily divide the training data by a linear hyperplane. One solution is to map data from its original space, referenced as input space, to a new space of larger dimension, called the feature space. The Figure 6 shows the difference between a hyperplane separator in the input space, Figure 6(a), and in the feature space, Figure 6(b).



(a) Nonlinear hyperplane (b) Linear hyperplane

Figure 6: Difference between a hyperplane in the 6(a) input space and in the 6(b) feature space.

However, the feature space may have a very high (even infinite) dimension, which may make computing extremely costly or impractical. The use of kernel functions, which perform a product in the input space itself and not in the feature space, help solve this problem. The Figure 7 shows what a hyperplane would look like using the input space.

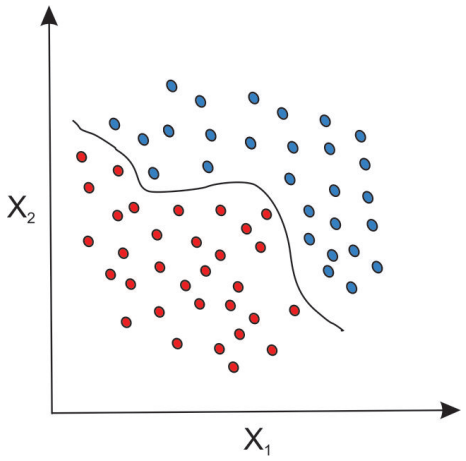


Figure 7: Example of a hyperplane separator in the input space using kernel function.

The Table 1 presents the most known and used kernel functions.

Table 1: Most used kernel functions.

Kernel	Function $K(x_i, x_j)$
Linear	$\langle x_i \cdot x_j \rangle$
RBF (Gaussian)	$exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$

3.5 Cross validation

One of the problems related to prediction models is the so-called overfitting, which occurs when the data is not completely accessible or the number of samples is relatively small, causing the model to be conditioned to the training data, thus failing to validation when different data is used. One of the alternatives for this type of problem is the application of the k-fold cross validation technique. This technique consists in the separation of k subsets of equal size, where of these, $k - 1$ subsets are used in the training of the model and the remaining subset as the foundation for the validation of the model. This process is repeated k times, using a distinct test subset at each iteration.

The Figure 8 shows an example of the division of the data set into $k = 5$ subsets at each iteration of the cross-validation process.

3.6 Grid search

In applying any of the previously described methods, it is necessary to have a parameter adjustment in order to achieve the best configuration for the execution of the processes. To search for the best configuration of parameters for the models, the technique called grid search is used, which consists of an exhaustive search in the space of parameters defined by the user. The model is then trained for each parameters setting and then evaluated through cross validation. From this validation, the configuration containing the parameters that produced the best results with the employed method is the one chosen (Bergstra & Bengio, 2012).

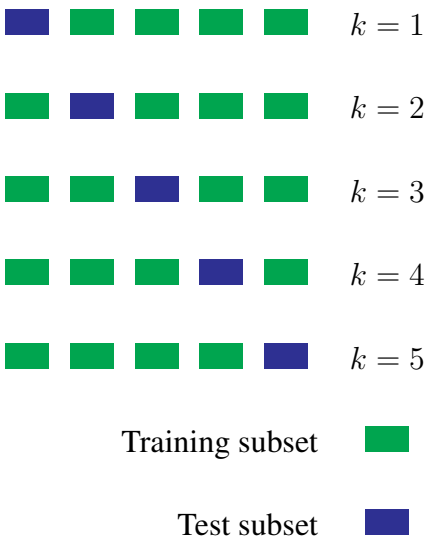


Figure 8: Division of data set into k = 5 subsets (folds).

3.7 Principal component analysis

Principal component analysis (PCA) was elaborated for the first time by (Pearson, 1901). PCA is a well known multivariate statistical technique and has been used as a pattern recognition technique for some time, obtaining excellent results (Tibaduiza *et al.*, 2016).

PCA is used for visualization of complex data. It analyzes the data table that contains the observations and which are usually described by dependent and usually correlated variables. The intention is to extract the essential patterns from the data and display this information as a set of new and orthogonal variables called principal components, which can be displayed as points in maps (Abdi & Williams, 2010). These new variables are linear combinations of the original variables and they are uncorrelated with each other.

The Figure 9 shows an example of how the PCA technique works.

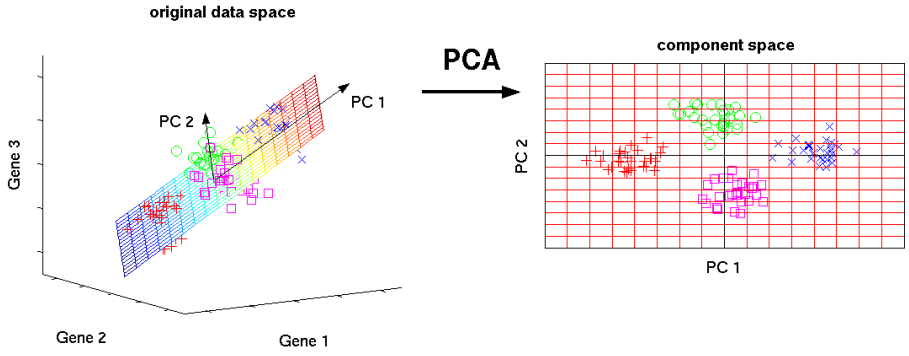


Figure 9: Example of a two clusters k-means (k = 2).

The importance of a variable in a PCA model is indicated by the size of its residual variance. The first principal component is needed to have the largest possible variance. The second

principal component must orthogonal to the first and have the largest possible variance remaining in data. The other components are computed in the same way (Abdi & Williams, 2010).

3.8 *K*-means

The idea behind using unsupervised machine learning in data mining is to seek to train the computer to find patterns across the data of a set, in order to cluster the ones with the most similarities. After the training, the computer would be able to identify which cluster a new data will come to belong to.

The k-means is an unsupervised machine learning method and it was first described in (Hartigan, 1975), where k stands for the number of clusters and must be defined previously.

The first step of k-means is to define k different centers, one for each cluster, and then associate the data to the center with less distance and therefore to its respective cluster. After this step, a new centroid is calculated for each cluster and the first step is then repeated. A loop is formed until the centroids are no longer altered.

The Figure 10 shows an example of clustering using the k-means method with two clusters.

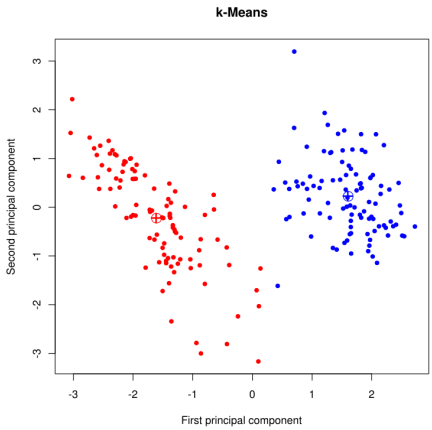


Figure 10: Example of a two clusters k-means (k = 2).

4 RESULTS AND DISCUSSION

4.1 Supervised methods

The methods were developed using the Python programming language together with the Scikit-learn library (Pedregosa *et al.*, 2011), which has tools for the development of machine learning methods.

To obtain a better generalization and therefore greater reliability of the results, each model was trained 30 times, where for each iteration the cross-validation technique, presented in Section 3.5, was used.

Table 2: Set of parameters of each method for the use of the grid search technique.

Model	Description of parameters	Variation of Parameters
KNN	Number of neighbors	[2], [3], [4], [5], [10]
	Weight function	uniform, distance
RF	Maximum depth of the tree	[5], [10], [15]
	Number of trees	[5], [10], [15]
	Maximum number of features	none, sqrt, log2
SVM	Penalty parameter C	[0.1], [0.25], [0.5], [1]
	Kernel type	linear, rbf
	Decision function	one-vs-rest, one-vs-one

In each of the 30 independent iterations, the original dataset was divided into five subsets ($k = 5$) and then the grid search technique was employed in the parameters which are shown in Table 2.

For all classification methods, the current signals were converted into symbolic data in the form of 4, 10, 16, 20 category histograms and used as input for the models.

The best configuration of the parameters obtained through the grid search for the KNN method was the number of neighbors to be considered equal to five and the weight function used in the classification is given according to the Euclidean distance, where closer neighbors will have a greater influence than neighbors which are further away. For RF, the number of decision trees was equal to 15, with a depth of 10 and the number of features to consider when looking for the best split is the number of features. For the SVM model, a linear kernel was used with a one-vs-one decision function and the penalty parameter $C = 0.1$.

The Table 3 shows the classification rates obtained by the supervised methods. The three models had good results, highlighting the SVM method, which obtained the best accuracy among them all.

When compared with other studies done, such as the one from (Aguiar *et al.*, 2014), the results obtained by this were not satisfactory, since, the classification rates were smaller. What may have contributed to the results that were lower than expected was the fact that the acquired data were unbalanced, as observed in Section 2. This scenario occurs because of the greater facility in obtaining current signals from a normal operation.

In an attempt to obtain better results, the original data for lack of adjustment, lack of lubrication, and malfunction of a component have been balanced to match the amount of data describing a normal operation. The data were then converted back into symbolic data and used as input to the classification models. The Table 4 shows the accuracy of the models for the new data set.

The results show a good improvement compared to those presented in the Table 3, and competitive results when compared with results obtained in other studies. Again, the SVM

Table 3: Classification accuracy for unbalanced data.

Model	Bins	Train accuracy	Test accuracy
KNN	4	0.970 (0.018)	0.946 (0.006)
	10	0.994 (0.012)	0.958 (0.005)
	16	0.997 (0.009)	0.959 (0.005)
	20	0.993 (0.013)	0.960 (0.004)
RF	4	0.968 (0.009)	0.946 (0.008)
	10	0.990 (0.007)	0.960 (0.006)
	16	0.987 (0.010)	0.956 (0.006)
	20	0.989 (0.005)	0.954 (0.006)
SVM	4	0.949 (0.003)	0.947 (0.006)
	10	0.966 (0.002)	0.957 (0.005)
	16	0.971 (0.003)	0.956 (0.006)
	20	0.974 (0.003)	0.962 (0.006)

Table 4: Classification accuracy for balanced data

Model	Bins	Train accuracy	Test accuracy
KNN	4	0.998 (0.002)	0.986 (0.003)
	10	1.000 (0)	0.993 (0.002)
	16	1.000 (0)	0.994 (0.002)
	20	1.000 (0)	0.994 (0.002)
RF	4	0.998 (0.0005216)	0.991 (0.003)
	10	1.000 (0.0003038)	0.995 (0.001)
	16	1.000 (0.000183)	0.997 (0.001)
	20	1.000 (0.0002066)	0.997 (0.001)
SVM	4	0.998 (0.0006622)	0.993 (0.002)
	10	1.000 (0.0002544)	0.999 (0.0005813)
	16	1.000 (0.0001382)	1.000 (0.0003709)
	20	1.000 (7.714e-05)	1.000 (0.0002653)

method, which obtained better results, with an approximate accuracy of 100% for a 10-category, 16-category, and 20-category histogram was highlighted.

4.2 Unsupervised methods

The K-Means unsupervised clustering method combined with Principal Component Analysis was used with the intention of grouping in clusters the data that have the highest similarity between themselves and, thus, seek to separate the data describing a normal operation of the switch machine from those that describe a failure.

The Figure 11 shows a comparison between the true data, the figures in the left column, and the clusters obtained through the *k*-means method, the figures in the right column.

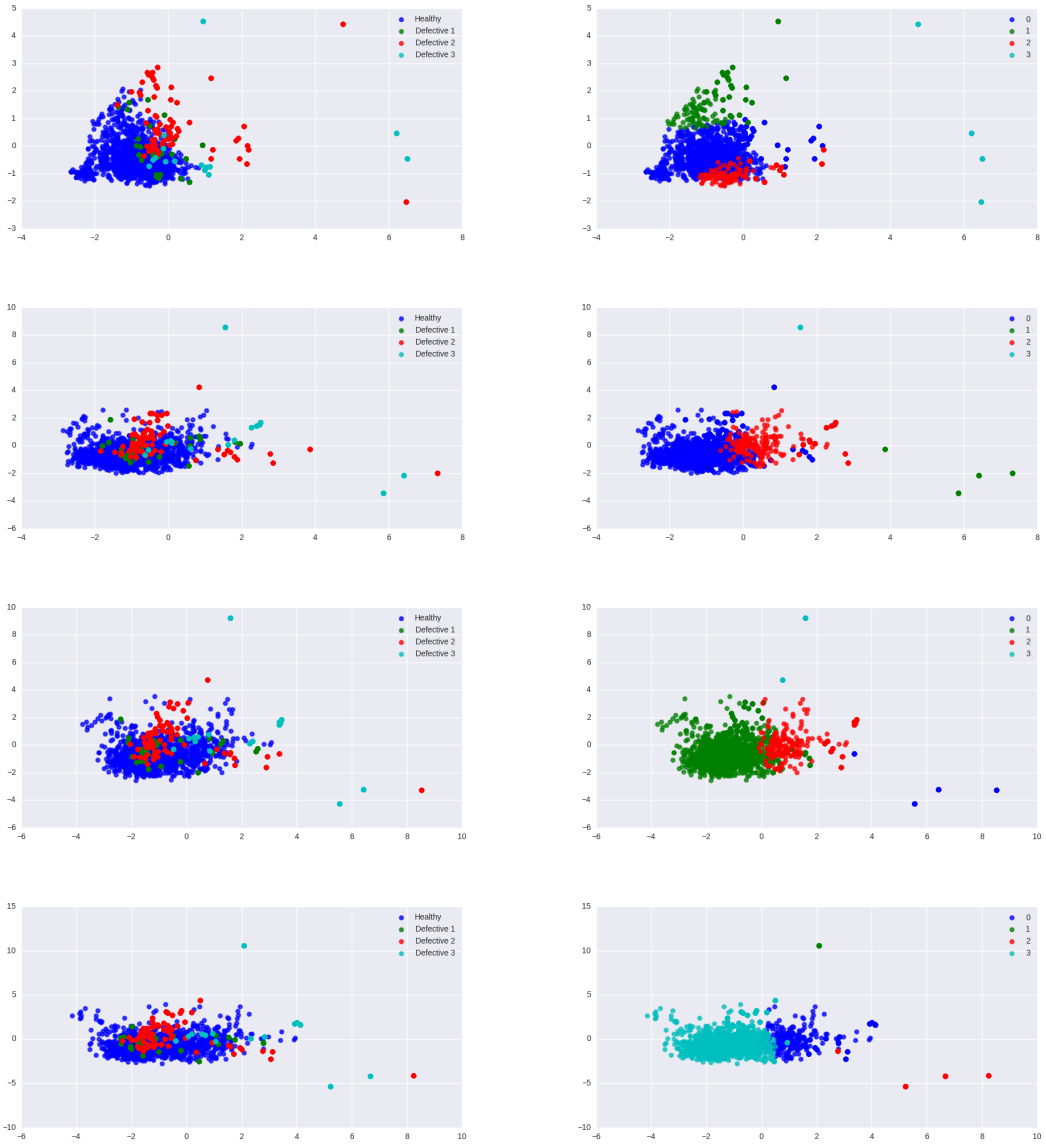


Figure 11: On the left, the true data and on the right, the clusters obtained through the *k*-means.

The k-means method divided the data into different clusters, however, it was not as previously expected as can be seen. The clusters do not represent the current states of the switch machine. This may have occurred because the data originated from different machines, different weather conditions, among other things.

5 CONCLUSION

The aim of this paper was to identify faults that may occur in switch machines using known computer intelligence techniques combined with symbolic data analysis for data processing, which helped to reduce their dimensionality. The use of cross-validation and grid search secured the best configuration and reliability for the methods applied here.

Initially, the data set used for the training of the proposed models was not balanced, that is, the classes that described a defective behavior of the switch machine had a significantly smaller number of samples than the class describing a normal operation. The results were satisfactory, but not so good when compared with the ones obtained by other papers.

After balancing the data, the models were trained again, using the new data set. The results were competitive and showed that the support vector machine had better accuracy when compared to the other models, being very effective in classifying the faults of a switch machine. Another information that can be drawn from the results is that the conversion of the classic data into a 10-category histogram is adequate to solve the classification problem using SVM.

However, the unsupervised method did not succeed in dividing the data into clusters that describe the current state of the switch machine. The fact that the data originated from different models of the machine may have interfered with the model's good performance. New research will be carried out in order to improve performance concerning data separation, using different clustering techniques, such as concentric clustering.

ACKNOWLEDGEMENTS

The authors would like to thanks UFJF, CNPq, CAPES and FAPEMIG for supporting the development of this research.

REFERENCES

Abdi, H. & Williams, L. J., 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*. vol. 2, n. 4, pp. 433–459.

Aguiar, E., Nogueira, F., Amaral, R., Fabri, D., Rossignoli, S., Ferreira, J. G., Vellasco, M., Tanscheit, R., Ribeiro, M., & Vellasco, P., 2014. *Classification of Events in Switch Machines Using Bayes, Fuzzy Logic System and Neural Network*. Springer International Publishing, Cham. pp. 81–91.

Aguiar, E. P., Fernando, M. d. A., Vellasco, M. M., & Ribeiro, M. V., 2017. Set-Membership Type-1 Fuzzy Logic System Applied to Fault Classification in a Switch Machine. *IEEE Transactions on Intelligent Transportation Systems*.

Aguiar, E. P., Nogueira, F. M. d. A., Amaral, R. P. F., Fabri, D. F., Rossignoli, S. C. d. A., Ferreira, J. G., Vellasco, M. M. B. R., Tanscheit, R., Vellasco, P. C. G. d. S., & Ribeiro, M. V., 2016. EANN 2014: a fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine. *Neural Computing and Applications*. vol. 27, n. 5, pp. 1175–1189.

Bergstra, J. & Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*. vol. 13, n. Feb, pp. 281–305.

Breiman, L., 2001. Random forests. *Machine learning*. vol. 45, n. 1, pp. 5–32.

Cury, A. & Crémona, C., 2012. Pattern recognition of structural behaviors based on learning algorithms and symbolic data concepts. *Structural Control and Health Monitoring*. vol. 19, n. 2, pp. 161–186.

Cury, A. & Crémona, C., 2012. Assignment of structural behaviours in long-term monitoring: Application to a strengthened railway bridge. *Structural Health Monitoring*. vol. 11, n. 4, pp. 422–441.

Cury, A., Crémona, C., & Diday, E., 2010. Application of symbolic data analysis for structural modification assessment. *Engineering Structures*. vol. 32, n. 3, pp. 762 – 775.

Eker, O., Camci, F., & Kumar, U., 2012. SVM based diagnostics on railway turnouts. *International Journal of Performability Engineering*. vol. 8, n. 3, pp. 289–298.

Fix, E. & Hodges Jr, J. L., 1951. Discriminatory analysis-nonparametric discrimination: consistency properties. Tech. rep.. California Univ Berkeley.

Han, J., Pei, J., & Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.

Hartigan, J. A., 1975. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA. 99th ed.

Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. vol. 2, n. 11, pp. 559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., *et al.*, 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. vol. 12, pp. 2825–2830.

Tao, H. & Zhao, Y., 2015. Intelligent fault prediction of railway switch based on improved least squares support vector machine. *Metallurgical and Mining Industry*. vol. 7, n. 10, pp. 69–75.

Tibaduiza, D. A., Mujica, L. E., Rodellar, J., & Güemes, A., 2016. Structural damage detection using principal component analysis and damage indices. *Journal of Intelligent Material Systems and Structures*. vol. 27, n. 2, pp. 233–248.

Vapnik, V. N., 1998. *Statistical learning theory*. vol. 1. Wiley New York.