A new method of fuzzy patches construction in Neuro-Fuzzy for malware detection

Andrii Shalaginov[1] **Katrin Franke**[2]

[1,2]Center for Cyber- and Information Security, Gjøvik University College, Norway

Abstract

Soft Computing is being widely used in Information Security applications. Particularly, Neuro-Fuzzy approach provides a classification with human-understandable rules, yet the accuracy may not be sufficiently high. In this paper we seek for an optimal fuzzy patch configuration that uses elliptic fuzzy patches to automatically extract parameters for the Mamdami-type rules. We proposed a new method based on χ^2 test of data to estimate rotatable patch configuration together with Gaussian membership function. This method has been tested on the automated malware analysis with accuracy up to 92%. Further on, it can find an application in Digital Forensics.

Keywords: malware detection, neuro-fuzzy, digital forensics, optimization

1. Introduction

Malware analysis and detection is an important task in the domain of Digital Forensics that has become more demanded with the increased number of committed cybercrimes. Per today this work is done mainly manually by analyst that requires reverse engineering and investigation of system artefacts. Modern Anti-Virus software include composed malware signatures that need considerable amount of time to analyse systems and software artefacts if the malware is sophisticated. At this point there have been already used Neuro-Fuzzy (NF) as a Soft Computing method in Digital Forensics that provides inexact solutions with possible understandable model to a human expert. Hybrid Intelligence model such that NF allows automated analysis of the malware properties that results in a construction of corresponding decision fuzzy rules. A NF can be constructed either as a cooperative or a hybrid model. Second model incorporates automated tuning of the fuzzy rules that considered to be more appropriate in unmanned analysis. In [1] Guillaume studied various hybrid models stating that NF is one of the most useful data approximation techniques. It was mentioned that the research of possible ways of optimization is required to deal with a great number of parameters in the model and preserve the interpretability. This papers is about how the construction of fuzzy rules can be done using more precise configuration of elliptic fuzzy patches.

The NF can produce either regression Sugeno-type or classification Mandami-type rules while learned. In this work we consider methods that focus on the Mamdami-type since this model gives understandable class label (for example, "benign" or "malicious") and possess a good performance in solving diverse problems [2]. Fuzzy rules are constructed on the predefined set of linguistic terms in each fuzzy variable. Goztepe in [3] described an expert system that uses predefined set of attack techniques as input fuzzy set based on triangular membership function. In the [4] by Zhang et al diagonal Mahalanobis distance was used to measure the degree of truth of manually constructed links of behavioural characteristics. In the study [5] Shafiq described how the Mamdani-type rules can be constructed manually, yet not in combination with Artificial Neural Networks (ANN). Instead the ANFIS system was given for malware analysis based on Sugeno-type fuzzy rules, which is out of our scope. The publication [6] shows a need for more sophisticated mechanisms to extract and define parameters of fuzzy sets to improve a classification performance. Drobics in [7] presented an inductive learning for deriving rules from Self-Organizing Map (SOM) with fixed set of equal fuzzy regions in each set. This way of fuzzy rules extraction makes it mandatory to ask for manual input from the field expert as well as linguistic definition of the regions.

Another challenge that exist is a type of a membership function (MF) to be used in the 2nd of the NF method. Kosko [8] at first defined Gaussian MF, yet then converted to triangular MF that complies with a projection of circumscribed rectangular around the elliptic fuzzy patch. Chi in [9] used combined triangular MF from SOM prototypes without involving a correlated features transformation. In the research [10] authors stated and proved that despite the faster approximation by common triangular MF in Mamdami-type inference system, the Gaussian-like methods show good fit and more stable results. In the work by Wang in 1992 [11] the Gaussian MF function was used to show how the Gaussian-like regions can be used to approximate the real-valued continuous function with help of Stone-Weierstrass theorem. Furthermore, in 1995 Kim [12] introduced extension of the univariate Gaussian MF into Gaussian sum approximation for function approximation. Moreover, it was shown that the such approximation better fits an inter-

polated function among different radial-basis functions. Therefore, we use modified Gaussian MF in this work to incorporate mentioned merits.

Properly adjusted parameters of fuzzy sets are important for maintaining both accuracy and understandability classification rules. However, it is not a trivial task to place these parameters especially when dealing with Big Data analytics. Therefore, in studied literature, the manual analysis is initially required to define each particular fuzzy set and then learning is used to tune the classification model. Though the Kosko established a theoretical foundation for the fuzzy sets allocation [13], it is still not used due to multiple assumptions when dealing with real-world data. So, such procedure has not been used for malware detection so far due to a challenging estimation of the rules parameters.

In this paper we consider one of the problems in Digital Forensics that is malware detection based on the similarities in software properties. Despite the fact that static signature-based malware detection is widely used it requires a precise definition of the malware properties. So, the challenge is to define similarities by means of specially constructed rules automatically without a help of analyst. Our method is based on the method defined by Dickerson and Kosko in 1996 [8] while improvement is done to establish more generalized parameters of the fuzzy rules based on the probabilistic modelling. Basically, we proposed a method that makes an optimal configuration of the elliptic regions of the fuzzy rules based on the probabilistic χ^2 test. Moreover, corresponding MF is constructed to incorporate information from hyperellipsoid.

The reminder of this paper is organized as following. The Section 2 gives an insight into NF method and fuzzy patches allocation using SOM. In this Section we will present Kosko work and the way how elliptic regions are constructed. Additionally, challenges with Kosko method are pointed out. Then, in the Section 3 the new way of getting parameters of elliptic regions is detailed. Additionally, the rotatable Gaussian MF is introduced replacing the projection-base triangular MF used by Kosko [13]. The experimental design and datasets, including characteristics, are described in the Section 4. Further, the Section 5 provides overview performance of the new proposed method with respect to simple rectangular and Kosko patches. At last, the discussions, conclusions and propositions for future work are given in the Section 6.

2. Unsupervised estimation of fuzzy patches by Kosko

In this Section we will concentrate our attention on the Kosko method for the fuzzy patches construction. Though the most simple method of fuzzy patches construction is rectangular patches, where the regions are defined according to the clusters extracted from SOM. In this case the patch configuration is defined by 1st and n^{th} order statistics in a cluster and then simple triangular MF is used. From the other side, Kosko method includes two stages. Initially, the parameters of the ellipsoid fuzzy rules are estimated by means of unsupervised SOM procedure. Then, supervised learning is done using a ANN model to tune the parameters of the classification model. Further on, we will concentrate our attention on improvement of the effectiveness of the rough placements of the fuzzy patches.

1. *Unsupervised data fitting into elliptic regions*

   As is it mentioned in the Kosko method the fuzzy rule patch Π^i, which corresponds to the space of input-output vector products $X_i \times Y_i$, can be considered as an elliptic region rather than rectangular to fit the input real data and reduce error [14].

   Definition 1: The input data sample $X_i = \{A \in R^M\}$, which is a collection of the features $a = \{x_0, \ldots, a_M\}$ can be characterized as a point in M-dimensional space. The whole set of N data samples is therefore contains in an M-dimensional ellipsoid (hyperelipsoid) with a radius α of general form $\sum_{n=0}^{M-1} x_i^2 = \alpha^2$. It expanded as a following generalized equation of hyperellipsoid due to non-uniformity of features and shifted center of origin in the centroids of the features c_i (non-zero mean):

$$\alpha^2 = (x - c)^T \ (x - c) \tag{1}$$

   where α defines a pseudo-radius of the fuzzy patch for orthogonal uncorrelated features.

2. *Ellipsoid rotation caused by features correlation*

   Since the features do not always possess the same statistical properties it will cause the hyperellipsoid to be rotated and strained with respect to features axes.

   Definition 2: To incorporate the correlation between the features we include it in a decomposed way:

$$(x - c)^T \ P \ \Lambda \ P^T \ (x - c) = \alpha^2 \tag{2}$$

   where $\Sigma^{-1} = P \ \Lambda \ P^T$ is positive definite symmetric inverted covariance matrix, then factorized by means of eigendecomposition into the diagonal matrix of eigenvalues $\Lambda = (\lambda_1, \ldots, \lambda_M)$ and orthogonal matrix of eigenvectors $P = (e_1, \ldots, e_M)$ that rotates the ellipsoid. It should be noted that according to the Eigen decomposition theorem [15] the covariance matrix is transformed into decomposition $\Sigma^{-1} = P \ \Lambda \ P^T$ since eigenvectors are orthogonal. This is the initial step of the Principle Component Analysis (PCA) [16], where eigenvectors defines the direction of the distribution and eigenvalues - the degree of variance (stretch) along the corresponding direction. Furthermore, the radius of each corresponding ellipsoid axis is equal to $\alpha/\sqrt{\lambda_i}$ due to $\lambda_i \cdot (x_i - c_i)^2 = \alpha^2$ [13] since for orthogonal matrix

$P^T = P^{-1}$, which will eliminate the eigenvectors values in the Equation 2.

3. *Triangular MF in elliptic input-output space*

In [8] Dickerson and Kosko used simplification of the hyperellipsoid by inscripting it into a rectangular region, which is projected on the axes to form parameters for triangular MF.

Definition 3: Triangular MF function is based on the projection of the hyperrectangular in which the hyperellipsoid is circumscribed.

$$\mu_j(X) = \begin{cases} 1 - \dfrac{|x_j - c_j|}{p_j}, & |x - c_j| \leq \dfrac{p_j}{2} \\ 0, & otherwise \end{cases} \quad (3)$$

where the μ_j defines the MF of j feature and the projection of circumscribed rectangular on i axis is

$$p_i = 2 \cdot \alpha \cdot \sum \frac{|cos\gamma_{ij}|}{\sqrt{\lambda_i}} \quad (4)$$

where angle between the i axis and j eigenvector e:

$$\gamma_{ij} = arccos(e_j(i)) \quad (5)$$

Since the unit eigenvector represents the vector of direction cosines between the principle axis of rotation and original features axis as it is eliminated from the characteristic polynomial $Ae - \lambda e = 0$. In Euclidean space the features vectors are mutually orthogonal as well as corresponding set of eigenvectors. Therefore, it feasible to use a set of eigenvectors as rotation matrix in eigendecomposition of the inversed covariance matrix Σ^{-1}.

So, this is the method developed by Dickerson an Kosko [8] for unsupervised estimation of hyperellipsoid parameters and fuzzy patches placement is presented. It can be seen that the definition of α^2 in the Equation 2 determines the efficiency of the method. So far this parameter have been defined empirically from data. Another complication in the Equation 3 is a triangular MF, which is used in fuzzy sets. The projection on the of the hyperrectangle circumscribed around the target hyperellipsoid are the parameters of this MF. We can see that such projection does not fit the data properly since the data might contain outliers. In the mentioned above researches the authors stated that such way of MF composition does not use the orientation of the ellipsoid and this might be improved by utilization of the ellipsoid patch itself in MF. The following Section gives a view on the probabilistic modelling in definition of parameter α^2. Moreover, this MF function is presented based on the multivariate distribution.

As was mentioned before the most common ways of definition of fuzzy patches is manual one. Usually, it was used defined set of fuzzy terms that spits interval in a number of equal intervals corresponds to each of the fuzzy term. This method model will produce K^M rules for K terms in each of M input features that makes it hardly possible to use all rules that require multiple evaluation against questioned application. However, this approach is not efficient and not scalable in case when the exact fuzzy terms are not known or can not be defined manually.

3. Optimization of elliptic fuzzy patches allocation

In this Section we will present an improvement of the Kosko method that was proposed in [8]. In particular we will focus on automated estimation of the elliptic fuzzy patches parameters and MF.

3.1. Construction of elliptic fuzzy patches based on multivariate distribution test

Since there are no exact information about the hyperellipsoid, it has to be determined from data distribution.

Proposition 1: The problem of estimation of the elliptic fuzzy patch Π^i parameters for real-data clusters can be reformulated as a parametric distribution test of fitting data into a distribution model.

At this point we make an assumption that the data can be described by means of normal distribution we can state that the cluster derived from SOM should fit elliptic Gaussian multivariate distribution according to Kosko method [13]. The data used further in the experiments complies with this assumption, however, for other data this assumption has to be proven. It means that by performing a χ^2 test the test of statement above is performed. This test is designed to measure how well the distribute data fits Gaussian distribution. Since this test is originally designed for the categorical data, we have to use χ^2 test for the variance as described in the chapter 12 of the [17] for $df = M - 1$ transforms to an equation for the M random variables:

$$\chi^2 = \frac{(M-1)S^2}{\sigma^2} = \sum_{i=0}^{M-1} \left(\frac{x_i - \bar{x_i}}{\sigma_i}\right)^2 \quad (6)$$

By considering the sample variance S^2 as variance of all elements in the particular data cluster and standard deviation σ^2 as a theoretical deviation in this cluster we can state that $\chi^2 \approx \alpha^2$ with some degree of confidence interval β. This challenge is related to a chance of outlier rather than fuzziness that was explained by Ross in the [18]. So, by introducing β we are able to control chance of the data to be in distribution to avoid possible outliers.

Proof: By considering the non-transformed unfolded equation of hyperellipsoid we will get:

$$\frac{(x_0 - c_0)^2}{\sigma_0^2} + \ldots + \frac{(x_{M-1} - c_{M-1})^2}{\sigma_{M-1}^2} = \alpha^2 \quad (7)$$

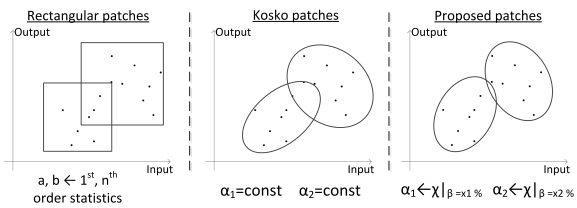

Figure 1: Examples of patches configuration in simple rectangular, Kosko and proposed methods

which is the same as following considering one feature as a fixed variable:

$$\sum_{i=0}^{M-1} \frac{(x_i - c_i)^2}{\sigma_i^2} = \alpha^2 \tag{8}$$

The χ^2 test in the Equation 9 makes it clear that the parameter α^2 in the Equation 2 can be estimated from contingency table and equal to the value of χ^2 for a particular β for a defined number of *df*:

$$\sum_{i=0}^{M-1} \left(\frac{x_i - c_i}{\sigma_i}\right)^2 = \alpha^2 = \chi^2|_\beta \tag{9}$$

This definition of elliptic region complies with the goodness of fit of data within an elliptic region that may illuminate outliers or error values. In case of significant number of rules, this should reduce it to get more specific fuzzy regions and reduce uncertainty by overlapping regions. Comparison between the simple rectangular, Kosko and proposed way of fuzzy patches contraction is given in the Figure 1.

To summarize, we use χ^2 test to find value of the parameter α^2 from the Kosko method. As was proved before, we come to a conclusion that $\chi^2 = \alpha^2$ in a defined β with specified *df* for the statistical model of the data. This gives an adoptable model that adjusts configuration of elliptic patches according to specified qualities of data that influence selection of β.

3.2. Modified Gaussian MF for correlated data

The triangular MF used in a simple rectangular and Kosko methods are not appropriate for this purpose since it does not incorporate all available information from constructed elliptic fuzzy patch. Moreover, the function should count on rotation of the patch and distance from the center.

Proposition 2: A radial basis Gaussian MF can be used instead of the triangular projection-based MF to provide a better fit to the data in elliptic fuzzy patch when rules combination is calculated.

Proof: The minimum principle used to define a rule's MF [8] is a combination of the following form of Cartesian product:

$$\mu_R = \mu_0(X) \wedge \mu_1(X) \cdots \wedge \mu_{M-1}(X) \tag{10}$$

where each MF function μ_i of each feature is defined as a triangular one 3. At this point, we replace the triangular MF by means of Gaussian function for the feature i:

$$\mu_{a_i} = s_i e^{-\frac{1}{2}\left(\frac{x_i - c_i}{\sigma_i}\right)^2} \tag{11}$$

where $s_i \in (0; 1]$ is a scaling constant and other parameters are the corresponding statistical propertied of ellipsoid projection on each feature space. Furthermore, the this is used in the rule's MF in the Equation 10 to for the overall MF in the Equation 12. In the [19] and [20] the overall membership grade is described in such way considering collection of different image bands. The authors propose to use M-th root to derive the overall MF. However, this will give a significant overlap between the rules that cause overfitting of the classification model.

$$\begin{aligned}\mu_R &= s_0 e^{-\frac{1}{2}\left(\frac{x_0 - c_0}{\sigma_0}\right)^2} \cdots \wedge s_{M-1} e^{-\frac{1}{2}\left(\frac{x_{M-1} - c_{M-1}}{\sigma_{M-1}}\right)^2} \\ &= \prod_{i=0}^{M-1} s_i e^{-\frac{1}{2}\left(\frac{x_i - c_i}{\sigma_i}\right)^2} = \prod_{i=0}^{M-1} s_i \cdot e^{-\frac{1}{2}(x-c)^T (x-c)}\end{aligned} \tag{12}$$

using the matrix form of hyperellipsoid in the Equation 2. The scaling factors s_i are not known and have to be defined empirically. However, we consider the product of Guassian MF functions in the Equation 10 as the multivariate distribution. From the other side Kim in 1995 [12] introduced a Gaussian sum approximation as a product of singular Gaussian MF. As result, we make the scaling factors s_i equal to 1.0 because the rule's MF should not be restricted to a magnitude of $\frac{1}{\sqrt{(2\pi)^M |\Sigma|}}$ in multivariate probability density function [21] as it is used in the [12]. Piegat in the book [22] mentioned that the angle between the axes of hyperellipsoid and features can be a possible solution to increase the precision of the MF. Gaussian MF was also employed for Mamdami-type model by Soto et al in [23] to achieve the lowest error for time series prediction. This book presented an example of 2-D Gaussian MF that incorporates also an angle α. Author mentioned that such model is going to have 5 degrees of freedom $(x_1, x_2, c_1, c_2, \alpha)$ for two features model in comparison to non-rotatable function and therefore it might be an obstacle to use it. Yet the set of angles from the Equation 5 is already known, which does not require additional computation steps. So, the generalized equation of the hyperellipsoid is used then in the derived Gaussian MF sum approximation that incorporates all available information from the elliptic region by means of covariance matrix:

$$\mu_R = e^{-\frac{1}{2}(x-c)^T \; P \; \Lambda \; P^T \; (x-c)} \tag{13}$$

Comparison between the MF functions for all three methods is given in the Figure 2. Though

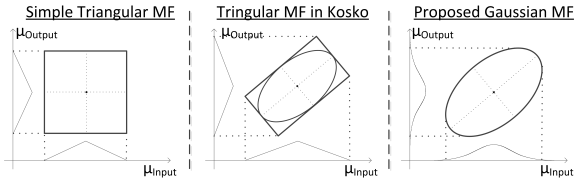

Figure 2: Comparison of MF in simple rectangular, Kosko and proposed methods

the proposed method also incorporates rotation angle and lengthiness of the ellipsoid. This can not be included in the triangular MF used for rectangular patches and Kosko patches.

So, we have presented how the triangular MF in the Equation 3 is replaced by modified Gaussian function to form a rule MF. This will provide a better fit into the data when the features are correlated and triangular MF based on the hyperellipsoid projections can not cover the model properly.

3.3. Proposed algorithm for malware detection

Based on the studied literature and problems with predefined set of fuzzy terms we propose following Hybrid NF methods based on the SOM. It ensembles unsupervised clustering of the similar applications and tuning of extracted fuzzy rules.

1. *Rules discovery procedure based on elliptic fuzzy patches*

(a) *Clustering based on the features similarities.* SOM is trained to convert M-dimensional feature vector into $2D$-lattice that consists of $H \times W$ nodes. After training each node $S_{i,j}$ includes cluster of similar data samples.

(b) *Construction of elliptic regions.* We make a hypothesis that the multivariate distribution defines the model [21] for the data clustered in the node $S_{i,j}$ in SOM (referring to the Figure of histograms from different features from the malware dataset):

$$g(x) = \frac{1}{\sqrt{(2 \cdot \pi)^M |\Sigma|}} \cdot e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})} \tag{14}$$

Based on the this assumption and the properties of Gaussian multivariate distribution there have to be defined corresponding parameters $\{\bar{c'}, \bar{\sigma'}, \sum\}$ for the given data. The distribution model represents an n-dimensional elliptic region or fuzzy patch for the cluster in Euclidean geometry.

We have assumed based on the data sample set that the features distribution is a Gaussian one. To test this we employ generalized Pearson χ^2 test for multidimensional data, where the χ^2 will reflect the probabilistic radius of the hyperellipsoid or Mahalanobis distance from the centroid of cluster to any point in the distribution:

$$\chi^2 = \sum_{n=0}^{M-1} \frac{(x_i(a_n) - E[a_n])^2}{E[a_n]} \tag{15}$$

where the $E(a_n)$ represents a theoretical expectation of the particular feature a_i and equal to sample's mean c_i. By means of ranging of the continuous variables the χ^2 statistics can be calculated based on the number of df, which corresponds to a number of dimensions M.

So, in this statistical model we define goodness of fit [17] of the data samples in the hyperelliptical region by means of χ^2 distribution test. It describes how well the distributed data samples fit defined multivariate distribution. The χ^2 chi distribution roughly is a sum of squared difference between all points in a given set. To determine the value of the χ^2 based on the β and df we use contingency table. Further, the squared radius of the hyperellipsoid is equal to the χ^2 considering the Equation 15.

(c) *Extracting the parameters of the fuzzy patches.* At this point we apply the first stages of Principle Component Analysis (PCA) to extract the set of eigenvalues $\bar{\lambda}$ and set of eigenvectors $\bar{\bar{v}}$ [16]. With help of PCA we rotate the original multidimensional distribution to remove the correlation. This is done since the distribution might have unequal deviations and directions different from the main axes.

Then, mentioned set of parameters has to be defined as a complete characteristics of the fuzzy patch. Therefore we can define each feature region as Π^i that defines fuzzy patch. This is done since at the current step it is hard to understand the heuristics behind this region, yet possible to understand similarities by such fuzzy patch definition.

(d) *Construction of membership function.* Each fuzzy patch is characterized by the MF that binds an input with an output. To increase precision of the fuzzy patches coverage and replace general rectangular patches the following MF is defined for the elliptic region and based on the two dimensional example [22]:

$$\mu_R = e^{-\frac{1}{2}(x-c)^T \; P \; \Lambda \; P^T \; (x-c)} \tag{16}$$

2. *Tuning of the rule-based classification model* After discovery all fuzzy patches, the Hybrid NF model is composed considering rules that were discovered. Further one the model fitting by Delta Learning rule in order to reduce error between the labelled data samples and predicted by the rule. Output of the ANN contains two nodes for the malicious and benign classes.

4. Experimental Design

In this Section we proposed the following experiments to show the answer on the question mentioned in the introduction. So, we performed the following experiments: (1) Estimation of the errors in training on the 2nd step of the NF, (2) Comparison with other rules-based methods and (3) Performance estimation. On the 1st stage of the NF, the number of the training iterations for SOM is equal to the size of the dataset, when the data are randomly selected for SOM training. Preliminary trials were performed on different sizes of SOM (3x3, 5x5, 10x10) to estimate the effect on the accuracy for different datasets. The smallest size 3x3 provided the best accuracy.

Finally, for the Kosko method we used fixed value of α for all experiments, while confidence interval $\beta = 95\%$ was used to illuminate outliers in proposed method. Number of df is chosen to be $3 * M - 1$ since for each of the dimension the statistical model will include mean, angle and spread around mean.

During the 2nd step, max of 100 iterations of the ANN and fixed learning rate 0.1 were chosen since showed good accuracy. The performance evaluation includes two domains: (1) regression-based accuracy of the model with Mean Absolute Error $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - d_i|$, Mean Absolute Percent Error $MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{y_i - d_i}{d_i}|\cdot 100\%$ and Relative Absolute Error $RAE = \frac{\sum_{i=1}^{N}|y_i - d_i|}{\sum_{i=1}^{N}|d_i - \bar{d}|}$, where y_i is an actual output of ANN and d_i - is a data sample class; (2) classification % of the data using min-max principle and derived rules samples in cross-validation.

Data sets. In order to prove the proposed methods we selected several datasets and implemented both stages of the NF with rectangular, Kosko and proposed configuration of the fuzzy patches. There have been selected several versatile dataset to demonstrate the proof of concept. At this point the bootstrap aggregation is used since SOM is a variant of ANN, which is unstable and produces the different clustering results during the each run. We randomly generated 10 samples B_i from the training set and the select the sample that possess the best classification accuracy on the cross-validation. After this, the sample B_i is going to be used in later experiments to show the results consistency.

The applicability of the methods is shown on the datasets that contains binary classified datasets from UCI Machine Learning Repository [1] that are described in the Table 1. There included dataset of manually constructed features derived from the static and dynamic analysis of malicious and benign applications.

[1] https://archive.ics.uci.edu/ml/datasets/

Table 1: The properties of the datasets used in the experiments

№	Dataset	Feat.	Samples	Norm.
1	Climate simul.	18	540	Yes
2	Fertility	9	100	Yes
3	Banknote auth.	4	1372	No
4	Mobile malware	36	596	No

Table 2: Performance comparison of the methods. "R" refers to number of extracted rules and "M" - method ("S" - simple rectangular, "K" - Kosko, "P" - proposed). №defines a corresponding dataset

№	R	M	Performance			
			MAE	MAPE,%	RAE	Acc, %
1	15	S.	0.71	37.04	4.48	23.33
		K.	0.91	45.64	5.74	91.48
		P.	0.03	3.15	0.19	97.03
2	9	S.	0.18	13.03	0.88	88.00
		K.	0.12	6.00	0.56	88.00
		P..	0.07	4.32	0.36	88.00
3	17	S.	0.10	5.65	0.20	96.13
		K.	0.44	22.36	0.90	92.63
		P.	0.06	6.28	0.12	100.00
4	18	S.	0.57	37.61	1.17	41.34
		K.	0.40	40.70	0.83	58.48
		P.	0.07	4.97	0.14	91.76

5. Results & Analysis

This Section devoted to an explanation of the performed tests and the comprehensive results to evaluate the proposed method from different points of view. The results of the methods are given in the Table 2 using clusters derived from 3x3 SOM.

The proposed configuration of the fuzzy patches fits well the data. The percentage of error shows only 3-5% error for the proposed method based on the datasets, while classification accuracy using min-max principle is always above Kosko method. Using the χ^2 for building fuzzy patches provides better way to fit data. So, we can summarize that both original models do not possess the same accuracy as the proposed one neither on the regression performance nor on the classification based on the fuzzy rules model. Moreover, the proposed method performs well on such datasets with considerable amount of features keeping the classification accuracy of around 90%. To see how well the patches describe the data, the MAE was observed during the learning on the 2nd step of NF form mobile malware dataset. The results are 0.618945, 0.392991, 0.094395 for simple rectangular, Kosko and proposed methods respectively. The Kosko method requires more iteration to tune the model, while new proposed method gives the best results and simple rectangular is not affected mostly by tuning. Therefore, we can state that proposed patches fit the data with the lowest possible error.

Table 3: Performance comparison of the new proposed method with existing numerical decision tree- and rule-based methods implemented in Weka. The regression metrics are "MAE" and "RAE, %". №defines a corresponding dataset

№	Proposed		J48		JRip	
	MAE	RAE	MAE	RAE	MAE	RAE
1	0.03	19.82	0.08	51.749	0.10	65.05
2	0.07	36.25	0.22	102.97	0.21	98.41
3	0.06	12.81	0.01	3.67	0.02	4.29
4	0.07	12.49	0.07	16.121	0.1	20.80

5.1. Evaluation with respect to other rule-based classification methods

Since the proposed method is rule-based classifier we decided to compare the accuracy with some tree-based and rule-based classifiers with a good performance. We consider decision trees since this model can be linearised into decision rules. At this point we tested four data in the Table 3 sets with the methods J48 (implementation of statistical classifier C4.5) and JRip (implementation of rule learner Repeated Incremental Pruning to Produce Error Reduction - RIPPER) that can be found in Weka.

The results are consistent considering that proposed method is based on the intermediate fuzzy components and C4.5 and RIPPER are based on the numerical values, which have to be more accurate.

5.2. Complexity

It can be noted from the Section 3 that new method requires more resources to learn from data and form a fuzzy logic model. In this Section we perform expensive analysis of the time required for learning and space to be occupied. The first aspect is given with respect to single- and multi-threaded applications that are used in modern computational systems. The estimation of the computational complexity is an important issues that to be considered in critical applications of real-time systems and Big Data analysis. In the Table 4 we present the time required to learn and to make a decision for Kosko, rectangular and proposed methods. The time measurements are given for a single-threaded and a multi-threaded learning from data. For the testing purpose we took mobile malware dataset and 15 rules derived from 3x3 SOM and results are presented in the Table 4. It has been measure execution of the implemented version in two modes.

When parallel optimization is used the required time of the proposed method is considerably less than others. The time complexity is an important issues that affects decision whether to use particular methods while dealing with Big Data since the data complexity will affect the execution time. To summarize, the proposed methods requires more time then Kosko and a simple rectangular, yet the accuracy is much better.

Table 4: Time required to learn models and inference new data for different amount of fuzzy rules, Seconds. "M" refers to used method

M	Learning		Inference ,10^{-6}	
	Sequent.	Parallel	Sequent.	Parallel
Rect.	0.43	0.57	10.37	1.57
Kosko	0.87	0.65	15.91	2.49
Prop.	0.42	0.21	353.11	47.62

Table 5: Size required to stored the rules for the banknote authentication dataset, Bytes. The measurements are: "Structure" - size of empty rule structure, "1 Rule "- size required to store a single rule using for mentioned earlier dataset with 4 features, "Model" - total size required to store all the classification rules

Arch.	Models	Structure	1 Rule	Model
32bit	Kosko	28	104	3224
	Proposed	24	232	7192
64bit	Kosko	56	200	6400
	Proposed	48	456	14592

The proposed structure of radial-basis rules need to store M^2 elements of the inversed covariance matrix + N centroids, while triangular MF-based model needs only $2 \cdot N$ centroids. It means that the size of the rules in Big Data analysis will converge to a number of samples, which means that if $M << N$ then size of the Kosko and Rectangular models will exceed the size of the proposed model making them more demanding in storage complexity.

Otherwise, the proposed rules will occupy more space on the small sample as pointed in the Table 5. The complete model of 31 rules with 4 features can be stored using 6 KBytes for triangular MF rules and 15 KBytes for the proposed MF. The size looks reasonable considering the capacities of modern computers. It makes possible to apply the rules on the embedded devices with significant resources limitation since the vectors stored in contingency memory and does not require multiple random access for inference. The implementation was done using C++ 11 with Boost, STL, Eigen and OpenMP. The test machine included 8-cores Intel i7-3632QM - 2.20GHz with 8GB DDR3.

6. Discussion & Conclusions

In this paper we investigated two ways of fuzzy patches construction: simple rectangular and Kosko. It was shown that these methods are good for simple datasets with uncorrelated features. However, due to the existence of the outliers and errors the classification accuracy drops consistently. We have proposed a new method for fuzzy patches construction in the NF method using χ^2 tests that

provides up to 92% against 58% for Kosko on the mobile malware dataset. It was proposed a new method based on the statistical properties of the distributions to construct the elliptic fuzzy patches. Furthermore, we use Gaussian elliptic MF to be more precise in defining the degree of truth of each fuzzy rule. It was noticed that the proposed method of elliptic regions construction requires less optimization and tuning on the 2nd step, which brings regression error down to 3-5%. We can make a conclusion that the proposed method posses better performance and the tuning on the 2nd step provides only little improvement of the model in comparison to the rectangular and Kosko methods. Moreover, this proposed is suitable for Big Data analysis when the amount of data instance is incredibly large and there are some complex relationships between features. The complexity tests show a tradeoff in space and speed with respect to considerable improvement in the accuracy. As a future work we see application of non-parametric distribution rather than Gaussian parametric to be more efficient and fit the data.

References

[1] S. Guillaume. Designing fuzzy inference systems from data: An interpretability-oriented review. *Trans. Fuz Sys.*, 9(3):426–443, June 2001.

[2] I. Iancu. *A Mamdani Type Fuzzy Logic Controller*. INTECH Open Access Publisher, 2012.

[3] Kerim Goztepe. Designing fuzzy rule based expert system for cyber security. *International Journal of Information Security Science*, 1(1):13–19, April 2012.

[4] Yichi Zhang, Jianmin Pang, Feng Yue, and Jinxian Cui. Fuzzy neural network for malware detect. In *Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on*, volume 1, pages 780–783, Oct 2010.

[5] M. Zubair Shafiq, Muddassar Farooq, and Syed Ali Khayam. A comparative study of fuzzy inference systems, neural networks and adaptive neuro fuzzy inference systems for portscan detection. In *Proceedings of the 2008 Conference on Applications of Evolutionary Computing*, Evo'08, pages 52–61, Berlin, Heidelberg, 2008. Springer-Verlag.

[6] Andrii Shalaginov and Katrin Franke. Automatic rule-mining for malware detection employing neuro-fuzzy approach. In *Norsk informasjonssikkerhetskonferanse (NISK)*, 2013.

[7] M. Drobics, W. Winiwater, and U. Bodenhofer. Interpretation of self-organizing maps with fuzzy rules. In *Tools with Artificial Intelligence, 2000. ICTAI 2000. Proceedings. 12th IEEE International Conference on*, pages 304–311, 2000.

[8] J. A. Dickerson and B. Kosko. Fuzzy function approximation with ellipsoidal rules. *Trans. Sys. Man Cyber. Part B*, 26(4):542–560, August 1996.

[9] Zheru Chi, Jing Wu, and Hong Yan. Handwritten numeral recognition using self-organizing maps and fuzzy rules. *Pattern Recognition*, 28(1):59 – 66, 1995.

[10] P.A.M. Romagnoli, N.H. Romero, and J.C.S. Mora. Equivalence between gaussian-like and mamdani fuzzy methods in the multi-objective design of electronic circuits. In *Electronics, Robotics and Automotive Mechanics Conference, 2008. CERMA '08*, pages 68–73, Sept 2008.

[11] L. X. Wang and J. M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *Trans. Neur. Netw.*, 3(5):807–814, September 1992.

[12] Hyun Mun Kim and J.M. Mendel. Fuzzy basis functions: comparisons with other basis functions. *Fuzzy Systems, IEEE Transactions on*, 3(2):158–168, May 1995.

[13] Bart Kosko. *Fuzzy Engineering*. Number v. 1 in Fuzzy Engineering. Prentice Hall, 1997.

[14] Bridget Bertoni. Multi-dimensional ellipsoidal fitting. Technical report, Southern Methodist University, 2010. accessed: 07.10.2014.

[15] Herve Abdi. *Encyclopedia of Social Networks and Mining*. Thousand Oaks (CA), 2007.

[16] Lindsay I Smith. A tutorial on principal components analysis. Technical report, Cornell University, USA, February 26 2002.

[17] Timothy C. Krehbiel Mark L. Berenson, David M. Levine. *Basic Business Statistics, 11/E*. Pearson, 2009.

[18] T.J. Ross. *Fuzzy Logic with Engineering Applications*. Wiley, 2009.

[19] Caiyun Zhang and Fang Qiu. Hyperspectral image classification using an unsupervised neuro-fuzzy system. *Journal of Applied Remote Sensing*, 6(1):063515–1–063515–14, 2012.

[20] F. Qiu. Neuro-fuzzy based analysis of hyperspectral imagery. In *Photogrammetric Engineering and Remote Sensing*, volume 74, pages 1235–1247, 2008.

[21] Undergraduate Advanced Data Analysis 36-402. Advanced data analysis from elementary point of view. Technical report, Department of Statistics, Carnegie Mellon University, 2013.

[22] A. Piegat. *Fuzzy Modeling and Control*. Studies in Fuzziness and Soft Computing. Physica-Verlag HD, 2001.

[23] Jesus Soto, Patricia Melin, and Oscar Castillo. Time series prediction using ensembles of neuro-fuzzy models with interval type-2 and type-1 fuzzy integrators. In *The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013*, pages 1–6, 2013.