Design and implementation of artificial intelligence for the prevention of cybercrime in the Republic of Congo based on machine learning: Approach based on artificial learning in cybersecurity for the detection of intrusions

Vivien Armel Eyangolo, Roch Corneille Ngoubou, Pierre Kafunda Katalay

Design and implementation of artificial intelligence for the prevention of cybercrime in the Republic of Congo based on machine learning: Approach based on artificial learning in cybersecurity for the detection of intrusions.

Vivien Armel EYANGOLO[1],Roch Corneille NGOUBOU[2],Pierre KAFUNDA KATALAY[3]

[1]Bircham International University, Madrid, Spain.

[2]Mathematics and computer science laboratory, University of Kinshasa, DRC.

[3]Computer Training Center-Computer and Research Center for the Army and Security, Brazzaville, Congo.

Abstract

Companies produce enough data in their operations and this data is stored in their information systems. These information systems operate in an insecure environment, that is to say less secure because of the communication channel used which is the Internet. It is exposed to several types of attacks such as: denial of service, SQL injection, password cracking, etc. This is how cyber security was developed to deal with these scourges of attacks which can alter the proper functioning of an organization. Cyber security is a set of processes aimed at protecting an information system against cyber attack.

Our approach in this article is to develop a security system based on artificial intelligence, more precisely on artificial learning.

The objective of artificial learning is to create predictive models from a training sample. In this article we propose a new approach, which consists of controlling the sensitivity of support vector machine using the perturbation method.

Keyword:Artificial learning, Predictive model, Support vector machine, Imbalanced classes, Sensitivity and Specificity.

Introduction

In our modern technological age, cybersecurityplays an essential role in theprotection against the ever-present dangers of cybercrime. These malicious attacks, such as hacks, ransomware and online fraud, highlight the importance of ensuring the protection of our sensitive data. Faced with the ever-evolving tactics employed by cybercriminals, maintaining cybersecurity is a constant battle. This requires the development of revolutionary solutions capable of effectively protecting the integrity, confidentiality and accessibility of informationin the field of virtual reality

Most learning systems assume that all the datasets used to learn are balanced. However, when it comes to real applications, this balance is not always verified. And in this case, the classifier cannot accurately detect the false positive and false negative.

In a two-class classification problem, when the training data of the majority class are much greater in number than those of the minority class, the algorithms allowing to obtain a minimum error rate will always tend to neglect the class minority, because of this disproportion. Which justifies the great connection between the two forms of asymmetry in supervised learning. Indeed, asymmetry comes in two main forms: class imbalance and cost asymmetry. Class imbalance concerns problems where one of the modalities of the target variable is much less represented than the others, which disrupts the learning algorithms. Cost asymmetry concerns cases where the costs of errors are not symmetrical. In this article, it is the asymmetry of classes that interests us. To deal with this problem, several algorithms have been developed, for example sampling, under sampling. In this article, we propose a new approach, which is an algorithmic approach which consists of perturbing the machine, so as to take into account the minority class.

Our problem is presented as follows: Given a learning sample in an asymmetric situation, how to create a machine capable of making optimal assignments, without being influenced by this class imbalance. To solve this problem we propose an algorithmic approach based on the disturbance of SVM by inserting two parameters at the level of the economic function, one of which is the cost of misclassification of positive examples and the other designates the cost of misclassification of negative examples.

This article is organized as follows:

- Section 1 shows how to determine model parameters when the data is balanced;
- in section 2 which is our contribution, we showed how to determine the model parameters when the data is unbalanced;
- Section 3 is the evaluation of our model.

I. Cyber Security

Several definitions of the term “cybersecurity” have been established at the national and international levels. For the purposes of this document, “cybersecurity” means all tools, policies, guidelines, risk management methods, actions, training, best practices, safeguards and technologies that can be used to protect the availability, integration and confidentiality of assets in connected infrastructures of government, private organizations and users. These assets include connected computing devices, personnel, infrastructure, applications, services, telecommunications systems and data in the cyber environment.

Before going any further, this is probably where we need to try to clarify the use of several expressions and terms commonly used almost interchangeably, such as:

- information system security (ISS) which can be understood as a set of measures implemented to achieve and maintain the cybersecurity state of an information system (IS);
- cybersecurity which is therefore the desirable state to achieve;
- cyber defense which also includes measures implemented to maintain a state of cybersecurity, but in the face of particularly marked adversity and within a very specific time frame.

Therefore, the protection of the Information System (IS) aims to identify, analyze and evaluate the risks that affect these assets (which can be hardware, software, business processes), and to take the necessary measures to control these risks. In order to properly protect itself from threats from online sources, a country or organization can take certain security measures. In this area, you have the choice between several levers. However, the implementation of such measures must be done in an informed and reasonable manner, compatible with the threat of pressure on the elements intended to be protected and the value attributed to them by their owners.

The first basic approach consists of setting up a measurement base based on a priori benchmarks applicable to the organization's environment. These standards can be associated with professional or activity-specific regulations. They can also be very basic lists of indicators. This approach focuses resources on implementing security measures, instead of defining a list of measures to be implemented. It is not very personal, but is designed to be easy to access, in particular to protect the simple IS from general and simple threats.

For its part, the risk analysis method makes it possible to resolve the problem from above by studying the threats and terrible events which particularly affect the IS studied, in order to better adapt to the security measures in place. Therefore, risks can have various origins or properties, but it is obvious that what people can now call cyber risks is a risk that puts pressure on countries, organizations and individuals. Risk analysis methods are particularly suited to complex systems subject to major threats. However, it must rely on a relatively diversified and specific skills base, which will allow the implementation of appropriate methodological tools. Depending on the objectives set for the exercise, it may require a substantial investment of resources.

These two methods complement each other: depending on the issues of the IS studied, the compliance method will constitute a basic foundation which can effectively complement the risk analysis work, and adapt more precisely and specifically to the IS concerned and its context. The business processes in which he participates and the risk scenarios that put him under pressure.

Cyber risk, the threat of exploiting one or more vulnerabilities in a digital system, is itself a risk. It has strategic importance, and its achievements can be fatal and dizzying for an organization. For all these reasons, if the notion of cyber risk historically refers to a very technical area, it is now obvious that it

will inevitably extend to areas that matter to managers and decision-makers, and will come more from all areas. At the functional level, any profession that uses digital technology to support its value chain or produces digital equipment or services will be affected by cyber risks. Digital security is, as we often say, everyone's business. This awareness and this collective and holistic involvement in the search for the state of cybersecurity is not trivial.

Taking preventive measures is obviously an important aspect of ensuring the security of the system network. However, this is not unique. Once the protection elements are deployed, it is also useful to dedicate resources to the monitoring system in order to detect possible security incidents that may arise. For this reason, security monitoring capabilities rely at least on software or hardware tools and data to guide them properly. In order to detect non-trivial attacks, manual analysis is almost essential. When capabilities are implemented as services for the entire organization or a group of organizations, larger infrastructure can be deployed and dedicated human resources can be hired to provide monitoring functions, analysis and possible answers.

A country that wishes to respond to a victim attack carries out a detailed analysis of the opportunity to use specific levers. It will be difficult to find a combination of levers expressed as precisely as possible and consistent with strategic political objectives, to maximize the effectiveness of the response while remaining within a precise framework of action, such as the framework of international law. In order to conduct the analysis and decision-making process as calmly as possible, during a real attack, it would be interesting if the country concerned could prepare in advance, which will help it direct its final response and, if necessary, where appropriate, to guide its implementation of operational planning.

II. Nonlinear Support Vector Machines [5, 7,10,12, 14,15,22]

The Linear Separator (decision function) defined by the SVM is given by:

$\boldsymbol{f}(\vec{\boldsymbol{x}})$=. +b, where: $\boldsymbol{wx}$

w is a vector perpendicular to the linear separator, called a weight vector;

b is called bias;

". » represents the dot product of the vectors and.$\boldsymbol{wx}\vec{\boldsymbol{x}}$ $\vec{\boldsymbol{w}}$*[5, 7]*

Assume that the data is nonlinearly separable. The determination of the decision function first involves a transformation of the data space into another characteristic () or representation space, possibly of high dimension, where the data becomes linearly separable.$\mathcal{F}$

*This approach is based on Cover's theorem in 1965 which indicates that a set of examples transformed nonlinearly into a higher-dimensional space is more likely to be linearly separable than in its original space. [***22]**

- Determination of Wide Margin Separator**[10, 12]**

Consider the applicationϕ : X → $\mathcal{F}$

$$x \rightarrow \phi(x)$$

With representation space of larger dimension than the data space$\mathcal{F}$X

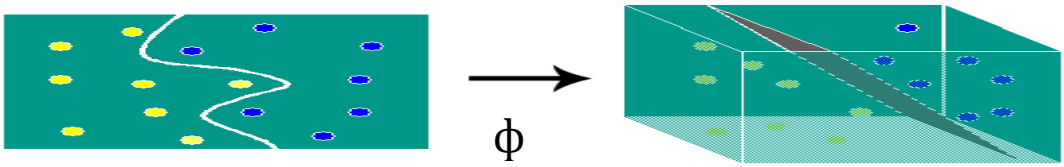

Figure 1 :*Data space transformation [10]*

The set of training data becomes:

E = , where: and Є {-1, +1}. [8,14]$\{(\phi(x_i), y_i)\}_{1\le i\le M} \phi(x_i) \in \mathcal{F}\ y_i$

Therefore, the optimization problem can be written as follows:

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ S/c \\ \forall i,\ y_i(w.\phi(x_i) + b) \geq 1 \end{cases} \quad (1)$$

$$L = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{M} \alpha_i [y_i(\omega.\Phi(x_i) + b) - 1] (2)$$

By solving the system

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \\ \frac{\partial L}{\partial b} = 0 \end{cases}$$

We obtain the result:

$$\begin{cases} w = \sum_{i=1}^{M} \alpha_i y_i \phi(x_i) \\ \sum_{i=1}^{M} \alpha_i y_i = 0 \end{cases} \quad (3)$$

Withw ∈ $\mathcal{F}$

By replacing (3) in (2) we obtain the dual of problem (1):

$$\begin{cases} \max \ \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \\ S/C \\ \forall i,\ \sum_{i=1}^{M} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases} \quad (4)$$

For certain characteristic spaces and associated applications, the scalar products are easily calculated using specific functions, called kernel functions such as:ϕ

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (5)$$

The interest of these kernel functions is to make it possible to calculate scalar products without having to explicitly transform the data by the function, therefore, without necessarily knowing this function.$\mathcal{F}$ ϕ ϕ

By integrating equation (1) into (2), we obtain the following dual problem:

$$\begin{cases} \max \ \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \, \alpha_j y_i y_j K(x_i, x_j) \\ S/C \\ \forall \mathrm{i}, \ \sum_{i=1}^{M} \alpha_i \, y_i = 0 \\ \alpha_i \geq 0 \end{cases} \qquad (6)$$

- Classification of New Data**[12, 15]**

From the above, we can therefore deduce the decision function, for the classification of new data:

$$f(x) = sgn\,(w.\phi\,(x\,) + b)$$

$$= sgn\,((\sum_{i=1}^{M} \alpha^*{}_{\mathrm{i}} y_i \phi\,(x_i\,)).\phi\,(x\,) + b)$$

$$= sgn\,(\sum_{i=1}^{M} \alpha^*{}_{\mathrm{i}} y_i \langle \phi\,(x_i\,), \phi\,(x\,)\rangle + b)$$

$$= sgn\,(\sum_{i=1}^{M} \alpha^*{}_{\mathrm{i}} y_i K(x_i, x) + b)$$

II.1 Soft margin [4,8,23,25]

In the case of soft margin, we proceed in the same way as we did above. Our contribution consisted of disrupting the economic function and relaxing the constraints by introducing error terms.ξ_i

ξ_iIndicates to what extent the example is on the wrong side or not.x_i

If =0 then is well classified$\xi_i x_i$

If ≥1 then is misclassified$\xi_i x_i$

Thus, in general, the optimization problem of Support Vector Machines can be written as follows [4,6,23]:

$$\begin{cases} \min \frac{1}{2}\|\mathrm{w}\|^2 + C\sum_{i=1}^{M} \xi_i \\ S/C \\ \forall \mathrm{i}, \ \mathrm{y_i}(\mathrm{w}.\phi\,(x_i\,) + \mathrm{b}) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \qquad (7)$$

The dual of the general optimization problem to be solved will therefore be:

$$\begin{cases} \max \ \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ S/C \\ \forall i, \ \sum_{i=1}^{M} \alpha_i y_i = 0 \\ 0 \leq \ \alpha_i \leq C \end{cases} \quad (8)$$

❖ Core Function

Consider a set X of observations in the training data. The Gram matrix of the kernel associated with this set is a square matrix of order M and general term: . [1.8, 25]$\{x_i\}_{1\leq i\leq M}$ $k(.,.)K(i,j) = k(x_i, x_j)$

Theorem 1[8]:

A function is a valid kernel if it is symmetric and positive definite.$\boldsymbol{kX \times X \rightarrow \mathbb{R}}$

In other words, a function is kernel if and only if:k

- $K(i,j) = K(j,i)$;
- $\sum_{i=1}^{M}\sum_{j=1}^{M} c_i c_j \ k(x_i, x_j) \geq 0, \ \forall \ c_1, c_2 \ldots c_M \in \mathbb{R}$

This last condition results in the fact that all the eigenvalues of the Gram matrix are positive and non-zero.

Proof: (see [8])

II.2 Examples of Some Kernel Functions [2,3, 17]

- Linear: $K(x_i, x_j) = x_i . x_j$
- Polynomial: or$K(x_i, x_j) = (x_i . x_j)^d (x_i . x_j + c)^d$
- RBF (Radial Basic Function): $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \ \gamma > 0$

III. Unbalanced Classes and SVM [20,24]

III.1. Performance Measurement of a Classifier

Performance evaluation is a very important step when doing supervised learning. We have said that imbalance poses a performance problem when it is not taken into account during the learning process. The question is how to tell that the model is efficient or not?
To measure the performance of a classifier we use the confusion matrix.

III.2. Indicators in the case of two classes

As we said above, real supervised learning situations are generally two-class problems where one of them is the class of interest. Very often, we notice that the examples of this so-called class of interest are in the minority. In the following, we consider the examples of the minority class as positive and those of the majority class as negative.

The basic tool for model evaluation is the confusion matrix.

Real class	Predicted class		Sum
	+	-	
+	True Positives (VP)	False Negatives (FN)	N_+
-	False Positives (FP)	True Negatives (TN)	N_-

Table II.1. Confusion Matrix [18,24]

III.3. Sensitivity and Specificity [20]

a) Sensitivity

The probability that a positive example is classified as positive by the model is what we call the sensitivity rate:

$$sen = \frac{VP}{VP + FN}$$

b) Specificity

The probability that a negative individual will be classified as negative is what we call the specificity rate:

$$spe = \frac{VN}{VN + FP}$$

By aggregating these two indices, we can obtain a single index:

a) *Youden index*[24]

$$youden = sen + spe - 1$$

The decision rule is as follows: "the model becomes better when the Youden index is close to 1".

b) *Likelihood ReportL*[19.24]

$$L = \frac{sen}{1 - spe}$$

The decision rule is: “The higher the likelihood ratio, the better the model”
Evidence(see [20]).

The following table gives certain results according to the likelihood ratio.

L	Contribution of the Model
10 and above	Important
5-10	Moderate
1-5	Weak
1	none

Table II.2. Contribution of the model according to the likelihood ratio. [24]

c) *Geometric Mean*

$$G = \sqrt{sen * spe}$$

This indicator is most used when the data is unbalanced.

Noticed :

These indicators allow, by focusing on the class of interest, to evaluate the quality of the prediction. The recall is equal to the sensitivity presented in the previous point. However, precision is the proportion of positive individuals among those who were classified as positive.

$$r = \frac{VP}{VP + FN}$$

$$p = \frac{VP}{VP + FP}$$

These two indicators are better when it comes to evaluating the performance of a model where one of the classes is the one of interest.
When and , then such a classifier is said to be perfect because it classifies all positive individuals well and does not classify a negative individual as positive.$r = 1 p = 1$

III.4. SVM in cases of unbalanced classes[9,11,16,20,21,24,]

The problem of unbalanced classes is becoming very frequent with the applications of Machine Learning algorithms in several fields, notably in telecommunications in the detection of telephone fraud, bioinformatics, text classification, voice recognition, intrusion detection and others. [9.21]

In telephone fraud detection, fraudulent credit card detection and intrusion detection, the imbalance is real, and when it is not taken into account, it leads to serious performance problems, and

thus generates very high costs when the classifier misclassifies elements of the minority class. In the case of intrusion detection for example, saying that an intrusion is normal access is very serious. [9.16]

Despite the importance attached to handling unbalanced data, most classifiers tend to only optimize accuracy without taking into account the relative distribution of each class. As a result, these classifiers poorly classify elements of the minority class when the data distribution is highly skewed. This poses a performance problem. To solve this problem, we propose an algorithmic solution which is our contribution which consists of disrupting the economic function by inserting two parametersC^+ AndC^-which are respectively the cost of misclassifications for positive examples and the costs of misclassifications for negative examples.[11, 20,24]

By assigning high misclassification cost for minority class compared to majority class (), the effect of class imbalance will be reduced $C^+ > C^-$

We therefore optimize the following mathematical program:

$$\min_{\omega,\xi_i} \frac{1}{2}\|\omega\|^2 + C^+ \sum_{i|y_i=1}^{N} \xi_i + C^- \sum_{i|y_i=-1}^{N} \xi_i \quad (9)$$

$$s/c \quad y_i(\omega^t x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0, i = 1, \dots\dots, N$$

$$L_p = \frac{\|\omega\|^2}{2} + C^+ \sum_{i|y_i=1}^{N} \xi_i + C^- \sum_{i|y_i=-1}^{N} \xi_i - \sum_{i=1}^{N} \alpha_i[y_i(\omega.x_i + b) - 1 + \xi_i] - \sum_{i=1}^{N} \mu_i \xi_i$$

Or$\alpha_i \geq 0 \; et \; \mu_i \geq 0$

By solving the system:

$$\begin{cases} \frac{\partial L_p}{\partial \omega} = 0 \\ \frac{\partial L_p}{\partial b} = 0 \\ \frac{\partial L_p}{\partial \xi_i} = 0 \\ \alpha_i[y_i(\omega.x_i + b) - 1 + \xi_i] = 0 \\ \mu_i \xi_i = 0 \end{cases} \quad (10)$$

We obtain the following dual program:

$L_D = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j k(x_i x_j) - \frac{1}{4c^+}\sum_{(i|y_i=+1)}^{N} \alpha_i^2 - \frac{1}{4c^-}\sum_{(i|y_i=-1)}^{N} \alpha_i^2 (11)$

Either$\epsilon^+ = \frac{1}{4c^+}$, $\epsilon^+ = \frac{1}{4c^-}$

The balance between sensitivity and specificity can be controlled using the following scheme:

The components of the diagonal of the kernel matrices are fixed as follows:

$$k(x_i, x_i) = \begin{cases} k(x_i, x_i) + \epsilon^+ \; pour \; y_i = +1 \\ k(x_i, x_i) + \epsilon^- \; pour \; y_i = -1 \end{cases}$$

Algorithm

Beginning

- *Introduce: C, Core*
- *Calculation*L_p
- *Initialize b and all to 0*α_i
- *Do Until KT (Kuhn and Tucker conditions) are satisfied:*
  - *Find the vector*ω
  - *Calculation*L_p
  - *Determine*α_i
  - *Calculate the bias b*

    Error : $E_i = f(x_i) - y_i = \left(\sum_{j=1}^{M} \alpha_j y_j K(x_j, x_i) + b\right) - y_i$

The complexity of this algorithm is:$O(n^2)$.

VI. Digital experimentation and results

The data on intrusion detection is truly unbalanced. There are many more negative examples (normal use) than positive examples (intrusion). In this experiment, we use data from KDD (Data Mining and Knowlegment Discovery) Cup 1999 [source: kdd.ics.uci.edu/databases/kddcup99/kddcup99.html*J*which are relatively complex data and suitable for testing the effectiveness of the modified SVM model on imbalanced data.

Indeed, this data was collected and distributed by the MIT laboratory with the sponsor of DARPA (Defense Advanced Research Projects Agency) and AFRL (Air Force Research

Laboratory) for the evaluation of research on intrusion detection. The KDD data is obtained from raw data from tcpdump, a command-line packet analyzer, based on simulations on the United States Air Force network over a nine-week period. This data concerns several attacks and is divided into five classes in total which are: normal, DOS (Denial of Service), R2L (Remote To Local Attack), U2R (User To Root Attack) and Probing Attack. And each record has 30 attributes shown in the following table. This learning database contains 4940000 examples for a size of 744 MB and 10% are used as training data. In our experiment, we want to treat the problem in a binary case, which is why we consider the DOS, R2L, U2R, Probing examples belonging to the same class (attack). Which means that we only have two classes: normal and intrusion.

No.	Attribute	Description
1	Duration	Connection duration in seconds
2	Protocol	Protocol type
3	Service	Network service for the destination
4	Flag	Connection status (normal, error
5	src_bytes	Data size in bytes from source to destination
6	dst_bytes	Data size in bytes from destination to source
7	land	1 if the connection uses the same port at the source as at the destination
8	wrong_fragment	Wrong number of fragments
9	Urgent	Number of urgent packets
10	Hot	Number of hot indicators
11	num_failed_logins	Number of incorrect access attempts
12	logged in	1 if accessed successfully and 0 otherwise
13	num_compromised	Number of compromised conditions
14	root_shell	1 if root shell obtained and 0 otherwise
15	su_attempted	1 if attempting a "su root" command and 0 otherwise
16	num_root	Number of root accesses
17	num_fil_creation	Number of file creation operations
18	num_shells	Number of command prompts or shells requested
19	num_access_files	Number of configuration file access operations
20	num_out_bound_cmds	Number of commands outside the ftp session
21	is_host_login	1 If access belongs to the hot list and 0 otherwise
22	is_guess_login	1 if invited 0 otherwise
23	Count	Number of connections to the same host as the current connection two seconds ago
24	srv_count	Number of connections to the same service as the current service two seconds ago
25	serror_rate	% of connections with SYN error
26	srv_serror_rate	% of connections with SYN error
27	rerror_rate	% of connections with REJ error
28	srv_rerror_rate	% of connections with REJ error
29	same_srv_rate	% of connections to the same service
30	diff_srv_rate	% of connections to different services

Table III.1. Attributes characterizing records**[kdd.ics.uci.edu/databases/kddcup99/kddcup99.html]**

Results Obtained

We implemented our algorithm ((9), (10)) in python and found the following results by considering 1830 example datasets.

	Attack	Normal	Total
Attack	57	243	300
Normal	3	1527	1530
Total	60	1770	1830

	Attack	Normal	Total
Attack	218	82	300
Normal	1	1525	1530
Total	219	1611	1830

By calculating the different parameters allowing the performance of the model to be measured, we obtain the following:

Classic SVM		**Our Algorithm**	
Se	Sp	Se	MS
0.19	0.99803922	0.72666667	0.9994641

Classic SVM	**Our Algorithm**
G_Means	**G_Means**
0.4354234	0.85216883

We note that, faced with unbalanced data, classical algorithms provide very poor results. We must therefore take this imbalance into account as we did in section 2, to improve the performance of the model. Our proposed approach gives better results.

Conclusion

Most real data is often unbalanced, and this poses a performance problem for the classifier which is more likely in this case to classify individuals from the positive (minority) class as negative, which constitutes a great danger. In this article, we solved this problem of imbalance of training games. We took the case of intrusion detection using the Machine Learning approach, we used the SVM with an algorithmic modification, and the KDD99 cup data. We found very satisfactory results. As future prospects we propose the use of ensemble methods which involve several classifiers and then combine them by majority vote. This way of doing things can also lead to good results since we use several experts.

Bibliography

1. BOUSQUET Olivier, Introduction to Support Vector Machines", Center for Applied Mathematics, Ecole Polytechnique Palaiseau, Orsay 2001
2. CAPO-CHICHI, Machine learning for business relationship detection, University of Montreal, 2012
3. FIESCHI Marius, Data Mining, data mining: Concepts and techniques, University of the Mediterranean Aix Marseille II, February 2006.
4. GINNY Mak, "the implementation of support vector machines using the sequential minimal optimization algorithm", McGill University, Montreal, Canada, 2000
5. GRAF Geraldine and Julien, Analytical CRM OLAP analysis tools and Data Mining, University of Fribourg, April 26, 2008
6. HAMOUI Fady, "Fraud detection and knowledge extraction", Montpellier II University, 2007
7. EL HASSANI Imane, SVR with boosting for long-term forecasting, Polytechnic School of the University of Tours, 2011-2012.
8. ISHAK Ben Anis, Selection of Variables by Support Vector Machines for Binary and Multiclass Discrimination in High Dimensions, thesis, University of the Mediterranean, 2007
9. JAYSHREE Jha, Intrusion Detection System using Support Vector Machine, International Journal of Applied Information Systems, 2013
10. KAFUNDA Katalay Pierre, Supervised Machine Learning based on Vector Machine Support for Churn analysis in a telecommunications company, Anales de la faculty des Sciences, volume 1/2017, UNIKIN, 2017
11. KIM Sungchul and HWANYO YU, SVM: Classification, Regression and Ranking, Springer-Verlag Berlin Heidelberg 2012.
12. LIAUDET Bertrand, Data Mining Course, EPF – 4, 5th year, Business and Project Engineering Option, 2002.
13. MARREF Nadia, "Incremental learning & Support Vector Machines", HADJ LAKHDAR-BATNA University, 2013
14. MILHEM Hélène, Machine Vector Support, Toulouse Institute of Mathematics, INSA Toulouse, France IUP SID, 2011-2012
15. ORCHARD, B., Yang, C. and Ali, M.: Innovation in Applied Artificial Intelligence: 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Springer, New York, 2004, 1272 p

16. PERNER, P.: Machine Learning and Data Mining in Pattern Recognition: 5th International Conference, Springer, New York, 2007, 913 p.
17. PLAT John, Fast Training of Support Vector Machines using Sequential Minimal Optimization, Microsoft Research, August 14, 2000
18. PREUX Ph, Data mining, Course notes, University of Lille 3, August 31, 2009.
19. RAKOTOMAMONJY Alain and GASSO Gilles, Wide Margin Separators, INSA Rouen-ASI Department, LITIS Laboratory, November 11, 2014
20. REHAN A. et al, Applying Support Vector Machines to Imbalanced Datasets, Springer-Verlag Berlin Heidelberg, 2004
21. RUNG-CHING Chen and Kai-Fan Cheng, Using Rough Set and Support VectorMachine for Network Intrusion Detection System, First Asian Conference on Intelligent information and Database Systems, 2009
22. SONGLUN Zhao, "Intrusion detection using Support Vector Machine enhanced with a feature-weight kernel", Computer Science, University of Regina, 2007
23. VAPNIK, VN: The Nature of Statistical Learning Theory, Springer, New York, 1995, 188p.
24. VEROPOULOS, C. CAMPBELL, N. CRISTIANINI, Controlling the Sensitivit y of Supp ort Vector Machines Department of Engineering Mathematics, Bristol Universit y, Bristol BS8 1TR, United King, 1999
25. WANG, L.: Support Vector Machines: Theory and Applications, Springer, New York, 2005, 431 p