Volume 2, No. 3, May-June 2011
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

A novel approach for intrusion detection using KNN classification and DS-Theory

Deepika Dave*
Lakshmi Narayan College of Technology,
Bhopal, Madhya Pradesh
India
deepikadave.mds@gmail.com

Dr. RK Pandey
UIT , BU
Bhopal, Madhya Pradesh
India
rk_pandey@yahoo.com

Prof. Vineet Richhariya
Lakshmi Narayan College of Technology
Bhopal, Madhya Pradesh
India
vineet_rich@yahoo.com

Abstract:Intrusion detection is a very challenging area of research in a current scenario. Now every day find a new pattern of intrusion and detection of this pattern are very challenging job. In this paper we have discuss a novel approach for intrusion detection using KNN classification and Dempester theory of evidence. Through these approaches gathered a new discovered pattern of intrusion and classify Category of pattern and apply event evidence logic with the help of DS- Theory. Finned pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of pattern of intrusion detection and improved the true rate of intrusion detection. we have also perform some experimental task with KDD99Cup and DARPA98 databases from MIT Lincoln Laboratory show that the proposed method provides competitively high detection rates compared with other machine-learning techniques and crisp data mining
Keywords:-intrusion Detection, KNN, DS-Theory

I. INTRODUCTION

With growing the usage of network the safe guarding of security has been come as challenges for user with malicious attacks. For these types of challenges IDS(Intrusion Detection System) are implemented ,these detect the intrusion as it occur in network or our system. An intrusion is break or misuses the systems. Here IDS detect an intruder those breaking a system or whoever those misusing our system resources.IDS identify suspicious patterns that may be indicate a attack from some one attempting to break in to a compromise a system. A set of malicious action identify by NIDS (Network Intrusion Detection System) that threaten the intrigity, confidentiality and availability of network resources. Traditionally Intrusion detection is categorized into main two category :(i) Misuse Detection and (ii) Anomaly Detection. (i) Misuse Detection: Searches a specific patterns or user behavior that matches a known intrusion scenarios. (ii) Anomaly Detection: For a normal network of behavior a new model is develop an anomaly detection it's also detect a new intrusion by evaluating a significant deviation from the normal behavior. In this paper we proposed the dempster theory, this work on event evidence and find the validity of data and reduce the rate of intrusion. Here we also proposed a class association rule mining, this rule mining is used to discover association rules in data set on set of attributes, the relationship between a dataset in association rule
mining is expressed by X=>Y,where X and Y both are set of attributes. This means that if a tuple satisfy the X, it is also satisfy tuple Y. The rest of this paper is organized as follows. In section II Section III, some related works are reviewed.Section IV deals with KNN classifier.section V Overview of Dempester Theory. Section V Ionclusion.

II.CLASSIFICATION METHOD BY FUZZY GNP – BASED CLASS ASSOCIATION RULES

Ci Chen * ,Shingo Mabu* Chuan Yue [1] devise in the filled of intrustion detection with the approach As the fuzzy GNP based class association approach is designed for databases containig both discrete and continuous attribute as Network Connection Database,secific classification method is describe as a follows: The defination of the matching degree between the continous attribute A_i in rule r with q_i and testing data connection with value ai is:

$$\text{MatchDegree}(q_i, a_i) = Fq_i(a_i) \quad (1)$$

Where, Fq_i represent the membership function for linguistic term q_i.

And the matching between rule r (p continuos and q discrete attributes) and new unlabeled connection d is defined as:

$$\text{Match}_r(d) = \frac{1}{p+q} \left(\sum_{i \in Ap} \text{MatchDegree}(q_i, a_i) + t \right). \quad (2)$$

where.
i: index of continuous attribute in rule r;
Ap: set of suffixes of continuous attribute in rule r ;
p: number of continuous attribute in rule r;
q:number of discrete attribute in r;
t: :number of discrete attribute in new unlabeled connection d satisfying rulr r;
$\text{Match}_r(d)$ ranges from 0 to 1. If $\text{Match}_r(d)$ equals to 1.0, rule r matches coonection data d competly. While $\text{Match}_r(d)$ equals to 0, rule r does not matche connection d at all.Then the average matching between connection data d and all the rules in a certain rule pool is defined as:

$$\text{MATCHr}(d) = \frac{1}{|Rp|} \sum_{r \in Rp} \text{Matchr}(d) \quad (3)$$

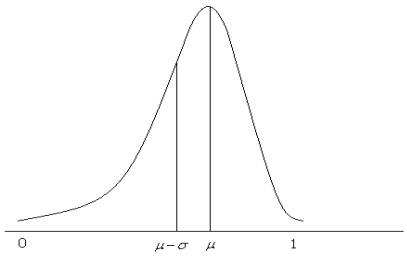
Where R_p is the set of suffixes of extrated important class association rule in a cetain rules pool.

A. Classifier for misuse detection

The average matching betwen connection data d and all the rules in the normal rule
in pool $MATCH_n(d)$ and the avearage matching between connection data d and all the rules in the intrustion rule pool $MATCH_i(d)$ are calculated and compared.
If $MATCH_r(d) \geq MATCH_r(d)$,connection data d is labaeld as normal. On the other hand if $MATCH_n(d) < MATCH_i(d)$ connection data d is labaeld as intrustion.
In summary, a new connection data is labeled according to their matching with normal and intrustion rule pools.Larger matching suggests the heigher possibilty of belonging to this class.

B. Classifier for anomaly detection

After getting matching between each connection data and rules in the normal rule pool. We can have the distribution of the matching with the mean value μ and standard deviation σ. Fig showes one example of the distribution.
In this testing peroid ,when a new unlabled connection data comes ,the matching between the data and the rules in normal rule pool is calculated. If $MATCH_n(d) < (\mu - k\sigma)$,label the connection as intrustion. On the hand,if $MATCH_n(d) \geq (\mu - k\sigma)$, label is normal. By adjusting parameter k, we can balance the PFR (Positive False Rate) and NFR(Negative False Rate).
In all, by using the improvrd Fuzzy GNP –based class association rule mining . we can find a large number of rules related to normal behaviour so as to explore the space of the normal connections. And any significant deviation from the normal space is viewed as an intrustion.



III. PROBABILISTIC CLASSIFICATION

Nannan Lu, Shingo Mabu, Wenjing LI [2] devise in the filled of intrustion detection with the Nannan Lu, Shingo Mabu, Wenjing LI [2] devise in the filled of intrustion detection with the approach as: After extacting a number of important class association rules including normal and intrustion, a classifier is constructed to classify new connection data into normal ,misuse and anomaly intrusion correctly. The key points probabilsitc classification concerns three aspects.First , the probabilty density function of the avrage matching degree of data with rules is used .Second, the probabilty that data is classified to anomaly intrustion also considerd.Third ,in order to improve the classification accuracy,weights are used to revise the probabiltyapproach as: After extacting a number of important class association rules including normal and intrustion, a classifier is constructed to classify new connection data into normal ,misuse and anomaly intrusion correctly. The key points probabilsitc classification concerns three aspects.First , the probabilty density function of the avrage matching degree of data with rules is used .Second, the probabilty that data is classified to anomaly intrustion also considerd.Third ,in order to improve the classification accuracy.

$$MatchDegree_k(Q_i, a_i) = F_{Qi}(a_i)$$

Where F_{Qi} represents the membership function of linguistic term Q_i. Then, the matching degree between data and rule r (including p continuous attributes and q discrete attributes) is defined as:

$$Match_k(d, r) = \frac{1}{p+q} \left(\sum_{i \in CA} MatchDegree_k(Q_i, a_i) + t \right), \quad (5)$$

Where, I is the suffix of continuous attributes in rule r; CA denotes the set of suffix of continuous attributes in rule r; p and q represent the number of continuous attribute and discrete attributes in rule r, respectively, and t is the number of matched discrete attributes in rule r with data. Then, the average matching degree can be defined as

$$m_k(d) = \frac{1}{|Rk|} \left(\sum_{k \in C} Match_k(d, r), \right. \quad (6)$$

where, R_k is the set of suffixes of the extracted rules in class k in the rule pool(normal rules or misuse rules). Finally, the marginal probability density function $f_1(x_1), f_2(x_2), \ldots f_K(x_k)$ can be generated by calculating the distribution of the average matching degree of training data d ε Dtrain(k) with r ε R_k, where, Dtrain(k) is the set of suffix of training data in class k. K=2 is used in this paper.

A. Building a Classifier

After creating the probability density function $f_K(x_k)$ of the average matching degree between training data d $\varepsilon D_{train}(k)$ and rule r ε R_k, the probability that new connection data d εD_{test} belongs to class k is represented as follow:

$$P_k(d) = \int_{mK(d)}^{1.0} f_K(x_K)dx_k \ldots \sum_{k \in C} f_K(x_K)dx_k \ldots \int_{m1(d)}^{1.0} f_1(x_1)dx_1, \quad (7)$$

where, Dtest is the set of suffix of testing data. Actually, the probability that d Sigma Dtest belongs to anomaly class is defined as:

$$P_0(d) = \sum_{k \in C} 1 - P_k(d) \quad (8)$$

Where, C is the set of suffix of classes having training data. In the case of two classes, the probabilities of the first class and the second class can be calculated by the following equations.

$$P_1(d) = \int_{m2(d)}^{1.0} f_2(x_2)dx_2 \int_{0}^{m1(d)} f_1(x_1)dx_1 \quad (9)$$

$$P_2(d) = \int_{0}^{m2(d)} f_2(x_2)dx2 \int_{m1(d)}^{1.0} f_1(x_1)dx_1 \quad (10)$$

Then, the probability that a new connection data belongs to anomaly class is calculated by $P_0(d) = 1 - \sum_{k \in C} P_k(d)$. Based on the calculation of these probabilities, d is assigned to the class with highest probability.

B. Revision of Probability

In intrusion detection, in order to balance positive false rate (PFR) negative false rate (NFR),we assigned the following weights to adjust $P_k(d)$ and $P_0(d)$

$$P_k(d) \leftarrow W_k P_k(d) / W_0 P_0(d) + \Sigma\, k\varepsilon C\, W_k P_k(d) \quad (11)$$

Where,W0 and Wk(Kε C) are the weight parameters

IV PROBABILISTIC CLASSIFICATION

KNN is a *non parametric lazy learning* algorithm. That is a pretty concise statement. When you say a technique is non parametric , it means that it does not make any assumptions on the underlying data distribution. This is pretty useful , as in the real world , most of the practical data does not obey the typical theoretical assumptions made (eg gaussian mixtures, linearly separable etc) . Non parametric algorithms like KNN come to the rescue here.

It is also a lazy algorithm. What this means is that it does not use the training data points to do any *generalization*. In other words, there is *no explicit training phase* or it is very minimal. This means the training phase is pretty fast . Lack of generalization means that KNN keeps all the training data. More exactly, all the training data is needed during the testing phase. (Well this is an exaggeration, but not far from truth). This is in contrast to other techniques like SVM where you can discard all non support vectors without any problem. Most of the lazy algorithms – especially KNN – makes decision based on the entire training data set (in the best case a subset of them).

There are various methods which can be used to determine nearest neighbour. Figure 3 shows the way in which decision is taken to decide the category of new point.

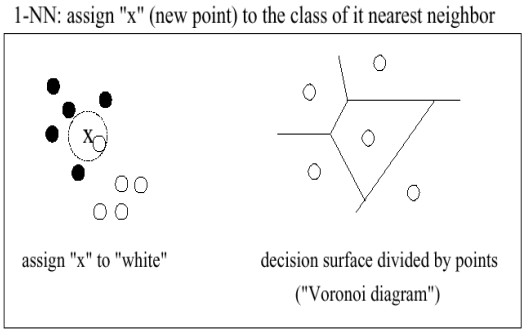


Figure.1 Decision of nearest neighbour

Figure 4 and 5 shows various methods for deciding the nearest neighbor.

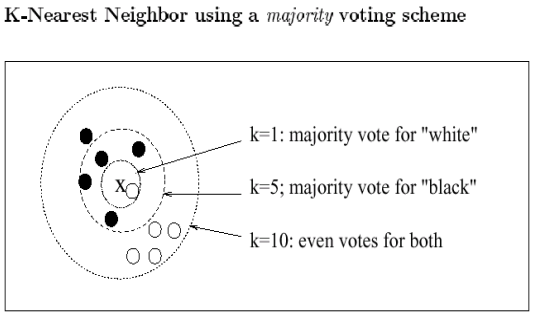


Figure.2 Majority voting scheme

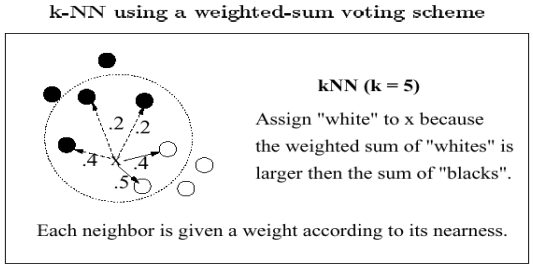


Figure.3 Weighted-sum voting scheme

k-NN is a kind of example-based text categorization algorithm. However,the determination of the *k* has not yet got good solution. Moreover, the good selection of *k* most similar texts also has bigger effect on categorization results. Also k-NN cannot effectively solve the problem overlapped category borders.

Statistical rules are used in general in the classi.cation of textual information, which include several tasks in Information Retrieval. It includes not only the determination of good documents in terms of relevance attending to user needs but also the classi.cation of documents into categories (topics) attending to prede.ned classes [18]. In the following, we include studies found in the literature about both the retrieval and the categorization tasks.

The use of rules for categorization comes from a process of classification of documents into different categories regarding their topics in order to optimize a posteriori retrieval process. One of the most relevant works of categorization using rules is the one of [20]. The general idea of this work is the discovery of classification patterns automatically for document categorization. The aim of the induction process is to .nd sets of decision rules to distinguish among di.erent categories which documents belong to. The attributes of the rules can be one word or a pair of words constructing a dictionary where an elimination process of the less frequent words is carried out. Finally, association rules have been also used for categorization [21], where the authors propose a solution for text categorization based on the application of the best generated association rules to build a classifier.

V.APPROACH

The Dempster –Shefer theory(DST) of evidence originated in the work of [3,4]on theory of probabilities with upper and lower bounds. It has since been extended by numerous

authors and popularized, but only to a degree, in the literature on Artificial Intelligence (AI) and expert systems, as a technique for modeling reasoning under uncertainty. In this respect it can be seen to offer numerous advantages over the more "traditional" methods of Statistics and Bayesian decision theory. Hajek [5] remarked that real, practical applications of DST methods have been rare, but subsequent to these remarks there has been a marked increase in the applications incorporating the use of DST. Although DST is not in widespread use, it has been applied with some success to such topics as face recognition [6], statistical classification [7] and target identification [8]. Additional applications centered on multi-source information, including medical diagnosis [9] and plan recognition [10]. An exception is the paper by cortes –Rello and Golshani [11], which although written for a computing science /AI readership does deal with the "knowledge domain" of forecasting and Marketing Planning. For those with even limited knowledge og these domains the paper appears rather naive , referring for example to rather naive. Referring for example to rather venerable old editions of standard texts such as[12].the aim of this paper is to suggest that there is a good deal of potential in the DST approach, which is as yet very largely unexploited. The origins of the mathematical theory of probability date back at least to the work of the eighteenth century scholar, The Reversed Thomas [13],whose work was published posthumously in 1763.it provides the foundations for the theory of statistical inference (involving both estimation and testing of hypotheses) and for techniques of design making under certainty. The roots of decision analysis lie in the 1930s and 1940s .Wald[14], included the "complete class theorem" ,which stated that any procedure in a statistical decision problem can be beaten or at least matched in performance by Bayesian procedure, defined as procedure based on the adoption of some set of prior probabilities . The fact that numerous statistical principles and techniques may be developed without using prior and posterior probability distribution involves no loss of generality, given that the special case of a uniform or rectangular prior distribution may be adopted. Decision analysis relies more on a subjectivist view of the use of probability , whereby the probability of an event indicates the degree to which someone believes it, rather than the alternative frequents approach .The latter approach is based only on the number of times an event is observed to occur .As savage [15,16] discusses ,the subjectivists have been responsible for much of the theoretical work into statically practice. He goes on to argue that the frequentists hold an un easy upper hand over their Bayesian / subjective colleagues in the domain of mathematical statistics. Bayesian statisticians may agree that their goal is to estimate objective probabilities from frequency data, but they advocate using subjective prior probabilities to improve the estimates [17] . french questions savages's theriry of subjective ezpected utility,which suggests that each of us has within us an exact subjective probability for each possible event in the small world (model) under consideration. For a much fuller discussion of subjective and frequentist approaches see the collection of papers in [18] who notes that the three defining attributes of the Bayesian approach are:

1.Reliance on a complete probabilistic model of the domain or "frame of discernment".

2.Willingness to aaccept subjective judgement as an expedient substitute for empirical data.

3.the use of Bayes Therom (conditionality) the primarly mechanism for updating belifes in ,light of new information. However,The Bayesian technique is not without its critics including among others Walleyl[19],as well as Caselton ans Luo[20] who discussed the difficulty arising when conventional Bayesian analysis is presented only with weak information sources. In such cases we have the "Bayesian domega of precision ",whereby the information concerning uncertain statistical parameters, no matter how vague, must be represented by conventional exactly specified ,probability distribution.

Some of the difficulties can be understood through the "principle of Insufficient Reason" as illustrated by Wilson [21].Suppose we are given a random device that randomly generates integer numbers between 1 and 6(its "frame of discernment") but with unknown chances. What is our belief in"1" being the next number? A Bayesian will use a symmetry argument, or the Principle of insufficient Reason to say that the Bayesian belief in a "1" being the next number, say P(1) should be 1/6. In general in a situation of ignorance a Bayesian is force to use this principle to evenly allocate subjective (additive) probabilities over the frame of discernment. To further understand the Bayesian approach, especially with the regard to representation of ignorance, consider the following example, similar to that in [21]. Let a be a preposition that; "I live in Kings Road, Cardiff".

How could one construct P(a), a Bayesian belief in a? Firstly we must choose a frame of discernment, denoted by Θ and a subset A of Θ representing the preposition a; then would need to use the Principle of Insufficient Reason to arrive at a Bayesian belief. The problem is there are number of possible frames of discernment Θ that we could choose, depending effectively on how many Cardiff roads can be enumerated. If only two such streams are identifiable, then $\Theta=\{x_1,x_2\}$,$A=\{x_1\}$.The "Principle of Insufficient Reason" then gives P(a), to be 0.5,through evenly allocating subjective probabilities over the frame of discernment. If it is estimated that there are about 1000 roads in Cardiff, then $\Theta=\{x_1,x_2,\ldots\ldots x_{1000}\}$ with again $A=\{x_i\}$ and other x_i 's representing the other roads. In this case the "theory of insufficient reason" gives P(A)=0.001. Either of these frames may be reasonable, but the probability assigned to A is crucially dependent upon the frame chosen. Hence once Bayesian belief is a function not only of the information given and one's background knowledge, but also of sometimes arbitrary choice of frame of discernment. To put the point another way, we need to distinguish between uncertainty and ignorance. Similar arguments hold where we are discussing not probabilities per se but weights which measure subjective assessments of relative importance. This issue arises in decision support models such as the Analytic

Hierarchy Process (AHP), which requires that certain weights on a given level of decision tree to unity see [22]

VI.CONCLUSION

A dempester evidence theory method is discussed in this paper, and the method is used to intrusion detection. Which solve the problem that traditional technique of intrusion detection, these techniques are not finding a new pattern of intrusion. And experiments prove that the method has the property of high classification accuracy. In the future we have implemented and simulated our proposed method and compare its result through existing method.

VII. REFERENCES

[I] JiaWei Han,Micheline Karnber, "Data Mining: Concept and Technology" [M]. China Machine Press, 2001.8

[2] Freund Y. "Boosting a Weak Learning Algorithm by Majority"[J]. Information and Computation, 1995,121(2):256-285

[3]The Dempster Shafer theory of evidence: an alternative approach to multicriteria decision modelling Malcolm Beynon, Bruce Curry*, Peter Morgan Cardi Business School, Colum Drive, Cardi, CF1 3EU, UK Received 1 December 1998; accepted 1 June 1999.

[4]Dempster AP. Upper and lower probabilities induced by a muilti-valued mapping. Ann Math Stat 1967;38:325-39.

[5] Hajek P. Systems of conditional beliefs in Dempster Shafer theory and expert systems. Int J General Systems 1994;22:113-24.

[6] Ip HHS, Ng JMC. Human face recognition using Dempster±Shafer theory. In: ICIP. 1st International Conference on Image Processing, vol. 2, 1994. p. 292-5.

[7] Denoeux T. A k-nearest neighbour classi®cation rule based on Dempster±Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 1995;25(5):804-13.

[8] Buede DM, Girardi P. A target identi®cation comparison of Bayesian and Dempster±Shafer multisensor fusion. IEEE Transaction on Systems, Man and Cybernetics- Part A: Systems and Humans 1997;27(5):569-77.

[9] Yen J. GERTIS: A Dempster±Shafer approach to diag-nosing hierarchical hypotheses. Commun ACM1989;32(5):573-85.

[10] Bauer M. A Dempster±Shafer approach to modeling agent preferences for plan recognition. User Modeling and User-Adapted Interaction 1996;5:317-48.

[11]Cortes-Rello E, Golshani F. Uncertain reasoning using the Dempster-Shafer method: an application in forecasting and marketing management. Expert Systems 1990;7(1):9-17.

[12] Kotler P. Marketing management: analysis, planning and control. Englewood Cliffs, NJ: Prentice Hall, 1980.

[13].Bayes T. An essay toward solving a problem in the doc- trine of chances. Phil Trans Roy Soc (London) 1763;53:370-418.

[14]. Wald A. Statistical decision functions. New York: Wiley, 1950.

[15] Savage LJ. The foundations of statistics. New York:Wiley, 1954 (2nd rev.ed., 1972 Dover).

[16] Savage LJ. The foundation of statistics reconsidered. In: Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability 1. Berkeley: University of California Press, 1961.

[17] Good IJ. Good thinking: the foundations of probability and its applications. Minneapolis: University of Minnesota Press, 1983.

[18]Shafer G, Pearl J. Readings in uncertain reasoning. San Mateo, CA: Morgan Kaufman, 1990.

[19] Walley P. Belief-function representations of statistical evidence. Ann Stat 1987;10:741-61.

[20] Caselton WF, Luo W. Decision making with imprecise probabilities: Dempster±Shafer theory and applications. Water Resources Research 1992;28(12):3071-83.

[21] Wilson PN. Some theoretical aspects of the Dempster± Shafer theory. PhD Thesis, Oxford Polytechnic, 1992.

[22] Saaty TL. The Analytic Hierarchy Process: planning, pri-ority setting, resource allocation. New York: McGraw- Hill, 1980.