퍼지 및 지능시스템학회 논문지 2001, Vol. 11, No. 7, pp. 633-640

Genetic Algorithm Application to Machine Learning

Myung-mook Han* and Yill-byung Lee**

* Department of Computer Science, Kyungwon University

** Department of Computer Science, Yonsei University

Abstract

In this paper we examine the machine learning issues raised by the domain of the Intrusion Detection Systems(IDS), which have difficulty successfully classifying intruders. These systems also require a significant amount of computational overhead making it difficult to create robust real-time IDS.

Machine learning techniques can reduce the human effort required to build these systems and can improve their performance. Genetic algorithms are used to improve the performance of search problems, while data mining has been used for data analysis.

Data Mining is the exploration and analysis of large quantities of data to discover meaningful patterns and rules. Among the tasks for data mining, we concentrate the classification task. Since classification is the basic element of humans way of thinking, it is a well-studied problem in a wide variety of application. In this paper, we propose a classifier system based on genetic algorithm, and the proposed system is evaluated by applying it to IDS problem related to classification task in data mining. We report our experiments in using these methods on KDD audit data.

Key words : Genetic Algorithms, Machine Learning, Data Mining, Intrusion Detection System

1. Introduction

A long-standing problem in the field of computer security is that of intrusion detection[1]. The purpose is to detect violations of security policy for a computer system. Of the many possible approaches to intrusion detection, there are mainly two types of intrusion detection techniques. Misuse detection uses patterns of well-known attacks or weak spots of the system to match and identify intrusions. Anomaly detection tries to determine whether deviation from the established normal usage patterns can be identified as intrusion. Detecting anomalous behavior can be regards as a binary valued classification in which measurements of system activity are used to produce a classification of the state of the system as normal or abnormal.

Data Mining(DM), also called knowledge discovery in databases, can be defined as efficiently discovering interesting patterns from large databases, and has been emerged as a promising new area for database research[2][3].

The knowledge by which is extracted DM can be applied in a wide range of domains including decision support of the company. DM methods can be grouped into classification, clustering, summarization. Classification has been identified as an important problem in the emerging field of DM. While classification is a well-studied problem, only recently has there been focus on algorithms that can handle large databases. In classification, we are given a set of example records, called a training set, where each record consists of several fields or attributes. Attributes are either continuous, coming from an ordered domain, or categorical, coming from an unordered domain. One of the attributes, called the classifying attribute, indicates the class to which each example belongs. The objective of classification is to build a model of the classifying attribute based upon the other attributes.

Genetic Algorithm(GA) [4][5] has been spotlighted as discovering the solution to solve the combinatorial optimization problems. GA is the search technique of a solution to imitate a process of biological evolution, in which the population of the plural coded gene to represent a candidate for the solution evolve gradually into the thing to be an object of optimization through the process of the change of generations and the selection. This GA has an advantage in searching efficiently the wide scope of a state space in order to explore the solution from lots of initial states.

Our research aims to develop an automated approach in building IDS. We are developing tools that can be applied to a variety data sources to generate intrusion detection models. That is, our approach is to apply data mining methods, based on GA, to the audit data to compute models that accurately capture the patterns of intrusions and normal activities.

The rest of the paper is organized as follows. In Sec.2, we introduce IDS, DM, and GA. In Sec.3, we discuss the simulation in IDS domain. Section 4 concludes the paper with a summary.

This paper was supported by a grant from the Post-Doc Program of Korea Research Foundation in 1997.

2. Intrusion Detection System and Data Mining

2.1 Intrusion Detection System(IDS)

An intrusion can be defined as any set of actions that attempt to compromise the integrity, confidentiality or availability of a resource. An intrusion attempts to subvert the security policy in force on a system in order to gain access to information, to alter the systems operation, or to deny the system to other users. Intrusions are enacted by users a user runs a program which executes actions in excess of that users authorization, for example. This would be classified as an intrusion. Similarly, a user who attempted to overload a system and prevent other users from working would also have initiated an intrusion. It is noted that this is still all interpreted relative to a security policy.

To detect and report intrusions, a system must have some form of intrusion detection system(IDS) installed. An IDS must identify, preferably in real time, unauthorized use, misuse, and abuse of computer systems.

As this is a reactive form of defense, it will not stop an intrusion from occurring. But without an IDS in place, the only evidence that an intrusion has occurred may be the disappearance of valuable files, or when the system halts. A human operator cannot be reasonably expected to filter megabytes of audit logs daily to detect potential intruders.

2.2 Data Mining(DM)

DM or Knowledge Discovery is the process which extracts automatically hidden knowledge from the large amount of data base. That is, DM can be explained as the extracting the hidden and predictive information from the large scale of data base. There are several processes to extract the useful information or knowledge from a raw and large scale data.

First of all, selection process selects or divides the data according to the certain category. Preprocessing process purifies the data and transformation process transform the data to use or search. After the patterns, a set of facts, are extracted from the data by the DM process. We finally interpret from identified patterns to the knowledge to be used the decision support in interpretation and evaluation process.

The quantity information and data to be stored in electronic structure have increased tremendously in last 20 years. The accumulation of data has been happened in explosive rate.

The quantity of information has increased two times in every 20 months in a world, the size and number of data base have increased in very fast rate. The increase of data is affected by the development of scientific data collection. There are several elements to be led the DM methods to the front line of business decision support.

These are included: the value not to be used in large scale data base, the enforcement of data base to aim a viewpoint of single customer, the concept of information or data warehouse from the data base integration, the decrease of the hardwares price, and deepening of competition in market place.

The name of DM has originated with the similarity between searching the valuable data of a large data base and mining the mountain to strike a valuable vein. The DM methods from the data base of a sufficient size and good in quality provide a following ability to create a new business competition power. That is, it discovers automatically a trend, an automatic forecast of action and a configuration to be hidden.

The DM methods have come out from the result of product development and of a long research process. These developments begin in storing the business data to computer initially, and the approach method to data has been developed continuously.

Recently it makes the technique that an user can search their data perfectly in real-time. DM go through the development process that the future data and the preceding information are came from the approach on past data and search. DM can supply the good application by a sufficiently mature techniques. That is, the collection of a vast data, the strong multiprocessor computer, and DM algorithms. Figure 1 shows that DM appears from the fields such as inductive learning, machine learning and statistics.

Inductive learning infers the information from the data, and is a model which constructs a process at the analysis of data base. Since it is able to predict the classes to be unseen objects, the similar objects are grouped to classes and the rules become formalize. Therefore induction is the inference of patterns. The quality of model to be made by the inductive learning method depends on the using that the model forecast the output of future situation. The problem is that most of environments have different states.

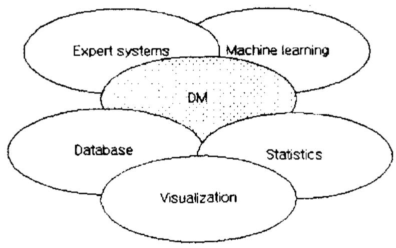

Fig. 1. The relation between DM and other fields

Since statistics has a firm and theoretical foundation, the result of statistics can not deny. But it is different to interpret a guide to user in regard to where and how the data analyses. DM is able to use the knowledge concerning experts data and advanced analysis technique of computer.

Machine learning is an automation of learning process and learning is to build rules based on the observation of transition and environmental states. Machine learning is a wide scope which includes learning by example, learning enforcement, and learning by supervisor. Learning algorithm inputs the set of data and information, and outputs the concepts to represent the result of learning. Machine learning examines the previous example and its result, and learns how to reproduct and how to generalize according to new case.

Data mining methods are grouped by the application classes to be used or the function to be implemented, and the representative function are classification, estimate, forecast, grouping, c-cluster, and description. Data mining uses a various techniques. Generally the more techniques use, the more correct results appear. This means that if one technique cannot find a valuable thing, another technique can find it.

The representative techniques are included in market basket analysis, memory based reasoning, cluster detection, link analysis.

2.3 Applying Genetic Algorithms to Data Mining

Genetic Algorithms(GAs) are efficient and independent search method and have been used to learning classifier rules[6]. Also GAs are applied to concept learning[7], feature selection[8], adjust of parameter[9], and construction of feature[10]. The fundamental process of GAs is shown as Figure 2.

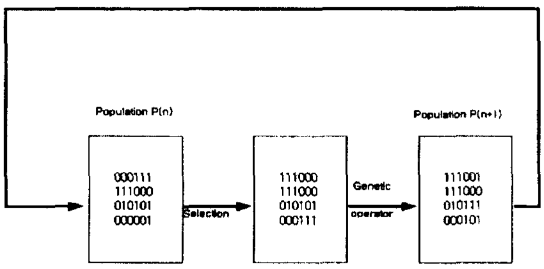

Fig. 2. The example of GA

The most of data mining system uses the modification of traditional machine learning algorithm. In machine learning, there are two purposes that it learns the complex system and makes the appropriate output of system. The machine learning based on genetic algorithm is called as GA machine learning or GBML(genetic based machine learning).

The aspect that machine learning method is basically different from the optimization problem is to seek a set of rules. Since the purpose of optimization problem is to seek the optimal solution, it is enough for one thing to converge the individual. But machine learning is not to seek a best rule, but to seek a set of rules cooperating each other. Generally, there are two approaches in GBML. The whole rule set is represented as an individual, and an individual of population is a rule set of candidate. And then it is natural that a new generation of a rule set is created by selection and genetic operator. That is, a traditional genetic algorithm is used and each entity in a population is a rule set which represents a complete solution of learning problem. This is called Pitt approach method[11].

At the similar time, Holland has developed a classifier system that an individual of population is a rule and a rule set is represented by a population. This method is called Michigan approach method[12]. The Michigan approach method has used a quite different evolution method, that a population consists individual rules and each rule represents a candidate solution of an overall learning task. That is, Pitt approach method is similar to evolutionary computation but Michigan approach method is to use a quite different method. Figure 3 shows the Pitt and Michigan approach.

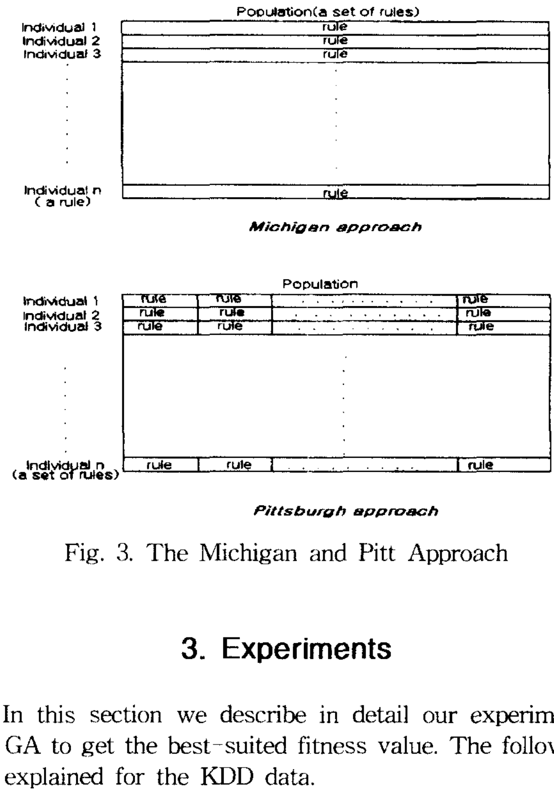

Fig. 3. The Michigan and Pitt Approach

3. Experiments

In this section we describe in detail our experiments in GA to get the best-suited fitness value. The following is explained for the KDD data.

3.1 Problem Domain for IDS

In this experiments, we used the following KDD data. This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99. The Fifth International Conference on Knowledge Discovery and Data Mining. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between bad connections, called intrusions or attacks, and good normal connections. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment [13].

The KDD data consists of information and Data files.

Information files:

Task description. This is the original task decription given to competition participants.

Data files:

kddcup.names : A list of features.
kddcup.data.gz : The full data set.
kddcup.data_10_percent.gz : A 10% subset of kddcup.data.gz.
kddcup.newtestdata_10_percent_unlabeled.gz : A 10% subset of kddcup.testdata.unlabeled.gz.
kddcup.testdata.unlabeled.gz : The test data.
kddcup.testdata.unlabeled_10_percent.gz : A 10% subset of kddcup.testdata.unlabeled.gz.
corrected.gz : Test data with corrected labels.
training_attack_types : A list of intrusion types.

Table1. Define the features by the binary

Name	Translation Value
Udp	001
Tcp	010
lcmp	011
Private	100
Domain u	101
http	110
Smtp	111
ftp data	1000
Auth	1001
SF	1010
Other	1100
Eco I	1101
Ecr I	1110
trlnet	1111

In above data, we select the 'corrected.gz'. A reason of using the data is the following.

▶ described data is certain data in all each individual which is separated with normal and abnormal.
▶ In this experiments, we must use the binary 0 and 1, that is, it is suitable data under pre-processing because of read the bit-value and execute computation.

The above table 1 defines the each case for data pre-processing.

Therefore, KDD data are preprocessed by the above method. We can the data pre-processing to KDD.

We must translate to bit type for using the Genetic Algorithm from any data. For bit value and symbol value take the length such thats attributes, flag the 1 to the position of relative value. But, they separate same length because they are continuous value. Simply, we can separate the data by n-bit or frequency of the data. The other side, we can use the fuzzy technique for the task. In this experiments, we used the first method. The data consist of 60 individuals that described with 43 attributes. The following is the used data define by the table1.

A1 : back,buffer_overflow,ftp_write,guess_passwd, imap,ipsweep,land,loadmodule,multihop,neptune,nmap,normal,perl,phf,pod,portsweep,rootkit,satan,smurf,spy,teardrop,warezclient,warezmaster.
A2 : duration: continuous.
A3 : protocol_type: symbolic.
A4 : service: symbolic.
A5 : flag: symbolic.
A6 : src_bytes: continuous.
A7 : dst_bytes: continuous.
A8 : land: symbolic.
A9 : wrong_fragment: continuous.
A10 : urgent: continuous.
A11 : hot: continuous.
A12 : num_failed_logins: continuous.
A13 : logged_in: symbolic.
A14 : mpromised: continuous.
A15 : root_shell: continuous.
A16 : su_attempted: continuous.
A17 : num_root: continuous.
A18 : num_file_creations: continuous.
A19 : num_shells: continuous.
A20 : num_access_files: continuous.
A21 : num_outbound_cmds: continuous.
A22 : is_host_login: symbolic.
A23 : is_guest_login: symbolic.
A24 : count: continuous.
A25 : srv_count: continuous.
A26 : serror_rate: continuous.
A27 : srv_serror_rate: continuous.
A28 : rerror_rate: continuous.
A29 : srv_rerror_rate: continuous.
A30 : same_srv_rate: continuous.
A31 : diff_srv_rate: continuous.
A32 : srv_diff_host_rate: continuous.
A33 : dst_host_count: continuous.
A34 : dst_host_srv_count: continuous.
A35 : dst_host_same_srv_rate: continuous.
A36 : dst_host_diff_srv_rate: continuous.
A37 : dst_host_same_src_port_rate: continuous.
A38 : dst_host_srv_diff_host_rate: continuous.
A39 : dst_host_serror_rate: continuous.
A40 : dst_host_srv_serror_rate: continuous.
A41 : dst_host_rerror_rate: continuous.
A42 : dst_host_srv_rerror_rate: continuous.

3.2 Expression of search space

The rule type to classify any class is a type with the exception of a class in data bit sequence. In this type, the rule express AND between permit the internal disjunction that define by OR. That is, a chromosome expresses the solution for the concept to be learned.

Assume that cannot express each chromosome is a set of any synthesis and cardinal of the synthesis is any numbers.

The Disjunctive Normal Form (DNF) is the best method, disjunctive set declare cross as much as possible the classification rules, to describe the concepts.

The left part of rules consists of connection of one or more tests that include the feature value. The right part of rules expresses the concept or classification, which is related by the left part. That is, if rules classify exactly to the elements of feature set, then the set of rules regarded as expression of unknown concept. If these rules permit any complex item of connected left then it must get very strength expression length which difficult to express with the string.

However, by reducing the complex degree of the elements of the conjunctions, we can use a string representation and standard GAs, with the only weak point that rules may be required to express concept. That is, we can get the expression with each element, which is reduced by connection of the test form.

For instance, a rule can have the following symbol type.

If (F1=large) and (F2=sphere or cube)
then it is a tool

In an above the rule, it miss (lose) not a generality that must get only one test under each feature, because of the left part is connected form of statement.

We can make the inner expression of fixed length to classify rule with it. In this paper, we use only a feature of the nominal value.

For instance, if the set of F1 feature values is {small, medium, large}, then the 011 would represent the test for F1 being medium or large.

In addition, assume that the feature F2 has the values {sphere, cube, brick, tube} and there are two classes tools and parts. Then the expression is the following for the problem with the two features.

F1	F2	class
111	1000	0

This rule is equal to If (F1=small or medium or large) and (F2=sphere) then it is a tool . It is noted that a feature that involving all is equal to dropping that term.

Therefore, this rule is equal to If (F2=sphere) then it is a tool.

3.3 The set of classification rules

Because the concept description is consist of one or more classification rules, we must describe the method that is used for the evolution of rule set by GAs.

In this paper, we make use of the Pittsburgh approach method and we report the result with it. That is, each element of sets is a string with any length that expresses the set of the fixed length rules. The number of rules under the specific individual can be unrestricted or limited by a user defined upper bound.

For instance,

F1	F2	class	F1	F2	Class
100	1111	0	011	0010	0

This is equal to the following.

If (F1=small) then it is a tool.
or
If ((F1=medium or large) and (F2=brick))
then it is a tool.

3.4 Genetic operator and Fitness function

Genetic operator modifies the elements in a set to generate new elements for measurement and valuation.

In the way with the tradition, the crossover and mutation are the very important operator. The crossover produces the new elements with the two elements by an exchange of the selected data part. The mutation flips random bits in a selected individual with a small probability. One foal was to get a cencept learning representation exploiting the genetic operators.

In this experiments, we were used the traditional method and executed bit-level mutations. The crossover was used to extended of traditional 2-points crossover for control the changed length of rule sets. In the standard crossover, there are no limits on where the crossover point may occur. The only consideration is that the corresponding crossover points on two parents match up semantically. That is, if one parent being cut on the boundary. Then the other parent should be cut on the boundary. So, one parent being cut on 5-bit, the other parents should be cut on the same point.

For instance, let's consider the following two rule sets.

F1	F2	Class	F1	F2	Class
100	**0100**	**0**	**011**	**0010**	**0**
010	0001	0	110	0011	0

I'd like you to pay attention to this fact that left cut point is offset two bits from the rule boundary, while the right cut point is offset one bit from the rule boundary. If we change the bits in the point that is cut, we can get the rule set of three rules, and the rule set of one rule.

F1	F2	Class	F1	F2	Class	F1	F2	Class
100	0001	0	110	0011	**0**	**011**	**0010**	**01**
010	**0100**	0						

After we choose the good expression, it is important to define the fitness function that compensates to individual entity of the correct. In this paper, we chose the fitness function including classification performance only. The degree of fitness of each rule set is calculated by inspecting the rule set in the present set of training example

$$\text{fitness(individual i)} = (\text{percent correct})^2$$

In each generation, all the chromosomes are evaluated by their degree of fitness, and the new individual entity group consists of the better chromosomes. And then, the operators are applied to the new individual entity group, and this process is repeated. In this paper, we express the thing, which are divided it into the preestimated data at the whole data.

$$\left(\frac{Number-of-Correct-Data}{Number-of-Train-Data}\right)^2$$

3.5 The concept learner based on GAs

We explain the system that is used in this paper and the concept learner of a gene algorithm base. The core of this system is GA which searches the space of the rule set for conducting rule in the given sets of positive and negative examples.

```
procedure GA;
            begin
              t = 0;
              initialize population P(t);
              fitness P(t);
              until(done)
               t = t + 1;
               select P(t) from P(t-1);
               crossover P(t);
               mutate P(t);
               fitness P(t);
              end.
```

Fig. 4. System and GA

P(t) expresses the group of rule sets. After we make the group returns to their nature at our discretion, each rule sets is evaluated by the fitness function, which is mentioned before. The rule sets are chosen, under a probability, in proportion to the degree of their fitness. The crossover and the mutation are applied to the surviving rule set, and a new group is formed. This cycle is continued until the perfect rule set is found in the given condition of the time and the space.

3.6 The Results of Experiment

To derive a solution to a problem, the GA creates an initial single population which has 60 individuals in this experiment. The individuals in the single population are selected by a roulette wheel method. That is, the probability that an individual is chosen is proportional to its fitness value. An individual with high fitness value may have several copies in the population. Crossover provides a powerful exploration ability by exchanging parts of two parents. The crossover rate, the probability that each mating pair executes crossover, was set 0.8 and the mutation rate was set 0.5 initially. The Fig.5 is the result of experiment.

The initial average value was 0.141 and the average value was 0.766 until 316 generations. The more the generation advances, the more the fitness value increases. Finally the value became 1 in the 500 generations.

In next experiment, we have tested the variation of crossover rate. That is, the crossover rate was decreased to 0.6 and the mutation rate had the same rate. Since the crossover operator has helped to search the best solution, the experiment has converged in the early generation. The initial value was 0.141 and the value has increased from 0.250 to 0.766 until 200 generations. The final value has shown in the 408 generations. The figure 6 is the result of this experiment.

By the adjustment of crossover rate we were able to obtain the promising individual in the early generations.

It is useful to consider crossover as playing two roles. These can be called the idea and the mechanics of crossover. The idea of crossover is the hope that building blocks from two individuals may be combined into an offspring whose fitness exceeds either parent. The of each of them.

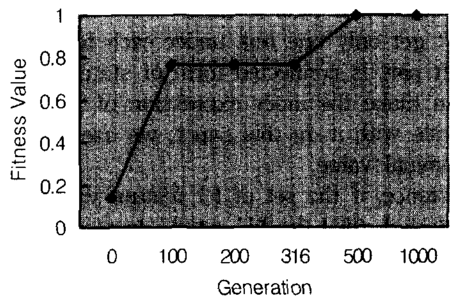

Fig. 5. Result of the first experiment

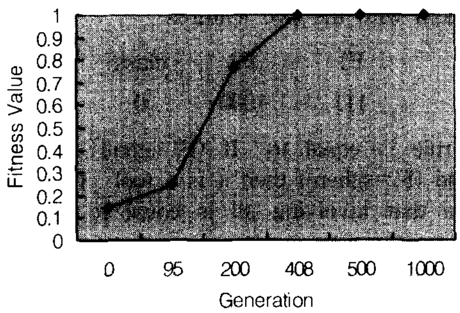

Fig. 6. Result of the crossover variation.

The mechanics of crossover is the process by which an attempt is made to implement this idea. All forms of crossover share a similar idea, but the mechanics vary considerably.

The previous experiments indicate that the higher rate one has converged rapidly in the early generation. It seems, that if there are relatively regular local regions within the fitness function, the higher rate crossover schemes can exploit faster and thus fasten the whole evolutionary optimization.

At the beginning of genetic search, there is a widely random and diverse population and crossover operator tends to perform widespread search for exploring all solution space. As the high fitness solutions develop, the crossover operator provides exploration in the neighborhood of each of them.

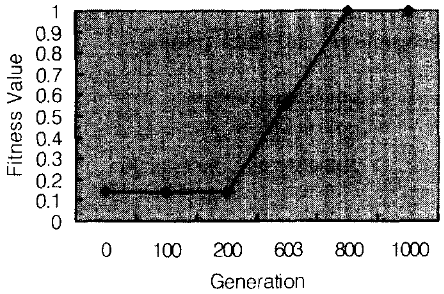

Fig. 7. Result of the mutation variation.

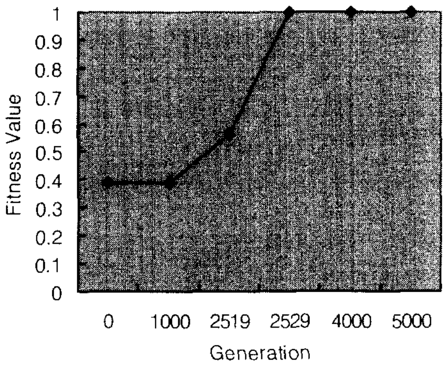

Fig. 8. Result of the mutation variation1

We have evaluated the influence of mutation. The mutation operator is the secondary instrument of variation in GAs and has the properties of local-search. That is, crossover is the major instrument of variation and innovation in GAs, with mutation insuring the population against permanent fixation at any particular locus and thus playing more of a background role.

As shown in Fig.7, the crossover rate was the same rate and the mutation rate was decreased to 0.1. According to the role of mutation, the experiment has converged in 800 generations. Note that mutation operator has the properties of local search techniques. Therefore, after the process has found the local region with crossover operator, mutation operator searches slowly the optimal solution within the region.

Finally, we have decreased to 0.07 of mutation rate. The Fig.8 has shown that the value became 1 in the late generation. Since the mutation operator searches the optimal solution gradually, the solution has been converged in 2529 generations. It noted that, while mutation and crossover have the same ability for disruption of existing schemes, crossover is a more robust constructor of new schema and that the power of mutation has came from a hill-climbing strategy.

4. Conclusion and Future work

Intrusion Detection is an important monitoring technique in computer security aimed at the detection of security breaches that cannot be easily prevented by access and information flow control techniques.

To detect and report intrusions, a system must have some form of intrusion detection system(IDS) installed. An IDS must identify, preferably in real time, unauthorized use, misuse, and abuse of computer systems.

In this paper, we examine the machine learning issues raised by the domain of information security and propose a data mining model for constructing intrusion detection systems. That is, our approach is to apply data mining methods, based on GA, to the audit data to compute models that accurately capture the patterns of intrusions and normal activities. During the process, we apply the GA-based classifier system to KDD data and find an encouraging result.

Our future work includes evaluating the perfect rule set that come out as an experiment result and developing a combined model that incorporate evidence from multiple base models. Also we must extend the association rules and frequent episodes algorithms to the proposed models.

References

[1] Denning, Dorothy E, An Intrusion Detection Model, *IEEE Transaction on Software Engineering*, Feb, 1987.

[2] Michael J.A.Berry and Gordon Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley Computer Publishing, 1997.

[3] Pieter Adriaans and Dolf Zantinge, *Data Mining*, Addison-Wesley, 1996.

[4] D.E.Goldberg, *Genetic Algorihms in Search, optimization, and Machine Learning*, Addison Wesley, 1989.

[5] J.H.Holland, *Adaptation in Natural and Artificial Systems*, Ph.D. thesis, Univ. of Michigan, Ann Arbor, Mich., 1975.

[6] S.Wilson, Classifier Systems and the Animat Problem, *Machine Learning*, vol. 2, 199-228, 1987.

[7] J.Bala, K.A.De Jong, and P.Pachowicz, Learning Noise Tolerant Classification Procedures by Integrating Inductive Learning and Genetic Algorithms , Proc. *First Intl Workshop on Multistrategy Learning*, Harpers Ferry, W.Va., pp.316-323, 1991

[8] F.Z.Brill, D.E.Brown, and W.N.Martin, Fast Genetic Selection of Features for Neural Network Classifiers, *IEEE Trans. Neural Network*, vol. 3, no. 2, pp.324-328, 1992.

[9] M.Botta, A.Giordana, and L.Saitta, Learning Fuzzy Concept Definition , *Proc. Second IEEE Intl Conf. Fuzzy*

Systems, San Francisco, Calif., pp.18-22, 1993.

[10] A.Giordana, G.Lo Bello, and L.Saitta, Abstraction in Propositional Calculus, *proc. Workshop Knowledge Compilation and Speed Up Learning*, Amherst, Mass., pp.56-64, 1994.

[11] Stephen F.Smith, *A Learning System Based on Genetic Adaptive Algorithms*, Ph.D. thesis, Univ. of Pittsburgh, 1980.

[12] John H.Holland, Escaping brittles: the possibilities of general purpose learning algorithms applied to parallel rule-based systems, *Machine Learning, an artificial intelligence approach, 2*. 1986.

[13] The KDD data URL is http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Myung-Mook Han

received the B.E. in ceramic engineering from Yonsei University, Seoul, Korea, in 1980, and the M.S. in computer science from New York Institute of Technology, New York, U.S.A., in 1987. He received Ph.D from Osaka City University, Osaka, Japan, in 1997. From 1997 to 1998 he was at Yonsei University as a Post-Doc. Researcher with a Korea Research Foundation fellowship. He is currently an assistant professor in the Department of computer science of the College of Engineering, Kyungwon University. His research interests include genetic algorithms, machine learning, data mining, and information security.

Phone : +82-031-750-5522
Fax : +82-031-750-5522
E-mail : mmhan@mail.kyungwon.ac.kr

Yill-byung Lee

received the B.S. in Electronic Engineering from Yonsei University, Seoul, Korea, in 1976, and M.S. in Computer Science from University of Illinois at Urbana in 1982. He received Ph.D. in Computer and Information science from University of Massachussetts at Amherst in 1986. Since 1986, he has been a faculty member of the Department of Computer Science at Yonsei University, where he is currently a Professor. His research interests are handwritten character recognition, pattern recognition, and data mining.

Phone : +82-2-2123-2713
Fax : +82-2-365-2579
E-mail : yblee@csai.yonsei.ac.kr