Can an Immuno-fuzzy Approach Detect Only a Few Non-self Cells Existed in an Enormous Amount of Self Cells?

Akira Imada
Brest State Technical University, Belarus

***Abstract*— In the context of Network Intrusion Detection, we test a lately reported technique which generates a set of fuzzy rules to recognize unknown abnormal patterns using a test-function, what we call a-tiny-island-in-a-huge-lake. Our concern is whether or not we can train the system only with a set of already known normal patterns. Yet another of our concern is what happens in an extreme case where a sample of abnormal patterns are extremely few comparing to the normal ones, and what if it eventually shrinks to zero, which is what they call a-needle-in-a-haystack.**

I. Introduction

This paper reports a snapshot of our on-going experiments in which a common target we call *a-tiny-island-in-a-huge-lake* is explored with different methods ranging from a data-mining technique to an artificial immune system. Our implicit interest is a network intrusion detection, and we assume data floating in the *huge lake* are normal while ones found on the *tiny island* are abnormal. Our goal here is twofold. One is to know *whether or not it is possible to train a system using just normal data alone.* The other is to study *a limit of the size of the detectable area*, when we decrease the size of the island eventually shrinking to zero, equivalently so-called *a-needle-in-a-haystack* (See Fig. 1) which is still an open and worth while tackling problem. To learn these two issues, a fuzzy rule extraction system with fixed triangle/trapezoid membership functions are exploited in this paper.

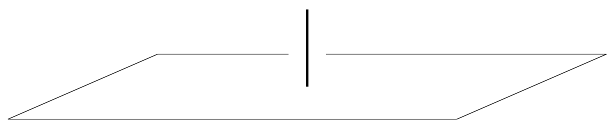

Fig. 1. A fictitious sketch of fitness landscape of *a-needle-in-a-haystack*. The haystack here is drawn as a two-dimensional flat plane of fitness zero.

When we think of a network intrusion detection, we have a large collection of normal patterns while the number of possible anomaly patterns we know is extremely few, which is of usual cases. Then our concern is how we can train a system with a lack of negative training data.

Furthermore, we usually don't know what do anomaly patterns look like in advance. It is usually too late when we know it. Hence, our second concern is whether or not we can train the system with only a set of normal patterns. So far, lots of such approaches have been proposed with very successful results. In other words, those intrusion detection systems train themselves with already known normal patterns and then it succeeds in warning when a so far unknown anomaly pattern is given, though we doubt it more or less.

"A tiny island in a huge lake" — this is a problem we'd proposed in which a few number of unknown patterns should be recognized or classified from an enormous number of known patterns [1], which is up to now still an open issue. Originally the problem described, in particular, a fitness landscape when we searched for a weight configuration which gives a network with fully-connected spiking neurons a function of associative memory. In this paper, we regard the problem from a view point of anomaly detection by an immuno-fuzzy approach. That is, we take it just a general pattern classification problem on the condition, however, that we have two classes one of which has an extremely few patterns while the other has almost infinite number of patterns. Or, we might as well take it a task of discrimination of a few of non-self cells as anomaly patterns from enormous amount of self cells from artificial immune system point of view.

In order to explore this issue, we formalize the problem as follows. We assume we have n-dimensional patterns all of whose real-valued co-ordinates lie in $[-1, 1]$. This construct our self/non-self space. Anomaly patterns have its co-ordinates each of whose value lies in $[-a, a]$, while all others are regarded as normal patters. We can control the difficulty of the task by changing the value of a. In the ultimate case in which the pattern whose co-ordinates are all zero, is called a-needle-in-a-haystack problem.

II. EXPERIMENT

A set of fuzzy rules is used to cover the non-self patterns. As already mentioned, self/non-self sells are represented by n-dimensional real valued vectors each of whose coordinate lies in $[-1, 1]$. That is, the self/non-self space is $[-1, 1]^n$, and a self/non-self pattern is represented by a vector $(x_1, \cdots, x_n)$ where $x_i \in [-1, 1]$. Then a fuzzy rule to detect non-self patterns is

$$\textit{If } x_1 \textit{ is } T_1, \cdots, \textit{ and } x_n \textit{ is } T_n \textit{ then } \mathbf{x} \textit{ is non-self}$$

where T_i is a fuzzy linguistic terms which is either one of {*Low, Low-Middle, Middle, Middle-High, High*}. Each of

T_i maps the x_i to a real value between 0 and 1, expressing the degree to how it is likely to the linguistic value. This is calculated by a membership function, which is defined here using fixed shaped triangular and trapezoidal fuzzy membership functions.

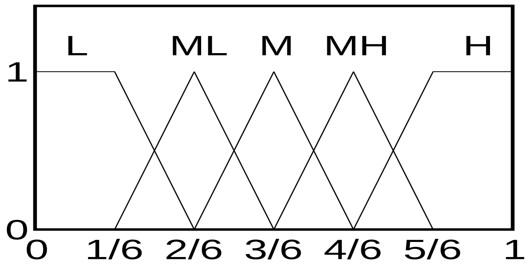

Fig. 2. Five fixed shaped membership functions each of which describes how likely a coordinate is either Low (L), Middle-Low (ML), Middle (M), Middle-High (MH), or High (H). Note that the coordinates in $[-1, 1]$ are translated into $[0, 1]$ with interpretation being intact.

Then a genetic algorithm evolves these fuzzy rules, with chromosomes being $(T_1, \cdots, T_n)$, starting with those chromosomes randomly created. To be more specific, our chromosome is made up of n integer genes whose value is chosen from $\{0, 1, 2, 3, 4\}$. The fitness of a rule is evaluated by applying the rule to all the self patterns $\mathbf{x} = (x_1, \cdots, x_n)$, one by one, and calculated as

$$\text{fitness}(R) = 1 - \max_{\mathbf{x} \in Self} \{ \min_{i=1,\cdots,n} \{\mu_{T_i}(x_i)\}\}$$

which implies how the rule covers the non-self space.

The size of self cells is fixed to 400; each GA run is iterated for 1000 generations; size population of chromosmes are 100, to be comparable with the original Dasgupt et al's experiment [6].

III. RESULTS & DISCUSSION

As a preliminary experiment, we tried a random search for the 20-dimensional needle by creating 5000 candidate strings of 20 binary bit at random, assuming all 0 string alone is the needle. The result is shown in Figure. 2, and we found it is still not such a difficult problem if we use a standard PC found everywhere nowadays.

Then what will happen if we exploit one of the lately reported more sophisticate methods?

As our goal is to find a rule to identify an island, that is whether or not the search point $\mathbf{x} = (x_1, \cdots, x_n)$ fulfills the condition that $-a < x_i < a$ for all $i = 1, \cdots, n$, all we need is one rule, that is,

IF x_1 *is Middle, and* $\cdots$*, and* x_n *is Middle*
THEN $\mathbf{x}$ *is on the island.*

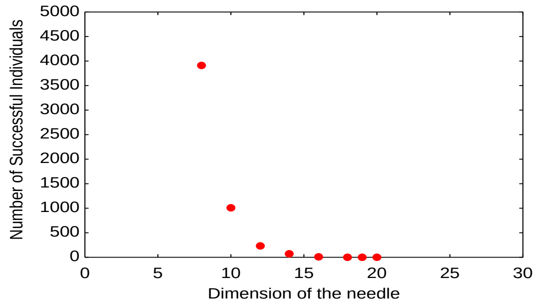

Fig. 3. The number of happened-to-be-the-needle out of 5000 random creations of the candidate.

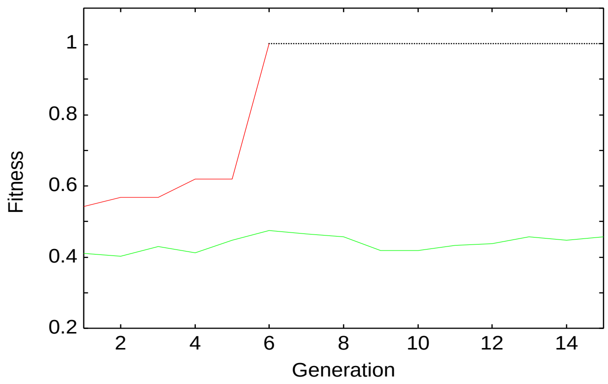

Fig. 4. Evolution to find the island in the 6-dimensional lake. Upper plot is when $x_i \in [0.25, 0.75]$ and lower plot is for a small island where $x_i \in [0.45, 0.55]$ $(i = 1, \cdots, 6)$.

And we found an evolution converges to a successful such chromosome

$$(2\ 2\ 2\ 2\ 2\ 2).$$

However, this holds only on the condition that the dimension is small and the island is fairly large. In Figure 4. we plot two cases of evolution. One is the island made up of the points $x_i \in [0.25, 0.75]$ and the other is $x_i \in [0.45, 0.55]$, both in the 6-dimensional lake. This discrepancy in the size of the lake depending on the dimensionality is around a critical condition for the success of the evolution. Our temporary goal of 20-dimensional *a-needle-in-a-haystack* is far beyond from this critical point. A trial of a search for the island of $x_i \in [0.45, 0.55]$ $(i = 1, \cdots, 20)$ in the 20-dimensional lake converged a chromosome

$$(0\ 1\ 3\ 4\ 0\ 4\ 4\ 1\ 4\ 4\ 1\ 1\ 4\ 1\ 4\ 3\ 0\ 0\ 0\ 2).$$

Alas, the curse of dimensionality.

As for the size of the island, we might explore the evolution in which shape of membership functions are also adaptively evolved, such as Gaussian one, expecting very narrow membership function corresponding to the linguistic term *MIDDLE*.

IV. CONCLUSION

We have described how we would be able to find a small island in a huge lake with a system training only using data in a lake. This is a metaphor in which a very few of unpredictable abnormal transaction patterns hidden in an enormous amount of normal patterns, in the context of network intrusion detection.

We need test samples to test the system we are designing. So far, so many such sample are proposed. For example, Ayara et al. [4] used randomly generated 8-bit binary patterns assuming they are normal patterns; Kim & Bentley [5] used Fisher's Iris Flower Database, which includes the data of three different species of Iris flower, by taking one set as a normal pattern set and the rest as an abnormal set. Then with a tenfold cross-validation method, it was claimed the best true positive rate reached 100% and the worst false positive rate was only 1%; or more directly, some others such as Gomez et al. (2004) [6] used the 1998 DARPA intrusion detection evaluation data-set prepared by MIT, also known as KDD cup 99 data-set.[1] We have also a newer such public domain data-set called KDD cup 2003 which was used in the data mining competition held in conjunction with the Ninth Annual ACM SIGKDD Conference.[2]

We assume three conditions for a set of database to be used for the purpose of train and test a system we are designing as follows.

(1) dataset is comprised of all possible patterns.
(2) Training should be only by normal samples
(3) Number of normal sample should be enormous while abnormal sample is only a few.

When an iris flower is normal then are others abnormal? None of the above mentioned examples fulfills all of the three conditions.

We are still neutral as for the possibility of designing an intrusion detection system using test samples in this way. Or rather negative if we are not beating around the bush.

We hope this article will play a role of call for challenges to this issue.

References

[1] A. Imada (2004) “How a Peak on a Completely-flatland-elsewhere can be Searched for? — A Fitness Landscape of Associative Memory by Spiking Neurons.” Proceedings of Advanced Computer Systems (ACS) and Computer Information Systems and Industrial Management Applications (CISIM), Vol.2, pp. 171–150.

[2] J. Gomez, F. Gonzalez, and D. Dasgupta (2003) ”An Immuno-Fuzzy Approach to Anomaly Detection” proceedings of the 12th IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 1219-1224.

[3] F. Provost, T. Fawcett, and R. Kohavi (1989) “The case against accuracy estimation for comparing induction algorithms.” Proceedings of international conference on machine learning, (San Francisco, Ca), pp. 445–453.

[4] M. Ayara, J. Timmis, R. D. Lemos, L. N. D. Castro, and R. Duncan (2002) ”Negative Selection: How to Generate Detectors.” Proceedings of 1st International Conference on Artificial Immune Systems pp. 89–98.

[5] J. Kim and P. J. Bentley (2001) “Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator.” Proceedings of the Congress on Evolutionary Computation.

[6] J. Gomez, F. Gonzalez, and D. Dasgupta (2003) ”An Immuno-Fuzzy Approach to Anomaly Detection” proceedings of the 12th IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 1219-1224.

[1] http://kdd.ics.uci.edu/databases/kddcup99
[2] http://www.cs.cornell.edu/projects/kddcup