Estimating the Detector Coverage in a Negative Selection Algorithm

Zhou Ji
St. Jude Children's Research Hospital
Memphis, TN 38105
zhou.ji@stjude.org

Dipankar Dasgupta
The University of Memphis
Memphis, TN 38152
ddasgupt@memphis.edu

ABSTRACT

This paper proposes a statistical mechanism to analyze the detector coverage in a negative selection algorithm, namely a quantitative measurement of a detector set's capability to detect nonself data. This novel method has the advantage of statistical confidence in the estimation of the actual coverage. Furthermore, unlike the existing analysis works of negative selection, it doesn't depend on specific detector representation and generation algorithm. Not only can it be implemented as a procedure independent from the steps to generate detectors, the experiments in this paper showed that it can also be tightly integrated into the detector generation algorithm to control the number of detectors.

Categories and Subject Descriptors: I.2 [Computing Methodologies]: Artificial Intelligence

General Terms: Algorithms, Performance, Reliability.

Keywords: Negative selection, detector coverage, hypothesis testing.

1. INTRODUCTION

Artificial negative selection is a soft computational paradigm inspired by natural immune system's self/nonself discrimination mechanism. It was designed by modeling the biological process in which T-cells mature in thymus through being censored against self cells[8]. As one of the earliest models of artificial immune systems (AIS), it is often simply called negative selection (algorithm) with the descriptive "artificial" dropped [4, 9, 10].

In a negative selection algorithm, a collection of detectors, usually called detector set, is used to check incoming data items to be normal (self) or not (nonself). Whether the detector set can detect all the anomalies, or how many of all the anomalies it can detect, is the main concern. The proportion of the nonself space that is covered or recognized by the detector set is the measure of its detection power. This proportion is called *detector coverage.* It is not trivial to determine the coverage quantitatively for a specific negative selection algorithm, or to decide the necessary number and distribution of detectors for a given coverage.

This paper describes a statistical method to estimate detector coverage. A real-valued negative selection algorithm, *V-detector*[16], is used as the platform to implement this idea. It proved to be an effective and beneficial supplement to the negative selection algorithm. It is possible to implement a similar mechanism in other representations like the popular binary string representation. The same methodology also applies to different algorithms in which the similar issue of proportion estimation exists.

This method is especially useful when the near-perfect coverage is not necessary and alternative method may be inaccurate about the coverage.

GECCO'05, June 25–29, 2005, Washington, DC, USA.

2. RELATED WORKS

Statistical methods were used in several works in AIS area. Some earlier works like [8, 5] used matching probability or failure probability to decide or evaluate the number of detectors. That method was based on the specific detector scheme. Furthermore, it was oriented to general analysis of the relation between the number of detectors and the coverage. A given detector set is not in the question. While other works may have focused on different aspects, e.g. lower bound for the fault probability [17], they still used statistics from the same point of view. On the other hand, this paper's proposal is to use statistical tools to estimate the coverage of any given detector set, or more generally, any detection mechanism in which individual point can be verified but there is no explicit algorithm to evaluate the rate of success.

Considering the multiple issues involved, the work described here evolved out of concepts from a few different fields.

2.1 Real-valued Negative Selection Algorithms

While more research in negative selection algorithms uses binary representation [1, 6], real-valued representation has its unique role. It is relatively less explored due to the fact that the search space is usually continuous and hard to analyze by enumerative combinatorics. Nevertheless, it is necessary for many applications that cannot be represented effectively in binary form. For those problems that are naturally real valued, real-valued representation makes it easier to interpret the results and usually results in more stable algorithm by maintaining affinity in representation space.

In real-valued negative selection algorithms, the detectors are represented as hyperspheres or hyper-rectangles. They are generated by the original method of randomly generating and then eliminating [11] or other methods like genetic algorithm [3]. Some algorithms involve resizing and redistribution of the initial detectors. Most of the works considered a unit hypercube in n-dimensional space, $[0, 1]^n$, as the universe space.

V-detector algorithm [16] generates the detector set in one run. It took advantage of variable size and avoided the need to redistribute detectors for the purpose of minimal overlap. Estimate of detector coverage was used as a criterion to decide when the number of detectors is adequate, comparing to the previous works of either presetting the number or deciding the number by rough geometric estimation. Although the estimation in [16] is rather simple statistically, it leads to the work in this paper.

2.2 Statistical Inference

Estimation of detector coverage in [16] is based on point estimation, which doesn't tell us how much or how likely the estimate may be different from the population proportion - in this case, the actual detector coverage.

To obtain a more informative estimate of population parameters, more formal statistical inference is desirable [14, 7]. The *Central limit theorem* in statistics says that if X is a random variable having finite mean μ and finite variance σ^2, then the probability distribution of the random variable $(\bar{x} - \mu)/(\sigma\sqrt{n})$ approaches the standard normal distribution as the sample size n becomes infinite [13]. In other words, although we don't get the same mean every time we repeat sampling of a fixed size, the distribution of these means is close to a normal distribution. The standard deviation of the sampling distribution for a given sample size is equal to the population standard deviation divided by the square root of the sample size.

The central limit theorem justifies using a normal distribution as an approximation for the distribution of $\bar{x}$ when n is sufficiently large. There are two apparent sources of error in using the normal distribution as an approximation of the binomial distribution: (1) The normal distribution is always symmetrical; the binomial distribution is symmetric only if the probability of one outcome, p, is 0.5. (2)The normal distribution is continuous; the binomial distribution is discrete. A rule of thumb taking into account both the problems of asymmetry and discreteness is to use the normal distribution approximation only if $np > 5$, $n(1 - p) > 5$ and $n > 10$.

There are alternative distributions that can be used to deal with the asymmetry problem, and there are mathematically strict corrections for discontinuity too. However, these are not the main issues in our application here even though the issue of asymmetry is in fact not negligible. Because the proportion of covered nonself points, p, is the variable to be considered, it is very likely that we need to consider a large p, for example, 90% or 99%. Fortunately, we can circumvent the issue with proper strategy in our proposed algorithm considering the fact that we care more about enough coverage than its exact value.

2.2.1 Confidence interval

The most basic statistical inference is point estimation, in which we use a sample statistic, for example mean or proportion, as the estimator of the population parameter. We need to know the probability, namely confidence level, that the population parameter falls within the range called *confidence interval* around the sample parameter [7, 14].

Generally, we can conclude that

$$\hat{p} - E < p < \hat{p} + E, \quad (1)$$

where p is the population parameter, $\hat{p}$ is the sample statistic,

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (2)$$

is the margin of error, n is the sample size, and $\hat{q} = 1 - \hat{p}$. In the case of estimating detector coverage, we are more interested in making a conclusion about the lower limit of coverage, $p > p_{min}$, where p_{min} is the minimum coverage we can presume with some certainty. So we can use a one-side confidence interval

$$p > \hat{p} - E, \quad (3)$$

where

$$E = z_{\alpha}\sqrt{\frac{\hat{p}\hat{q}}{n}}. \quad (4)$$

To ensure the assumption that the binomial random variable is approximately normally distributed with the mean $\mu = np$ and standard deviation $\sigma = \sqrt{npq}$, we should have $np \geq 5$, $nq \geq 5$.

In Equation (1), $z_{\alpha/2}$ is the z score for a confidence level of $1-\alpha/2$ - the positive standard z value that separates an area of $\alpha/2$ in the right tail of the standard normal distribution curve. For a standard normal distribution, the probability of $-z_{\alpha/2} \leq x \leq z_{\alpha/2}$, $P(-z_{\alpha/2} \leq x \leq z_{\alpha/2}) = 1 - \alpha$. The probability of $x \leq z_{\alpha/2}$, $P(x \leq z_{\alpha/2}) = 1 - \alpha/2$. Similarly, z_α in Equation (4) is where $P(x \leq z_\alpha) = 1 - \alpha$.

2.2.2 Hypothesis Testing

Hypothesis testing is another approach of statistical inference also based on Equation (3). It fits our purpose better because the goal here is to know when so we can stop generating or including more detectors.

In conducting a statistical hypothesis test, we need to identify the *null hypothesis*. We assume that Type I Error (rejecting the true null hypothesis) is more costly than Type II Error (accepting a false null hypothesis).

The normal procedure of hypothesis testing involves the following steps:

1. State the null hypothesis and alternative hypothesis. The null hypothesis is the statement that we'd rather take as true if there is not strong enough evidence showing otherwise.

2. Determine the cost associated with the two types of decision-making errors.

3. Choose the significant level, α. That is the maximum probability we are willing to accept in making Type I Error. Typical values are 0.05 or 0.01.

4. Collect the data and compute the sample statistic. To test based on proportion we can use z score

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}.$$

5. Reject or accept the null hypothesis. The traditional method is to check whether the test statistic is in critical region ($z > z_\alpha$) or not. If $z > z_\alpha$, we reject the null hypothesis. An alternative way is to use a p-value test, which is easier [14].

2.3 Learning Theory Point of View

One of the frameworks of learning, Probably Approximately Correct learning (PAC learning), was proposed so that learning can be analyzed mathematically [2, 12]. In terms of PAC learning, successful learning of an unknown target concept should entail obtaining, with high probability, a hypothesis that is a good approximation of it. Self/nonself discrimination algorithms, including the negative selection algorithm discussed in this paper, can be examined with this framework. Accuracy, or how good the approximation is, is described by ϵ: the hypothesis returned h should satisfy $error(h) \leq \epsilon$. Confidence, or the chance we can correctly obtain the hypothesis h, is described by σ: the probability of returning h is at least $1-\sigma$. Probability of distribution on the instance space plays an important role in the language of PAC learning

What is most relevant to our discussion here is the basic assumptions we must make about the negative selection algorithm and the training data it takes. Previous works in this area, especially those were not based on binary representation and did not assume all self features are present in training data, are hard to compare with one another shoulder-to-shoulder due to the lack of equivalent assumptions.

In the following analysis, we assume

- Both self and nonself points appear in some bounded n-dimensional real space. For simplicity, let us assume it is $[0, 1]^n$.
- Some finite number of self samples are provided as input. They are randomly distributed over the self region.
- The training data is noise free - meaning all the self samples are real self point. This is not necessary in principle, but used to simplify the discussion.
- To evaluate the detection performance, the testing data are finite number of random points over the entire space in question described above. Each of those points can be verified to be self or nonself.

Considering the framework of PAC learning, a negative selection algorithm is a PAC algorithm under certain conditions. Although we assume that adequate self samples are available over the entire self region, specific distribution is not required. On the other hand, test data over the entire space are used to evaluate the performance but definitely not mandatory to use the algorithm.

3. ALGORITHM

3.1 Coverage - Proportion - Probability

Definition 1. The *detector coverage of a given detector set* is defined as the ratio of the volume of the nonself region that can be recognized by any detector in the detector set to the volume of the entire nonself region.

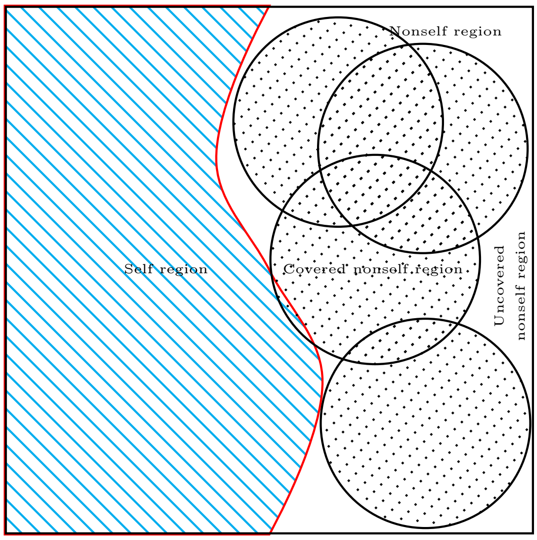


Figure 1: Negative selection algorithm using real-valued representation: Different regions

Generally, it can be written as

$$p = \frac{\int_{\vec{x} \in D} d\vec{x}}{\int_{\vec{x} \in \overline{S}} d\vec{x}},$$

where $\overline{S}$ is the set of nonself points and D is the set of nonself points that are recognized by the detectors. In the case of 2-dimensional continuous space, it is reduced to the ratio of the area covered to the area of the entire nonself region

$$p = \frac{\iint_{(x,y) \in D} dxdy}{\iint_{x \in \overline{S}} dxdy}.$$

If the space in question is discrete and finite, it can re-written as

$$p = \frac{|D|}{|\overline{S}|},$$

where $|A|$ denotes the cardinality of a set A.

Figure 1 illustrates the three regions in the question: self region, covered nonself region, and uncovered nonself region in a 2-D diagram. The area without hatched shade on the right-side of the diagram is $\bar{S}$ and the dotted area cover by circular detectors are D.

In statistical term, the points of the nonself region are our population. Generally speaking, the population size is infinite. The probability of each point to be covered by detectors is a binomial distribution. The detector coverage is the same as the proportion of the covered points, which equals to the probability that a random point is a covered point. Assuming all the points from the entire nonself region are equally likely to be chosen in a random sampling process, the probability of a sample point being recognized by the detectors is thus equal to p. For a sample of fixed size, the proportion of covered points is

$$\hat{p} = \frac{\left|\hat{D}\right|}{\left|\hat{S}\right|},$$

where $\hat{S}$ is the sample; and $\hat{D}$ is the set of sample points that are recognized by the detectors. $|\hat{S}|$ is thus the sample size. $\hat{p}$ is the sample statistic that is the point estimate of the population proportion.

3.2 Integration of Hypothesis Testing and Detector Generation

The main idea of this method is to finish generating detectors when the coverage is close enough to the target value. This contrasts with other works that replies on the number of detectors to provide enough coverage.

The original *V-detector* [16] has a simple estimate to stop the detector generation procedure. Random points are generated to be detector candidates. If it is a nonself point but not covered, a new detector is generated on it. If it is a covered nonself point, it is discarded as a candidate but the attempt is recorded in a counter which will be used to estimate the coverage. If the counter of consecutive attempts that fall on covered point reaches a limit m, the generation stage finishes with the belief that the coverage is large enough. m is not preset. It is decided by the target coverage.

$$m = \frac{1}{1-\alpha}, \tag{5}$$

where α is the *target coverage*, a control parameter. Equation (5) is explained as following. If there is 1 uncovered point in a sample of size m', the point estimate of proportion of uncovered region is $\frac{1}{m'}$, and the estimate of coverage is

$$\alpha' = 1 - \frac{1}{m'}. \tag{6}$$

If in fact there is 0 uncovered point in a sample of size m', we have a better than average chance that the actual coverage is larger than α'. Because m is decided by Equation (5), when we see m consecutive points that are all covered, we can estimate that the actual coverage is more likely to be at least α. As mentioned before, that is based on point estimation without a confidence interval. Comparing with the new algorithm, we call that method "naïve estimate".

To extend to more strict statistical inference, estimating with a confidence interval directly does not fit the problem as well as hypothesis testing because our goal to make a decision of adding more detector or not. What makes this paper's method different from traditional statistical inference is that the testing can be done as part of the detector generation algorithm. Although it may be implemented as a relatively independent module, we still have to face a dilemma: the detector coverage or the proportion to be estimated is actually changing during the detector generation. So we need to design a process in which the hypothesis testing happens only when we temporarily stop adding new detectors. Otherwise, the testing will be meaningless. At the same time, we also try to reuse the random samples we use in hypothesis testing as the candidate detectors. This doubles the advantage of an algorithm of hypothesis testing integrated in *V-detector*.

In the case of estimating coverage, the null hypothesis would be "The coverage of the non-self region by all the existing detectors is below percentage p_{min}." If we accept the null hypothesis, we would include more detectors. If the null hypothesis is actually false, the cost of a Type II Error would be more unnecessary detectors. On the other hand, if we reject the null hypothesis by mistake, we would end up with lower than actual coverage. The latter, so called Type I Error, is exactly our concern. The significant level α is the maximum acceptable probability that we may make a Type I Error - end up with fewer than needed detectors. We need a fixed sample size to do the hypothesis testing. If the conclusion is that we need more detectors, we take all the sample points as detector candidates. This largely saves the cost of the entire algorithm.

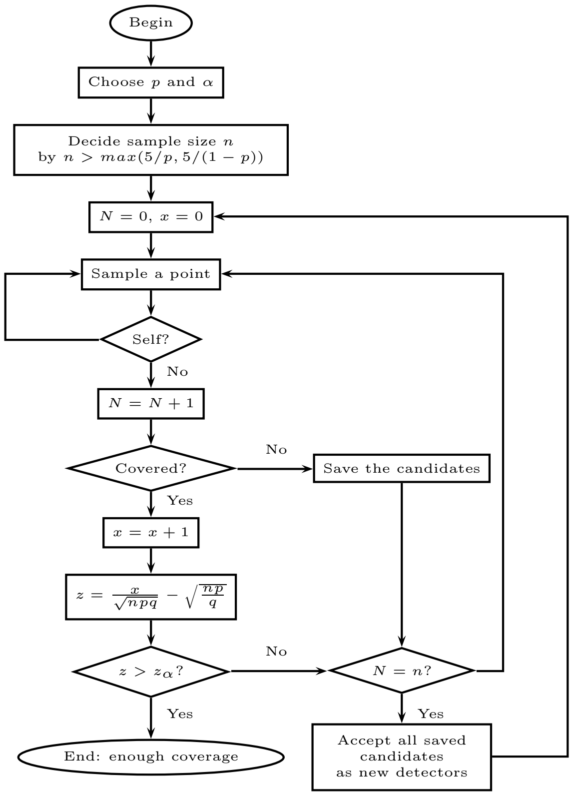


Figure 2: *V-detector* generation algorithm with statistical estimate of coverage

Figure 2 shows the diagram of the modified *V-detector* that uses hypothesis testing to estimate the detector coverage.

To guarantee the assumption $np \geq 5$ and $nq \equiv n(1-p) \geq 5$ is valid, we can choose sample size by

$$n > max(5/p, 5/(1-p)).$$

If there is x points covered, $\hat{p} = x/n$, where n is the sample size, we have

$$z = \frac{x}{\sqrt{npq}} - \sqrt{\frac{np}{q}}.$$

During the procedure to test more points, x will either increase (when the point is covered) or stay unchanged (when the point is uncovered). So does z. Before the procedure finishes for all n points, if z based on the tested points is larger than z_α, it is enough to reject the null hypothesis and claim enough coverage. At that point, the test can be stopped. Since the ultimate conclusion from the procedure is either rejection or acceptance of the null hypothesis, not the estimate of p and confidence interval, it is not necessary to finish trying to get a "better" answer.

Table 1: Shapes of self area

Type of Shape	Geometric Parameters
Cross	thickness and location of the cross
Ring	outer and inner radius
Stripe	width
Intersection	cross size and location, circle radius
Pentagram	size (radius of circumscribed circle)

If the the assumption $nq > 5$ is in fact invalid because the real p is larger than the p we used, then the actual coverage is more than what we want to test. Our confidence in the coverage is not comprised in this case. If the assumption $np > 5$ is in fact invalid because p is so small, the hypothesis test will pass only when it could pass the actual non-normal distribution. Because the probability curve skew to the left side (origin side), z_α would be smaller than z_α for normal distribution. If z does not pass this skewed z_α, it will not pass normal distribution's z_α either: $z \le z_\alpha|_{p<5/n} \le z_\alpha$.

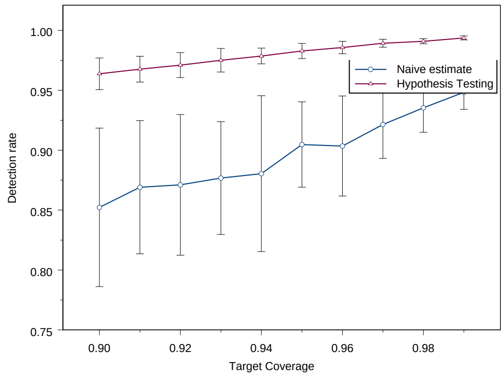


Figure 6: Influence of target coverage

4. EXPERIMENTS AND DISCUSSION

To test the algorithm described in the previous section, experiments were carried out using 2-dimensional synthetic data. Over the unit square $[0, 1]^2$, various shapes are used as the 'real' self region in these experiments. They belong to one of the five types listed in Table 1, which also shows the geometric parameters that extend each type to different sizes or variations. Figure 3 shows the basic shapes of the five types of self region.

A fixed number of random points from the self region are used as the self sample to generate the detector set. Another number of random points, in which some are self, some non-self, are used to test the detection performance of the detector set. Figure 4 shows examples of training data (self samples) and test data. (a) is a self sample of 100 points. (b) is a self sample of 1000 points. (c) is 1000 test data including both self points and nonself points. It can be predicted from this figure that the number of training data will have obvious influence on the detection results.

There exist two versions of *V-detector* algorithms. The earlier version treats each training data point (self sample) individually [16]. We call it point-wise *V-detector*. A later version brought out a new advantage of negative selection algorithm so that it is able to detect the boundary of self region. We call it boundary-aware *V-detector* [15].

Figure 5 shows the detector-covered area using these two different numbers of detectors (boundary-aware algorithm, hypothesis testing, 99% target coverage). (a) is 100 point; (b) is 1000 points. When other control parameters are different, e.g. using point-wise algorithm, the covered area will not be the same as in Figure 5, but the number of detectors still plays an important role.

The influence of control parameters and difference of strategies were explored with more experiments. From the data side, the difference in results may come from: number of sample points and different shapes (including size) of the self region. From the algorithm side, the difference may come from: target coverage, significant level of hypothesis testing, estimate methods (naïve estimate or hypothesis testing), self threshold, and *V-detector* strategy (point-wise or boundary-aware). The results we want to compare are detection rate, which is the main concern, and false alarm rate and number of detectors. Significant level α is set to be 0.1 in the results reported in this paper.

Figure 6 shows some results of detection rate for target coverages from 90% through 99%. The number of sample points is 1000. The boundary-aware algorithm was used. The self region is a pentagram whose radius of circumscribed circle is 1/3. The plot shows the mean of 100 repeated tests; standard deviation is shown as error bar on the graph. Results obtained with naïve estimate and hypothesis testing are plotted together to compare. Hypothesis testing has a small but consistent advantage over the naïve method.

Figures 7 and 8 show the detection rate and false alarm rate, respectively, comparing 100 points and 1000 points of the self sample. The boundary-aware algorithm plus hypothesis testing was used. The difference in detection rate is rather small, but the false alarm using 100 points is significantly higher. On the other hand, if the point-wise algorithm is used, the false alarm rate can be controlled over a range of self thresholds, but the detection rate of 100 points will be much lower. It is not surprising that the number of self sample points has a major affect on detection performance. That has little to do with detector generation and detection process. Similarly, false alarms (false positive) also mainly come from the definition of self that is totally based on these discrete samples.

Figure 9 shows the difference between the point-wise and boundary-aware *V-detector* when all the other settings are the same. Figure 10 shows the false alarm rate. Although the boundary-aware algorithm has higher false alarm at very low self threshold, it is not an issue generally. The difference in the two strategies' performance is related to the fact that the concept of 'self' here is defined by the discrete self points. The boundary-aware *V-detector* improves the result obviously when detecting the boundary of self region is important. That advantage largely depends on the unique process of negative selection algorithm.

Figure 11 shows the detection rate results of different shapes of self regions for a range of self threshold. Totally 10 different shapes are shown in this figure including all the five types in Figure 3 plus their complementary shapes. The results are consistent without major difference. 100 self points were used to train in those results. When 1000 points were used, the difference were even smaller.

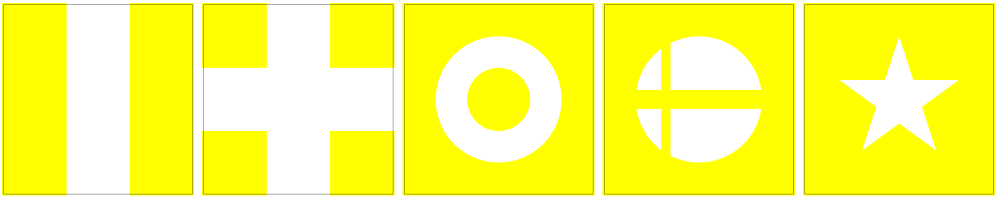
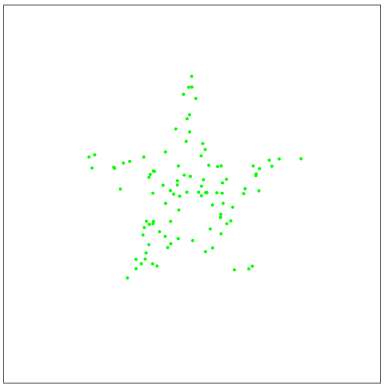
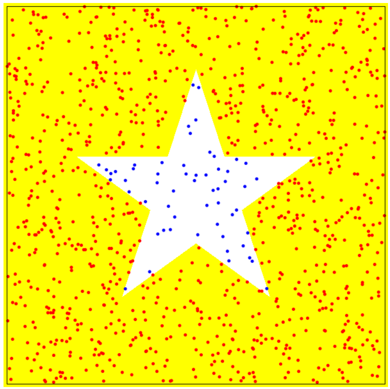


Figure 3: Different types of shape



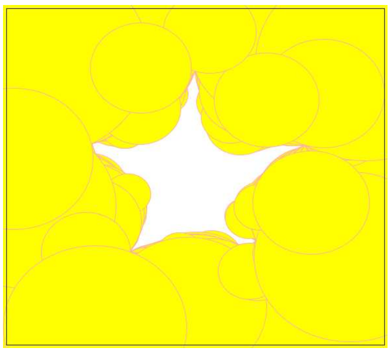
(a) 100 points of self sample

(b) 1000 points of self sample

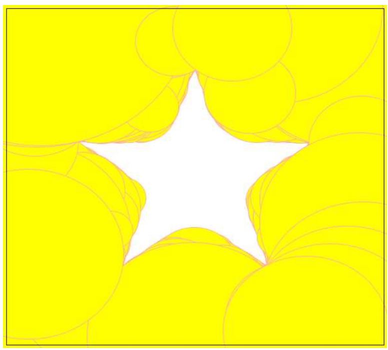


(c) 1000 points of test data

Figure 4: Self samples and test data



(a) Trained with 100 points



(b) Trained with 1000 points

Figure 5: Detector-covered area

Table 2: Performance difference between naïve estimate and hypothesis testing

	detection rate/σ	false alarm rate/σ	number of detectors/σ
naïve estimate	95.18%/1.35%	0.72%/0.86%	19.98/2.86
hypothesis testing	99.35%/0.17%	3.41%/0.96%	500/0

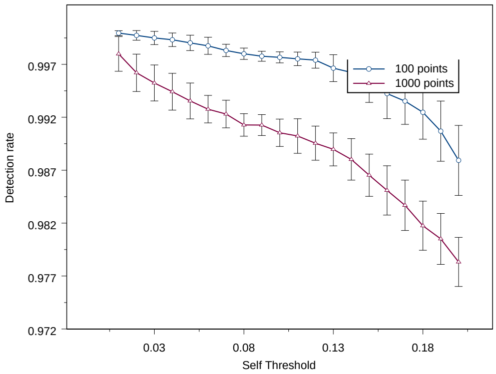


Figure 7: Detection rate of different training size

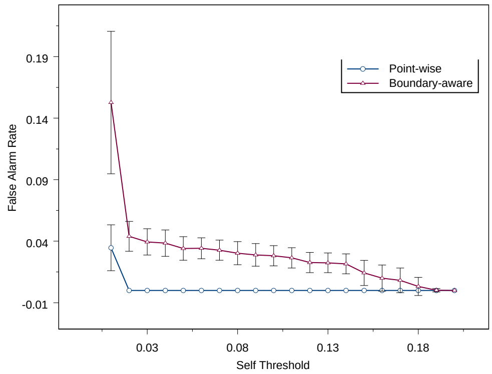


Figure 10: Two strategies in *V-detector*: False Alarm Rate

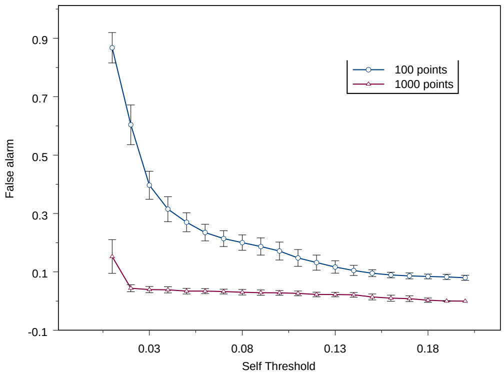


Figure 8: False alarm rate of different training size

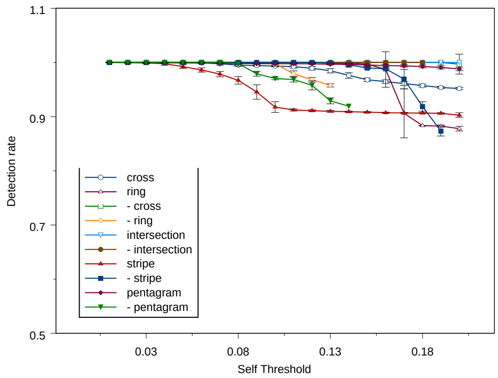


Figure 11: Detection rate for various shapes of self region

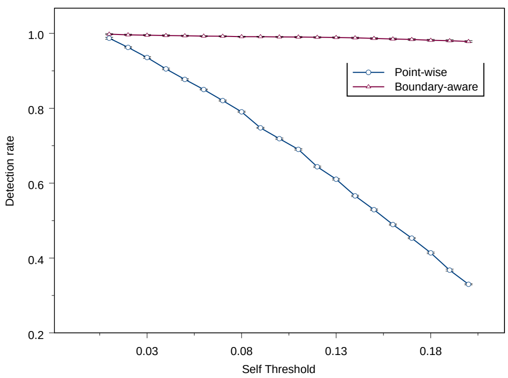


Figure 9: Two strategies in *V-detector*: Detection Rate

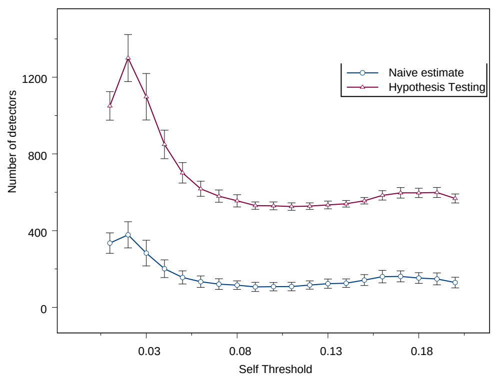


Figure 12: Number of detectors (Ring shape, 100 training points)

Figure 12 shows the number of detectors used for ring shape of self region generated using 100 training points. At least two characteristics are noteworthy. One, the number is near constant as long as the self threshold is larger than 0.05; two, hypothesis testing resulted in more detectors. The reason for the second inclination is that the detector candidates must be processed in groups of proper sample size required by hypothesis testing. That disadvantage is limited and will not scale with the number of training points or other parameters.

Table 2 again highlights the difference between naïve estimate and hypothesis testing. The results were from the following setting: pentagram-shaped self region; boundary-aware strategy; 1000 self sample points; target coverage 99%; self threshold 0.05. The numbers are the mean of 100 repeated tests and the standard deviation σ is also tabulated with the corresponding variables.

5. CONCLUSIONS

A statistical approach is investigated to analyze the detector coverage in a negative selection algorithm. It makes the algorithm more reliable. An effective strategy was developed for implementation.

The unique feature of a real-value negative selection algorithm, *V-detector*, makes it a prefect platform for hypothesis testing: (1) Generation of detectors in one run makes the algorithm more stable and easier to use; (2) Coverage estimate makes negative selection more reliable and saves the need of detector adjustment; (3) Hypothesis testing is a major development over the earlier naïve estimation.

Another advantage of this method is that it applies to any detector schemes and detection mechanisms as long as it is verifiable whether a sample point is covered or not. In *V-detector*, it is even better since it can be implemented partly as a byproduct of the generation process without adding much extra computational cost.

The strategy can be extended to different representations. For example, extension to binary representation will make this method applicable to variety of applications.

Many issues in the performance of negative selection algorithms are based on the properties of the data. For the comparison and analysis of negative selection algorithms to be more meaningful, it is important to develop a framework concerning the fundamental assumptions and to categorize the type of data to be processed.

6. ACKNOWLEDGMENTS

This work was supported in part by NIH Cancer Center Support Core Grant CA-21765 and the American Lebanese Syrian Associated Charities (ALSAC).

7. REFERENCES

[1] M. Ayara, J. Timmis, R. de Lemos, L. de Castro, and R. Duncan. Negative selection: How to generate detectors. In J. Timmis and P. J. Bentley, editors, *Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS)*, volume 1 of *89-98*, University of Kent at Canterbury, September 2002. University of Kent at Canterbury Printing Unit.

[2] F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin (New Series) of the American Mathematical Society*, 39(1):1–49, October 2001.

[3] D. Dasgupta and F. Gonzalez. An immunity-based technique to characterize intrusion in computer networks. *IEEE Transactions on Evolutionary Computation*, 6(3):1081–1088, June 2002.

[4] L. N. de Castro and J. Timmis. *Artificial Immune System: A New Computational Intelligence Approach*. Springer, 2002.

[5] P. D'haeseleer, S. Forrest, and P. Helman. An immunological approach to change detection: Algorithms, analysis, and implications. In *Proceedings of the 1996 IEEE Symposium on Computer Security and Privacy*, 1996.

[6] F. Esponda, E. S. Ackley, S. Forrest, and P. Helman. Online negative databases. In G. N. et al, editor, *Proceedings of Third International Conference on Artificial Immune Systems (ICARIS 2004)*, pages 175 – 188, September 2004.

[7] J. L. Fleiss. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 1981.

[8] S. Forrest, A. Perelson, L. Allen, R., and Cherukuri. Self-nonself discrimination in a computer. In *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, Los Alamitos, CA, 1994. IEEE Computer Society Press.

[9] A. A. Freitas and J. Timmis. Revisiting the foundation of artificial immune systems: A problem-oriented perspective. In *Proceedings of Second International Conference on Artificial Immune System (ICARIS 2003)*, 2003.

[10] S. M. Garret. A survey of artificial immune system: Are they useful? *Evolutionary Computation*, in press, 2004.

[11] F. Gonzalez, D. Dasgupta, and L. F. Nino. A randomized real-value negative selection algorithm. In *Proceedings of Second International Conference on Artificial Immune System (ICARIS 2003)*, September 2003.

[12] D. Haussler. Probably approximately correct learning. In *National Conference on Artificial Intelligence*, pages 1101–1108, `citeseer.ist.psu.edu/haussler90probably.html`, 1990.

[13] C. A. Hawkins and J. E. Weber. *Statistical Analysis - Applications to Business and Economics*. Harper & Row, Publishers, New York, 1980.

[14] R. V. Hogg and E. A. Tanis. *Probability and Statistical Inference*. Prentice Hall, 6th edition, 2001.

[15] Z. Ji. Influence of training data's interpretation strategy on real-valued negative selection algorithm. Technical report, The University of Memphis, June, 2004.

[16] Z. Ji and D. Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In *LNCS 3102, Proceedings of GECCO*, pages 287–298, 2004.

[17] S. Wierzchon. Discriminative power of the receptors activated by k-contiguous bits rule. *Journal of Computer Science and Technology*, 1(3):1–13, 2000.