VorAIS: A Multi-Objective Voronoi Diagram-based Artificial Immune System

Luis Martí[1,2] Arsene Fansi-Tchango[3] Laurent Navarro[3] Marc Schoenauer[1]
[1] TAO, INRIA, LRI/CNRS/UP-Sud, Université Paris-Saclay, Orsay, France
[2] Universidade Federal Fluminense, Av. Gal. Milton Tavares de Souza, Niterói (RJ) 24210-346 Brazil
[3] ThereSIS, Thales Group, 1 avenue Augustin Fresnel, 91767 Palaiseau, France
{luis.marti,marc.schoenauer}@inria.fr;
{arsene.fansi-tchango,laurent.navarro}@thalesgroup.com

ABSTRACT

This paper introduces the Voronoi diagram-based Artificial Immune System (VorAIS). VorAIS models the self/non-self using a Voronoi diagram that determines which areas of the problem domain correspond to self or to non-self. The diagram is evolved using a multi-objective bio-inspired approach in order to conjointly optimize various classification metrics (accuracy, recall and specificity). VorAIS is experimentally validated, first on standard classification problems, then on the well-known NSL-KDD dataset for anomaly detection where it favorably compares with other AIS approaches.

Keywords

Artificial Immune Systems, Anomaly Detection, Voronoi Diagrams

1. PRELIMINARIES

Artificial immune systems (AISs) are bio-inspired algorithms that have been derived from existing theories of the functioning of biological immune systems. AISs are particularly appealing as they capture the ability of the biological system of telling apart normal body cells from pathogens.

Using inspiration from representations that had been proposed for evolutionary shape design, this paper proposes VorAIS, that uses a Voronoi-based representation for self- and non-self parts of the search space. Such representation offers a flexible and compact alternative to some common representations used in AIS such as hyper-spheres and hyper-rectangles.

Such partition of the space into self- and non-self-spaces is very similar to the way one can represent shapes in some real space, and there has been several propositions for such representations for shapes, including hyper-rectangles [4]. Other alternatives that have been proposed there for shapes include the so-called Voronoi representation, that is central to this work.

2. VORONOI DIAGRAMS-BASED AIS

In the context of classification, the target phenotypes are partitions of the parameter space into positive and negative examples (in the case of two classes), and can hence also be represented by Voronoi Diagrams with Boolean labels associated to its sites.

The genotype of Voronoi representations is a variable length list of Voronoi sites $(S_1, \dots, S_p)$, with $p \in [1, P_{Max}]$, where each site is defined by its n coordinates in $\mathcal{E}$. Each site S has an associated label $S.ell$ that determines how a point that falls within it is classified.

The crossover operator for Voronoi representation should not simply exchange some Voronoi sites between both parents, but should respect the locality of the representation. This is achieved by the geometric crossover, that draws a random hyperplane in the space, and exchanges the Voronoi sites from both sides of the hyperplane.

Several mutation operators can be designed for such a variable-length representation:

- At the individual level, a Voronoi site can be added, at a randomly chosen position, with a random label; or a randomly chosen Voronoi site can be removed.
- At the site level, Voronoi sites can be moved around in the space – and the well-known self-adaptive Gaussian mutation has been chosen here, inspired by Evolution Strategies; or the label of a Voronoi site can be changed.

When applying AIS to problems like classification and, in particular, anomaly detection, no single measure can capture the different desired properties simultaneously. Accuracy seems the best choice in the general case, as one wants to correctly identify all examples. But when dealing with anomalies, the dataset is generally highly imbalanced, as normally there are fewer anomalous instances than 'normal' ones. If only the classification accuracy is used, the error contribution of the anomalies will be reduced and hence the model will be biased to not regard them. On the other hand, neither recall nor specificity alone, if maximized, will lead to a satisfactory classifier. This naturally leads to the idea of using several objectives together, in a multi-objective approach, that will empower the algorithm with the capacity to address all the requirements of the task at the same time. Multi-objective evolutionary algorithms (MOEAs) are a competent approach for this task.

VorAIS consolidates the previous components as an algorithm that constructs a classification model. The algorithm

GECCO'16 Companion July 20-24, 2016, Denver, CO, USA

ACM ISBN 978-1-4503-4323-7/16/07.
DOI: http://dx.doi.org/10.1145/2908961.2909027

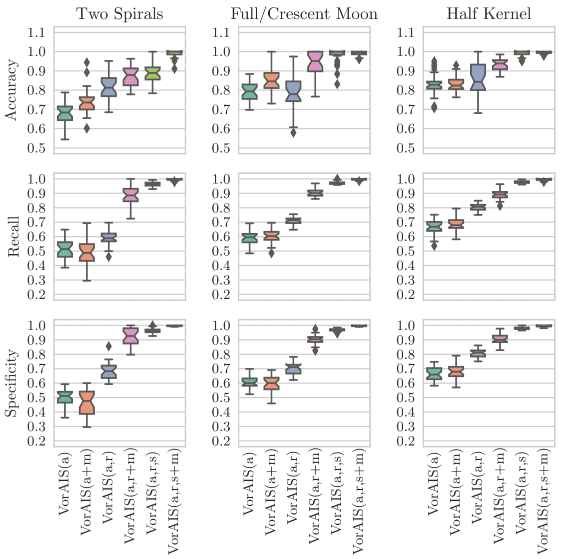

Figure 1: Box plots of the experimental evaluations on the 3 classification benchmarks.

goes by creating an initial random population $\mathcal{P}_0$ of n_{pop} individuals. At a given iteration t, individuals in the population $\mathcal{P}_t$ are then mutated and mated using operators described above and thus producing an offspring population $\mathcal{P}_{\text{off}}$ that consists of n_{off} individuals.

At this point, individuals that have not yet been evaluated are presented with the dataset and the values of the different objective functions are calculated. From the union of $\mathcal{P}_t$ and $\mathcal{P}_{\text{off}}$, the best n_{pop} are selected using the non-dominated sorting selection of NSGA-II [2].

3. EXPERIMENTS

One of the main questions regarding VorAIS is at what point a multi-objective affinity function would actually generate better results at an admissible cost. Experiments use a three classification benchmarks: the two spirals, the full moon/crescent moon and the half densities problems. They have the advantage that they can be visualized in 2D while not being excessively simple.

Clarifying these matter is the central issue of these experiments. To this end, we create different instances of VorAIS: (*i*) VorAIS(a), a single-objective classification accuracy one; (*ii*) VorAIS(a,r), a two-objective version that computes accuracy and recall and (*iii*) VorAIS(a,r,s) that extends the previous by including specificity as a third objective. In order to also study the influence of the mating operator we created versions of the above instances that included the mating operator, named VorAIS(a+m), VorAIS(a,r+m) and VorAIS(a,r,s+m), respectively.

The results of this experiments are shown as box plots in Figure 1. It can be inferred form those plots that the three-objective form of VorAIS yielded the best results. Similarly, it is also evident that including the mating operator improves substantially the outcome of the experiments.

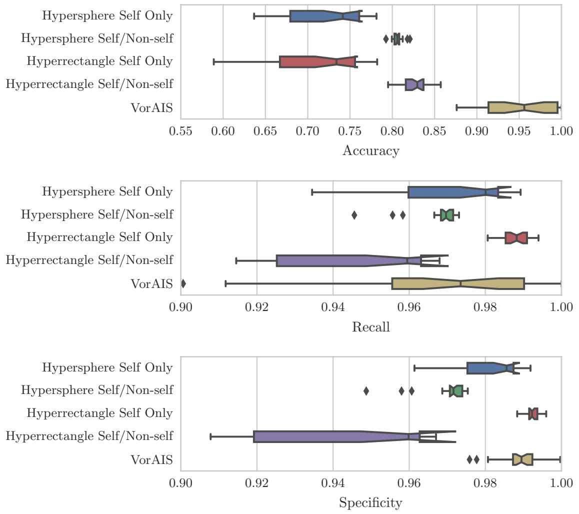

Figure 2: Box plots of the experimental evaluations of implemented approaches on NSL-KDD.

We also evaluated the proposed algorithm, in particular VorAIS(a,r,m+m), within this context and compare the performances obtained with respect to existing approaches. To this end, we consider the NSL-KDD dataset [1] that we describe next. For comparison purposes with other AIS approaches, we implemented the negative selection algorithm (NSA) as proposed by [3] using, on one hand, both variable-sized hyper-spheres (NSA_{sp}), and hyper-rectangles (NSA_{re}) on the other. Figure 2 shows the outcome of the experiments as box plots for easy visualization.

4. CONCLUSION

In this paper we have introduced VorAIS, a multi-objetive artificial immune system that relies on Voronoi diagrams for its representation. It has been devised with the problem of anomaly detection in mind. VorAIS representation is optimized by means of a bio-inspired process that mutates and mates Voronoi diagrams guided by multiobjective principles[1].

5. REFERENCES

[1] NSL-KDD dataset. http://www.unb.ca/research/iscx/dataset/iscx-NSL-KDD-dataset.html. Accessed: 2016-02-03.

[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA–II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[3] Z. Ji and D. Dasgupta. Real-valued negative selection algorithm with variable-sized detectors. In K. Deb, editor, *Proceedings of the 2004 Genetic and Evolutionary Computation (GECCO 2004) Part I*, pages 287–298, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[4] M. Schoenauer. Shape representation for evolutionary optimization and identification in structural mechanics. In G. Winter, J. Périaux, M. Galán, and P. Cuesta, editors, *Genetic Algorithms in Engineering and Computer Science (EUROGEN95)*, pages 443–464, 1995.

---

[1] Work funded by the project PIA-FSN-P3344-146479.