

Collaborative Bio-Inspired Algorithms

Lecture 9: Engineering Artificial Immune Systems

Prof Jon Timmis

November 2, 2010

Outline

AIS Framework

Representation

Affinity Measures

Algorithms

Applying to simple pattern recognition

Bias

Summary

Simple Framework for Engineering AIS

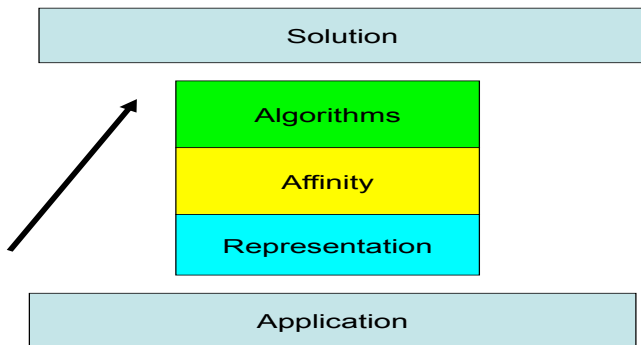


Figure: General framework for engineering AIS [1]

Shape Space

- ▶ Need to represent the data. This can be done in many ways.
- ▶ Known as *shape space*, many different types binary, real value etc. . . .

Representation

- ▶ Very common to mix up representations

	Description	Date	Flight	Country	From	To	Price (£)
Antibody (Ab ₁):	Business	1996	212	Brazil	Campinas	Greece	546.78
Antibody (Ab ₂):	Holiday	2000	312	U.K.	London	Paris	102.35
Antigen (Ag):	Holiday	2000	212	U.K.	London	Greece	546.78
Match Ag - Ab ₁ :	0	0	1	0	0	1	1
Match Ag - Ab ₂ :	1	1	0	1	1	0	0

Figure: More complicated shape space

Affinity Measures

- ▶ Computationally, the degree of interaction of an antibody-antigen or antibody-antibody can be evaluated by a distance or affinity measure
- ▶ The choice of affinity measure is crucial
 - ▶ It alters the shape-space topology
 - ▶ It will introduce an inductive bias into the algorithm
 - ▶ It needs to take into account the data-set used and the problem you are trying to solve

How similar is something?

In a binary shape space, can use, for example, Hamming distance, or r-contiguous matching

Mutation

Need to add in diversity. This can be achieved in many ways via mutation.

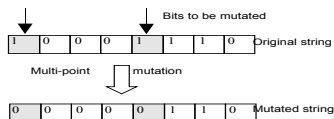


Figure: Mutation of a simple binary representation

Clonal Selection

input : S = set of patterns to be recognised, n the number of worst elements to select for removal

output: M = set of memory detectors capable of classifying unseen patterns

begin

 Create an initial random set of antibodies, A

forall *patterns in S do*

 Determine the affinity with each antibody in A

 Generate clones of a subset of the antibodies in A with the highest affinity. The number of clones for an antibody is proportional to its affinity

 Mutate attributes of these clones inversely proportional to its affinity. Add these clones to the set A , and place a copy of the highest affinity antibodies in A into the memory set, M

 Replace the n lowest affinity antibodies in A with new randomly generated antibodies

end

end

Algorithm 1: Generic Clonal Selection Algorithm, after [3]

Clonal selection: Simple case study

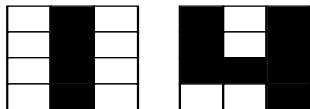


Figure: Generate a simple set of detectors capable of identifying a pattern

Clonal selection: Simple case study

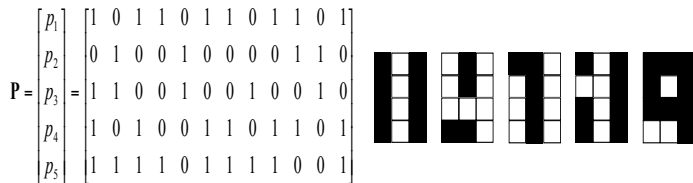


Figure: Decide on how to represent the problem

Clonal selection: Simple case study

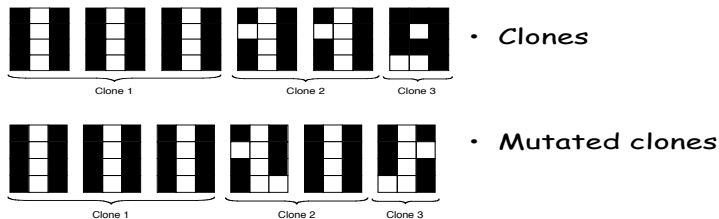


Figure: Use mutation and selection to evolve a good set of solutions

Inductive Bias

- ▶ Can affect choice of representation and affinity functions [2]
- ▶ It is any (explicit or implicit) bias favouring one hypothesis over another;
- ▶ All learning algorithms have it, otherwise it would only perform rote learning
- ▶ Bias is domain dependant
- ▶ Therefore, can not always say A is better than B

Inductive Bias

- ▶ **Representation Bias**
 - ▶ Associated with the knowledge representation of the algorithm
- ▶ **Preference Bias**
 - ▶ Associated with the evaluation function employed
 - ▶ Use distance to measure affinity (this is important for AIS in particular)
- ▶ Make sure you pick the right representation . . .

Choice of affinity function

- ▶ Choice of function should take into account the data being mined as they will all have a bias . . .
- ▶ Binary Representation
 - ▶ Typically employ Hamming or r -contiguous rule
 - ▶ Argued that r -contiguous is more biologically plausible, therefore, use it . . . not so.
 - ▶ This assumes an ordering within the data that may not exist and will introduce a positional bias
 - ▶ In data mining, quite common not to have unordered sets, representing the data when doing classification.
 - ▶ Therefore, a measure that takes into account position is not needed.
- ▶ Compare two different affinity measures . . .

Distance measures

- ▶ Continuous Representation
- ▶ Vast majority of AIS use Euclidean .. Because ... ?
- ▶ Also is Manhattan. They will produce different results ... They have different inductive biases and are more effective for different data sets

$$Dist(Ab, Ag) = \sqrt{(\sum (Abi - Agi)^2)} \text{ (Euclidan)}$$

$$Dist(Ab, Ag) = \sum |Abi - Agi| \text{ (Manhattan) How do they differ?}$$

Distance measures

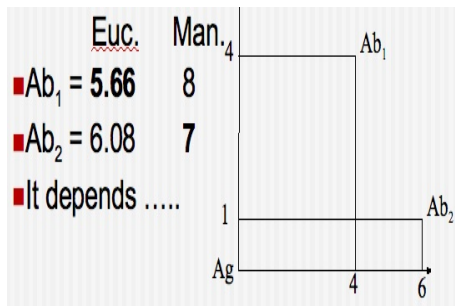


Figure: Difference in metrics

What did we cover?

- ▶ Basic engineering framework: representation, affinity, algorithms
- ▶ Clonal selection algorithm
- ▶ Remember, many different types of the clonal selection algorithms
- ▶ Careful choice of representation and affinity measures.



L. N. de Castro and J. Timmis.

Artificial Immune Systems: A New Computational Intelligence Approach.

Springer, 2002.



A. Freitas and J. Timmis.

Revisiting the foundations of artificial immune systems for data mining.

IEEE Trans. Evol. Comp., 11(4):521–540, 2007.



J. Timmis, P. S. Andrews, N. Owens, and E. Clark.

An interdisciplinary perspective on artificial immune systems.

Evolutionary Intelligence, 1(1):5–26, 2008.