

## Application du système immunitaire artificiel pour la reconnaissance des chiffres

Hiba KHELIL<sup>1</sup>, Abdelkader BENYETTOU<sup>1</sup>, Abdel BELAÏD<sup>2</sup>

<sup>1</sup>Laboratoire SIMPA

Université des Sciences et Technologie d'Oran – Mohammed BOUDIAF

BP 1505 Mnaouer Oran - ALGERIE

<sup>2</sup>LORIA-CNRS

Campus Scientifique BP 239

54506 Vandoeuvre-Lés-Nancy FRANCE

[hibakhelil@yahoo.fr](mailto:hibakhelil@yahoo.fr) [benyettou@simpa-usto.dz](mailto:benyettou@simpa-usto.dz) [abelaid@loria.fr](mailto:abelaid@loria.fr)

### Résumé

*La reconnaissance automatique de l'écriture occupe un espace important dans la recherche scientifique car elle offre une facilité d'utilisation dans différents domaines d'application : domaine bancaire, postal, le e-commerce... De nombreuses méthodes ont été utilisées pour la reconnaissance d'écriture, dans cet article nous présenterons des méthodes inspirées du système immunitaire naturel que nous appliquerons pour la reconnaissance des chiffres. Des résultats satisfaisants ont été notés durant les expériences d'un taux maximal de 95% en vu d'amélioration par hybridation avec des méthodes d'optimisations.*

**Mots-clés:** Reconnaissance des chiffres, système immunitaire naturel (NIS), système immunitaire artificiel (AIS), antigène, cellules B mémoire.

### 1. Introduction

Dans le but de résoudre des problèmes complexes, des idées inspirées à partir de mécanismes naturels ont été exploitées pour développer des heuristiques inspirées de la nature, le système immunitaire artificiel (AIS) est un paradigme récent qui tente de capturer des caractéristiques intéressantes des systèmes immunitaires naturels (NIS), comme la mémorisation, la reconnaissance de formes, l'apprentissage, et les capacités d'adaptation [1] [2] [3] [4] [5], la détection d'intrusion dans les réseaux informatiques [6], la robotique [7], l'apprentissage machine [8], etc.

Dans cet article, on donnera quelques notions de base sur le NIS, ensuite on présentera par simulation quelques algorithmes immunitaires pour la reconnaissance des formes [9] [10] [11] [13] [14]. On discutera les résultats préliminaires obtenus par l'application du AIS pour la reconnaissance des chiffres suivis d'une conclusion.

### 2. NIS : Le système immunitaire naturel

Le NIS est un système complexe qui peut être vu de différents angles : molécules, cellules et organes. Le système doit protéger le corps des entités dangereuses appelées antigènes [12]. Les éléments de base du NIS sont les globules blancs ou lymphocytes. Pour pouvoir identifier les autres molécules, des lymphocytes particuliers (cellules B) produisent des récepteurs, appelés anticorps (paratope), responsables à reconnaître des antigènes. Le paratope se lie à une partie spécifique de l'antigène appelée épitope. Le degré de cette liaison ou affinité est très fort si seulement les deux formes sont complémentaires.

Ainsi, la reconnaissance d'un antigène par une cellule B est en fonction de l'affinité entre les anticorps de la cellule B et cet antigène. Les cellules B qui reconnaissent mieux l'antigène vont être proliférées en se clonant, selon le principe de la sélection clonale [2]. Les clones subissent alors des mutations somatiques qui vont promouvoir leur variation génétique. Lorsque la population atteint la maturité, les clones se différencient en cellules mémoires et cellules plasma. Cette expansion clonale confère au système immunitaire sa mémoire. D'un autre côté, les cellules B avec une faible affinité seront mutées, ou détruites par sélection négative.

### 3. AIS : Le système immunitaire artificiel

L'AIS est un modèle qui englobe à la fois, des principes mathématiques et biologiques, car le NIS offre des caractéristiques intéressantes comme la mémorisation et l'apprentissage qui seront utiles dans le domaine de la reconnaissance des formes et autres. Dans la session qui suit, on présentera chronologiquement quelques algorithmes du AIS proposés initialement par A. Watkins dans sa thèse de Master à l'université de Mississippi en 2001 [9], ensuite une amélioration par A.

Watkins, J. Timmis et L. Boggess en 2004 [10] validée en 2005 dans la thèse PhD de A. Watkins à l'université de Kent [11].

## 4. AIRS : Système immunitaire artificiel pour la reconnaissance des formes

L'algorithme d'apprentissage immunitaire artificiel, proposé dans [9], nécessite l'utilisation d'un ensemble d'antigènes comme données d'apprentissage dont le système doit produire un ensemble d'anticorps qui sera utile pour la phase de classification. Le déroulement de l'apprentissage (AIRS) est divisé en quatre parties dont chacune est une introduction à la suivante.

### 4.1. Etape d'initialisation

Cette partie est une préparation des données utiles pour le bon déroulement de l'apprentissage dont on va extraire les données d'apprentissage (antigènes) qui seront par la suite normalisées sur l'intervalle [0, 1].

Par la suite un seuil d'affinité est calculé à partir de cet ensemble d'antigènes, représentant l'affinité moyenne entre tous les antigènes deux à deux (voir Eq.1):

$$seuil\_d'affinité = \frac{\sum_{i=1}^n \sum_{j=i+1}^n affinité(ag_i, ag_j)}{n(n-1)/2} \quad (1)$$

Avec  $ag_i$  et  $ag_j$  deux antigènes et  $affinité(ag_i, ag_j)$  retourne la distance euclidienne entre ces deux antigènes.

La dernière étape consiste à initialiser l'ensemble des cellules mémoires (anticorps) et la population des ARB<sup>1</sup> (Artificial Recognition Ball), par tirage aléatoire des exemples de l'ensemble d'apprentissage (antigènes).

### 4.2. Etape d'identification de cellules B et génération des ARBs

Une fois que l'initialisation est achevée, cette étape aura lieu pour chaque nouvel antigène introduit. Une cellule mémoire, nommée  $mc_{match}$ , est sélectionnée à partir de l'ensemble des cellules B ( $MC$ ). Elle est tirée de telle sorte qu'elle ressemble<sup>2</sup> le plus à l'antigène en cours de traitement (la plus grande valeur de stimulation) selon Eq.2.

$$mc_{match} = \arg \max_{mc \in MC} stimulation(ag, mc) \quad (2)$$

Une fois que la cellule  $mc_{match}$  est sélectionnée, elle servira pour générer les nouveaux ARBs (clonage). Cet ensemble sera additionné à l'ensemble total des ARBs généré par l'ensemble des antigènes préalablement traité.

Le nombre de clones de la cellule  $mc_{match}$  sera limité par la formule (3):

$$nbr\_clone = hyper\_clonal\_rate * clonal\_rate * stimulation(mc_{match}, ag_i) \quad (3)$$

Notons que chaque ARB, généré par  $mc_{match}$ , est muté selon l'algorithme de mutation décrit dans [9]. La mutation est un opérateur d'exploration de l'espace de recherche alors que le clonage participe à la survie de la cellule  $mc_{match}$ .

### 4.3. Etape de compétition des ressources et développement des cellules mémoire candidates

Cette partie est complémentaire de la précédente, car elle complète les informations des ARBs générés en calculant leurs ressources selon (4) entre chaque anticorps avec l'antigène en cours de traitement. Ces ressources sont mises à jour durant l'apprentissage et chaque ARB n'ayant pas de ressources est supprimé de l'ensemble des ARBs.

$$ressources = stimulation(ag, anticorps) * clonal\_rate \quad (4)$$

Les étapes 4.2 et 4.3 sont exécutées pour un antigène donné jusqu'à ce que la condition  $si \geq seuil\_d'affinité$  soit vérifiée, avec

$$s_i = \frac{\sum_{j=1}^{|AB_i|} ab_{j,stim}}{|AB_i|}, ab_{j} \in AB_i \quad (5)$$

Lorsque la condition d'arrêt est atteinte, l'ensemble des ARBs obtenus est prêt pour être introduit à l'étape finale de l'apprentissage (4.4).

### 4.4. Etape d'introduction des cellules mémoires

Cette étape consiste à choisir à partir des ARBs la cellule candidate qui convient le plus à l'antigène que  $mc_{match}$  en terme de similarité. La cellule candidate sera additionnée à l'ensemble des cellules mémoires seulement si elle retourne une valeur de stimulation plus élevée que  $mc_{match}$ , sinon la cellule  $mc_{match}$  sera retirée de l'ensemble si son affinité avec la cellule candidate ne dépasse pas la valeur  $AT * ATS$ <sup>3</sup>.

Dans cette partie, l'algorithme d'apprentissage présenté est celui d'un AIS destiné pour la reconnaissance des formes (AIRS); d'autres modifications portées à cet algorithme seront présentées dans les prochaines sessions (5 et 6).

## 5. AIRS2 : AIRS amélioré

Cet algorithme ne diffère pas trop de l'original (AIRS), si ce n'est au niveau de la mutation et de la compétition des ressources où les auteurs A. Watkins, J. Timmis et L. Boggess [10] ont introduit la notion de stimulation anticorps-antigène dans la mutation, et ne

<sup>1</sup> Les ARBs sont un magasin de cellules B contenant chacun des informations sur la cellule (anticorps, ressource, classe).

<sup>2</sup> La ressemblance est calculée par la formule de stimulation :  $stimulation(ag, mc) = 1 - affinité(ag, mc)$

<sup>3</sup> Avec  $AT$  est le seuil d'affinité calculé à partir de 4.1 et  $ATS$  est une constante dans l'intervalle [0,1]

prennent en considération durant le traitement que des anticorps de même classe que l'antigène.

## 6. L'apprentissage d'algorithme AIRS et AIRS2 par usage de facteur

D'autres modifications ont été portées aussi sur AIRS et AIRS2 dans la dernière étape d'apprentissage (introduction des cellules mémoires) plus précisément sur le critère d'introduction de la cellule.

$$\begin{aligned}
 CandStim &\leftarrow Stimulation(ag, mc_{candidate}) \\
 MatchStim &\leftarrow Stimulation(ag, mc_{match}) \\
 CellAff &\leftarrow affinity(mc_{candidate}, mc_{match}) \\
 \text{if } (CandStim > MatchStim) & \quad (7) \\
 \quad \text{if } (CellAff < AT * ATS) & \\
 \quad \quad MC &\leftarrow MC - mc_{match} \\
 \quad \quad MC &\leftarrow MC \cup mc_{candidate}
 \end{aligned}$$

Le code présenté dans (7) représente une partie de la dernière étape d'apprentissage d'AIRS et AIRS2. Les modifications portées sur cette partie consistent d'ajouter un facteur (voir la condition 8) qui servira à éliminer la cellule  $mc_{match}$  très ressemblantes à la cellule  $mc_{candidate}$ .

$$\text{if } (CellAff < AT * ATS + factor) \quad (8)$$

Avec  $factor$  est une valeur calculée en fonction de nombre de classe  $np$  donné dans l'équation (9):

$$factor = AT * ATS * dampener * \log(np) \quad (9)$$

$ATS$  et  $dampener$  se sont deux paramètres entre 0 et 1.

Cette modification est portée pour qu'il n'ait pas beaucoup de cellules B très ressemblantes, dans l'idée de minimiser la population de cellule et relaxer le traitement aux prochaines itérations. La session 9.2 représente l'évolution des cellules B pour chaque méthode de AIS.

## 7. Classification

A la fin de l'apprentissage, les cellules B générées seront prêtes pour la phase de classification durant lequel le nouvel antigène sera affecté à la classe la plus adéquate en utilisant le principe de k-means entre l'antigène et l'ensemble des noyaux des cellules B de chaque classe.

## 8. Etude sur la base des chiffres

Pour la construction de la base de données; des exemples sont pris manuellement, numérisés par le scanner et un petit programme développé pour réduire la

dimension des chiffres scannés, pour obtenir à la fin une représentation binaire des chiffres. Voici quelques étapes indiquées pour la construction de la base des chiffres :

### 8.1. Réduction de dimension de représentation

Après l'obtention des chiffres numérisés, il est préférable de réduire la dimension de représentation. Cette réduction est faite selon un principe de compression, indiqué comme suit:

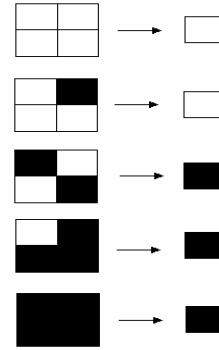


Figure 1 : Méthode pour la compression d'image

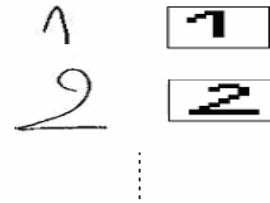


Figure 2 : Exemple sur la réduction d'espace de représentation des chiffres en image

### 8.2. Codage des données

Une fois que la compression est établie, l'étape suivante consiste à charger ces images dans des structures de données exploitables par le programme d'apprentissage et test. La figure 3 montre l'étape de codage:

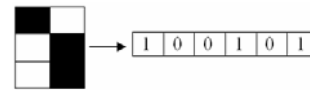


Figure 3 : Codage d'image binaire

## 9. Etude expérimentale

Pour une étude comparative sur la reconnaissance des chiffres, on a appliqué les quatre méthodes en utilisant le langage C++ sous l'environnement linux version Mandriva 2006. L'apprentissage s'est déroulé pour un nombre de générations  $G = 50$ , pour obtenir à la fin un ensemble de cellules B pour chaque classe d'antigènes (chiffre); ces cellules B seront utiles comme nous avons vu dans la phase de classification.

### 9.1. Paramètres d'apprentissage et résultats de classification

L'apprentissage nécessite de fixer les paramètres d'apprentissage suivant : *hyper\_clonal\_rate*, *clonal\_rate* et *mutation\_rate* utiles pour le calcul du nombre de clones, les ressources d'ARBs et la mutation. Les valeurs des paramètres sont données dans la table 1 :

Paramètres	Type	Valeur
Hyper_clonal_rate	Entier	30
Clonal_rate	Entier	20
Mutation_rate	Réel [0,1]	0.25

**Table 1 :** Paramètres d'apprentissage

A la fin d'apprentissage, les cellules B obtenues seront utiles pour la phase de test. Dans ce travail nous avons pris en considération le principe de distance euclidienne pour l'algorithme de classification KMeans. Les taux de reconnaissance pour chaque méthode appliquée sont donnés dans la table 2 :

	Expérience 1		Expérience 2	
	%Train1	%Test1	%Train2	%Test2
AIRS	95	86	85	84
AIRS2	95	84	88	76
AIRS + Fact	95	86	86	76
AIRS2 + Fact	95	84	88	83

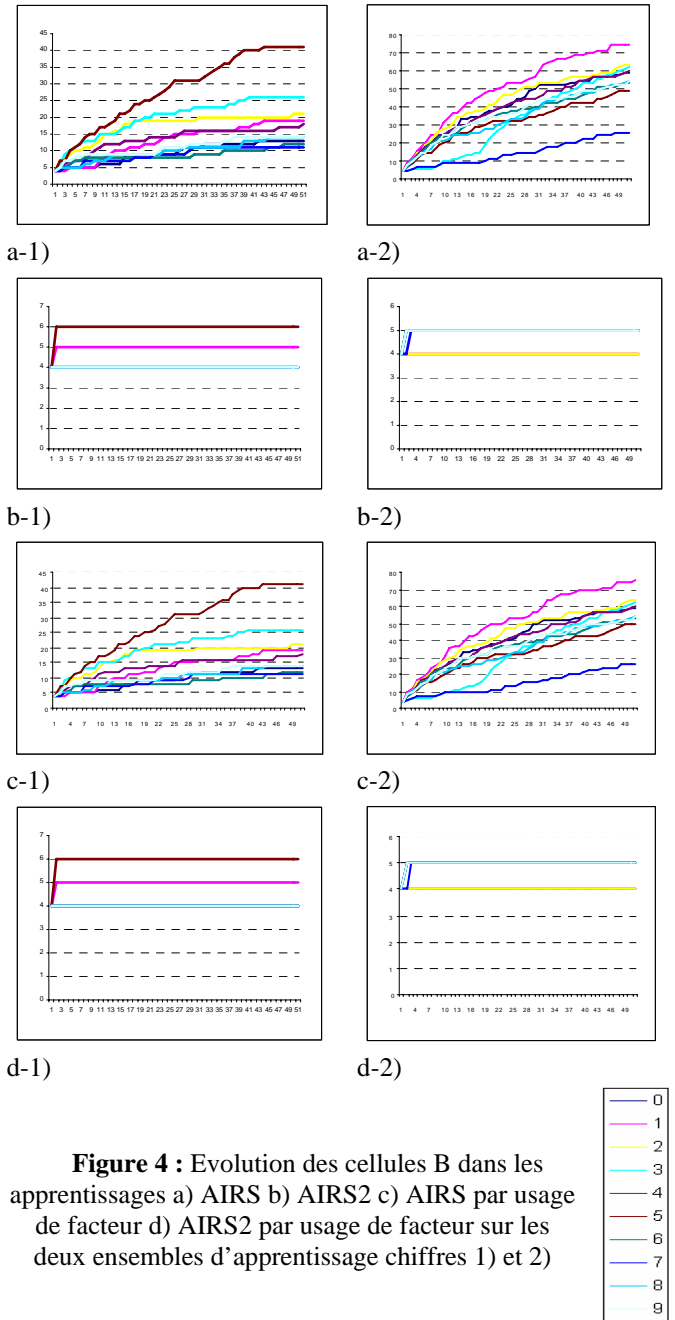
**Table 2:** Taux de reconnaissance des chiffres pour deux expériences sur deux ensemble d'apprentissage et test différents.

Selon les résultats obtenus, on remarque que le facteur n'a presque aucune influence sur le taux de reconnaissance ainsi que les résultats obtenus dans AIRS2 et AIRS2 par usage de facteur sont inférieurs que AIRS et AIRS par usage de facteur pour un nombre d'itérations identique. La discussion de ces résultats sera portée dans 9.2.

### 9.2. Evolution des cellules B et discussions des résultats

Dans l'étape d'initialisation, les ensembles de cellules B sont initialisés équitablement, dont ils subiront une évolution à travers les itérations et au fur et à mesure que les antigènes seront traités. Ce qui fait qu'à la fin d'apprentissage on obtient un nombre de cellules B non équitable pour chaque classe, cette différence est due au degré de ressemblance des exemples d'antigènes car pour des antigènes non hétérogènes il est difficile que les cellules B (produites durant l'apprentissage) de les reconnaître pour cela un nombre de cellules relativement élevé sera produit pour que la stimulation antigène-anticorps soit importante (faible affinité).

Les figures suivantes représentent l'évolution des cellules B durant l'apprentissage des systèmes immunitaires pour la reconnaissance des chiffres présentés dans cet article :



**Figure 4 :** Evolution des cellules B dans les apprentissages a) AIRS b) AIRS2 c) AIRS par usage de facteur d) AIRS2 par usage de facteur sur les deux ensembles d'apprentissage chiffres 1) et 2)

Selon ces figures, nous remarquons que l'évolution des cellules B augmente au fur et à mesure des itérations, seulement pour le cas b) et d) l'évolution n'est pas trop rapide voir stable car durant l'apprentissage il y'a moins de cellules B utilisées, car (comme indiqué dans 5) le système prend en considération que celles de même classe que l'antigène en cours.

En contre partie, l'évolution des cellules dans c) et d) est surtout logarithmique ce qui n'est pas le cas pour la reconnaissance des chiffres car le corpus d'apprentissage est assez réduit ; par contre dans [15] l'évolution en logarithme est très remarquable car cela revient à l'usage du facteur calculé à partir de  $\log(np)$  (voir Eq.9).

Le nombre réduit de cellules dans AIRS2 et AIRS2 par usage de facteur peut rendre le système rigide en comparant avec AIRS et AIRS par usage de facteur, ce qui peut être expliqué par le taux de reconnaissance inférieur ce qui est le cas pour la reconnaissance de la parole dans [15]. Il est possible que b) et d) représentent un avantage car pour un nombre de cellules très réduit on est arrivé à des résultats peu moins de a) et c) il est possible d'avoir un taux plus important pour un nombre d'itération plus élevé que 50 générations.

## 10. Conclusion

Ce travail a consisté à appliquer une approche immunitaire bio-inspirée pour la reconnaissance des chiffres; de ce fait on s'est intéressé à l'immuno-computing qui est une approche très récente en informatique bio-inspirée. On a présenté les différentes étapes d'algorithmes d'apprentissage ainsi que les résultats obtenus pour la reconnaissance des chiffres. Cette application peut être utile dans différents domaines, exemple au niveau bancaire où l'agent peut scanner seulement le chèque au lieu de saisir le montant demandé, etc.

Les résultats préliminaires obtenus sont encourageants, seulement certaines applications ne tolèrent pas la moindre erreur provoquée par le système, de ce fait il est souhaitable d'optimiser cette méthode en donnant plus d'importance aux paramètres d'apprentissage en appliquant des principes génétiques pour le choix de bons paramètres d'apprentissage et sur un corpus plus grand.

## 11. Références

- [1] A. Secker, A. Freitas et J. Timmis "AISEC: An artificial immune system for e-mail classification". Proc. Congress on Evolutionary Computation, Canberra, Australia, 2003, pp. 131-139.
- [2] M. Lingjun, V. D. P. Peter et W. A. Haiyang, "Comprehensive benchmark of the artificial immune recognition system (AIRS)". Proc. advanced data mining and applications ADMA, Chine, 2005, vol. 3584, pp. 575-582.
- [3] A. Deneche, S. Meshoul et M. Batouche, "Une approche hybride pour la reconnaissance des formes en utilisant un système immunitaire artificiel". Proc. graphic computer science, Biskra, Algeria, 2005.
- [4] A. Deneche, "Approches bio-inspirées pour la reconnaissance de formes", Master computer science in Mentouri university, Constantine, Algeria, 2006.
- [5] D. Goodman, L. Boggess et A. Watkins, "Artificial immune system classification of multiple class problems". Journal of Intelligent Engineering Systems Through Artificial Neural, 2002.
- [6] J. Kim et P. Bently, "Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator". Proc. Congress on Evolutionary Computation, South Korea, 2001, vol. 2, pp. 1244-1252.
- [7] J.H. Jun, D.W. Lee, K.B. Sim, "Realization of cooperative and swarm behavior in distributed autonomous robotic systems using artificial immune system". Proc. IEEE- int. Conf. of Man and Cybernetics, New York, 1999, vol. 6, pp. 614-619.
- [8] J. Timmis, "Artificial immune systems: a novel data analysis technique inspired by the immune network theory", PHD thesis, Wales university UK, 2000.
- [9] A. Watkins, "AIRS: A resource limited artificial immune classifier", Master thesis, university of Mississippi, 2001.
- [10] A. Watkins, J. Timmis et L. Boggess, "Artificial immune recognition system (airs): an immune inspired supervised learning algorithm", journal of Genetic Programming and Evolvable Machines, 2004, vol. 5, pp. 291-317.
- [11] A. Watkins, "Exploiting immunological metaphors in the development of serial, parallel, and distributed learning algorithms", PhD, Kent university, 2005.
- [12] P. Emilie, "Organisation du system immunitaire felin", thèse de doctorat à l'école national de Lyon, France, 2006.
- [13] A. Watkins, J. Timmis, "Artificial immune recognition system (airs): revisions and refinements". Proc. first International Conference on artificial immune system ICARIS, Kent university, 2005, pp. 173-181.
- [14] A. Watkins et L. Boggess, "A new classifier based on resources limited artificial immune systems", Proc. Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Honolulu, HI, USA, Kent University, 2005, pp. 1546-1551.
- [15] H. Khelil et A. Benyettou, "Artificial Immune System For Speech Recognition", International Arab Conference on Information Technology, ACIT'2007, Syria, 2007.