

L'algorithme des Mouches et l'approche Parisienne pour la classification non supervisée des images

FIZAZI IZABETENE Hadria GHARDAOUI Sarah

Université des sciences et technologie d'ORAN
Laboratoire SIMPA, équipe imagerie et décision

hadriafizazi@yahoo.com
ghasar85@gmail.com

Résumé : La classification non supervisée consiste à rechercher des groupes homogènes dans un ensemble de données. Ce problème est donc d'une grande complexité et l'utilisation des algorithmes d'approximation est inévitable. Parmi ces algorithmes nous nous sommes intéressés à l'Algorithme des Mouches pour faire une classification non supervisée des images. Ce nouvel algorithme qui bénéficie de la bonne codification des systèmes immunitaires et des opérateurs de variation des algorithmes évolutionnaires a été combiné avec l'Approche Parisienne. Cette dernière fait en sorte que la population entière constitue la solution, dont les principaux intérêts sont la représentation des individus de taille réduite ainsi qu'une convergence rapide.

Mots Clés : classification non supervisée, image, algorithme des mouches, approche parisienne, algorithme évolutif.

1. Introduction

L'actualité dans le domaine de la classification non supervisée des images tourne autour des approches évolutives. Plusieurs algorithmes ont été appliqués tel que les algorithmes génétiques dans [7] et les systèmes immunitaires artificiels dans [8]. Ces deux derniers ont prouvé leurs performances mais chacun d'eux présente un défaut ; d'où vient l'idée d'associer la puissance des algorithmes génétiques et de la codification des classes des systèmes immunitaires artificiels. Cela a été déjà appliqué avec succès dans quelques domaines tel que la robotique et la stéréovision [6]. Dans ce papier nous allons voir qu'il est même possible de l'utiliser pour résoudre le problème de la classification non supervisée des images.

2. Classification non supervisée (segmentation)

La classification non supervisée consiste à rechercher des groupes (classes) homogènes dans un ensemble de données. Les données appartenant à la même classe doivent être les plus proches entre elles. Dans le cas contraire, les données appartenant à des classes différentes ; sachant que le nombre de classes et leurs significations ne sont pas connus [1][2].

3. Approche Parisienne et algorithme des mouches

Les principaux intérêts de l'algorithme des mouches avec l'approche parisienne sont la représentation réduite des individus et une convergence rapide, d'où leur utilité et leur popularité en traitement temps réel que nécessite certaines applications robotiques. Nous allons voir qu'il est cependant possible de les utiliser pour résoudre le problème de classification non supervisée.

3.1. Approche Parisienne

Dans la plupart des approches évolutives, chaque individu de la population représente une solution potentielle au problème traité. Cette approche s'est révélée efficace pour les algorithmes génétiques, mais présente un problème si la solution encodée en individu est de grande taille. Manipuler une population d'un nombre important d'individus qui sont eux-mêmes volumineux peut s'avérer très coûteuse en espace et en temps. Ce problème a rendu difficile l'utilisation des approches évolutives dans des environnements où les contraintes de ressources et de temps sont grandes, notamment dans les systèmes embarqués et la robotique.

L'approche Parisienne, explore une nouvelle possibilité de codage de la solution. L'individu n'est donc plus une solution, mais un fragment de celle-ci, et il faut combiner tous les individus pour construire une solution.

Il y a deux conditions nécessaires à la mise en œuvre de cette approche :

- La solution peut être décomposée en éléments séparés.

- Les individus (fragments de solution) peuvent être jugés séparément par une fonction de fitness (ou fonction objectif), et cette fonction doit être représentative de la « qualité » de la solution générale. [6]

Par la suite nous verrons comment le problème de classification non supervisée peut satisfaire ces conditions.

3.2. L'algorithme des Mouches :

L'algorithme des mouches [5] est un algorithme évolutif particulier. Il présente les mêmes opérateurs de variation que les algorithmes évolutionnaires (croisement et mutation), et y ajoute la migration.

4. Application de l'Approche Parisienne et algorithme des mouches à la classification

Pour adapter l'algorithme des mouches avec l'Approche Parisienne au problème de la classification non supervisée des images, d'abord il faut satisfaire les deux conditions citées plus haut : à savoir coder les individus de façon à ce que toute la population soit une solution potentielle au problème, et trouver une fonction qui évalue ces individus de façon à ce que la qualité de la classification soit bonne.

4.1. Initialisation de la population:

Pour la codification, on a exploité celle des Systèmes Immunitaires. On crée aléatoirement une population de 'u' anticorps, qui vont être les individus de la population P_t . Chaque individu (encodé avec des nombres réels positifs) représente une classe. La taille de la population est déterminée par le nombre maximal de classes.

4.2. Croisement

Il est important de noter que la méthode de croisement choisie dépend de la représentation des individus (dans notre travail la représentation est réelle). Ainsi, pour calculer les nouveaux individus (les enfants) P_β il suffit de choisir un point de coupure (c'est-à-dire une position dans la chaîne) à partir de laquelle les données sont échangées entre les deux parents.

4.3. Mutation

La mutation est le deuxième opérateur de variation utilisé juste après le croisement. Elle consiste à modifier les individus issus du croisement.

Dans le cas d'une représentation réelle, la manière la plus simple est d'ajouter à chaque composante de l'individu une réalisation d'une loi normale centrée (d'espérance nulle et d'écart-type fixé par l'utilisateur).

4.4. Migration

La migration consiste à introduire un nombre prédéfini de nouveaux individus P_γ créés aléatoirement dans la population courante, afin d'assurer la diversité de la population.

4.5. Evaluation des individus et sélection pour le remplacement

La dernière étape de l'algorithme évolutionnaire est l'évaluation des individus et leurs sélections pour le remplacement [3]. En ce qui concerne l'évaluation, la fonction utilisée est la suivante :

$$Sim(a_j) = 1 - \frac{\sum_{i=1}^n d(a_j, o_i)}{n} \quad (1)$$

Où ' o_i ' les objets à classer, ' a_j ' anticorps et ' d ' est une distance normalisée quelconque.

Après avoir obtenu une population P' (les parents P_t , ainsi que de P_β enfants à partir du croisement de parents, P_α des mutations des enfants et P_γ issu de la migration. il nous faut une stratégie pour sélectionner les 'u' individus qui feront partie de la population de la génération suivante (c'est-à-dire à la prochaine itération). Plusieurs remplacements existent, Celui qui a été jugé efficace pour notre application est le remplacement élitiste et qui a pour principe de garder les parents qui sont plus performants que les enfants, sans faire de distinction entre parents et enfants et de garder les μ meilleurs individus parmi les P' .

Algorithme des mouches :

1. Initialiser une population P_t de μ individus.
 2. Tant que (condition d'arrêt = Faux) faire
 - Créer β individus par croisement à partir de P_t pour former la population P_β
 - Créer une population P_α avec ces individus mutés.
 - Générer aléatoirement γ individus pour former la population P_γ
 - $P' \leftarrow P_t + P_\beta + P_\alpha + P_\gamma$
 - Evaluer tous les individus de P' grâce à une fonction de fitness et choisir les μ meilleurs ($P_t \leftarrow$ meilleurs individus de P')
- Fin tant que.

Une fois la condition d'arrêt est vérifiée (nombre de générations atteint), on aura en sortie les 'u' meilleurs individus représentatifs des classes. A ce stade, on calcule la distance entre chaque pixel de l'image et les 'u' meilleurs individus, ainsi l'affectation sera à l'individu le plus proche.

4. mise en œuvre

L'application de l'algorithme des mouches, comme tout algorithme évolutif, est influencé par des facteurs tel que la forme de codage (binaire ou nombre réel), la taille de la population, la fonction de remise en forme, les opérateurs génétiques (croisement, mutation)... Dans ce qui suit nous allons discuter les paramètres qu'on a fixé en expliquant les raisons, ainsi que ceux qu'on va varier.

5.1. Discussion des paramètres

-pour fixer le nombre maximal de classes on estime (de façon visuelle) le nombre de classes constituant l'image, et par la suite on le multiplie par deux.

- le modèle de croisement utilisé est la généralisation de la méthode de croisement de chaînes binaires avec un seul point de coupure pour sa simplicité d'implémentation.

-Pour ce qui concerne l'écart type de la loi normale (mutation), il est important de noter que malgré la différence existante entre les individus, leur performance soit a peu près égale. Ce phénomène induit une convergence prématurée. Pour palier ce risque il vaut mieux réduire la variance en utilisant la loi normale centrée réduite (espérance=0, écart type = 1).

- un autre paramètre à fixer, le nombre d'individus migrés. Ce paramètre influe sur la convergence de l'algorithme en cas de migrer un nombre élevé d'individus ; dans le cas contraire, la diversité de la population s'affaiblit. Pour cette raison, il est préférable de migrer 'u' individus.

-le nombre de générations 'It' est le seul paramètre qui reste à varier pour savoir ce que cela donne.

Remarque. Il est important de noter que l'individu est un vecteur de trois éléments (RVB).

5.2. Discussion des résultats

Nous allons appliquer l'algorithme des mouches avec l'approche parisienne pour la classification non supervisée des images. Les images utilisées sont illustrées par la Figure 1



Figure 1. Images utilisées image1, image 2, image3 (de gauche à droite)

5.2.1. Essai 1

Comme premier essai on va appliquer notre algorithme sur les trois images tout en fixant le nombre de génération à 1. Les résultats obtenus sont données par Figure2.



Figure 2. Images résultantes de la classification non supervisée par les mouches (It=1)

	Clas1	Clas2	Clas3	Clas4	Clas5	Clas6	Clas 7 et 8
Img 1	36	2	175	43	0	0	0
Img 2	537	381	3574	20142	0	4607	0
Img 3	36544	0	815	14317	49222	0	0

Tableau1. Résultats de la classification It=1

Le tableau1 montre que le nombre de classes constituant l'image est reconnu correctement pour les trois images malgré les l'existence de petites confusions qui ne dépasse pas les deux pourcent, ainsi quelques pixels mal classés qu'on peut remarquer dans les images classifiées.

Cette confusion peut être causée par l'initialisation aléatoire de la population. C'est pour cette raison, dans les essais qui suivent on va garder la même population initiale tout en variant le nombre de générations.

5.2.2. Essai 2

Dans cet essai on va s'intéresser qu'à la première image avec une fixation du nombre de générations à 10 puis à 40

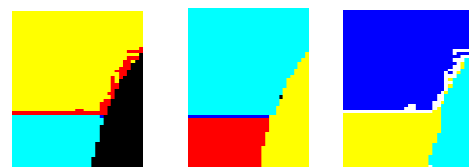


Figure 3. Résultats de la classification avec It=1, It=10, It=40 (de gauche à droite)

Img1	Clas1	Clas2	Clas3	Clas4	Clas5	Clas6
It =1	376	294	49	897	0	1
It =10	1	899	327	353	0	37
It =40	69	396	0	282	870	0

Tableau2. Résultats de la classification en variant le nombre de générations.

La Figure 3 et le tableau2 Ces résultats montrent que la variation du nombre de générations influe sur les résultats obtenus de deux façons différentes; de plus on augmente le nombre de générations on obtient une meilleure classification (It = 10) mais arrivant à un certain nombre de générations (It = 40) la performance de notre algorithme commence à dégrader. Cette dégradation est due à l'opérateur de migration qui consiste à intégrer des individus de manière uniforme à chaque génération.

5.2.3. Essai 2

Et puisque notre algorithme a été conçu dont le but est de réduire le temps d'exécution ainsi que la taille en espace mémoire, notre curiosité nous a mené à le comparer (en terme de performance et de temps d'exécution) avec un algorithme appliqué au niveau de notre laboratoire (colonies de fourmis "AntClust") [9] pour la classification non supervisée des images.

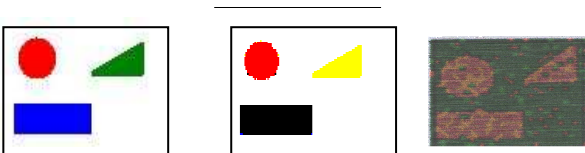


Figure 4. Image originale, classification par les mouches, classification par AntClust (de gauche à droite).

Cette figure montre que, la qualité de la classification en utilisant l'algorithme des mouches est nettement meilleure qu'en utilisant l'algorithme AntClust, sachant que l'exécution de ce dernier est trop lente (4heurs pour cette image) en la comparant avec celle de notre algorithme (1 seconde).

6. Conclusion

L'algorithme des mouches avec l'Approche Parisienne est une méthode évolutive récente qui présente l'avantage d'une convergence rapide et d'une population de taille réduite en espace mémoire. Les résultats obtenus se sont révélés globalement meilleurs que ceux des algorithmes évolutionnaires utilisant des codifications classiques. Néanmoins, il existe quelques classes non significatives ainsi que des petites confusions.

Ce problème peut être résolu en variant le nombre de générations, mais cette variation peut mener à une divergence causée par l'opérateur de migration en introduisant des individus de manière uniforme à chaque génération ce qui les empêche de se regrouper en classes ; donc il serait certainement plus efficace de faire en sorte que cet opérateur introduise des individus autour des groupes et non pas de manière uniforme.

Bibliographie

- [1] Admane, L. et Benatchba, K. et Koudil, M. et Siad, L. et Maziz, S. "AntPart : an algorithm for the unsupervised classification problem using ants." Laboratoire LMCS. INI. 2006.
- [2] Belaïd, A. "Implémentation d'une méthode de classification non supervisée de données à l'aide des AG évolutionnaires ". Mémoire d'ingénieur. INI. 2008.
- [3] Bhandari, D. et Murthy, C. A. et Pal, S. K. « Genetic algorithm with elitist model and its convergence ». *International Journal of Pattern Recognition and Artificial Intelligence*, 10(6). Pages : 731-747. 1996.
- [4] Glickman, M. et Balthrop, J. et Forrest, S. "A Machine Learning Evaluation of an Artificial Immune System". *Evolutionary Computation Journal*, Vol 13, n° 2, pages 179-212. 2005.
- [5] Lamali, M. "Approches évolutives pour la classification non supervisée de données" Ecole nationale Supérieure d'Informatique (ESI), Oued-Smar, Alger, 2009.
- [6] Louchet, J. et Guyon, M. et Lesot, M-J. et Boumaza, A. "L'algorithme des mouches dynamiques : guider un robot par évolution artificielle en temps réel ". *Extraction des Connaissances et Apprentissage : Apprentissage et Evolution*, vol.1- N.3. Pages 115-130. 2001.
- [7] Yang, F.Y. et Lohmann, P. et Heipke, C. "Genetic Algorithms For Multi-Spectral Image Classification"

Institute of Photogrammetry and GeoInformation, Leibniz University of Hannover, *springer*2006.

- [8] Zhang, L. et Zhong, Y. et Li, P. "An Unsupervised Artificial Immune Classifier for Multi/Hyperspectral Remote Sensing Imagery" *IEEE Transactions On Geoscience And Remote Sensing*, VOL. 44, NO. 2, FEBRUARY 2006
- [9] Mahiddine, A. Benamara, N. "Segmentation Des Images Par Colonies De Fourmis Approche Antclust" mémoire d'ingénieur département informatique usto2009.