

# Résumé de l'article "Substructural Niching in Estimation Distribution Algorithms"

Justine Lebrun

14 mars 2006

## 1 Résumé

Les auteurs présentent une nouvelle méthode de niche de sous-structure (eq à buildings blocks) pour les algorithmes d'estimation de distributions.

Leur méthode s'appuie sur trois points :

- L'identification de sous-structure et le problème de la décomposition
- La fitness d'estimation d'une sous-structure
- La préservation d'une niche de sous-structure

Ils utilisent la méthode du tournoi (RTS) pour évaluer leur méthodes.

## 2 Introduction

L'article situe le problème en rappelant la difficulté à construire des algorithmes génétiques (GA) capable de passer à l'échelle. Les algorithmes à estimation de distribution (EDA) ont, d'après l'article, remplacé efficacement certains GA opérateur classique de variation.

L'un des composant de l'EDA requiert une méthode de niche efficace. La plupart des méthodes de niches précédentes maintenaient la diversité au niveau individuel et non sur les sous-structures. Ils sont souvent peu adaptatif au niveau de la taille et de la distribution du problème, de plus ils n'exploitent pas le travail sous-jacent des algorithmes de liaisons.

Leur méthode utilise :

Pour l'identification d'une sous-structure → le modèle probabiliste construit par un EDA modifié.

Pour l'estimation d'une sous-structure → celle proposé par un des auteurs dans papier précédent.

Pour la préservation d'une niche → différentes méthodes ultérieures envisagées du moment qu'elles préservent les sous-structure hautement qualifiées dans les proportion désirée de manière stable

## 3 Algorithme génétique compacte étendu (eCGA)

Cette section présente une variation de l'EDA qui consiste à remplacer le traditionnel opérateur génétiques de variation et évolutionnaire en construisant un modèle probabiliste sur les solutions prometteuse et s'appuyer sur ce modèle pour générer de nouveau candidat.

### 3.1 Le modèle probabiliste

**Représentation** Les auteurs utilisent un modèle probabiliste de type MPM (Marginal product Model/Modèle des marginales à Posteriori).

**Paramètres de sélection de classe** eCGA s'appuie sur MDL (minimum description length). Il fait la somme de deux composantes :

Complexité du modèle

$$C_m = \log_2(n) \sum_{i=1}^m (2^{k_i} - 1)$$

Complexité de compression de la population

$$C_p = n \sum_{i=1}^m \sum_{j=1}^{2^{k_i}} -p_{ij} \log_2(p_{ij})$$

**Méthode de recherche de classe** La méthode commence par traité

Le modèle est représenté par un vecteur de probabilité contenant la proportion d'individu parmi ceux sélectionné qui ont une valeur de 1 (ou 0) pour chaque variable.

La méthode unifie les deux partitions qui ont le plus de rapport, jusqu'à qu'il n'y ait plus augmentation de la valeur "metric" (donné par MDL).

## 4 Niche de sous-structure

Les méthodes s'appuyant sur maintenance de la diversité individuellement dépendent fortement de la distribution des niches.

Comme Golberg ils pensent que l'identification, la décomposition et la mélange des sous-structures sont déterminant pour l'efficacité d'un algorithme génétique scalable.

### 4.1 Les 3 points de la méthodes

**Identification de Sous-structure.** Utilisation du modèle probabiliste construit par ecGA. D'après les auteurs d'autres techniques sont possibles.

**Estimation de sous-structure.** L'estimation d'une sous structure est le résultat de la différence entre la moyenne des fitness de tous les éléments contenant cette sous-structure et celle de tous les éléments.

**Préservation des niches.** N'importe quelle méthode du moment qu'elle permet de préserver des sous-structures en quantité désirée.

### 4.2 Quelques vérifications expérimentales

Les auteurs veulent vérifier l'exactitude et la stabilité de leur méthode. Pour cela il la teste sur quelques fonctions deceptives.

## 5 Expériences d'évaluation

### 5.1 Description

Le principe est de comparer sur divers plans (stabilité, effet de la population ) les résultats de leur méthode implémentée dans un eCGA avec ceux d'une autre méthode. Elle aussi est implémentée dans eCGA.

types problèmes fonctions deceptives de type m-4 trap.

méthode comparative RTS (Restricted Tournoi Selection)

## 5.2 Résultats

### Stabilité

Les résultats montrent une stabilité nettement plus élevée en faveur de leur méthode. Elle est toujours stable après 100e génération contre un éparpillement commençant dès la 20e génération chez RTS.

### Effet de la taille de la population

Une population de plus grande taille est requise pour maintenir l'optimum global sur plus grande période pour RTS. Cet effet est connu et analysé par Mahfoud pour les algorithmes traditionnels. RTS suit l'équation de Mahfoud pour la taille minimum de la population pour maintenir au moins une copie de l'optima. Leur méthode au lieu de suivre cette progression linéaire affiche une courbe à tendance logarithmique.

## 6 Conclusion

En conclusion, leur méthode fonctionne et l'idée clé d'une méthode de niche au niveau des sous-structures est de préserver celle qui conviennent le mieux et dans des proportions désirées pendant la recherche.

Ils ont aussi testé leur méthode sur une classe de problème. Leur tests sont concluants. Et de plus la méthode de niche sur sous-structure serait efficace avec des problèmes d'optimisation, hiérarchiques et dynamiques.

## 7 Discussions sur l'article

Les auteurs emploient un algorithme avec modèle probabiliste, RTS ne serait-il pas plus efficace avec un autre algorithme ?

Deuxièmement, les auteurs n'explicitent pas le choix de RTS comme méthode concurrente.

Ils ne précisent pas non la complexité de leur méthode or pour estimer une sous-structure ils utilisent la fitness des éléments.

En résumé, quelques tests supplémentaires auraient du être effectués pour évaluer la comparaison entre RTS et leur méthode afin de valider

## 8 Liste de documents annexes intéressants non cités par l'article

**généralités(dont les problème perceptif)**

SLIDE Université Utrecht : [http://www.cs.uu.nl/docs/vakken/ea/PMEA\\_4.pdf](http://www.cs.uu.nl/docs/vakken/ea/PMEA_4.pdf)

ARTICLE de V.J. Gillet, P.Willett et J.Bradshaw "Identification of Biological Activity Profiles Using Substructural Analysis and Genetic Algorithms"

### **Marginal Product Model**

THESE de A.Jalobeanu "Modèles, estimation bayésienne et algorithmes pour la déconvolution d'images satellitaires et aériennes"

ARTICLE de G.Harik "Linkage Learning via Probabilistic Modeling in the ECGA"

## **Niching**

THESE de J.Brownlee “Parallel niching genetic algorithms a crowding perspective”

ARTICLE de J-P. Watson “A performance Assesment of Modern Niching Methods for parameter Optimization Problèms”