

Adrien Maudet

Master Pro Informatique, Parcours II

Module Fouille de Données et Apprentissage : Présentation d'article

Article : Resource-Limited Genetic Programming: Replacing Tree Depth Limits

Résumé :

Cet article nous propose une méthode alternative à utiliser dans le cadre de la programmation génétique concernant des arbres : remplacer la méthode de limitation sur la profondeur des arbres par une limitation sur l'ensemble des ressources utilisé par toute la population

1. Introduction

Les auteurs présentent succinctement la programmation génétique sur des arbres et la problématique de l'augmentation de la taille des ressources utilisées par la population due à l'utilisation de certains opérateurs, qui entraînerait une stagnation du processus évolutif.

Après avoir présenté brièvement plusieurs méthodes pour contrer ce problème, l'article s'attarde sur la méthode la plus couramment utilisée, et celle défendue par les auteurs

2. Tree Depth Limits

Brève présentation de cette limitation qui donne une borne à la profondeur de chaque individu, et liste de ses inconvénients :

- Aucune limitation avant que la limite de profondeur ne soit atteinte
- Peut empêcher de trouver la solution en cas de problème très complexe
- Peut renvoyer une solution d'une certaine profondeur alors qu'il existe des solutions de moindre profondeur
- Utilisation restreinte à la programmation génétique sur des arbres

Une solution généralement avancée est de rendre cette profondeur limite dynamique en l'augmentant si nécessaire et en la diminuant si possible.

3. Limited Natural Resources in GP

Les auteurs présentent leur méthode : limiter les ressources au niveau de la population. Est considéré comme unité de ressource le nombre de nœuds de chaque individu, directement lié aux ressources utilisées pour évaluer un individu.

Comparaison avec une limitation de ressource au niveau biologique où les individus sont en compétition pour son exploitation. Cela soulève une interrogation : le lien qui semble évident pour les auteurs entre efficacité et taille de l'individu. (cf. Mon avis)

A chaque nouvelle génération les nouveaux individus survivent si il reste suffisamment de ressources disponibles pour exister, sur le principe du meilleur résultat, premier servi.

Cela entraîne théoriquement une diminution de la population (les individus tendant à s'améliorer et donc à devenir plus « gros »)

Un autre aspect soulevé est l'utilisation des ressources restantes, deux méthodes sont proposées :

- permettre grâce à ces ressources aux individus de la génération précédente de survivre.
- ne pas les utiliser.

4. Expérimentation

Les auteurs présentent ensuite une expérimentation sur un exemple précis :

- Régression sur 21 valeurs du polynôme $x^4 + x^3 + x^2 + x$, sur l'intervalle $[-1, 1]$
- Population initiale de 500 individus (constitué par la variable x et les opérateur $+$, $-$, $*$, $/$, \sin , \cos , \log , \exp)
- 50 générations
- Seul l'opération de Cross-Over est utilisée

L'expérience est appliquée 50 fois sur ces différentes méthodes :

- Pas de limitation : *None*
- Limitation sur la profondeur des arbres (17) : *Depth*
- Limitation sur les ressources globales (14500) avec allocation des ressources restantes au meilleur spécimen de la génération précédente : *Rsteady*
- Limitation sur les ressources globales sans allocation des ressources restantes : *Rlow*

On observe une sélection sur la reproduction, mais pas sur la survie (en dehors de la sélection propre à la limitation sur les ressources).

La valeur 14500 est obtenue en considérant le point où l'utilisation par génération de ressources avec la méthode sans limitation est égale à l'utilisation cumulée de ressource par la méthode avec limitation sur la profondeur des arbres. Même si effectivement le but recherché d'avoir une même utilisation de ressources entre les méthodes semblent réussie (cf. résultat) avec cette limitation, son fondement théorique n'est pas explicité et n'est pas évident.

5. Résultat

Intérêt porté sur 5 courbes :

- Total des ressources utilisées par génération
- Total cumulé des ressources utilisées par génération
- Taille de la population
- Taille moyenne des individus
- Evaluation des performances (fitness)

Les auteurs notent les observations suivantes :

- *Rsteady* et *Rlow* ont tendance à exploiter plus de ressources que *Depth* avant l'atteinte de la limitation, où la tendance s'inverse nettement
- Le total cumulé de ressources est similaire pour toutes les méthodes sauf *None*
- La population de *None* et de *Depth* reste proche de son origine (500), alors que pour *Rsteady* et *Rlow*, la population décroît fortement
- Parallèlement la taille des individus augmente partout, mais de manière nettement moins significative dans le cas de *None*
- Enfin du point de vue des performances, à l'exception du *None*, les trois autres techniques ont des résultats similaires

6. Conclusion et travaux à venir

- L'utilisation d'une limite sur les ressources globales ne diminue pas les performances par rapport à une limite sur la profondeur des arbres
- Elle ne les améliore pas non plus
- Elle résout par contre certain problème propre à la limite sur la profondeur (ceux décrits dans la partie 2) : pourtant l'expérience ne s'attarde sur aucun de ces aspects et aucun n'est prouvé.
- Reste le problème d'une augmentation non contrôlée avant l'atteinte de la limite. Les auteurs envisagent d'utiliser une limite variable (tout comme pour la limite sur la profondeur), ce qui fera l'objet de leurs études à venir

Mon avis

- L'idée de base semble intéressante
- Du point de vue théorique, je m'interroge sur le lien qui semble évident pour les auteurs entre efficacité et taille de l'individu. Dans une certaine mesure on peut effectivement souhaiter complexifier les résultats approchants pour coller encore plus aux données, mais ce point nécessiterait quand même d'être approfondie.
- Ce que montre concrètement l'expérimentation, c'est une nette diminution importante du nombre d'individus, en faveur des individus les plus « gros » : on observe donc une favorisation de l'exploitation, au détriment de l'exploration. L'absence d'interrogation sur ce sujet est problématique, d'autant plus que les résultats ne sont pas meilleurs que pour la limite sur la profondeur. Peut être une expérimentation sur un nombre plus important de génération pourrait être envisagé afin de voir l'influence de cette diminution sur un plus long terme ?
- L'expérimentation aurait aussi gagné à être appliquée sur plusieurs exemple.
- Aucune conclusion concernant les allocations de ressources restantes.