Indice d'Optimalité pour la Coloration de Graphe et Comptage de Solutions

Alexandre Gondran, Laurent Moalic

Indice d'Optimalité pour la Coloration de Graphe et Comptage de Solutions

Alexandre Gondran[1] Laurent Moalic[2]

[1] *ÉNAC, École Nationale de l'Aviation Civile*, Toulouse, France
alexandre.gondran@enac.fr

[2] *IRIMAS, Université de Haute Alsace*, Mulhouse, France
laurent.moalic@uha.fr

1 Problématique

Sur les instances de grande taille d'un problème NP-difficile, les méthodes exactes ne permettent pas de trouver une solution optimale. Elles permettent au mieux de trouver une borne inférieure et/ou une borne supérieure à la valeur optimale. Les métaheuristiques quant à elles, ne permettent de trouver que des solutions admissibles de "bonne qualité", c'est-à-dire, dans le cas d'un problème de minimisation, une borne supérieure à la valeur optimale. Le gap d'optimalité est l'écart entre ces deux bornes. L'optimalité n'est prouvée que lorsque ce gap est nul. Malheureusement pour les instances de grande taille d'un problème NP-difficile, ce gap est souvent important. Que faire dans cette situation ?

On peut évidemment améliorer les méthodes exactes et heuristiques utilisées, mais si $P \neq NP$ on trouvera toujours des instances où le gap est important. Que peut-on faire de mieux uniquement avec des heuristiques ?

On montre qu'une heuristique ne fournit pas seulement une borne supérieure au problème mais peut apporter des informations supplémentaires comme le nombre de solutions admissibles différentes ayant le plus faible coût trouvé (nombre de meilleures bornes supérieures trouvées différentes). Notre approche est alors fondée sur l'observation suivante : pour un problème de minimisation, le nombre de solutions admissibles diminue avec la valeur de la fonction objectif. En d'autres termes, plus le nombre de solutions admissibles ayant le même coût diminue, plus on se rapproche de l'optimalité.

D'abord, on démontre dans cette présentation l'exactitude de cette observation pour les problèmes de coloration de graphe et de recherche de clique maximum. Nous présentons ensuite une nouvelle approche qui permet de qualifier ou non une solution (trouvée par une heuristique) comme étant une potentielle solution optimale. Notre approche est appliquée au problème de coloration de graphe [1].

2 Coloration de Graphe

La coloration du graphe $G = (V, E)$ consiste à attribuer une couleur à chacun de ses sommets issu de V de telle manière que deux sommets reliés par une arête $e = (v_1, v_2) \in E$ aient des couleurs différentes. Soit un entier positif k, on appelle k-coloration de G, une coloration utilisant k couleurs. $\chi(G)$, le nombre chromatique de G, est le nombre minimal de couleurs nécessaire pour colorier G. Trouver $\chi(G)$ est un problème NP-difficile. On montre que si $i(G) > \mathcal{N}(G, k)$, avec $i(G)$, le nombre de stables de G [1] et $\mathcal{N}(G, k)$, le nombre de k-coloration de G différentes, alors k est le nombre chromatique de G.

1. Un stable de G est également appelé *ensemble indépendant* ; c'est un ensemble de sommets de G non deux à deux adjacents.

3 Indice d'optimalité

On définit [1] une procédure, appelée *indice d'optimalité* pour la coloration de graphe : soit G un graphe et $k > 0$ un entier positif, que l'on suspecte d'être le nombre chromatique de G. L'approche est fondée sur les cinq étapes suivantes :

1. Construire un échantillon de k-colorations de G : on exécute l'algorithme mémétique $HEAD$[2][2] sur G autant de fois que nécessaire pour obtenir t k-colorations. La taille de cet échantillon est égale à t. On fixe en général t à 1000.
2. Compter le nombre de k-colorations différentes dans l'échantillon. Ce nombre est égale à p. Évidement $0 \leq p \leq t$.
3. Estimer une borne supérieure du nombre de k-colorations différentes noté $UB(p, t)$; cette borne supérieure est fonction de t et p.
4. Calculer $i(G)$ ou au moins une borne inférieure si $i(G) > 10^6$, avec un algorithme exact.
5. Si $i(G) > UB(p, t)$, alors on dit qu'il y a **indice d'optimalité** pour que k soit égal à $\chi(G)$: les solutions de l'échantillon ont de grandes chances d'être optimales.

4 Comptage de Solutions

Il y a de nombreuses façons d'estimer $UB(p, t)$, la borne supérieure du nombre de k-colorations différentes dans G. Après une revue des méthodes exactes et approchées, nous avons opté pour une borne statistique définie sur un ensemble de graphes *test* pour lesquels le nombre exact de k-colorations est calculable. On prend :

$$UB(p,t) = \begin{cases} p + p^{\alpha^{\frac{t+p}{t}}} & \text{if } p < t \times 0.99 \\ +\infty & \text{otherwise} \end{cases}$$

avec $\alpha = 1.01$. On étend la validité de cette borne pour des graphes dont le nombre exact de k-colorations n'est pas calculable. Cette borne n'est donc pas définie de façon exacte, ce qui rend notre approche non exacte et c'est pourquoi on parle de l'indice d'optimalité et non d'optimalité au sens strict.

5 Tests et Perspectives

Notre approche est testée sur les benchmarks DIMACS (plus de 100 graphes) et RCBII (plus de 2000 graphes de densité et taille variées) et montre qu'il n'y a aucun *faux positif* : il n'y a pas de graphes pour lesquels on prouve l'indice d'optimalité égal à k alors $k > \chi(G)$[3]. Nous montrons par ailleurs l'indice d'optimalité à $k = 47$ pour le graphe DSJC500.5 de DIMACS.

Notre approche peut être généralisée à d'autres problèmes comme la recherche de la clique maximum. De plus, d'autres méthodes de comptage de solutions sont proposées qui permettraient d'appliquer notre méthode à un plus grand nombre de graphes. En effet, notre méthode ne s'applique que si le nombre de colorations optimales n'est pas trop élevé (<1000000).

Références

[1] A. Gondran and L. Moalic, "Optimality Clue for Graph Coloring Problem," 2018, preprint : hal-01952598 ; arXiv :1812.07734.

[2] L. Moalic and A. Gondran, "Variations on memetic algorithms for graph coloring problems," *Journal of Heuristics*, vol. 24, no. 1, pp. 1–24, Feb 2018.

2. Code disponible sur [github.com/graphcoloring/HEAD]

3. Nous proposons d'ailleurs une récompense correspondant à son poids en bouteilles de vin (ou jus de pomme) pour la première personne trouvant un faux positif à notre approche.