

Une méthode à base d'apprentissage par renforcement pour le problème de tournées de véhicules avec contrainte de capacité

Ali Yaddaden¹, Sébastien Hasrispe², Michel Vasquez²

¹ Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, 38000, Grenoble, France
ali.yaddaden@grenoble-inp.fr

² EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France
{sebastien.harispe, michel.vasquez}@mines-ales.fr

Mots-clés : *apprentissage par renforcement, apprentissage profond, problèmes de tournées, méthode à deux phases, procédure Split*

1 Introduction

L'apprentissage automatique, notamment les modèles d'apprentissage profond et les techniques d'apprentissage par renforcement, s'est avéré efficace pour résoudre des problèmes d'optimisation combinatoire difficiles sans recourir à une définition manuelle d'heuristiques de résolution. Dans ce cadre, les problèmes de tournées de véhicule sont largement utilisés pour évaluer l'efficacité de nouvelles approches, notamment le problème de tournées de véhicules avec contrainte de capacité (CVRP) [2, 1].

Le CVRP vise à construire des itinéraires de coût total minimal. Pour chaque itinéraire, un véhicule de capacité C part du dépôt, réalise un nombre de livraisons satisfaisant les demandes d'un sous-ensemble de clients et retourne au dépôt. Tous les clients doivent être visités une et une seule fois.

Pour le CVRP, les approches par apprentissage construisent les solutions candidates étape par étape, en choisissant à chaque itération soit de visiter un client, soit de retourner au dépôt pour se réapprovisionner, jusqu'à ce que tous les clients soient servis (à l'instar des méthodes constructives classiques telles que Savings ou Nearest Neighbors). Or, le choix du retour au dépôt peut être crucial pour la qualité des solutions candidates qui résultent des méthodes d'apprentissage ; dans la plupart des cas un plus grand nombre de retours au dépôt implique l'augmentation du coût de la solution candidate.

Par ailleurs, il existe, dans la littérature du CVRP, des méthodes à deux phases qui distinguent l'affectation des clients à un véhicule, et l'ordre dans lequel ce dernier doit les parcourir. En fonction du séquençage de ces deux opérations, nous distinguons entre *cluster-first route-second*, et *order-first split-second* [3].

Dans ce travail, nous introduisons une méthode en deux phases basée sur la stratégie *route-first split-second* pour le CVRP. Notre approche est fondée sur une combinaison de réseaux neuronaux profonds entraînés par apprentissage par renforcement et de l'algorithme "*Split*"[3].

2 Une méthode hybridant apprentissage et *Split*

Dans notre approche, nous proposons d'entraîner un réseau de neurones profond pour la résolution du CVRP sur la base de l'approche *order-first split-second*. Le réseau de neurones est entraîné à produire une représentation indirecte des solutions candidates sous forme de tours géants, définissant l'ordre de parcours des clients. L'extraction d'une solution réalisable est reportée sur l'algorithme *Split* via la résolution d'un problème de plus court chemin sur un graphe auxiliaire défini sur la base du tour géant.

Set	RL-OFSS	RFCS	Nearest Neighbor	Sweep	Savings
A	2.87	14.52	23.82	47.20	34.07
B	2.39	10.50	20.43	23.10	39.50
E	2.38	14.70	24.28	48.58	31.01
F	5.60	11.78	27.26	71.92	47.78
P	2.13	14.99	22.50	42.35	24.28
M	4.46	16.76	29.44	108.14	40.46
X	8.88	15.34	18.99	108.74	29.82

TAB. 1 – Moyennes des gaps à l’optimal par set d’instances de CVRPLib (%)

Pour une instance X du CVRP, notre réseau de neurones définit une politique stochastique donnant la probabilité de générer un tour géant sous la forme d’une séquence $Y = [y_1, \dots, y_n]$ (n étant le nombre de clients). La formule des probabilités composées permet d’exprimer cette politique par un produit des probabilités de sélection d’un client à l’itération t :

$$P_\theta(Y|X) = \prod_{t=1}^n p_\theta(y_t|y_0, \dots, y_{t-1}, X)$$

où p_θ est un réseau de neurones de paramètres θ .

Nous observons que *Split* agit comme un oracle qui évalue la qualité d’un tour géant en retournant le coût de la meilleure solution associée à celui-ci. Ainsi, nous pouvons entraîner le réseau de neurones via l’apprentissage par renforcement à minimiser l’espérance des coûts des tournées ($\mathcal{L}(\theta) = \mathbb{E}_{Y \sim P_\theta(\cdot|X)}[\text{Split}(Y, X)]$). Cette minimisation implique de trouver les paramètres θ qui donnent les tours géants dont la solution du CVRP associée est de coût minimal, par le biais de l’algorithme de descente de gradient. Ce dernier est donné par l’algorithme REINFORCE : $\nabla_\theta \mathcal{L}(\theta) = \mathbb{E}_{Y \sim P_\theta(\cdot|X)}[\text{Split}(Y, X) \nabla_\theta \log P_\theta(Y|X)]$

3 Résultats et perspectives

Nous avons testé notre modèle sur différents ensembles d’instances de CVRPLib. Nos résultats sont exposés dans le Tableau 1, qui illustre une comparaison des moyennes des gaps à l’optimal de notre réseau de neurones (RL-OFSS) avec des heuristiques classiques du CVRP : Route-First Cluster-Second (RFCS), *Nearest Neighbor*, *Savings*, et *Sweep*. Nos résultats montrent que le modèle de réseaux de neurones donne de meilleurs résultats avec des gaps à l’optimal de moins de 10 %. Ceux-ci sont encourageants, bien qu’encore loin des performances des méthodes de l’état de l’art du CVRP. Dans nos futurs travaux, nous envisageons d’autres schémas d’hybridation entre cette approche et d’autres méthodes de résolution afin d’en améliorer les performances.

Références

- [1] Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems ! In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.
- [2] Mohammadreza Nazari, Afshin Oroojlooy, Lawrence Snyder, and Martin Takác. Reinforcement learning for solving the vehicle routing problem. *Advances in neural information processing systems*, 31, 2018.
- [3] Christian Prins, Philippe Lacomme, and Caroline Prodhon. Order-first split-second methods for vehicle routing problems : A review. *Transportation Research Part C : Emerging Technologies*, 40 :179–200, 2014.