Approximer le coût de tournées de surveillance du réseau de transport gazier à l'aide d'apprentissage automatique

Nadia Ghernaout^{1,2}, Martin Cousineau³, Christelle Guéret¹, David Rivreau⁴

¹ Univ Angers, LARIS, SFR MATHSTIC, F-49000 Angers, France
{nadia.ghernaout, christelle.jussien-gueret}@univ-angers.fr

² Research and Innovation Center for Energy (RICE), GRTgaz
nadia.ghernaout@grtgaz.com

³ Department of Logistics and Operations Management, HEC Montréal, Montréal, Canada
martin.cousineau@hec.ca

⁴ UCO, LARIS, SFR MATHSTIC, F-49000 Angers, France
rivreau@uco.fr

Mots-clés : *apprentissage automatique, approximation, CARP, arc routing*

1 Introduction

GRTgaz est un gestionnaire de transport de gaz en France. Afin d'assurer la sécurité et la performance du réseau de transport de gaz, l'entreprise doit effectuer des tournées de surveillance le long des 33 000 km de canalisations. Celles-ci consistent à parcourir l'ensemble du réseau afin de détecter de potentiels dangers pouvant porter atteinte à l'intégrité des infrastructures gazières. Les tournées de surveillance sont périodiques et peuvent être effectuées à l'aide de différents modes tels que la voiture et l'avion. L'entreprise doit construire les tournées de chaque semaine de l'année, en prenant en compte les différentes fréquences de passage par portions de canalisations ainsi que différents modes de surveillance. Ce problème est défini comme un *Periodic Capacitated Arc Routing Problem (PCARP) with a multimodal fleet*. Il s'agit d'un problème NP-difficile qui implique des prises de décisions aux niveaux tactique et opérationnel. Afin d'accélérer les méthodes de résolution, il peut être intéressant de pouvoir calculer rapidement une approximation des coûts des tournées de chaque semaine une fois le mode de transport et l'ensemble des tronçons à surveiller fixés. Dans [2, 3] les auteurs présentent différentes fonctions d'approximation pour estimer la distance totale de problèmes de tournées de véhicules avec fenêtres temporelles, et contraintes de capacités. Dans [1, 4] les auteurs proposent de nouvelles variables permettant de décrire des instances de problèmes et développent des modèles d'apprentissage pour prédire la longueur d'une tournée pour le problème du voyageur du commerce et le problème de tournées de véhicules. Ces méthodes d'approximations n'ont jamais été étudiées pour des problèmes de tournées sur les arcs. Ainsi nous introduisons des nouvelles variables spécifiques aux problèmes de tournées sur les arcs et développons des modèles de régression linéaire pour la prévision des coûts de tournées.

2 Description du problème

Le territoire de GRTgaz est divisé en plusieurs secteurs. Chaque secteur $s \in S$ dispose d'une équipe de techniciens qui peut effectuer jusqu'à 5 tournées en voiture par semaine. Un avion peut effectuer jusqu'à 7 tournées par semaine sur l'ensemble du territoire.

Nous notons $X^v(s)$ un ensemble de noeuds accessibles en voiture (respectivement X^p en avion) et $E^v(s)$ un ensemble d'arêtes (resp. E^p) pondérées par des temps de passage t_{ij}^v (resp. t_{ij}^p). Un sous-ensemble d'arêtes $R^v(s) \subset E^v(s)$ (resp. $R^p \subset E^p$) associées à un temps de surveillance

$t_{ij}^{v'} > t_{ij}^v$ (resp. $t_{ij}^{p'} > t_{ij}^p$) doivent être inspectées. Enfin chaque secteur dispose d'un ensemble de dépôts $D^v(s)$ (resp. D^p) où débutent et s'achèvent chaque tournée. Contrairement aux tournées en voiture, les dépôts de début et fin de tournée des avions peuvent être différents. Le coût horaire est de c^v euros pour les voitures et c^p euros pour les avions. L'objectif est de fournir une méthode d'approximation des coûts des tournées pour chaque semaine.

3 Description de la méthode

Notre objectif est de construire pour chaque mode un modèle d'apprentissage automatique prédisant le coût des tournées d'une semaine. Pour cela nous construisons une base d'apprentissage constituée de tournées labellisées et associées à des variables explicatives. Des modèles de régression permettent d'apprendre la relation entre l'ensemble des variables et le coût.

La première étape à la construction d'une base d'apprentissage consiste à générer des observations. Pour cela nous tirons aléatoirement ou à l'aide de clusters un certain nombre d'arêtes requises et résolvons les problèmes de tournées associés sous forme de Programmes Linéaires en Nombres Entiers (PLNE). Ces modèles diffèrent selon le mode de surveillance utilisé. Le modèle de tournées en voiture est un CARP. Le modèle de tournées en avion est un *CARP with Intermediate Facilities (CARP-IF)* si l'on considère que l'avion fait une seule tournée par semaine avec des arrêts en différents dépôts intermédiaires. Afin d'enrichir la base d'apprentissage plusieurs caractéristiques spatiales sont extraites, elles représentent les variables explicatives du modèle. Certaines variables peuvent être spécifiques aux tournées en véhicules ou en avion. Une partie des variables est dite fixe et fournit des informations globales sur le secteur : aire et périmètre des enveloppes convexes formées par l'ensemble des arêtes, distance moyenne au dépôt, position du dépôt dans le secteur. L'autre partie dépend des arêtes requises à surveiller, elles décrivent la taille de celles-ci, l'éloignement par rapport aux dépôts, les éloignements par rapports aux arêtes, ainsi que leur position dans le graphe. Les variables pertinentes sont sélectionnées à l'aide de méthodes de régularisation telles que le LASSO. Enfin, différents modèles de régression tels que des régressions linéaires multiples, des modèles de forêts aléatoires ou de XGBoost sont entraînés sur la base d'apprentissage et évalués par validation croisée. Nous retenons des modèles de régressions linéaires multiples pour leur performance et simplicité.

4 Conclusions et perspectives

Cette approche a été testée sur des instances de GRTgaz. La base d'apprentissage inclue des tournées contenant jusqu'à 200 arêtes requises. Nous utilisons le solveur CPLEX 22.1 et programmons en Python pour la résolution des problèmes de tournées. Les prévisions sur la base de test atteignent une erreur moyenne de 6% pour des tournées en voiture ou en avion. Notre prochaine étape consiste à intégrer ces modèles d'approximation au niveau tactique dans un modèle de planification et adopter une stratégie de ré-entraînement de ces modèles.

Références

- [1] Fabian Akkerman and Martijn Mes. Distance approximation to support customer selection in vehicle routing problems. *Annals of operations research*, pages 1–29, 2022.
- [2] Miguel Andres Figliozzi. Planning approximations to the average length of vehicle routing problems with time window constraints. *Transportation Research Part B : Methodological*, 43(4) :438–447, 2009.
- [3] Daniel Nicola, Rudolf Vetschera, and A Dragomir. Total distance approximations for routing solutions. *Computers & Operations Research*, 102 :67–74, 2019.
- [4] Jussi Rasku, Tommi Kärkkäinen, and Nysret Musliu. Feature extractors for describing vehicle routing problem instances. In *OASICS*. Dagstuhl Publishing, 2016.