

# Co-clustering sous contraintes pour identifier les profils à risque : application à la sinistralité des flottes de véhicules d'entreprises

Romain Lorenzi, Clément Fauré, Dimitri Petroviche

GAC Technology, Lyon, France  
rlorenzi, cfaure, dpetroviche@gac-technology.com

## 1 Contexte

Pour un gestionnaire de flotte, la compréhension du comportement de ses véhicules est le point central d'une gestion optimale. L'identification de profils de véhicules/conducteurs permet notamment de traiter cette problématique. Cette pratique, appelée profiling, consiste à trouver des individus qui partagent plusieurs caractéristiques en commun et se comportent de la même façon vis-à-vis d'une cible d'étude. Afin de tirer le maximum de cette méthode, il est possible de limiter l'extraction des profils à ceux qui sont les plus intéressants. Ces derniers seront appelés "profils extrêmes" ou "profils à risque" et correspondent à des groupes d'individus ayant des habitudes extrêmement différentes des autres.

Nous présentons ici un algorithme permettant de répondre à ces besoins, nommé EPR pour Extreme Profiles Recognition. Son objectif est d'identifier des profils de véhicules ayant une sinistralité extrême (inexistante ou très importante) tout en garantissant des résultats variés et représentatifs des données. Il s'inscrit dans la catégorie des algorithmes de co-clustering sous contraintes. Ces dernières concernent notamment la cible d'étude : seuls les profils avec une sinistralité particulièrement haute ou basse sont gardés.

Nous avons isolé quatre composantes principales dans les profils à retourner :

- Les caractéristiques qu'ont en commun tous les véhicules du profil (parmi les variables de la base de données).
- Le nombre de véhicules composant le profil.
- La valeur de la cible d'étude associée au profil.
- La pureté : parmi les véhicules appartenant au profil, le pourcentage de ceux ayant la même valeur de la cible d'étude que celle du profil.

L'objectif est d'identifier des profils dont la pureté est proche des 100% tout en garantissant un nombre de véhicules suffisamment grand et un nombre de caractéristiques en commun raisonnable pour que l'information rendue soit pertinente et facilement compréhensible.

## 2 Extreme Profiles Recognition

Notre étude s'appuie sur une base de données orientée sinistralité. La cible d'étude est un score de sinistralité calculé à partir de plusieurs critères (nombre d'accidents corporels et maté-