

CONVENTION CIFRE



Entreprise : Renault SAS
Direction - Service : DEA-MS - Market Intelligence
Tuteur entreprise : Antoine DE BOSSOREILLE

Laboratoire : Institut de Mathématiques de Marseille – UMR7373
Directeur de thèse : Badih GHATTAS

TITRE : Réduction du volume des données d’usages clients automobiles

1- Contexte

Le thème de la thèse s’inscrit dans une perspective de conception performante et robuste de Groupe Moto-Propulseur (GMP) d’automobile. Produire des véhicules moins polluants et consommant moins d’énergie est un enjeu majeur de l’industrie automobile d’aujourd’hui, qui passe notamment par la réduction de masse du véhicule. Cet allègement implique le dimensionnement au plus juste des systèmes tout en respectant les objectifs de fiabilité. Or, le contexte projet impose toujours plus de contraintes en termes de réduction du nombre de boucles de conception et de développement, de diminution du nombre de prototypes physiques et d’essais, extrêmement coûteux. C’est pourquoi la simulation numérique des usages clients, qui permet de représenter la variabilité statistique du comportement de nos clients dans les dimensionnements GMP, est un enjeu important.

Aujourd’hui Renault dispose déjà d’une expérience pour simuler les usages des clients sur un projet véhicule en développement. L’équipe simulation usages clients utilise un volume important de données temporelles enregistrées sur véhicules existants (plus de 4000 clients réels pour un total de 6 millions de km) et les traite avec des données non temporelles, afin de les projeter sur de nouveaux projets en développement. Ces données non temporelles proviennent aujourd’hui d’enquêtes clients, relevés garage et véhicules connectés, en particulier électriques.

L’objectif principal du travail de recherche est la mise au point de méthodes d’apprentissage statistique (clustering, régression, sous-échantillonnage) adaptées aux objectifs finaux, afin de modéliser des usages clients (via des séries temporelles) avec un volume de données contenu tout en conservant une représentativité statistique. Ces données d’usages clients pourront ensuite être plus facilement utilisées (temps de calcul, volume de données) pour des simulations GMP / véhicule sur les technologies en développement.

2- Développements actuels

Les méthodes statistiques classiques ne nous permettent pas aujourd’hui de trouver de corrélation entre l’usage client et l’occurrence de phénomènes physiques complexes au niveau du GMP.

Par exemple, dans le cas du développement du filtre à particule essence (GPF), nous n’avons pas pu établir un lien statistique entre la mise en défaut du filtre à particule (chargement de suies maximum) et les différents indicateurs des usages clients qui sont extraits des enregistrements

temporels. Nous avons donc conservé la totalité des enregistrements temporels plutôt qu'une sélection réduite de 100 clients pour les différentes simulations de GPF.

Un verrou scientifique est apparu au niveau du choix de critères de similarité afin de réaliser un clustering pertinent de séries temporelles longues.

Par ailleurs, l'identification des critères les plus pertinents afin d'effectuer une sélection des clients représentatifs statistiquement en regard de la problématique finale étudiée n'a pas encore été possible.

4 – Les données et les sorties des modèles

La base de données actuelle contient environ 4000 clients ; chaque client est représenté par plusieurs variables temporelles (dont on retiendra une pour commencer, la vitesse des véhicules), observées à une fréquence de 10Hz pendant 15 jours, soit environ 13 millions de mesures pour une variable par client. Ces variables servent d'entrées au modèle de simulation, auxquelles s'ajoutent d'autres paramètres, telle que la température extérieure et les caractéristiques statiques du véhicule.

Le modèle de simulation renvoie en sortie plusieurs courbes (à la fréquence de 10 Hz), telles que la vitesse et la masse de suie. On s'intéressera surtout à la masse de suie, et en particulier la valeur maximale.

Le modèle de simulation a été exécuté sur un ensemble de 700 clients, et on analyse la distribution du maximum de suie observée sur cet échantillon. Les simulations ont été réalisées pour différentes valeurs de la température extérieure.

4 – Travaux attendus

Les travaux prévus dans le cadre de la thèse consisteront à répondre aux besoins suivants :

- Analyse des méthodes statistiques appliquées aujourd'hui à la construction de bases d'usages clients.
- Etat de l'art sur les techniques de classification et d'apprentissage qu'on pourrait utiliser pour la recherche de critères et de réduction du volume de données appliquées aux usages clients automobiles et aux signaux temporels.
- Recherche et mise au point de méthodes statistiques et d'apprentissage permettant de traiter la réduction des données d'usages clients. Il s'agit de l'optimisation de la méthode d'échantillonnage au vu de critères de distributions à déterminer.
- Quantification des incertitudes liées à l'utilisation de ces méthodes en comparaison avec l'analyse statistique massive.

Les moyens qui seront mis à disposition du candidat:

Données citées dans les sections précédentes, logiciels de l'entreprise (Matlab, Spotfire, Python, ...), stations de calcul standards et haute performance (88 cœurs et RAM de 512 Go) disponibles en local.

Si nécessaire, il sera également possible d'étudier l'accès aux moyens HPC (serveurs de calculs parallèle) et à l'infrastructure BigData (Datalake et Hadoop).

5- Compétences requises du futur thésard

- **Connaissances spécifiques:** Statistiques, Machine Learning, Big Data, R/Python, Culture automobile.
- **Formation souhaitée:** Ecole d'ingénieur généraliste ou spécialisée en statistiques, ESTACA, INSA, Centrale, ENSAI, et masters en statistiques ou data science.

- **Aptitudes personnelles souhaitées**: Sens du contact pour la collecte d'information et la compréhension des problématiques de l'ingénierie.

6- Localisation

Le suivi en entreprise aura lieu en région parisienne : Guyancourt (78) et/ou Lardy (91). Le laboratoire d'accueil sera à Marseille. Un système d'alternance sera mis en place entre l'entreprise et le labo.

7- Candidature

Envoyer CV et lettre de motivation à antoine.de-bossoreille@renault.com,
badihghattas@gmail.com