

OFFRE DE STAGE

Intitulé du stage : Modèles de graphes pour la prédiction des retards et de leur propagation sur le réseau ferroviaire

Durée du stage : 6 mois

Démarrage envisagé : Mars 2021

Contexte :

SNCF exploite quotidiennement un important volume de données de nature spatio-temporelle. Des retards peuvent apparaître puis se propager dans le réseau ferroviaire selon l'état du trafic, pénalisant le confort des voyageurs et la gestion opérationnelle de la circulation. Ainsi, la précision des prédictions des temps de retard figure parmi les objectifs principaux des transporteurs SNCF (TGV, Transilien, TER et Intercités).

L'apparition de retards et leur propagation dépend de facteurs de différentes natures. L'apparition des retards est avant tout un phénomène **non-stationnaire**, c'est-à-dire dont le comportement varie dans le temps. Il dépend par ailleurs des **trains** mis en circulation, et notamment de leur capacité à absorber les fortes affluences en gare, reconnues comme source de retards. Pour des raisons analogues, les **gares** traversées ou desservies influent également sur l'apparition et la variation des retards. Certains **événements extérieurs** au réseau ferroviaire (incident météorologique, grève, panne...) contribuent également au phénomène. Enfin, la **structure maillée du réseau ferroviaire** conduit naturellement les retards à se propager et à impacter d'autres trains (goulot d'étranglement, partage de voies, cantonnement...). Cette dernière composante constitue à la fois l'une des principales originalités du comportement des retards dans le réseau ferroviaire, mais aussi l'une des principales difficultés dans la conception d'algorithmes de prédiction des retards.

Bien que le sujet ait été largement étudié par la communauté scientifique, aucun algorithme de prédiction, inspiré ou non de l'apprentissage statistique, n'apporte des performances opérationnelles entièrement satisfaisantes, ni ne tire parti efficacement des facteurs explicatifs cités plus haut. Plusieurs projets de recherche visant à combler cette lacune ont été successivement initiés par SNCF, dont certains de concert avec le laboratoire CMLA de l'ENS Paris-Saclay. Les premiers résultats sont prometteurs et invitent à approfondir cette approche. L'équipe du Plateau IA, rattachée à la Direction de l'Innovation et de la Recherche SNCF, propose dans cette optique un stage co-encadré par le CMLA.

Direction Innovation & Recherche

1 avenue François Mitterrand
93210 La Plaine Saint-Denis

Mission :

Les travaux du stage se dérouleront en trois temps. Le premier temps (environ 2 mois) sera consacré à l'étude de l'état de l'art en matière de contrôle des transports publics et de prédiction des retards par apprentissage statistique. Cette phase offrira également l'occasion au stagiaire de se familiariser avec les travaux antérieurs réalisés sur le sujet par le CMLA et par SNCF. La seconde partie du stage (environ 3 mois) sera dédiée à la mise en œuvre d'un modèle de prédiction exploitant les facteurs explicatifs précédemment mentionnés, avec pour objectif de dépasser les performances opérationnelles du modèle en vigueur à SNCF. Enfin, le stage se conclura par une étape de synthèse des travaux, durant laquelle le stagiaire devra produire un rapport de stage ainsi qu'un code clair, structuré et documenté. Les modalités pratiques pourront être ajustées en fonction des attentes du CMLA et des contraintes universitaires du stagiaire.

Objectifs :

- Comprendre les problématiques inhérentes au système ferroviaire
- Se familiariser avec les algorithmes de prédiction à l'état de l'art et ceux précédemment explorés par SNCF et le CMLA
- Concevoir un algorithme de prédiction capable d'exploiter les caractéristiques temporelles et structurelles (trains, gares, maillage du réseau ferroviaire) des retards
- Comparer les performances obtenues à celles du modèle en vigueur à SNCF (*baseline*)
- Contribuer à l'amélioration de l'information aux voyageurs et de la gestion opérationnelle du réseau ferroviaire

Activités prévues :

- Exploration des données et des modèles de prédiction disponibles
- Choix des indicateurs de performance clefs (*KPI*) pour la prédiction des retards
- Définition d'un algorithme de prédiction des retards exploitant leurs caractéristiques temporelles et structurelles (trains, gares, maillage du réseau ferroviaire)
- Implémentation claire, structurée et documentée de l'algorithme
- Comparaison des performances du modèle conçu à celles de la *baseline* SNCF
- Points hebdomadaires sur l'avancement du stage et comptes rendus réguliers
- Rédaction d'un rapport de stage structuré et détaillé
- Soutenance interne de fin de stage à la Direction de l'Innovation et de la Recherche

Compétences :

- Connaissances approfondies en apprentissage statistique et en modèles de graphes
- Programmation Python méthodique et efficace, connaissance élémentaire de la POO
- Connaissance d'un *framework* d'apprentissage statistique (Keras/Tensorflow, Pytorch, scikit-learn...) et des outils d'exploration de données (ex : Pandas)
- Maîtrise d'un système standard de versioning de code (idéalement Git)

L'autonomie, la rigueur, les capacités rédactionnelles, la curiosité et la spontanéité seront autant de qualités très appréciées. Une connaissance des systèmes ferroviaires est un plus, mais n'est absolument pas indispensable.

Direction Innovation & Recherche

1 avenue François Mitterrand
93210 La Plaine Saint-Denis

Contacts :

M. Tom ROUSSEAU (SNCF)
tom.rousseau@sncf.fr

M. Argyris KALOGERATOS (CMLA)
argyris.kalogeratos@cmla.ens-cachan.fr

Gratification / Avantages :

- Gratification fixée en respect de la législation en vigueur et en fonction du niveau d'études
- Carte de circulation 2nde classe valable sur l'ensemble du réseau national
- Accès subventionné aux restaurations d'entreprise