

Utilisation des données de téléphonie mobile pour évaluer les déplacements et la fréquentation sur les axes ferroviaire en Belgique.

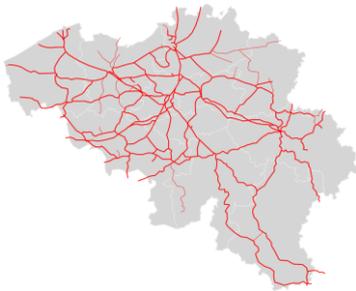


Figure 1: Réseau ferroviaire belge

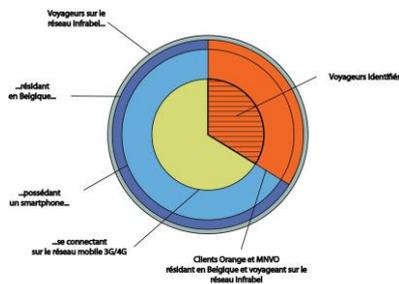


Figure 2: Redressement

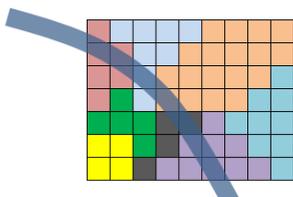


Figure 3: Part de voie ferrée

Contexte

L'hyper-connectivité de notre société actuelle génère un nombre de données énorme. Ces grandes bases de données, les « Big Data », peuvent constituer une source d'information très utile pour de nombreuses études, notamment dans les domaines du transport et de la mobilité. Dans ce cadre, Stratec collabore avec Orange Belgique et exploite leurs données de téléphonie mobile brutes afin d'évaluer les déplacements et la fréquentation sur le réseau ferroviaire belge (voir Figure 1).

Enjeux

Il existe de nombreux défis liés au traitement et à l'analyse des « Big Data » issus de la téléphonie mobile pour déterminer des déplacements. Si l'on dispose d'un très large échantillon comparé à des enquêtes traditionnelles, il est toutefois nécessaire de le traiter et d'en ressortir les informations exploitables. En effet, afin de pouvoir déterminer un déplacement, il est, par exemple, impératif d'avoir un nombre minimum de connexions au cours de celui-ci. Le pré-traitement de ces données est donc une étape très importante afin de ne préserver que les informations qui seront utiles à l'analyse. Cela permet également de réduire considérablement la taille de ces bases de données déjà très volumineuse en espace mémoire. En ce qui concerne l'analyse, il est essentiel de développer des algorithmes performants qui permettent de déterminer le plus précisément possible : si un utilisateur effectue un déplacement ; son origine et sa destination ; son mode de transport ; etc.

Méthodologie et algorithme

• Echantillon et redressement

Les données brutes d'Orange Belgique se composent de milliards d'enregistrements de positions de cartes SIM. En fonction de la zone de couverture d'une antenne il est possible de connaître avec un certain degré de précision la position d'une carte SIM (on utilise le barycentre de la zone captée par une antenne). Pour des raisons de fiabilité et de connexions suffisantes, seules les cartes SIM d'origine belge se trouvant dans un smartphone se connectant au réseau 3G/4G ont été retenues. Afin d'avoir des chiffres de fréquentations représentatifs sur l'ensemble de la Belgique, il est nécessaire de redresser les résultats obtenus. Ce facteur de redressement prendra en compte : la part de marché d'Orange ; la part de résidents belges ; le taux de pénétration de la téléphonie mobile ; la part de marché des smartphones ; le taux d'activité 3G/4G (voir Figure 2).

• Travail préliminaire

Suite à la discrétisation du territoire basée sur les zones de couvertures des antennes fournies par orange, nous en déduisons une série d'indicateurs et d'évaluateurs qui permettront d'opérer le choix entre un trajet en train ou un trajet en voiture. Pour chacune des zones nous calculons son intersection avec une zone tampon (typiquement 30 m) ajoutée de part et d'autre des voies ferrées et routières (voir l'illustration Figure 3). Cette opération permet de calculer une probabilité qu'une carte SIM captée par une antenne soit sur une voie routière ou ferroviaire. Un calcul similaire est effectué pour calculer la part de gare ferroviaire contenue dans une zone de couverture d'une antenne.



Figure 4: 6h – 9h

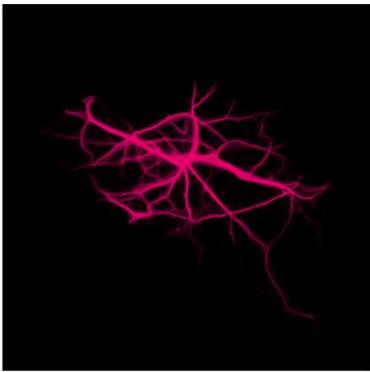


Figure 5: 11h – 14h



Figure 6: 16h – 19h

- **Traitement de la base de données**

La base de données fournie par Orange pour une durée de 14 jours correspond un volume d'environ 1 To. Il est nécessaire de traiter cette base afin de d'alléger la masse de données afin de réduire le temps d'analyse. Les données sont filtrées et triées au moyen de routines programmées dans le langage Python. Les connexions sont d'abord triées par identifiant de personnes et date de connexion au réseau. Ensuite ce fichier est filtré de telle sorte à supprimer les connexions inutilisables (carte SIM non belges, uniquement 2G, antennes inconnues, carte SIM inconnues, ...) et inutiles (s'il existe plus de deux connexions successives à la même antenne, seul la première et la dernière sont sauvegardées).

- **Détection de déplacements**

Avant de développer l'algorithme de choix sur le mode de déplacement, une phase préliminaire consiste à détecter des déplacements dans la liste des positions successives associées à une carte SIM. Des critères telles que la vitesse de mise en mouvement, la vitesse d'arrêt, la longueur et la durée du trajet.

- **Recherche de trajets en train**

Une fois un déplacement repéré nous lançons un algorithme qui va devoir déterminer si un trajet est plus probable d'avoir été effectué en train ou en voiture. L'algorithme décisionnel va attribuer des points en fonctions d'une série de critères tels que le nombre de gares visitées, les distances totales et successives effectuées sur le rail/route, etc. Une fois que tous les critères ont été traités, chaque trajet aura une note qui permettra de catégoriser le trajet soit en mode rail soit en mode route.

Visualisations et exploitation des résultats

Une fois que les trajets ont été attribués soit au mode rail soit au mode route, il est possible de représenter ces trajets sur voie ferrée en reliant tous les barycentres des zones associées aux antennes successive auxquelles s'est connecté le voyageur. En superposant tous ces trajets avec une transparence, il est possible d'évaluer la densité du trafic sur voie ferrée. La fréquentation du réseau ferroviaire est représentée respectivement sur la *Figure 4* pour la période 6h-9h, sur la *Figure 5* pour la période 11h-14h et sur la *Figure 6* pour la période 16h-19h.

Ces résultats, après avoir été redressés, peuvent être facilement intégrés à un logiciel de modélisation de réseaux de transport (par exemple PTV VISUM) sous la forme de **matrices origine-destination (gare-gare et commune-commune)**. Une fois le réseau affecté, on peut faire tourner des simulations et obtenir énormément d'informations telles que le nombre de montées et de descentes aux différentes gares, la fréquentation temporelle et spatiale du réseau, les distances parcourues, etc.

Conclusions

Cette approche innovante de l'analyse des données de téléphonie mobile dans le but d'obtenir des données tangibles en matière de transport donne des résultats très convaincants pour les déplacements longues distance. Cette approche originale et unique en Belgique (à ce jour et à notre connaissance) présente divers atouts comparés aux comptages en gare tels que le coût et la quantité d'information.

Nous sommes convaincus que cette approche est amenée à gagner de plus en plus de précision dans le futur. D'une part parce que le taux de pénétration des smartphones ne cessera d'augmenter, d'autre part parce que de multiples améliorations et développement sont encore possible afin d'améliorer la performance, la fiabilité et la robustesse de l'algorithme. Nous pensons également qu'il sera possible d'extraire beaucoup d'autres informations tels que des modes de transport dans les grandes villes, les motifs de déplacements, etc.