

Reconnaissance de l'activité à partir des déplacements d'utilisateurs

Par Charles Guin-Vallerand, Professeur agrégé d'informatique @ TELUQ & LICEF

15 octobre, 2018



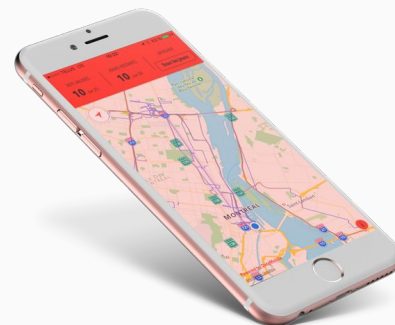
Contexte

La Ville de Montréal a fait développer une application pour téléphone intelligent appelée MTL Trajet (<https://ville.montreal.qc.ca/mtltrajet>). Celle-ci récupère les déplacements de citoyen-participants afin de mieux connaître la mobilité des citoyens de Montréal.



Fonctionnement :

- Récupère (*log*) les informations provenant du GPS des téléphones (lat/long/alt, vitesse, orientation)
- Récupère des informations socio-démographiques sur les participants : lieux de domicile et de travail/étude, type d'occupation, moyens de transport, etc.



Problématique

À partir des données récupérées par MTL Trajet, comment :

1. Découvrir automatiquement le lieu de domicile et de travail/étude des utilisateurs
2. Comment inférer/reconnaître les activités des utilisateurs à partir des déplacements pour :
 - a. Mieux connaître les activités et habitudes des citoyens
 - b. Aider au développement des services et infrastructures
 - c. Support au développement économique
 - d. Recommander du contenu/service/produit en ligne

Problème 1

Inférence du
domicile et du lieu
de travail

Problème 1 : Inférence du domicile et du lieu de travail (1/6)

Objectif :

- À partir des données GPS récupérées (lat/long/alt, speed, bearing), estimer de façon +/- *naïve* le lieu de résidence et le lieu de travail

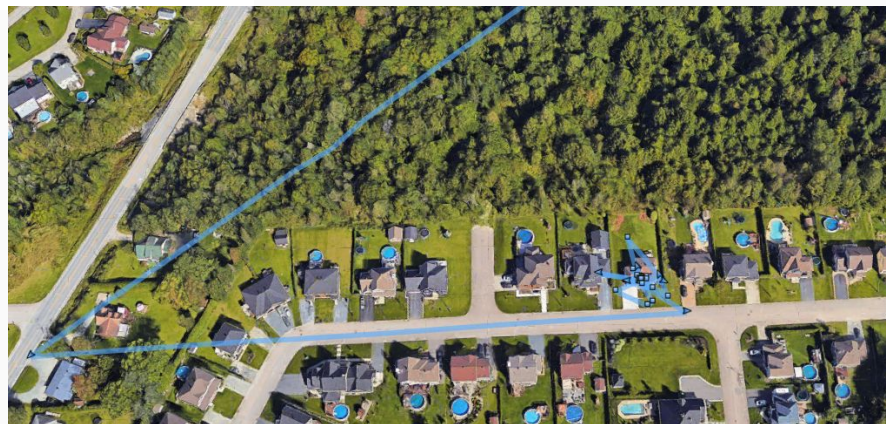
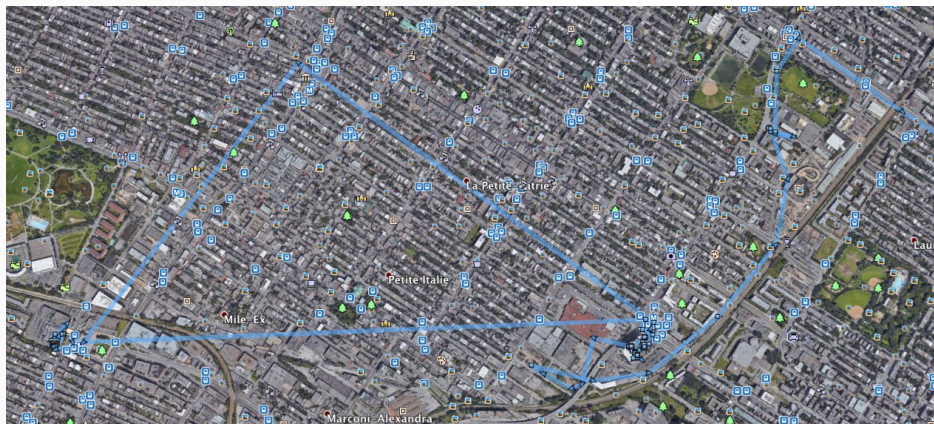
Questions ?

1. Quelles sont les hypothèses de travail?
2. Comment réduire la complexité du problème?
3. Quels sont les algorithmes pouvant être utilisés ?
4. Comment passer d'une position lat/long/alt vers une adresse physique?

Problème 1 : Inférence du domicile et du lieu de travail (2/6)

Exemple de traces :

Location Id	Latitude	Longitude	Time	Speed	Bearing
601	48.4210453	-71.0572413	Mon Nov10..	0.74	76.5
602	48.4210453	-71.0572532	Mon Nov10..	0.79	76.5
603	48.4210453	-71.0572567	Mon Nov10..	0.24	76.5



Problème 1 : Inférence du domicile et du lieu de travail (3/6)

Quelles sont les hypothèses de travail?

1. Domicile : les utilisateurs y sont majoritairement entre 18h et 7h (la semaine).
2. Travail/étude : les utilisateurs y sont majoritairement entre 7h et 18h.
3. La zone associée au domicile est de l'ordre de moins de 50 m
4. La zone associée au travail peut varier grandement (selon le travail)

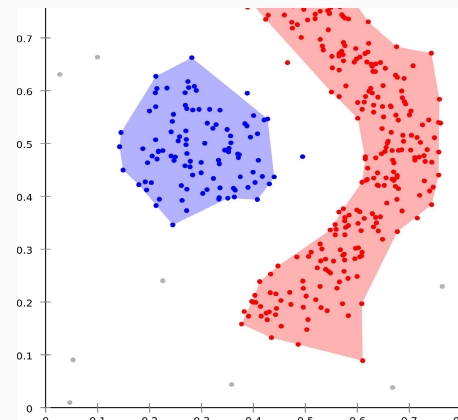
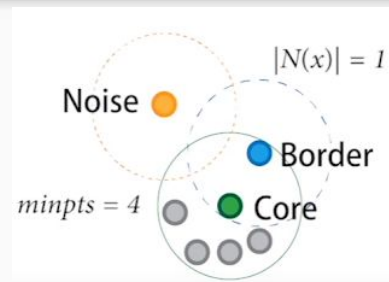
Comment réduire le problème ? :

- Utiliser les hypothèses pour filtrer les données :
 - Domicile : données semaine entre 18h et 7h.
 - Travail : données semaine entre 8h et 17h

Problème 1 : Inférence du domicile et du lieu de travail (4/6)

Quels sont les algorithmes pouvant être utilisés ? : Nous voulons identifier une densité de points lat/long, alors :

- Density-Based Spatial Clustering of Applications with Noise (DBSCAN) :
 - Lecture :
 - <http://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>
 - WEKA :
http://weka.sourceforge.net/doc/packages/optics_dbScan/weka/clusterers/DBSCAN.html
 - Python :
<http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>
 - Deux paramètres :
 - Epsilon : distance maximale pour appartenir à un cluster (ex. 50 m)
 - MinPts : nombre de pts minimum dans le cluster (ex. 1 données/30 sec, donc min. 720)
 - Utiliser le centroid du cluster comme lieu de résidence ou de travail



Problème 1 : reconnaissance du domicile et du lieu de travail (5/6)

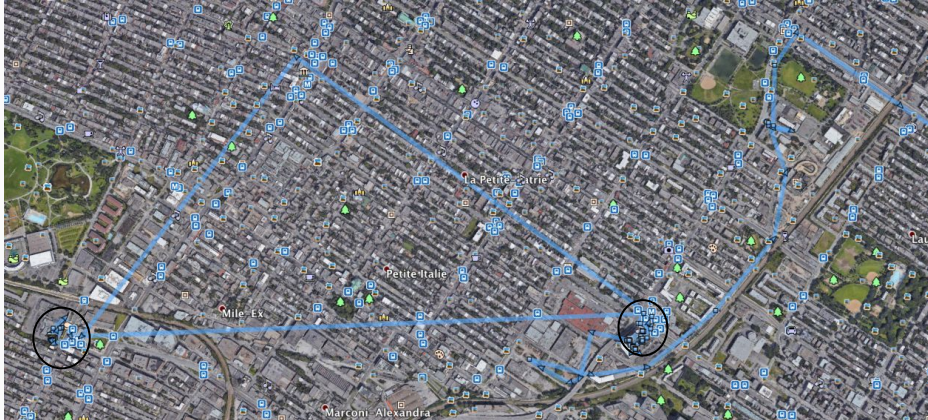
Comment passer d'une position lat/long/alt vers une adresse physique?

- Sous Android OS :
 - Utiliser le “reverse geo coding” :
<https://developer.android.com/training/location/display-address.html#java>
- Avec OpenStreetMap API :
 - Utiliser Nominatim : https://wiki.openstreetmap.org/wiki/Nominatim#Reverse_Geocoding
- En Javascript :
 - Utiliser le “reverse geo coding” :
<https://developers.google.com/maps/documentation/javascript/examples/geocoding-reverse>

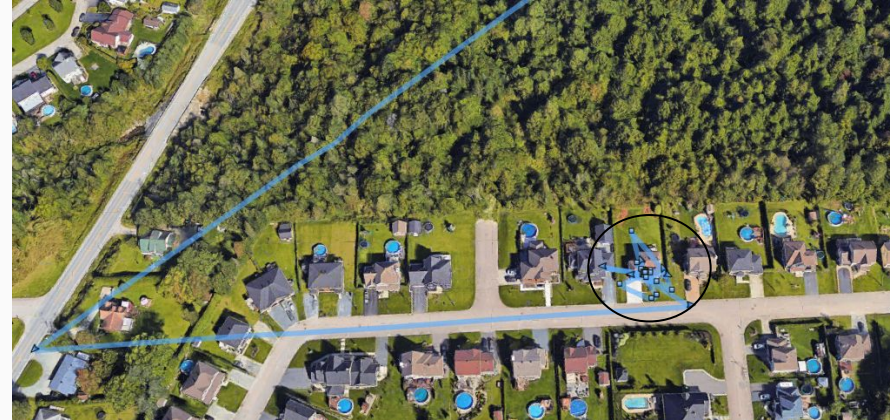
Problème 1 : Inférence du domicile et du lieu de travail (6/6)

Résultats :

2 clusters tout dépendant du
epsilon et minPts



1 cluster



Problème 2

Reconnaître les
activités
d'utilisateurs à partir
de leurs
déplacements

Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (1/7)

Objectif :

- Reconnaître trois types d'activités : déplacement, activité à l'arrêt et activité en mouvement (ex. magasinage) à partir des traces GPS

Questions ?

1. Quelles sont les hypothèses de travail?
2. Comment réduire la complexité du problème?
3. Quels sont les algorithmes pouvant être utilisés ?
4. Comment associer des données géospatiales (ex. centres de cluster) à une activité possible ?

Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (2/7)

Quelles sont les hypothèses de travail?

1. Trois types d'activité :
 - a. Arrêt : vitesses basses ou nulles et peu de variances dans l'orientation
 - b. Déplacement : vitesses élevées et peu de variance dans l'orientation
 - c. Activité avec déplacement : vitesses peu élevées et grandes variances dans l'orientation

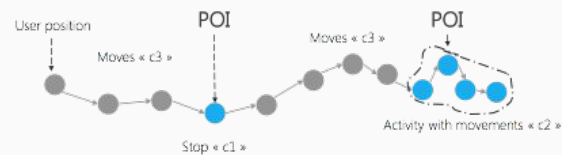
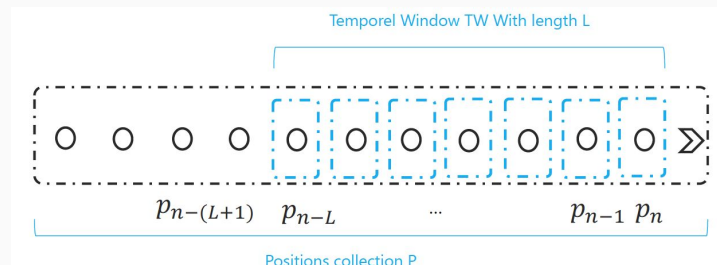


Location Id	Latitude	Longitude	Time	Speed	Bearing	Variance
601	48.4210453	-71.0572413	Mon Nov10..	0.74	76.5	49.7
602	48.4210453	-71.0572532	Mon Nov10..	0.79	76.5	88.36
603	48.4210453	-71.0572567	Mon Nov10..	0.24	76.5	115.7

Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (3/7)

Comment réduire la complexité du problème?

1. Découper les données en fenêtres temporelles d'une longueur L (ex. 10 min.)
 - Peut être fait hors ligne ou en ligne ("buffer")
2. Inférer les activités dans la fenêtre temporelle, puis accumuler les activités reconnues qui sont semblables (ex. déplacement, arrêt)

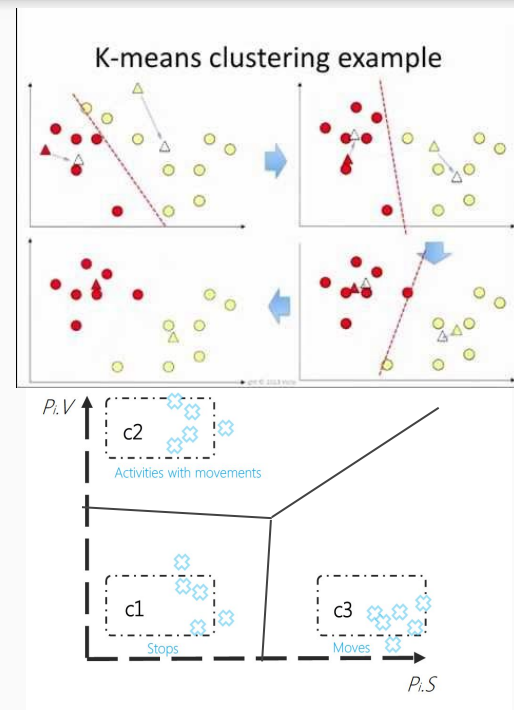


Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (4/7)

Quels sont les algorithmes pouvant être utilisés ?

Trois types d'activité, donc possiblement trois clusters à trouver, donc :

- K-Means : algorithme qui partitionne les données séquentiellement en K clusters de type Voronoi, en calculant la moyenne avec K centre de cluster aléatoire (au début) :
 - https://www.youtube.com/watch?v=_aWzGGNrcic
 - <http://weka.sourceforge.net/doc.dev/weka/clusterers/SimpleKMeans.html>
 - <http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- Faire le clustering sur la variance de l'orientation (bearing) et de la vitesse sur les données de la fenêtre de temps



Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (5/7)

Quels sont les algorithmes pouvant être utilisés ? (suite)

Family Activity
Coordination Dataset
de HCII, CMU

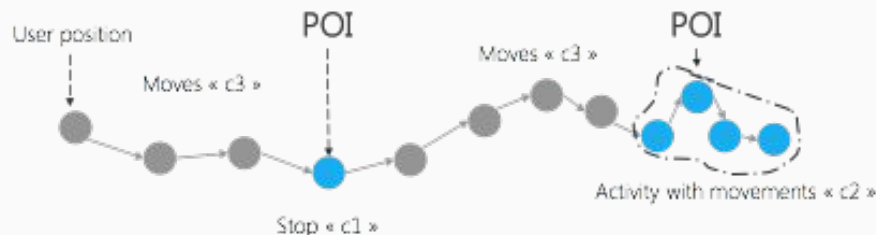


Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (6/7)

Quels sont les algorithmes pouvant être utilisés ? (suite)

Avec les résultats de K-Means :

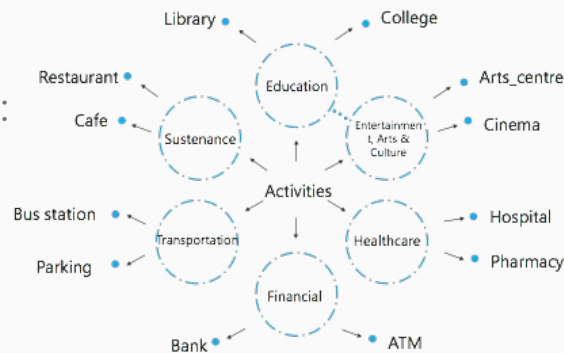
- classifier les données dans les trois types d'activité (ajouter des labels).
- Accumuler les clusters. Par exemple, si nous avons 4 clusters de type déplacement de suite (temporellement), alors les accumuler.



Problème 2 : reconnaître les activités d'utilisateurs à partir de leurs déplacements (7/7)

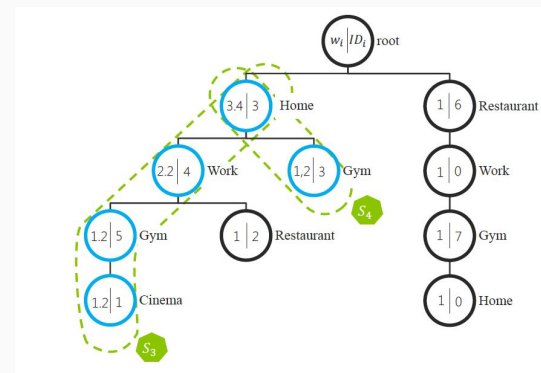
Comment associer des données géospatiales (ex. centres de cluster) à une activité possible ?

- Interroger des bases de données géospatiales pour en retirer les “points of interest” (POI) les plus près du centre des clusters, exemple :
 - OpenStreetMap Overpass :
https://wiki.openstreetmap.org/wiki/Overpass_API#Around
 - Google Places :
<https://developers.google.com/places/web-service/search>
- Utiliser une taxonomie pour lier le POI à des catégories. Associer la catégorie à un type d'activité. Ex. centre sportif = activité sportive, centre commercial = magasinage, etc.



Pour aller plus loin ?

- Possible de prédire avec des modèles bayésiens (ex. Chaîne de Markov)!
 - Prediction of next destinations from irregular Patterns. Journal of Ambient Intelligence and Humanized Computing, 1-13.
<https://link.springer.com/article/10.1007/s12652-017-0519-z>
- Ajouter de la logique temporelle ... matin vs. soir vs. fin de semaines ...
- Ajouter de la micro-reconnaissance d'activités à partir d'autres types de données : accéléromètres, son, image/flux vidéo, etc.
- Ajouter des données provenant des utilisateurs (apprentissage avec renforcement)



Qu'est-il dans d'autres scénarios ?

- Activité des animaux (déplacements, aires de repos, aires de reproduction, etc.) ?
- Activité industrielles ?
- Mouvements de foule ?
- ...

Merci de votre attention !

Questions ?

Travaux supportés par :

