

TP1 : attaques par inférence sur traces de mobilité

Durée du TP : 4h.

Date limite de remise du TP : vendredi 5 février 2016 à 23h59.

Le tp se fait par groupe de 2 étudiants (au maximum) et les fichiers à remettre doivent être envoyés au plus tard vendredi 5 février à minuit par courriel à l'adresse électronique "sgambs@irisa.fr".

Rappel : tout plagiat est formellement interdit et bien qu'il soit naturel que vous puissiez parfois discuter avec vos camarades oralement sur comment attaquer ou résoudre un problème, il est formellement interdit d'échanger des fichiers de code.

Description :

Le but de ce premier TP est de vous familiariser avec l'analyse et les attaques par inférence sur des données géolocalisées. Plus précisément, vous allez recevoir deux fichiers représentant le déplacement de deux taxis de San Francisco pendant plusieurs mois. Il vous faudra analyser ces traces de mobilité en essayant d'extraire des informations telles que les "hotspots" (points de la carte très fréquentés) et si c'est possible les points d'intérêts concernant les chauffeurs de Taxi dont vous avez reçu les traces. Pour cela, vous utiliserez Google Earth pour visualiser les traces de mobilité des taxis mais vous avez évidemment le droit d'utiliser d'autres sources publiques d'information (tel que YahooMaps ou Google Streetview).

Partie 1 : visualisation des données

Vous allez recevoir deux fichiers de mobilité de deux taxis (dont les noms sont remplacés par des pseudonymes) en format kmz. Vous devrez les convertir en format kml avant de commencer à les utiliser. Pour la conversion de fichier, utiliser le site suivant : <http://www.gpsvisualizer.com/>

Vous devrez aussi télécharger et installer Google Earth pour visualiser les traces de mobilité. Familiarisez vous ensuite avec les données avant de passer à la partie suivante.

Partie 2 : attaque par inférence

Vous devrez implémenter une attaque par inférence sur données géolocalisées dans ce TP. Afin d'implémenter cette attaque, vous êtes libres d'utiliser le langage de programmation de votre choix. Cette attaque devra d'identifier les points d'intérêts d'un individu en utilisant un algorithme de clustering du type k-moyennes (voir les pages 23, 24, 72 et 73 du cours 3 pour plus de détails) pour essayer de trouver des groupes de localisation dont la localisation moyenne ou médianes pourrait correspondre à un point d'intérêt (par exemple la maison). Cet algorithme retournera en sortie un fichier texte contenant une représentation des différents groupes (clusters) découverts.

Travail à rendre :

Suite à ce TP, vous devrez remettre un court rapport qui résume ce que vous avez découvert en analysant les traces de mobilité des deux taxis (préciser leurs pseudonymes dans le rapport rendu). En particulier, vous devrez pour chaque taxi

décrire son comportement de mobilité, les points d'intérêts qu'il visite fréquemment ainsi que par exemple si vous pensez avoir identifier le lieu d'habitation de ce taxi ou d'autres informations personnelles. Vous devrez décrire brièvement les problèmes et défis que vous avez rencontré lors de votre analyse et aussi les idées possibles que vous auriez pour les surmonter (mais vous n'êtes pas obligé de les implémenter). Enfin, vous remettrez aussi un fichier contenant le résultat de l'attaque par inférence.