

Anonymisation de trajectoires d'appels mobiles à l'aide de modèles en grilles et de chaînes de Markov

Françoise Fessant*, Fabrice Clérot *
Koray Ozbek **,**

*Orange Labs, Lannion
francoise.fessant, fabrice.clerot@orange.com,
**Université Limoges
koray.ozbek@hotmail.fr

Résumé. Dans cet article, nous proposons une méthodologie pour anonymiser un ensemble de trajectoires individuelles d'appels mobiles. L'objectif est de publier un ensemble de trajectoires anonymes construit à partir de l'ensemble initial, qui protège contre le risque de ré-identification. En d'autres termes, on ne doit pas pouvoir lier une trajectoire dans les données publiées à un individu présent dans la base originale. La solution proposée associe la segmentation des trajectoires d'appels à l'aide de modèles en grille, à un modèle générateur basé sur les chaînes de Markov, pour produire des trajectoires d'appels synthétiques qui peuvent ensuite être publiées à la place des trajectoires originales. L'utilité des données synthétiques, ainsi que le niveau de protection offert sont évalués à travers différents indicateurs statistiques.

1 Introduction

La multiplication des équipements connectés, omniprésents dans notre vie quotidienne et l'adoption des services basés sur la géolocalisation, rendent possible le traçage numérique d'une grande partie de nos activités et déplacements. Tweeter au sujet d'un événement, rechercher un itinéraire avec le système de navigation de son véhicule, appeler quelqu'un avec son téléphone mobile, payer avec sa carte bancaire sont des exemples de situations qui génèrent automatiquement des grandes masses de données qui sont collectées et stockées dans des bases de données. L'analyse de ces données est l'objet d'un intérêt soutenu de la part de différents acteurs : compréhension des mobilités humaines pour le transport (gestion du trafic routier) ou les collectivités locales (fréquentation des sites, (Jiang et al., 2017)), le géomarketing (Giannotti et al., 2009) ou l'aide aux politiques publiques (Blondel et al., 2012). Ces analyses reposent essentiellement sur la collecte de trajectoires de mobilité, c'est à dire l'historique horodaté des positions géographiques successives visitées par un individu. Cependant, l'exploitation ou la publication de telles données pose des questions relatives à la protection des données à caractère personnel. En effet, connaître les déplacements d'une personne peut permettre d'inférer par exemple ses lieux de résidence ou de travail, son identité, ses centres d'intérêts, ses habitudes, ses préférences politiques et sa santé voire une déviation par rapport à son comportement habituel (Gambis et al., 2011).