

Prédiction des comportements touristiques par minage des motifs et des règles séquentielles

Sonia Djebali, Guillaume Guerard

Pôle Universitaire Léonard de Vinci, Research Center, Paris La Défense, France
{sonia.djebali,guillaume.guerard}@devinci.fr

Mots-clés : *aide à la décision, automate stochastique, minage des données.*

1 Introduction

Dans le monde d'aujourd'hui, le tourisme est devenu un des domaines très lucratifs pour l'économie. L'industrie du tourisme est considérée comme l'une des industries les plus importantes et ayant la croissance la plus rapide [3]. Le tourisme international représente actuellement 7% des exportations mondiales de biens, avec un bénéfice estimé à 1500 milliards de dollars. Le nombre total de touristes arrivant dans le monde est passé de près de 25 millions en 1950 à plus de 1323 milliard en 2017, selon l'Organisation mondiale du tourisme, UNWTO¹.

La prévision des visites et de leur ordre chronologique (nommé "comportement") des touristes et la connaissance de la motivation du voyage jouent un rôle clé dans le marketing touristique. Cette utilisation des données permet de créer une demande et d'aider les touristes à prendre des décisions [5].

Le tourisme a bénéficié des réseaux sociaux tels que TripAdvisor, Booking, Facebook, Instagram, etc. Grâce au nombre croissant d'outils numériques, de nouveaux comportements ont émergés. Lorsque des photos sont ajoutées sur le Web, de nombreuses informations sont incluses telles que la localisation, des données sur l'utilisateur et des commentaires. Les images et les photos représentent donc une vue touristique et sociologique [1, 2].

Sur la base d'informations géolocalisées et temporelles de photographies sur Instagram, nous proposons une approche originale pour comprendre et prédire les comportements des touristes en analysant des séquences de lieux visités au cours des voyages effectués par un échantillon choisi de touriste (en fonction de l'âge, la nationalité, etc).

2 Notre approche

En se basant sur la définition du tourisme de l'UNWTO nous définissons un séjour comme une séquence (ordonnée) de lieux visités par un touriste. A partir des data Instagram récolté nous avons construit un dataset de comportements touristiques.

Cette approche (voir Figure 1) est basée sur l'exploration de données, en particulier l'exploration de motifs [6] et de règles séquentielles [4]. L'idée principale est de trouver les lieux les plus visités et l'itinéraire le plus fiable emprunté afin de prédire des séjours et circuits touristiques.

Suite à ce traitement des données, un automate probabiliste (non stochastique) est construit en fonction des règles et de la confiance de ces dernières (voir Figure 2). La confiance d'une séquence de règle représente la proportion de chemin partant d'un noeud A vers un noeud B sur le nombre total de chemin partant de A .

Il est alors possible d'effectuer une marche aléatoire sur cet automate afin de prédire des comportements touristiques, tout en analysant la confiance et la conviction de ces derniers. La conviction mesure le taux d'erreur attendu d'une séquence, une conviction de 1 correspond à

1. World Tourism Organization UNWTO. *UNWTO Tourism Highlights : 2018 Edition*

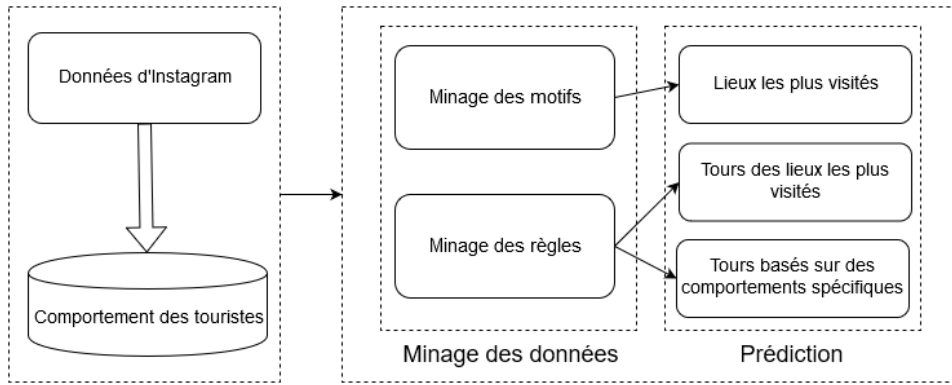


FIG. 1 – Démarche sur l'analyse des données.

une séquence qui serait purement aléatoire. Une grande conviction montre que la séquence est très fréquente. Par exemple le chemin {Musée du Louvre, Pont Neuf, Cathédrale Notre-Dame de paris} aura une confiance de 0.13 et une conviction moyenne de 7.61.

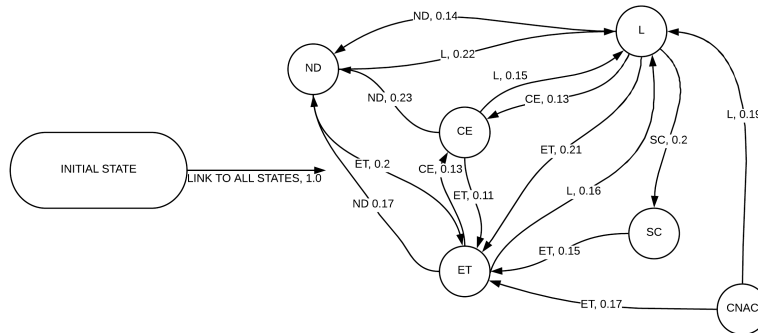


FIG. 2 – Automate probabiliste par analyse des règles séquentielles.

Références

- [1] Richard M Chalfen. Photograph's role in tourism : Some unexplored relationships. *Annals of tourism research*, 6(4) :435–447, 1979.
- [2] G Chareyron, S Cousin, J Da-Rugna, and D Gabay. Touriscope : map the world using geolocated photographs. In *IGU meeting, Geography of Tourism, Leisure and Global Change*, 2009.
- [3] Chris Cooper and C Michael Hall. *Contemporary tourism*. Routledge, 2007.
- [4] Philippe Fournier-Viger, Roger Nkambou, and Vincent Shin-Mu Tseng. Rulegrowth : mining sequential rules common to several sequences by pattern-growth. In *Proceedings of the 2011 ACM symposium on applied computing*, pages 956–961. ACM, 2011.
- [5] Roger March and Arch G Woodside. *Tourism behaviour : Travellers' decisions and actions*. Cabi, 2005.
- [6] Qiankun Zhao and Sourav S Bhowmick. Sequential pattern mining : A survey. *ITechnical Report CAIS Nanyang Technological University Singapore*, 1 :26, 2003.