

À la recherche du quartier idéal

Nelly Barret*, Fabien Duchateau*, Franck Favetta*,
Maryvonne Miquel*, Aurélien Gentil**, Loïc Bonneval**

*LIRIS UMR5205, Université de Lyon, France
prénom.nom@liris.cnrs.fr

**Centre Max Weber, Université de Lyon, France
prénom.nom@univ-lyon2.fr

Résumé. Le choix d'un quartier est primordial lors d'un achat ou d'une location immobilière. Or, il est fréquent de ne pas connaître la ville où l'on arrive (e.g., mutation professionnelle) et la sélection d'un quartier pertinent devient alors un véritable défi. Dans cet article, nous présentons un outil qui facilite la comparaison entre quartiers. Nous exploitons plusieurs indicateurs pour différencier les quartiers et plusieurs algorithmes permettent soit de recommander un quartier, soit de regrouper des quartiers similaires.

1 Introduction

Avec l'émergence du web, les systèmes de recommandation sont devenus des outils incontournables pour découvrir de nouveaux objets, en particulier pour les sites de e-commerce (Ricci et al. (2011)). La multiplication des contenus multimédias a également contribué à ce phénomène, comme le montre le challenge Netflix de 2009¹. Cet engouement s'est propagé à d'autres domaines, que ce soit la recommandation d'articles scientifiques (Beel et al. (2016)) ou de lieux (Bobadilla et al. (2013)). Malgré ces avancées, peu de travaux portent sur les mobilités géographiques et résidentielles. Si celles-ci sont courantes (e.g., premier travail, mutation professionnelle, alternance ou stage), il s'avère dans bien des cas que les personnes qui déménagent ne connaissent pas leur future ville d'habitation et ses quartiers. Comme le choix d'un quartier a un impact significatif sur la vie des personnes, nous proposons un nouvel outil de comparaison entre quartiers permettant d'accompagner les personnes dans leur mobilité.

La recommandation de quartiers soulève différents problèmes. Premièrement, la notion de quartiers (i.e., définition, délimitations, perception) est floue et sujette à différentes interprétations (Authier et al. (2007)). Un second problème concerne la collecte de données. Les systèmes de recommandation de lieux exploitent essentiellement les réseaux sociaux géolocalisés, et les "likes", "checkins" ou évaluations servent de données pour construire et comparer des profils d'utilisateurs (Wang et al. (2013); Li et al. (2016)). Ces informations, indisponibles en grand nombre et pour chaque quartier, sont non pertinentes dans notre contexte

Ce travail a été réalisé au sein du LABEX IMU (ANR-10-LABX-0088) de l'Université de Lyon, dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) de l'Etat Français.

1. Compétition Netflix, <http://www.netflixprize.com/>

Un outil pour la recommandation et le regroupement de quartiers

(e.g., un quartier où l'on aime sortir n'est pas forcément un quartier où l'on souhaite habiter). Enfin, un dernier défi consiste à identifier les informations à recueillir auprès des utilisateurs pour lancer une recommandation. Des plateformes comme Airbnb² ou Rue89 Lyon³ offrent un service de découverte d'un quartier idéal, mais restent limitées en terme de villes (une vingtaine pour Airbnb) ou en terme de critères (en général une dizaine). Le système de recommandation immobilière de Yuan et al. (2013) a été proposé pour quelques villes de Corée du Sud. Il se concentre davantage sur la recommandation de logement que sur celle de quartier, ce dernier n'étant qu'un critère parmi d'autres. Trois profils d'utilisateurs ont été définis, et le moteur trouve des recommandations par *case-based reasoning*, i.e., par comparaison avec des cas résolus. Un tel système est difficilement généralisable pour l'ensemble d'un pays.

Afin de dépasser ces différentes limites, nous avons développé un nouveau prototype de système de recommandation et de regroupement de quartiers pour la France : VizLIRIS. Nous détaillerons d'abord le coeur de ce prototype (intégration de données et fonctionnement des algorithmes) pour ensuite présenter et démontrer son utilité à travers deux scénarios.

2 Un outil de comparaison entre quartiers

Le prototype inclut d'abord une étape de préparation (collecte des données) avant d'appliquer des algorithmes de comparaison ou de similarité entre quartiers.

Intégration de données. Pour collecter une quantité suffisante de données, nous utilisons l'unité géographique de l'INSEE appelée IRIS (Ilots Regroupés pour l'Information Statistique). Un IRIS est une maille du territoire, découpé de façon relativement homogène (même nombre d'habitant.e.s). Il est défini par un code, son contour (ensemble de coordonnées géographiques), mais aussi par plus de 800 indicateurs bruts (e.g. le nombre de boulangeries, la répartition par type de logement ou par catégorie socio-professionnelle) répartis dans une vingtaine de sources INSEE. Ces informations sont complétées par d'autres sources pour les prix moyens au m² (au niveau de la commune ou d'un arrondissement). Les indicateurs collectés pour chaque IRIS sont très spécifiques et doivent être regroupés en indicateurs synthétiques pour pouvoir être traités de façon systématique sur un grand nombre d'IRIS. Nous avons donc une trentaine d'indicateurs synthétiques (e.g., loisirs, commerces de proximité). Le voisinage d'un IRIS est également pris en compte pour se rapprocher du concept de quartier (un IRIS étant généralement plus petit qu'un quartier). Les mappings requis par cette intégration ont été développés à la main, et les données intégrées sont stockées au format GeoJSON.

Comparaison de quartiers. Pour l'élaboration du prototype, nous nous sommes placés dans le cas hypothétique où les clients recherchent un quartier similaire à celui dans lequel ils habitaient auparavant. Huit algorithmes sont utilisés pour recommander ou regrouper des quartiers similaires. Ils prennent en entrée un ou plusieurs quartiers de départ (e.g., domicile actuel, autres quartiers appréciés) et un point d'arrivée (e.g., lieu du nouvel emploi). Les quartiers candidats sont ceux situés dans un rayon donné autour du point d'arrivée. L'algorithme **similarité cosine** permet, dans notre contexte, de calculer directement le degré de ressemblance entre

2. Airbnb locations, <http://www.airbnb.fr/locations>

3. Datavilles, <http://www.rue89lyon.fr/?p=88778>

le quartier de départ et les quartiers candidats. Notre **algorithme de l'écart-type** construit un profil "utilisateur" à partir de plusieurs quartiers de départ. Notre hypothèse est qu'un écart-type faible, pour un indicateur, signifie que ce dernier est important. Chaque indicateur du profil a donc un poids calculé selon l'inverse de son écart-type. Ensuite la mesure cosinus est appliquée entre le profil construit et les candidats. Les **algorithmes de clustering** (Birch, Kmeans, Mean Shift, Mini Batch KMeans et Spectral Clustering) peuvent être appliqués dans le cas de la recommandation et dans le cas d'un regroupement. Dans le premier cas, les quartiers candidats sont répartis dans des *clusters*, puis on détermine le cluster le plus proche pour le quartier de départ et les candidats appartenant au cluster choisi sont recommandés. Les algorithmes de clustering utilisés pour le regroupement permettent de détecter directement les quartiers les plus similaires au quartier contenant le point d'arrivée. Enfin l'**algorithme de type SVM** (one-class SVM) exploite plusieurs quartiers de départ comme des exemples positifs pour l'apprentissage. Un vecteur de caractéristiques représentatif est construit à partir des coefficients calculés pendant l'apprentissage, et ce vecteur est utilisé comme "quartier de départ fictif" d'un autre algorithme (e.g. similarité cosinus ou clustering).

3 Scénarios d'utilisation

Le prototype VizLIRIS est développé en Python et utilise la librairie scikit-learn⁴ pour les algorithmes de recommandation et regroupement ainsi que la librairie gdal⁵ pour les aspects cartographiques. Plusieurs paramètres (e.g., rayon de recherche, nombre de recommandations) sont réglables au niveau de l'interface. Dans la suite, nous présentons deux cas d'utilisation.

Scénario 1 : recommandation de quartiers. Alice vit à Montpellier dans le centre ville animé de l'Écusson, et elle déménage sur Lyon suite à une mutation professionnelle. Ne connaissant pas sa future ville, il lui est difficile de choisir un quartier où elle se plaira. Dans l'outil VizLIRIS, elle indique sur une carte son point de départ (centre de Montpellier) et choisit une zone d'arrivée de 3 kilomètres à Lyon. La figure 1a montre les recommandations suggérées, parmi lesquelles le *vieux Lyon*, les pentes de la Croix-Rousse ou la Part-Dieu / Brotteaux (avec animation, types de logement et commerces comme critères principaux).

Scénario 2 : regroupement de quartiers. Bob est sociologue et étudie les populations de Paris. La fonctionnalité *regroupement* de VizLIRIS lui permet de trouver les quartiers les plus similaires dans une zone donnée. La figure 1b montre les groupes obtenus avec l'algorithme *Birch* dans un rayon de 3 kilomètres. Bob remarque par exemple que plusieurs zones partagent les mêmes caractéristiques de population, par exemple les berges (couleur violet foncé) ou la périphérie (couleurs bleu ciel et rouge).

Ces deux scénarios illustrent les capacités de VizLIRIS à recommander ou regrouper des quartiers. En terme de perspectives, il est envisagé d'ajouter des justifications pour aider à comprendre les résultats, de faciliter la sélection d'un algorithme et de développer un algorithme pour détecter les quartiers en regroupant les IRIS adjacents similaires.

4. scikit-learn, <http://scikit-learn.org/>

5. Geospatial Data Abstraction Library, <http://www.gdal.org/>

Un outil pour la recommandation et le regroupement de quartiers

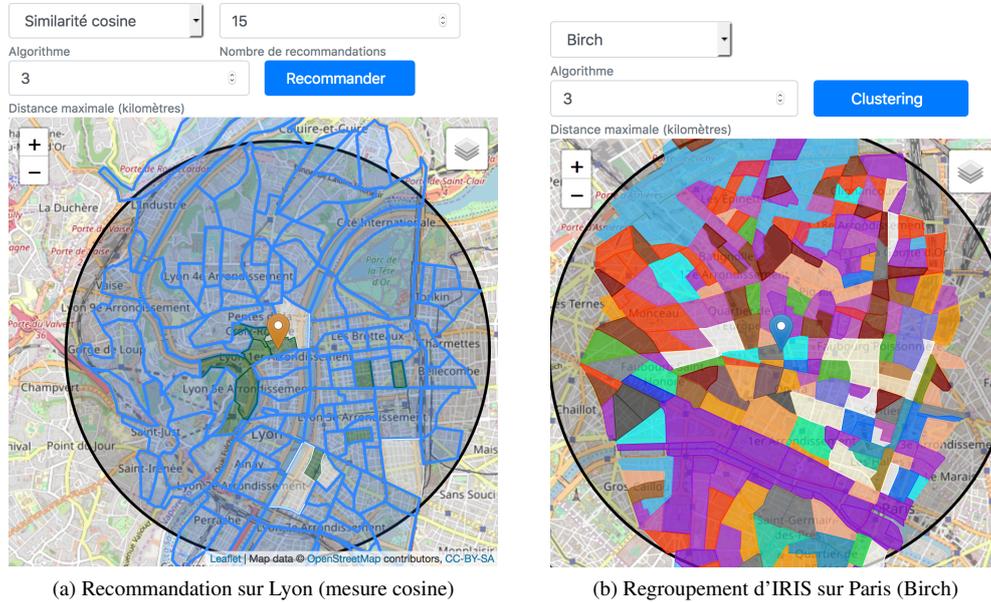


FIG. 1 – Captures d'écran du prototype VizLIRIS

Références

- Authier, J.-Y., M.-H. Bacqué, et F. Guérin-Pace (2007). *Le quartier*. La Découverte.
- Beel, J., B. Gipp, S. Langer, et C. Breiting (2016). Research-paper recommender systems : a literature survey. *International Journal on Digital Libraries* 17(4), 305–338.
- Bobadilla, J., F. Ortega, A. Hernando, et A. Gutiérrez (2013). Recommender systems survey. *Knowledge-Based Systems* 46, 109 – 132.
- Li, H., Y. Ge, R. Hong, et H. Zhu (2016). Point-of-interest recommendations : Learning potential check-ins from friends. In *SIGKDD*, pp. 975–984. ACM.
- Ricci, F., L. Rokach, B. Shapira, et K. Paul B. (2011). *Recommender systems handbook*. Springer.
- Wang, H., M. Terrovitis, et N. Mamoulis (2013). Location recommendation in location-based social networks using user check-in data. In *SIGSPATIAL*, pp. 374–383. ACM.
- Yuan, X., J.-H. Lee, S.-J. Kim, et Y.-H. Kim (2013). Toward a user-oriented recommendation system for real estate websites. *Information Systems* 38(2), 231 – 243.

Summary

When moving in a new city, the choice of a future neighborhood is a challenging issue. This paper presents a tool for comparing neighborhoods according to a set of indicators and based on 8 algorithms for recommending or clustering.