

Exploration Visuelle de Données Spatio-Temporelles Brutes

Romain Vuillemot

Université de Lyon, LIRIS
romain.vuillemot@gmail.com,
<http://romain.vuillemot.net/>

Résumé. Les données spatio-temporelles sont omniprésentes, et pour les comprendre dès leur collecte, il est nécessaire de développer des outils agnostiques vis à vis de leur type, volume et distribution. Nous proposons l'utilisation de graphiques standards (histogrammes, scatterplots, etc.) coordonnés pour l'analyse visuelle de ces données à ce stade très en amont sans hypothèse initiale. Ces graphiques ont la particularité de ne pas être dépendant des caractéristiques du jeu de données et passent à l'échelle facilement. Nous montrons un exemple d'implémentation ce type d'interface appliqué à des centaines de milliers de trajectoires aériennes de manière dynamique et dans le navigateur web.

Contexte. Les activités humaines telles que les déplacements en vélo et en avion, mais aussi les mouvements de pointeurs sur un écran d'ordinateur, génèrent de grands volumes de données qui ont une position (x, y) évoluant au fil du temps (t) . L'analyse de ces données permet de comprendre la dynamique temporelle d'événements dans un certain espace, et le comportement d'objets qui se déplacent dans plusieurs lieux (physiques ou virtuels). Le temps et l'espace constituant un index fort de ces données, qu'il est aussi difficile de dissocier, notamment lors de leur représentation : autrement dit les données spatio-temporelles doivent être représentées en points ou trajectoires dans un plan cartésien. Celles-ci peuvent être contextualisées avec des métadonnées géographiques (ex : plan d'un bâtiment), mais il est difficile de combiner des dimensions supplémentaires ou dérivées (vitesse, accélération, etc.). Ainsi les possibilités de mapping visuel sont limitées à une ou deux dimensions supplémentaires (en utilisant la couleur, des formes et l'animation). Il est donc nécessaire d'utiliser l'interaction utilisateur pour explorer les données et les représentations. Peuquet (2002) et Andrienko et al. (2011) ont proposé des frameworks complets permettant de systématiser l'exploration et la représentation de ces données. Peuquet propose la décomposition en *what+why=when*, *what+when=why* et *when+why=what* afin de décomposer et recombinaison les données pour répondre à certaines questions. Andrienko suggère une taxonomie de tâches élémentaires et de recherche de relation, aussi bien directes qu'indirectes.

Visualisations simples et génériques. Notre motivation de design pour explorer les données (x, y, t) repose sur la nécessité de disposer d'outils d'analyse ne faisant pas d'hypothèse sur les données collectées. Le but étant de pouvoir identifier d'éventuelles anomalies, mais aussi tendances comme des corrélations. Ces étapes sont préalables à toute phase de traitement et modélisation de données, comme la classification ou la prédiction. La Figure 1 illustre le type d'approche que nous avons déjà implémenté, et les détails de l'interface sont les suivants :

- ① **Vue centrale.** Affichage de la position (x, y) et les trajectoires sous forme d'arcs. Cette vue sert essentiellement d'aperçu sans nécessairement être interactive. D'autres

représentations que l'espace cartésien peut être utilisés, comme les Space-Time cubes avec une vue 3D de l'espace de données.

- ② ③ **Vues de distributions.** Au moyen d'histogrammes, il s'agit d'afficher les propriétés de chacune des dimensions de manière indépendante. L'utilisation de ces graphiques simples permet aussi de nettoyer les données Kandel et al. (2011). Des scatterplots peuvent aussi être utilisés, mais leur nature bi-variable nécessiteraient d'afficher toutes les permutations comme sous forme de matrice Elmqvist et al. (2008).
- ④ **Détails des données.** Il s'agit d'afficher sous forme de tables les données brutes et les valeurs exactes pour toutes les dimensions.
- **Coordination des vues.** Les vues précédentes doivent être rendues interactives afin de permettre le filtrage et la composition de requêtes complexes. Ces requêtes doivent être réalisées avec un temps de réponse quasiment immédiat (< 100ms) afin de garder l'attention de l'humain. Cette coordination dynamique a certes été proposée il y a plus de 25 ans avec HomeFinder Williamson et Shneiderman (1992), mais de nouveaux challenges se posent désormais avec de grands volumes de données, à la fois en termes de rapidité d'affichage et de progressivité du rendu si il n'est pas immédiat.

Nous avons testé cette interface illustrée figure 1 sur différents jeux de données (trajectoires d'avions et de vélos, motion capture 3D d'un humain, position curseur sur un écran, position d'individus dans un bâtiment) qui ont des propriétés suffisamment diverses pour mettre de valider la généralité de l'approche. Nos perspectives de travaux visent donc à étudier comment ce type d'interfaces peuvent être améliorés afin de supporter de manière complète les tâches d'exploration de Andrienko et al. (2011) et Peuquet (2002). En particulier nous sommes intéressés par les problèmes inverses en utilisant l'interaction avec la sortie de l'interface pour explorer les paramètres de configuration d'entrée similaire à Vuillemot et Perin (2015).

Références

- Andrienko, G., N. Andrienko, P. Bak, D. Keim, S. Kisilevich, et S. Wrobel (2011). A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing* 22(3), 213–232.
- Elmqvist, N., P. Dragicevic, et J. Fekete (2008). Rolling the Dice : Multidimensional Visual Exploration using Scatterplot Matrix Navigation. *IEEE Transactions on Visualization and Computer Graphics* 14(6), 1539–1148.
- Kandel, S., A. Paepcke, J. Hellerstein, et J. Heer (2011). Wrangler : Interactive Visual Specification of Data Transformation Scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, New York, NY, USA, pp. 3363–3372. ACM.
- Peuquet, D. J. (2002). *Representations of Space and Time*. Guilford Press.
- Vuillemot, R. et C. Perin (2015). Investigating the Direct Manipulation of Ranking Tables for Time Navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, pp. 2703–2706. ACM.
- Williamson, C. et B. Shneiderman (1992). The Dynamic HomeFinder : Evaluating Dynamic Queries in a Real-estate Information Exploration System. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '92, New York, NY, USA, pp. 338–346. ACM.

Annexe



FIG. 1 – Aperçu de l'interface générique d'exploration de données spatio-temporelles que nous proposons, appliquée avec un jeu de données de trajectoires d'avions, affichées sous forme de points et d'arcs dans la vue principale ①. Des vues latérales descriptives d'attributs permettent le filtrage ② et peuvent être combinées entre elles ③. Une table affiche les détails de chaque dimension ④. Après quelques cliques, l'utilisateur découvre que la plupart des vols en retard le sont le soir. L'interface utilise les jeux de données de ASA Data Expo et la bibliothèque DC.JS (<https://dc-js.github.io/dc.js/>) autour de Crossfilter (<http://square.github.io/crossfilter/>) Le jeu de données affiche 231 038 vols américains au premier trimestre 2001.