
Représentation Symbolique de Trajectoires Géolocalisées

Soda Marème Cissé¹, Olivier Ridoux¹, Peggy Cellier²

1. Irisa, Université Rennes 1

263, avenue du Général Leclerc, 35000 Rennes
soda.mareme.cisse@irisa.fr,olivier.ridoux@irisa.fr

2. Irisa, INSA de Rennes

263, avenue du Général Leclerc, 35000 Rennes
peggy.cellier@irisa.fr

RESUME. Énormément de trajectoires géolocalisées sont produites et disponibles, mais leur exploitation pour en tirer des informations intéressantes reste difficile. L'obstacle qui nous intéresse ici est la mauvaise qualité de certaines de ces données : échantillonnage très irrégulier et précision médiocre. Nous proposons de les transposer dans des domaines symboliques pour pouvoir effectuer des traitements tels que l'application de méthodes de fouille de données : extraction de motifs, règles d'association, clustering, etc. Cette transposition devra se faire en évitant le plus possible de fixer des valeurs seuil a priori afin que la même méthode s'applique quelle que soit la nature des objets mobiles.

Dans ce papier, nous proposons d'utiliser la méthode SAX (Symbolic Aggregate approXimation) pour la représentation symbolique de trajectoires d'objets mobiles. SAX permet de représenter une série de valeurs numériques unidimensionnelles par une série de valeurs symboliques déterminées automatiquement par rapport à la distribution des valeurs numériques. Nous utilisons dans ce papier deux méthodes pour appliquer SAX sur les trajectoires d'objets mobiles (données multidimensionnelles). Dans la première, les valeurs numériques sont des vitesses par rapport à des positions, et dans la seconde ce sont des positions par rapport au temps.

ABSTRACT. Lot of geolocalized trajectories are produced and available but their exploitation to extract interesting information remains difficult. The obstacle we address here is the poor quality of some of these data: very irregular sampling and poor precision. We propose to transpose them into symbolic representations to perform treatments such as the application of data mining methods: extraction of patterns, association rules, clustering, etc. This transposition should avoid as much as possible to rely on a priori threshold values so that the same method applies regardless of the nature of moving objects. In this paper, we propose to use the SAX method (Symbolic Aggregate approXimation) for the symbolic representation of moving objects trajectories. SAX is used to represent a series of one-dimensional numerical values by a series of symbolic values automatically determined in relation to the distribution of numerical values. We present two applications of SAX to moving objects trajectories

(multidimensional data). In the first application, numerical values are instantaneous speeds relative to positions and in the second application, these are positions relative to time.

MOTS-CLES : TRAJECTOIRES GEOLOCALISEES, SAX, COURBES DE HILBERT.

KEYWORDS: GEOLOCALIZED TRAJECTORIES, SAX, HILBERT CURVES.

1. Introduction

De plus en plus de trajectoires géolocalisées sont disponibles grâce aux outils électroniques tels que les Smartphones, les tablettes, etc. Une trajectoire est définie par (Spaccapietra et al., 2008) comme l'évolution de la position d'un objet qui se déplace dans l'espace durant un intervalle de temps donné afin d'atteindre un objectif donné. Les objets mobiles qui suivent ces trajectoires peuvent être des personnes, des animaux, des véhicules, etc. Les données de trajectoires sont disponibles sous forme de coordonnées géographiques avec des estampilles temporelles. Elles sont parfois imparfaites car imprécises, mal échantillonnées, etc. Leur aspect bruité conduit à un prétraitement consistant à corriger les imprécisions.

Des solutions ont été proposées pour le nettoyage de ces imprécisions en les supprimant ou en les réduisant (Meratnia et de By, 2004), (Lashley, 2006), (Marketos et al., 2008). Dans nos travaux, l'aspect du nettoyage qui nous concerne particulièrement est le ré-échantillonnage qui corrige l'échantillonnage très irrégulier des données auquel nous avons souvent eu à faire face. Notre objectif général est d'appliquer des méthodes de fouille de données symboliques : règles d'association, extraction de motifs, analyse de concepts formels, etc. Nous pensons que traiter les données symboliquement fournit des résultats plus faciles à interpréter par rapport à une représentation numérique.

Un modèle d'abstraction des trajectoires a été proposé par (Spaccapietra et al., 2008). C'est un modèle qui permet de segmenter la trajectoire en des épisodes de « Move » (déplacement) et de « Stop » (arrêt). Des solutions se basant sur ce modèle ont été proposées pour abstraire les trajectoires en les segmentant (Yan, 2011), (Alvares et al., 2007) mais aussi pour y ajouter plus de sémantique. (Zheng et al., 2008) et (Buchin et al., 2010) ont aussi proposé des solutions pour la segmentation de trajectoires. La plupart des méthodes de segmentation proposées définissent un seuil pour pouvoir distinguer par exemple les phases où l'objet mobile est quasi en arrêt et celles où il est en déplacement. Ceci peut conduire à définir plusieurs seuils suivant le type d'objet mobile et le type d'application et c'est ce genre de seuil a priori que nous souhaitons éviter. On souhaite aussi distinguer plusieurs types de déplacement (« Move ») comme un déplacement lent (« Slow »), un déplacement rapide (« Fast ») et un déplacement très rapide (« Very Fast »).

Nous proposons d'appliquer SAX (Symbolic Aggregate approXimation) sur les trajectoires d'objets mobiles dans le but de les discrétiser pour pouvoir ensuite appliquer des méthodes de fouille de données afin d'extraire des motifs, des trajets récurrents en évitant de définir des seuils a priori.

Ce papier s'articule comme suit. Dans la section 2 nous présentons les trajectoires d'objets mobiles en général. Dans la section 3, nous présentons SAX. Dans la section 4 nous présentons les courbes de Hilbert qui permettront de faire la correspondance de données multidimensionnelles vers des données unidimensionnelles. Dans la section 5 nous présentons l'application de SAX sur les trajectoires d'objets mobiles avec deux approches différentes notamment sur des vitesses et sur des coordonnées et dans la section 6 nous présentons les expériences qui ont été conduites. Enfin, dans la section 7 nous concluons et discutons de nos travaux futurs.

2. Trajectoires géolocalisées

2.1. Définition

Une trajectoire géolocalisée représente le chemin que suit un objet mobile durant un intervalle de temps donné. On représente une trajectoire comme suit :

$T = \{(x_1, y_1, t_1), \dots, (x_n, y_n, t_n)\}$ où (x_i, y_i) sont des coordonnées géographiques et t_i une estampille temporelle.

2.2. Ré-échantillonnage de trajectoires géolocalisées

Les données étant souvent mal échantillonnées, on passe par une phase de ré-échantillonnage des données. Par exemple, sur un jeu de données que nous utilisons, les trajectoires sont capturées par des balises disposées dans des remorques de camions autonomes énergétiquement et jamais éteintes pendant plusieurs années. Les données capturées sont envoyées sur un serveur pour un traitement hors ligne. Ces balises ont des périodes de silence de plusieurs heures alors que la période d'échantillonnage nominale est de 5 minutes environ. Nous pensons que la raison de ces silences est que les balises doivent fonctionner en autonomie totale pendant plusieurs années, en particulier en autonomie énergétique. Pour cela, elles ne font qu'un essai de consultation des satellites de géo-positionnement plutôt que de les réinterroger jusqu'à obtenir un signal correct ainsi certaines positions sont perdues. Le ré-échantillonnage consiste à faire une interpolation des positions manquantes en définissant un intervalle de temps régulier entre les prises de positions. Pour ce faire, on a conçu un algorithme permettant d'interpoler les positions manquantes en calculant le barycentre entre deux positions :

- Si la prochaine position qui doit être interpolée se situe entre la dernière position réellement mesurée p_i et la position suivante réellement mesurée p_{i+1} , on calcule le barycentre entre les deux positions réellement mesurées ;
- Tant qu'on n'a pas dépassé l'estampille temporelle de la position p_{i+1} , on fait toujours le calcul de barycentre cette fois entre la dernière position interpolée et p_{i+1} .

3. SAX

SAX (Symbolic Aggregate approXimation) est une méthode proposée par (Lin et al., 2003) pour représenter symboliquement une série de données numériques par des chaînes de symboles. Le but est de discrétiser et de réduire des séries temporelles pour appliquer des techniques de fouille de données sur ces données numériques. Dans SAX, une série temporelle de longueur n $S = s_1, \dots, s_n$ de valeurs numériques est donnée en entrée et une chaîne de symboles : $\hat{S} = \hat{s}_1, \dots, \hat{s}_w$ de dimension w , $w \ll n$ sur un alphabet (ensemble de symboles utilisés dans la discrétisation) de taille a est produite en sortie.

Dans la phase de production de la chaîne de symboles SAX, les valeurs numériques sont d'abord normalisées pour produire une nouvelle série de longueur n $S' = s'_1, \dots, s'_n$. La normalisation consiste à produire une séquence de valeurs numériques de moyenne (μ) de 0 et une déviation (σ) de 1 telle que :

$$s'_i = \frac{s_i - \mu(S)}{\sigma(S)}$$

La distribution de la série normalisée est divisée en a intervalles qui divisent la distribution en quantiles égaux. Un tableau de *apoints* séparant les intervalles, appelés des « *breakpoints* » est créé (Lin et al., 2003) pour associer à chaque quantile un symbole. Puis la série normalisée est divisée en w fenêtres de taille $\frac{n}{w}$. La moyenne (nécessitant le ré-échantillonnage régulier effectué au préalable) des valeurs numériques normalisées constituant chaque fenêtre est calculée et sert d'index dans le tableau de « *breakpoints* » pour déterminer le symbole qui lui correspond dans l'alphabet et qui correspond à un intervalle de valeurs numériques. Le calcul de ces moyennes est appelé PAA (*Piecewise Aggregate Approximation*). C'est une méthode proposée par (Keogh et al., 2001) et qui produit une nouvelle série : $\bar{S} = \bar{s}_1, \dots, \bar{s}_w$. Chaque moyenne est calculée comme suit :

$$\bar{s}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} s_j$$

Les symboles de toutes les fenêtres concaténés dans l'ordre de la série vont représenter un « *mot* » SAX. La correspondance entre un segment PAA et un symbole SAX est faite comme suit. Supposons que l'alphabet est $\{a, b, c, \dots\}$. Les valeurs de segments PAA qui sont en dessous du plus petit breakpoint sont représentées par le symbole a , les valeurs supérieures au plus petit breakpoint et inférieures au deuxième plus petit breakpoint sont représentées par le symbole b et ainsi de suite. La figure 1 donne pour un ensemble de valeurs numériques divisées en fenêtres un exemple d'affectation des valeurs PAA (les moyennes des fenêtres) avec les symboles SAX. La taille de l'alphabet est de 3 ici ($\{a, b, c\}$).

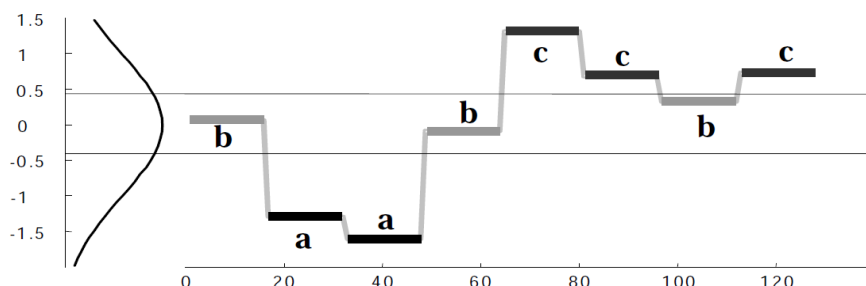


Figure 1. Distribution des symboles par rapport aux valeurs (Lin et al. 2003)

Nous utilisons la méthode SAX pour représenter symboliquement les données numériques dans le but de les segmenter mais aussi d'en extraire des patterns réguliers.

3.2. SAX et les trajectoires géolocalisées

SAX est utilisée dans plusieurs domaines en particulier sur les données spatio-temporelles qui nous concernent. SAX peut être appliquée pour étudier les trajectoires d'objets mobiles. (Bakalov et al., 2005a, 2005b) utilisent SAX pour identifier des paires de trajectoires ayant des comportements similaires.

Les trajectoires géolocalisées étant de nature multidimensionnelle et SAX étant une méthode prenant des valeurs unidimensionnelles indexées par le temps, on utilise d'abord dans un premier temps une série de vitesses instantanées de l'objet mobile sur chaque instant de la trajectoire pour appliquer SAX. Et dans un deuxième temps, nous utilisons les courbes de Hilbert pour transformer les positions (paires de coordonnées -latitude, longitude) bidimensionnelles en des positions unidimensionnelles.

4. Courbes de Hilbert

Une courbe de Hilbert est une courbe fractale continue remplissant le plan. Il en existe une grande variété mais celle-ci a été définie en 1890 par David Hilbert (Bader, 2013). C'est une courbe qui passe par tous les points de l'espace sur lequel elle est définie lorsqu'elle est itérée indéfiniment. En pratique, on n'itère pas indéfiniment, et on stoppe le raffinement une fois une définition désirée atteinte. Dans ce cas, la courbe ne passera pas par tous les points de l'espace, mais seulement par tous les points d'une grille ; c'est seulement à la limite qu'elle remplit le plan. D'une façon générale, ces courbes distordent les positions et les distances puisqu'elles projettent des positions quasi-continues sur une grille discrète et

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

qu'elles déterminent un ordre de parcours de cette grille. La courbe de Hilbert a la propriété d'être de celles qui distordent le moins les positions. Ceci justifie son utilisation pour faire la correspondance entre des données multidimensionnelles vers des données unidimensionnelles.

4.1. Définition

Soit la fonction $h : [0,1] \rightarrow [0,1]^2$, à chaque paramètre t position sur une courbe de Hilbert, il lui est associé un unique couple de coordonnées (x,y) position d'un point de la grille dans le plan. Par contre, plusieurs couples (x,y) peuvent avoir comme paramètre un même t .

4.2. Génération

À l'étape initiale, c'est une courbe qui relie les centres de 4 sous-carrés d'un carré initial (carré de Hilbert de côté 1). À l'étape suivante, la courbe obtenue à l'étape initiale (reliant les quatre centres) est dupliquée en 4 copies avec des orientations différentes et chacune est reliée avec sa voisine par un segment. À chaque itération, la courbe obtenue à l'itération précédente est copiée quatre fois et chaque copie orientée différemment. Le processus peut être répété indéfiniment ou un nombre fini de fois qu'on appellera *l'ordre* de la courbe. La figure qui suit montre les trois premières itérations de la courbe.

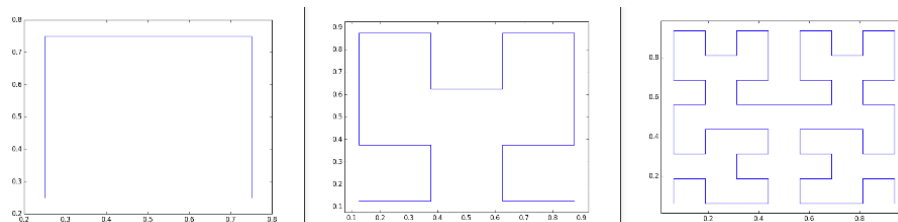


Figure 2. Trois premières itérations de la courbe de Hilbert de gauche à droite ordre 1, 2, 3

Quatre opérateurs (Bader, 2013) sont appliqués sur les paramètres t pour obtenir les différentes orientations de la courbe et pour calculer les points images des paramètres t . Les opérateurs sont appliqués en fonction de la représentation en base 4 du paramètre t . La représentation en base 4 permet de numéroter les sous-carrés de 0 à 3 et de pouvoir déterminer dans quel sous-carré on se trouve à chaque itération.

Ci-dessous, les quatre opérateurs :

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

- H_0 : la courbe est orientée dans le sens des aiguilles d'une montre à travers une réflexion sur la diagonale principale.
- H_1 : La courbe est laissée telle quelle, seule une translation à droite est effectuée.
- H_2 : La courbe est laissée telle quelle, seule une translation à gauche est effectuée.
- H_3 : La courbe est orientée dans le sens contraire des aiguilles d'une montre.

Exemple : Soit $t=0.125$ en base 10, en base 4, $t=0_4.02$, et $(x,y)=H_0H_2$

L'algorithme 1 détaille le calcul de l'image d'un paramètre t donné en fonction des quatre opérateurs et l'algorithme 2 permet de calculer un paramètre t pour un point (x,y) donné.

Algorithme 1 : Algorithme pour le calcul de (x,y) en fonction de t

Entrées : paramètre t , ordre de la courbe d ,
Sortie (x,y) image de t
1: Fonction $h(t, d)$
2: $(x,y)=(0.5,0.5)$;//initialisation de (x,y)
3: $tampon=t$;
4: $tableau$; // contient le nombre de chiffres en base 4 du paramètre en fonction de d
5: **pour tout** i allant de 0 à d faire
6: $q=partieEntiere(4*tampon)$;
7: $tampon=4*tampon-q$;
8: ajouter($q, tableau$) ;
9: **fin boucle**
10: **pour tout** i allant de 0 à $tailleTableau$ faire
11: **si** $elementTableau(i)=0$ **alors** $(x,y)=(y/2, x/2)$;// H_0
12: **si** $elementTableau(i)=1$ **alors** $(x,y)=(x/2, y/2+0.5)$;// H_1
13: **si** $elementTableau(i)=2$ **alors** $(x,y)=(x/2+0.5, y/2+0.5)$;// H_2
14: **si** $elementTableau(i)=3$ **alors** $(x,y)=(1.0-y/2, 0.5-x/2)$;// H_3
15: **fin boucle**

Algorithme 2 : Algorithme pour le calcul de t en fonction de (x,y)

Entrées : (x,y) , eps (precision)
Sortie : t paramètre de Hilbert
1: Fonction inverseH((x,y) , eps)
2: **si** $(eps>1)$ **alors** retourner 0 ;
3: **si** $(x<0.5)$ **alors**
4: **si** $(y<0.5)$ **alors** retourner $(0+inverseH(2*y, 2*x, 2*eps))/4$;
5: **sinon** retourner $(1+inverseH(2*x, 2*y-1, 2*eps))/4$;
6: **sinon si** $(y>=0.5)$ **alors** retourner $(2+inverseH(2*x-1, 2*y-1, 2*eps))/4$;

7: **sinon** retourner $(3+inverseH(1-2*y,2-2*x,2*eps))/4$;

4.3. Correspondance coordonnées géographiques et courbes de Hilbert

Pour appliquer SAX sur les coordonnées géographiques, on passe d'abord par une transformation de ces coordonnées bidimensionnelles en des coordonnées unidimensionnelles. Pour ce faire, on passe par trois étapes :

1. Les coordonnées initiales qui sont en degrés décimaux (système de coordonnées WGS-84) sont converties en coordonnées Lambert 93.
2. Les coordonnées Lambert 93 sont converties en coordonnées appartenant à $[0,1]^2$ dans un carré de Hilbert défini pour contenir la zone étudiée.
3. La fonction inverse de Hilbert (*Algorithme 2*) permettant d'obtenir à partir des coordonnées dans $[0,1]^2$ les paramètres de Hilbert t dans $[0,1]$ est appliquée.

Ainsi à la fin de la transformation, pour chaque paire de coordonnées (*latitude, longitude*) sur la trajectoire, on obtient un paramètre t .

La figure 3 illustre l'application des courbes de Hilbert sur une carte réelle. On veut reproduire une ligne entre Rennes et Nice en appliquant sur les coordonnées en degrés décimaux les différentes transformations (Lambert 93, coordonnées dans le carré $[0,1]^2$) ensuite on fait le chemin inverse (coordonnées dans $[0,1]^2$, Lambert 93, coordonnées en degrés décimaux). Les ordres 1, 5 et 10 sont représentés. Ces expériences montrent le rapport entre l'ordre de la courbe de Hilbert et la distorsion qu'elle introduit.

Sur la figure 3, on remarque qu'à l'ordre 1 on a juste 3 points (au départ 100 points sur la ligne d'origine) du fait que le carré est divisé en 4 sous-carrés et que chaque point va se trouver au centre du sous-carré dont il est le plus proche. Comme aucun point n'est proche du quatrième sous-carré, on a juste 3 points représentant les 3 sous-carrés. A l'ordre 5, on a plus de sous-carrés et donc plus de points plus proches de la courbe originelle, et à l'ordre 10 la ligne est reproduite car chaque point est très proche d'un ou de plusieurs centres de sous-carrés. Donc c'est un ordre suffisant et facile à calculer qui correspond à une définition de 1024×1024 .

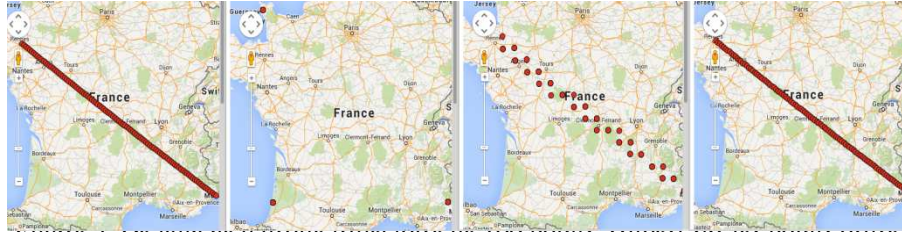


Figure 5. De gauche à droite trajectoire de 100 points, suivi sur les points ordre 1, ordre 5 et ordre 10

5. SAX et trajectoires d'objets mobiles

5.1. SAX et vitesses instantanées

L'application de SAX sur les vitesses d'objets mobiles permet de déterminer sur la trajectoire les sous-trajectoires où l'objet mobile est en arrêt (« stop »), ou en déplacement (« move ») lent (« slow »), rapide (« fast ») ou très rapide (« very slow »).

On définit $V = \{v_1, \dots, v_n\}$ une série où chaque v_i représente la vitesse instantanée au point p_i de coordonnées (lat_i, lon_i) . La série V va représenter les valeurs de la série temporelle S . Après application de la méthode PAA (Keogh et al., 2001) sur la série V , chaque valeur de la nouvelle trajectoire de longueur w appartient à un intervalle de valeurs représenté par un symbole SAX.

5.2. SAX et courbes de Hilbert

(Oates et al., 2013) utilisent les courbes de Hilbert pour faire le mapping entre les coordonnées géographiques bidimensionnelles et des paramètres de Hilbert (unidimensionnels). La trajectoire $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ est transformée en une nouvelle trajectoire $T_i = \{t_1, \dots, t_n\}$ où chaque t_i est le paramètre de Hilbert de (x_i, y_i) dans la trajectoire d'origine.

6. Expérimentations

6.1. Expérimentations avec les vitesses instantanées

Les expérimentations sont faites avec des données recueillies par des balises placées dans des remorques de camions pendant un mois. Ces trajectoires de camions ont un échantillonnage nominal de 300 secondes environ toutefois des intervalles de temps de plusieurs heures peuvent être observés entre deux prises consécutives.

Nous choisissons de ré-échantillonner les trajectoires en fonction de la période d'échantillonnage nominale.

6.1.1. Paramètres d'expériences

Le but est de faire ressortir grâce à SAX les positions qualifiées de point d'arrêt (exemple : parkings) et les positions qualifiées de point de déplacement (exemple : routes).

Pour ce faire, on a utilisé les paramètres suivants :

- Trajectoire sur 28 jours irrégulièrement échantillonnée (remorque de camion)
- Période de ré-échantillonnage : 300 secondes (5 minutes)
- Taille alphabet pour appliquer SAX : 11 $\{a,b,c,d,e,f,g,h,i,j,k\}$
- Nombre de fenêtres PAA : 806

6.1.2. Résultats

Après avoir appliqué SAX sur la trajectoire ré-échantillonnée, on obtient une chaîne de symboles où chaque symbole appartient à un intervalle de valeurs. Le tableau suivant montre chaque symbole avec son intervalle correspondant. Les valeurs négatives correspondant aux symboles a, b c et pour une partie de l'intervalle d viennent du fait que SAX calcule les intervalles en approchant la distribution réelle par une distribution normale. Celle-ci couvre donc des valeurs négatives qui ne correspondent à rien dans la réalité. Les symboles a, b et c ne seront donc jamais utilisés car la vitesse du camion est toujours supérieure ou égale à 0 et le symbole d ne le sera que pour des épisodes d'arrêt (« stop ») ou de progression très lente (vitesse ≤ 1.57 km/h).

Tableau1. Correspondance symbole SAX intervalle de valeurs

Symbole	Intervalles de valeurs (vitesses instantanées)
a	$]-\infty, -13.79[$
b	$[-13.79, -7.12[$
c	$[-7.12, -2.31[$
d	$[-2.31, 1.57[$
e	$[1.57, 5.30[$
f	$[5.30, 8.71[$
g	$[8.71, 12.44[$

h	[12.44,16.32[
i	[16.32,21.13[
j	[21.13,27.80[
k	[27.80,∞[

Pour affecter aux positions les symboles suivant les vitesses, dans un premier temps on prend toutes les positions comprises dans une fenêtre PAA, on calcule leur barycentre puis on affecte à chaque barycentre arrondi symbole de la fenêtre qui le contient. Dans un deuxième temps on regroupe chaque position avec ses différents symboles (le même symbole peut se répéter, et une même position peut être visitée à plusieurs vitesses différentes) et enfin on élimine les doublons.

Les barycentres des positions peuvent être visionnés sur une carte (figure 4). On fait correspondre à chaque couleur un symbole ou un ensemble de symboles : points blancs (d), points bleus (e), points jaunes (f,g), points verts (h,j) et points rouges (j,k)

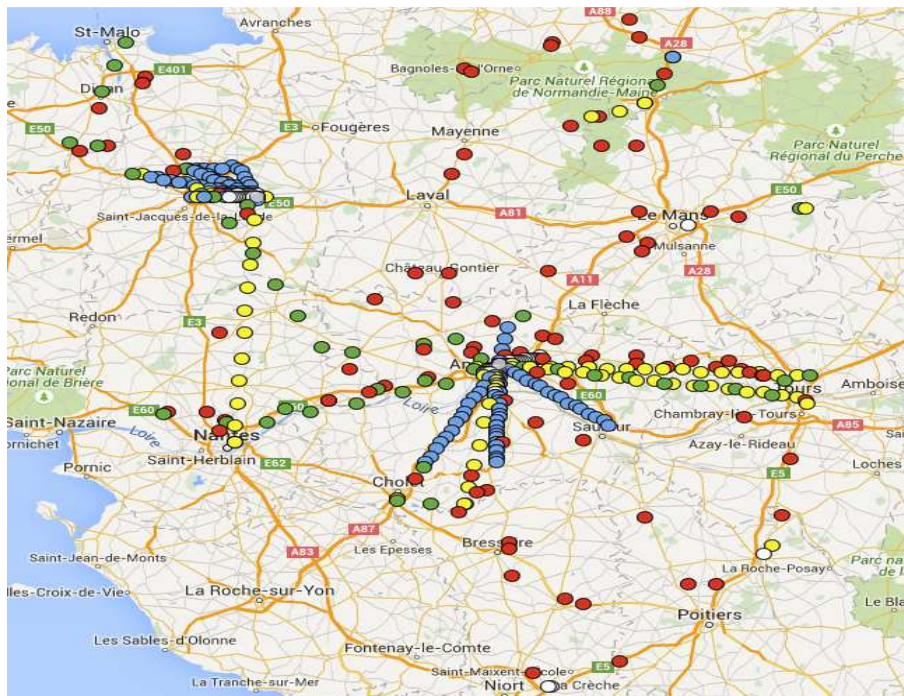


Figure 4. Visualisation de trajectoire sur une carte

Observation : On remarque quelques points blancs indiquant les endroits où l'objet mobile est en arrêt. Des lignes de points (bleus, jaunes, verts) peuvent être observées indiquant l'effet du ré-échantillonnage.

Conclusion : Le peu de nombre de points blancs observés est dû au fait que si le mobile est en arrêt c'est presque toujours la même position qui est détectée et en arrondissant on obtient donc une seule position. Les lignes droites bleues et jaunes correspondent aux positions fictives résultant de l'interpolation.

En créant un contexte pour l'analyse de concepts formels (Ganter et Wille, 1999) on obtient des règles d'association intéressantes comme $d \rightarrow e$ avec un support de 0.045 (support minimal 0.01) et une confiance de 0.27 (confiance minimale 0.20) qui montre que là où l'objet mobile est en arrêt, il est aussi en déplacement lent.

6.2 Expérimentations avec les positions

6.2.1. Paramètres d'expériences

Le but ici est de faire ressortir grâce à SAX, les zones qui sont géographiquement proches.

Pour ce faire, on a utilisé les paramètres suivants :

- Trajectoire sur 28 jours irrégulièrement échantillonnée (remorque de camion)
- Période de ré-échantillonnage : 300 secondes (5 minutes)
- Taille alphabet pour appliquer SAX : 8 $\{a,b,c,d,e,f,g,h\}$
- Nombre de fenêtres PAA : 806
- Carré choisi pour appliquer Hilbert : le « carré » englobant la France
- Ordre de la courbe de Hilbert 10

6.2.2. Résultats d'expériences

Pour pouvoir visualiser l'effet de l'application des courbes de Hilbert sur les trajectoires, on a affecté comme dans l'expérience précédente une couleur à chaque symbole SAX (figure 5) : points blancs (a), points bleus (b), points verts (c), points jaunes (d), points rouges (e), points marrons (f), points gris (g) et points noirs (h). Chaque symbole appartient à un intervalle de valeurs comprises dans $[0,1]$.

Observation : On observe que les zones les plus proches géographiquement ont une même couleur. L'effet du ré-échantillonnage se voit aussi par les lignes droites (rouge, marron).

Conclusion : Les sous-trajectoires récurrentes ou qui sont très proches vont avoir la même couleur. On entend par sous-trajectoires récurrentes, les parties de la

trajectoire entière où l'objet mobile passe plus d'une fois. Les couleurs sont regroupées car les zones proches géographiquement partagent la même couleur.

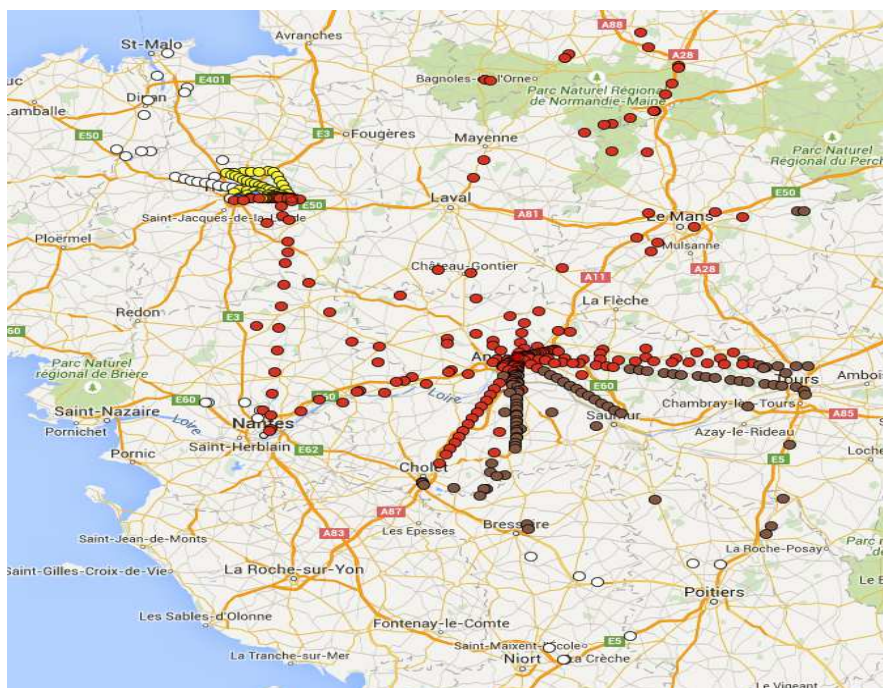


Figure 5. Visualisation d'une trajectoire après application des courbes Hilbert et SAX

7. Conclusion et perspectives

Dans cet article nous avons appliqué SAX sur des trajectoires d'objets mobiles afin de les représenter symboliquement. La représentation est faite d'abord sur des vitesses instantanées de l'objet mobile à chaque instant sur la trajectoire pour capturer des informations sur les positions où l'objet mobile est plutôt en déplacement ou en arrêt et étiqueter ses positions de route ou de parking. Ensuite la seconde application de SAX porte sur les positions transformées en des positions unidimensionnelles grâce aux courbes de Hilbert pour pouvoir détecter les sous-trajectoires récurrentes sur la trajectoire entière.

Dans ces travaux on a appliqué Hilbert sur une surface plane restreinte mais il est envisageable de l'appliquer sur une surface sphérique recouvrant par exemple toute la surface terrestre.

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

Nous envisageons d'extraire à partir de ces représentations, des motifs fréquents avec un algorithme se basant sur SEQUITUR (Nevill-Manning et Witten, 1997) consistant à la production de grammaires hors contexte à partir d'une chaîne de caractères (obtenue grâce à la représentation SAX dans notre cas). Chaque motif sera ici une règle grammaticale composée de symboles terminaux et non terminaux.

Bibliographie

- Bader M. (2013). *Space Filling Curves: An Introduction with Applications in Scientific Computing*, Springer-Verlag, Berlin.
- Bakalov P., Hadjieleftheriou M., Keogh E., Tsotras V. J. (2005). Efficient trajectory joins using symbolic representations. *In Proc. of the Int. Conf. on Mobile Data Management*, MDM'05, p. 86-93, 2005.
- Bakalov P., Hadjieleftheriou M., Tsotras V. J. (2005). Time relaxed spatiotemporal trajectory joins. *In Proc. of the Annual ACM Int. Workshop on Geographic Information Systems*, GIS'05, p. 182-191, 2005.
- Ganter B. et Wille R. (1999). *Formal Concept Analysis: Mathematical Foundations*, Springer-Verlag.
- Keogh J. E., Chakrabarti K., Pazzani M. J., Mehrotra S. (2001). Dimensionality Reduction for Fast Similarity Search in Large Time Series Database. *Knowl. Inf. Syst.*, vol. 3, n° 3, p. 263-286.
- Lashley M. (2006). *Kalman Filter Based Tracking Algorithms for Software GPS Receivers*. Master's Thesis, Auburn University, the Netherlands.
- Lin J., Keogh E., Lonardi S., Chiu B. (2003). A symbolic representation of time series, with implication of streaming algorithms. *In Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, DMKD '03, p. 2-11, New York, NY, USA, 2003.
- Meratnia N., R. A. de By, (2004). Spatiotemporal Compression Techniques for Moving Point Objects, *In EDBT '04: Proc. of the Int. Conf. on Extending Database Technology*, 2004.
- Nevill-Manning C. G., Witten I. H. (1997). *Identifying Hierarchical Structure in Sequences: A Linear-Time Algorithm*. *J. Artif. Intell. Res.*, vol. 7, p. 67-82
- Oates T., Boedihardjo A. P., Lin J., Chen C., Frankenstein S., Gandhi S., (2013). Motif Discovery in Spatial Trajectories using Grammar Inference. *In Proc. of the Int. Conf. on Information and Knowledge Management*, CIKM'13, p. 1465-1468, 2013. ACM
- Spaccapietra S., Parent C., Damiani M. L., Fernandes De Macêdo J. A., Porto F., Vangenot C. (2008). *A Conceptual View of Trajectories*. *Data Knowl. Eng.*, vol. 65, p. 126-146.
- Yan Z. (2011). *Computing and Understanding Mobility Data*. Phd's Thesis, Ecole Polytechnique Federale de Lausanne.