Calcul de similarité sémantique entre trajectoires

Moreau Clément [1], **Devogele Thomas** [1], **Etienne Laurent** [1]

Laboratoire d'Informatique Fondamentale et Appliquée de Tours
64, avenue Jean Portalis, 37200 Tours France
{clement.moreau,thomas.devogele,laurent.etienne}@univ-tours.fr

RÉSUMÉ. *La compréhension de la mobilité, qu'elle soit physique au sens spatial ou virtuelle au sens navigation web, soulève de nombreux enjeux en terme de surveillance des individus, d'aménagement du territoire ou de recommandation d'activité.*
Ayant accès aujourd'hui à de nombreuses ressources sur le caractère contextuelle de cette mobilité, une des préoccupations actuelles est de réussir à dégager des groupes d'individus similaires, relativement à leur mobilité. Pour se faire nous proposons dans cet article, un modèle de trajectoire sémantique enrichie par des ontologies au niveau des données contextuelles et permettant de calculer la similarité entre chaque épisode de mobilité de l'individu.
Par la suite, une distance d'édition est définie afin d'évaluer de manière fine et contextuelle la similarité entre deux trajectoires sémantiques.

ABSTRACT. *Understanding mobility, whether physical in the spatial sense or virtual in the web navigation sense, raises many challenges in terms of monitoring individuals, spatial planning or activity recommendations.*
Having access today to many resources on the contextual nature of this mobility, one of the current concerns is to succeed in identifying groups of similar individuals with regard to their mobility.
To do this, we propose a semantic trajectory model enriched by ontologies at the level of contextual data and allowing us to calculate the similarity between each episode of individual mobility. Subsequently, an editing distance is used to evaluate in a fine and contextual way the similarity between two sequences from a semantic point of view at the trajectory level.

MOTS-CLÉS : *Trajectoire sémantique, séquence sémantique, mesure de similarité sémantique, distance d'édition*

KEYWORDS: *Semantic Trajectory, Semantic Sequence, Semantic Similarity Metric, Edit Distance*

DOI:10.3166/RIG.00.1-24

1. Introduction

Comprendre la mobilité humaine, qu'elle soit physique au sens de la mobilité spatiale, ou bien virtuelle dans le cadre de la navigation entre pages Internet, est un enjeu majeur dans divers champs d'application tels que la recommandation d'activités, la surveillance d'individus ou encore l'optimisation de flux. Malgré son apparente complexité et sa diversité, González *et al*. (2008) et Song *et al*. (2010) montrent que la mobilité humaine possède un haut degré de régularité spatiale et temporelle ; ce qui indique que les déplacements individuels suivent des modèles reproductibles simples. Cependant, et même si ce résultat souligne le caractère prédictible des trajectoires humaines, certains auteurs comme Renso, Trasarti (2013); Yan (2009) défendent l'idée que l'axe spatio-temporel pur d'un déplacement, dépourvu de contextualisation, ne suffit pas à fournir une explication satisfaisante quant au sens et à la finalité du déplacement lui-même. En outre, les trajectoires nécessitent un enrichissement de connaissances contextuelles, c'est-à-dire d'informations externes, permettant d'interpréter le déplacement d'une façon haut niveau telles que les points d'intérêts, les événements en cours, la météo, le trafic routier, etc. Cette définition est empruntée à Parent *et al*. (2013) où une formalisation de ce concept de trajectoire sémantique est défini. Ainsi, toute information pouvant être récupérés depuis des médias sociaux (Twitter, Foursquare) [Hu, Ester (2013)], calendriers (i.e agendas) [Tu *et al*. (2017)] ou de bases de données géographiques (e.g OpenStreetMap) [Yan *et al*. (2011)] sont alors qualifiés de *données contextuelles*. Il est possible de considérer également comme donnée contextuelle l'activité effectuée par l'utilisateur lorsqu'elle est connue ou qu'elle peut être inférée à l'aide des données mobiles (accéléromètre, gyroscope, baromètre) [Shoaib *et al*. (2015)] pour l'activité physique. Beber *et al*. (2017) fournissent également une méthodologie complète sur l'inférence d'activité utilisateur au sein des trajectoires.

Ainsi, ces différentes sources de données potentielles doivent pouvoir venir *annoter* les données GPS initiales dans le but d'enrichir *sémantiquement* la trace spatio-temporelle brute (i.e raw data). Cependant, au vu des différents champs d'applications englobant la notion de mobilité et des possibilités d'ajout de données contextuelles en provenance de sources hétérogènes, il est primordial de concevoir un modèle conceptuel de la trajectoire, que l'on qualifiera alors de modèle de *trajectoire sémantique* générique et flexible. Ce modèle doit être flexible dans le sens où il doit être possible à l'utilisateur d'ajouter facilement toute sorte de données pouvant enrichir la trajectoire : des données issues d'une base contextuelle, comme précisé ci-dessus, ou par de simples annotations textuelles par chaîne de caractères ; il doit également être générique, c'est-à-dire facilement transposable à un type d'application différent. Comme mentionné précédemment, le concept de mouvement (i.e de trajectoire) peut s'appliquer aussi bien à une vision spatio-temporelle classique, qu'à celle de la navigation Internet sur différentes pages web par exemple. Dans ce cas, en dehors du cadre spatial, ces mouvements sont considérés comme des *séquences sémantiques ordonnées* selon une relation d'ordre de précédence.

Plus encore, ce type de structure séquentiel peut-être généralisé à tout type de données catégorielles ordonnées : évolution comportementale, séquence musicale via

une portée de notes, mutation d'une maladie ou plus généralement évolution des états d'une certaine entité considérée. Il faut préciser cependant que de tels états doivent pouvoir être comparables et comparés entre eux, nous revenons par la suite sur ce point mais pouvons d'ores et déjà proposer certains travaux sur des métriques de comparaison [Harispe (2014); Deza, Deza (2009); Cilibrasi, Vitanyi (2007)].

Au sein des systèmes d'information géographique (SIG), une application typique est la comparaison de trajectoires dans le but d'établir des groupes d'individus ayant des comportements similaires ; un processus de clustering peut alors être envisagé afin de déterminer ces classes d'individus semblables. Cependant, un tel procédé exige l'utilisation d'une métrique afin de pouvoir comparer les individus entre-eux. Il est alors nécessaire de disposer d'une mesure de similarité afin de pouvoir comparer les différentes instances de trajectoires sémantiques entre-elles. Plus encore, une telle métrique permettrait de s'approprier de nombreuses techniques issues de la fouille de données, notamment celles de l'apprentissage non supervisé.

Dans cet article, un modèle pour les trajectoires sémantiques appelé *APM* (Activity, POI and Move model) est présenté. Ce dernier se base sur la notion d'activité selon la vision de la Time Geography de Hägerstrand (1970). Un point important de ce modèle est la présence d'ontologie au niveau des données contextuelles/annotation.

La figure 1 schématise une trajectoire sémantique selon APM. Afin de faciliter la compréhension, les informations sémantiques sont représentées par des émoticônes. Une trajectoire sémantique est un ensemble ordonné d'épisodes qui peuvent être vus comme des activités mobiles ou statiques. Chaque épisode est décrit par un segment spatio-temporel capturé par GNSS (Global Navigation Satellite System) (e.g GPS) auxquel est ajouté des données contextuelles et annotations. Ces données contextuelles sont représentées selon des concepts issus de différentes ontologies. Ces ontologies servent ensuite à calculer la similarité entre la composante sémantique des épisodes. Par exemple, sur la figure 1, les épisodes de e_1 à e_4 traduisent le comportement suivant : (e_1) l'individu téléphone dans sa maison ; toujours en téléphonant, il quitte son domicile puis mène un déplacement, à pied (e_2) ; il raccroche son téléphone et continue son déplacement à pied (e_3) ; il arrive à un restaurant où il boit un café (e_4). Ainsi, si deux épisodes e_i et e'_i décrits dans le niveau sémantique par deux ensembles de concepts C_i et C'_i semblables, alors ces deux épisodes sont similaires d'un point de vue sémantique.

De plus, nous proposons une métrique afin d'évaluer la similarité de séquences sémantiques ordonnées, soit ici la composante sémantique d'APM. Cette métrique se base sur la distance d'édition développée, entre autres, dans Wagner, Fisher (1994). Cette dernière est enrichie de nouveaux opérateurs afin de saisir certaines spécificités comme le contexte ou la similarité sémantique des éléments de la trajectoire.

Ces propositions apportent quelques réponses aux problèmes complexes identifiés par Ferrero *et al*. (2016) qui sont : Comment représenter des informations contextuelles et hétérogènes dans des trajectoires sémantiques (*Représentation sous plu-*

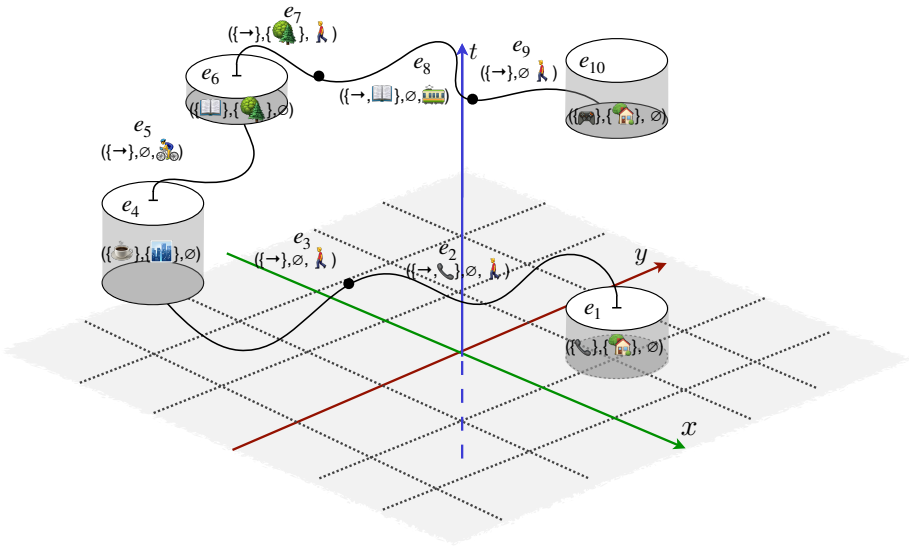


FIGURE *1. Un exemple de trajectoire sémantique d'APM*

sieurs aspects) ? Et, comment enrôler les dimensions spatiale, temporelle et sémantique dans une même métrique pour créer des clusters de trajectoires sémantiques (*Similarity Analysis and Data Mining*) ?

Notre contribution est développée de manière suivante dans la suite du papier :

1. Dans la première partie, un modèle générique de trajectoire sémantique (APM) axé sur la notion d'épisodes est proposé. Ce modèle hybride inclut des ontologies spécifiques pour évaluer de manière fine la proximité entre les données contextuelles venant annoter la trajectoire.

2. Dans la seconde partie, une métrique permettant la comparaison de séquences sémantiques est présentée. Cette mesure est basée sur la distance d'édition et l'utilisation d'une mesure de similarité entre les épisodes sémantiques d'APM afin d'établir une matrice des coûts. Cette matrice peut-être créée à l'aide d'ontologies (i.e. graphe de concepts) dont sont issues les données contextuelles. De plus, nous proposons l'enrichissement de certains opérateurs classiques d'édition et l'ajout de nouveaux opérateurs afin de quantifier plus précisément la proximité de deux séquences sémantiques.

2. Modélisation des trajectoires sémantiques

2.1. État de l'art

Plusieurs modèles de trajectoires sémantiques existent dans la littérature. À notre connaissance, Spaccapietra *et al*. (2008) furent les premiers à combiner les dimensions spatiale, temporelle et sémantique. La trajectoire sémantique est ainsi représentée par une séquence d'arrêts et de déplacements (Stops and moves).

Un modèle idéal de la trajectoire sémantique doit être suffisamment flexible pour que l'utilisateur puisse ajouter facilement des informations sémantiques. Dans ce sens, Yan *et al*. (2010) conceptualisent la trajectoire sémantique comme une séquence d'épisodes sémantiques. Ce modèle est concrétisé dans Yan *et al*. (2011) où un framework, nommé SeMiTri, est proposé. Une réalisation majeure sur la modélisation des trajectoires sémantiques est présentée dans Parent *et al*. (2013) où le concept de trajectoire sémantique est formalisé ; le modèle CONSTaNT introduit par Bogorny *et al*. (2014) vient parachever le travail théorique débuté par Parent *et al*..

Par sa nature qui s'inspire de la Time Geography (via la notion de but et de sous-trajectoire est assez proche de la notion d'épisode) et sa richesse d'expression sémantique, nous soutenons le modèle CONSTaNT. Cependant, la complexité du modèle de Bogorny et al. pose la difficulté de la comparaison des instances de CONSTaNT. La métrique MSM proposée dans Furtado *et al*. (2016) suggère certaines pistes pour mesurer la similarité de séquences multi-dimensionnelles comme les instances de CONSTaNT. Celle-ci se présente comme l'agrégation de métriques issues des dimensions spatiale, temporelle et sémantique. Cependant, les distances relatives aux dimensions considérées : Euclidienne (Spatiale), Jaccard modifié (Temps) et triviale (Sémantique) manquent de finesse, notamment sur le point sémantique.

La trajectoire sémantique est aussi souvent considérée dans le cadre d'une approche par fouille de motifs fréquents (Pattern mining). Dans l'article de Gianotti *et al*. (2007) par exemple, bien que les auteurs ne proposent pas explicitement de modèle de trajectoire sémantique, l'algorithme T-Pattern défini afin de déterminer les séquences de régions fréquemment visitées dans un ordre spécifié et avec des temps de transition similaires fournit un mode de représentation minimal de la dimension sémantique au sein d'une trajectoire. Les motifs fréquents sont alors représentés comme des séquences de POIs, semblables à des annotations/données contextuelles de premier niveau. Cette idée est reprise par Zhang *et al*. (2014) où des ensembles de POIs similaires son considérés ; les POI regroupés sont alors ceux partageant la même catégorie sémantique (contrainte sémantique) et spatialement proches (contrainte spatiale). De plus, les transitions entre deux groupes de POIs doivent être d'une durée inférieure à un seuil donné (contrainte temporelle). Ainsi, un motif à grain fin est une séquence de groupes de POIs qui correspond au moins à un nombre minimal de sous-séquences. De plus, il existe quelques modèles orientés pattern mining qui s'inspirent de la Time Geography dans leur conception de la trajectoire sémantique. Par exemple, Zheng *et al*. (2013) considèrent la trajectoire sémantique comme une trajectoire d'ac-

tivités, c'est-à-dire une séquence où des points spatiaux sont potentiellement associés à un ensemble d'activités. Dans l'article de Valdés, Güting (2018), un langage de requêtes adapté aux trajectoires sémantiques est présenté et implémenté dans un framework d'interrogation DBMS Secondo. Ce langage est basé sur une représentation de la trajectoire comme une séquence d'activités comme pour la Time Geography.

Enfin, il convient de noter que la trajectoire sémantique peut également être représentée par des ontologies. Par exemple, le système Athena développé par Baglioni *et al*. (2009) qui annote les trajectoires et les modèles avec des objectifs prédéfinis en utilisant la connaissance du domaine codée dans une ontologie. Des concepts similaires peuvent être trouvés dans [Noël *et al*. (2017)]. Cependant, ces modèles se concentrent sur une description ontologique de la trajectoire elle-même et non des instances qui la composent. Dans cette perspective, on trouve le framework Baquara[2] développé par Fileto *et al*. (2015) qui offre une possibilité pour l'enrichissement sémantique et l'analyse des trajectoires à l'aide du Linked Open Data. Par exemple, les données de médias sociaux géo-référencées, les données de mouvements peuvent être annotées avec des concepts et objets. De plus, un modèle ontologique définit la trajectoire afin de structurer et extraire des données de mouvement à plusieurs niveaux.

2.2. Modèle APM et notions de concept

Le modèle APM (Activity, Poi et Move) présenté ci-dessous est basé sur les points forts des modèles présentés précédemment. En outre, il reprend la richesse d'expression des modèles basés sur les activités/buts, la possibilité d'établir des modèles séquentiels à travers la notion de séquence d'épisodes et l'idée d'enrichissement sémantique par les ontologies.
Par exemple, la trajectoire sémantique de la figure 1 est représentée à l'aide d'APM comme une séquence $TS = \langle e_i | i \in [\![1, 10]\!] \rangle$ où $e_i = (e_i^{sem}, T_i)$ et e_i^{sem} est décrit plus amplement dans le paragraphe ci-après. T_i est la trajectoire spatio-temporelle (i.e Trace) de sorte que $T_i = \langle (p_1, t_1), ...(p_k, t_k) \rangle$ où t_i est une estampille temporelle et $\forall i \in [\![1, k-1]\!], 0 \leq t_i < t_{i+1}$ et p_i est un point de coordonnées $(x, y) \in \mathbb{R}^2$. Dans sa forme sémantique, cette trajectoire peut être considérée comme un *séquence sémantique* $\langle e_i^{sem} | i \in [\![1, 10]\!] \rangle$ où chaque $e_i^{sem} \in \Sigma$ représente un symbole sémantique d'un alphabet Σ. Ce symbole e^{sem} est analogue à la notion d'épisode dans sa dimension sémantique et correspond à un triplet de concepts où chaque concept est dérivé d'une ontologie.

Trois ontologies sont utilisées dans le modèle APM. Une ontologie d'*Activité* avec des concepts comme téléphoner, lire, manger, etc. représentant les activités pouvant être pratiquées par un utilisateur. Une ontologie des *POI* (Point Of Interest) décrit les différentes catégories de lieux. Les concepts comme maison, parc, magasin, etc. sont inclus dans l'ontologie. Enfin, pour les moyens de déplacement, une ontologie *Déplacement* est utilisée pour décrire les moyens de transport. Plusieurs concepts d'une

même ontologie peuvent décrire une activité. Par exemple, l'épisode e_2^{sem} de la figure 1 est décrit par les deux concepts de l'ontologie d'activité : "se déplacer" et "téléphoner". Si, le concept est inconnu i.e ne peut-être inféré, le symbole $\emptyset$ est employé.
Comme la fonction d'un lieu peut changer avec le temps, il est important de distinguer le concept d'activité du concept de lieu. Cependant, par expérience, il faut noter qu'il existe une forte corrélation entre le lieu et l'activité qui s'y déroule (par exemple, un "restaurant" est un lieu où la probabilité de l'activité "manger" est très élevée). Ces dernières propriétés peuvent être prises en compte si le concept d'activité est inconnu. D'après Aggarwal, Ryoo (2011); Beber *et al*. (2017) l'activité de l'utilisateur peut généralement être déduite par le concept de lieu.

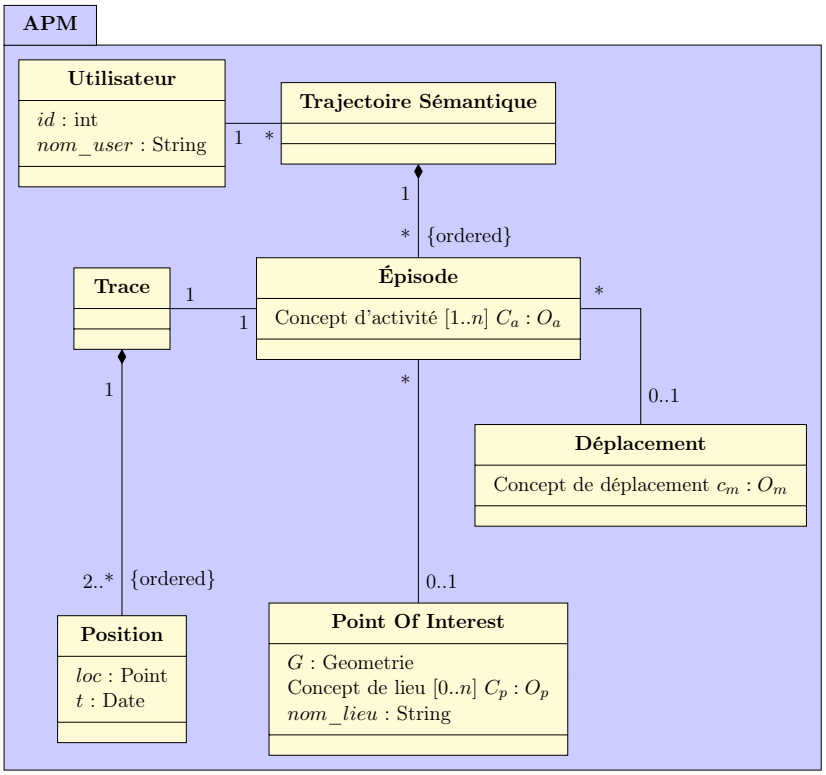


FIGURE 2. *Le modèle APM pour les trajectoires sémantiques*

2.3. Le modèle APM pour les trajectoires sémantiques

Comme précisé en introduction, le modèle APM s'inspire en partie du concept de la Time Geography d'Hägerstrand et de la notion d'activité. La figure 2 présente le modèle conceptuel d'APM. Celui-ci utilise un ensemble d'ontologies $\mathcal{O}$ tel que $\{O_a, O_p, O_d\} \subseteq \mathcal{O}$ où O_a est une ontologie des activités, O_p est une ontologie des lieux (i.e POI) et O_d est une ontologie des moyens de déplacement. Ces ontologies sont utilisées pour évaluer la composante sémantique de la trajectoire alors que la trace porte les dimensions spatiale et temporelle. Il existe de nombreuses ontologies centrées sur les activités quotidiennes comme Eurostat (2019) ou des hyper ontologies globales comme le Linked Open Data. Néanmoins ces ontologies, standardisées mais très génériques, demeurent généralement éloignées des contextes d'étude spécifiques requis.

À l'aide de ces notations, nous définissons certaines notions importantes à propos d'APM :

DÉFINITION 1. — (ALPHABET SÉMANTIQUE)
Un alphabet sémantique Σ est un ensemble de symboles sémantiques e^{sem} représentant les épisodes sémantiques. Ainsi, $\Sigma = (\mathcal{P}(O_a)\backslash\{\emptyset\}) \times \mathcal{P}(O_p) \times (O_m \cup \{\emptyset\})$, ou autrement dit $a^{sem} = (C_a, C_p, c_m) \in \Sigma$. Où $\mathcal{P}$ désigne l'ensemble des parties d'un ensemble, et C_a et C_p sont deux ensembles de concepts et c_m est un singleton.

DÉFINITION 2. — (SÉQUENCE SÉMANTIQUE)
Une séquence sémantique $S = \langle e_1^{sem}, e_2^{sem} ..., e_n^{sem} \rangle$ est une séquence ordonnée de symboles sémantiques telle que $S \in \Sigma^n$.

DÉFINITION 3. — (ÉPISODE SÉMANTIQUE)
Soit un symbole sémantique $e_i^{sem} \in \Sigma$ et une trace GPS $T_i = \langle (p_1, t_1), ... \ , (p_k, t_k) \rangle$ où $\forall i \in [\![1, k-1]\!], 0 \leq t_i < t_{i+1}$ avec $p_i \in \mathbb{R}^2$. Un épisode e_i est un couple de la forme $e_i = (e_i^{sem}, T_i)$.
Ainsi, la dimension temporelle, i.e l'intervalle de temps d_{e_i} de l'épisode e_i est portée par la trace T_i tel que $d_{e_i} = [t_1, t_k]$. On note $\Delta d_{e_i} = |t_k - t_1|$ la durée de l'épisode e_i. Enfin, e_i est potentiellement attaché à la géométrie G d'un POI tel que $T_i \subseteq G$.

À l'aide de ces définitions, nous pouvons définir la notion de trajectoire sémantique selon APM.

DÉFINITION 4. — (TRAJECTOIRE SÉMANTIQUE)
Une trajectoire sémantique $TS = \langle e_1, e_2 ..., e_n \rangle$ est une séquence d'épisodes. Il est possible de projeter TS selon différentes dimensions :

– Dimension sémantique, portée par la séquence sémantique S qui est la dérivée sémantique de TS que l'on notera $TS^{(S)} = S$.

– Dimension spatio-temporelle, portée par la trace $T = \langle T_1, ..., \ T_n \rangle$ qui est la dérivée spatio-temporelle de TS que l'on notera $TS^{(T)} = T$.

La proposition suivante détaille la méthode de segmentation de la trajectoire sémantique selon APM :

PROPOSITION 5. — *Soit une séquence sémantique S, alors* $\forall i \in [\![1, n-1]\!], e_i^{sem} \neq e_{i+1}^{sem}$. *En d'autres termes, pour tout symbole sémantique* $e_i^{sem} = \left(C_a^{(i)}, C_p^{(i)}, c_m^{(i)}\right)$ *dans S, il existe au moins un composant du triplet* e_{i+1}^{sem} *tel que ce même composant diffère dans* e_i^{sem}.

PROPOSITION 6. — *Soit un épisode* e_i *et un seuil* $\epsilon > 0$, *si* $\exists (x, y) \in \mathbb{R}^2, \forall p \in T_i, d((x, y), p) \leq \epsilon$, *alors* e_i *est un épisode statique où l'activité pratiquée est fixe (selon un seuil* ϵ*).*
Où d *est la distance Euclidienne ou d'Haversine.*

Les notions préliminaires exposées, nous pouvons nous pencher sur la figure 3 qui propose un exemple simple de deux trajectoires sémantiques selon APM.

EXEMPLE 7. — Dans la figure 3, la trajectoire sémantique rouge est nommée TS_1 et la bleue est appelée TS_2. Les épisodes e_i statiques sont représentés par un cylindre de hauteur Δd_{e_i} et de rayon ε.

La séquence sémantique S_1 de TS_1 est telle que :
$S_1 = \langle$({"jouer de la musique"}, {"maison"}, $\emptyset$),
({"manger"}, {"maison"}, $\emptyset$),
({"lire", "se déplacer"}, $\emptyset$, "tram"),
({"écouter de la musique", "se déplacer"}, {"parc"}, "à pied"),
({"écouter de la musique", "se déplacer"}, $\emptyset$, "à pied"),
({"danser"}, "scène"}, $\emptyset$),
({"se déplacer"}, $\emptyset$, "trotinnete"),
({"écouter de la musique"}, {"maison"}, $\emptyset$)$\rangle$
et S_2 telle que :
$S_2 = \langle$({"faire du théâtre"}, {"scène"}, $\emptyset$),
({"lire", "se déplacer"}, $\emptyset$, "tram"),
({"manger"}, {"restaurant"}, $\emptyset$),
({"prendre un café", "lire"}, {"restaurant"}, $\emptyset$),
({"se déplacer"}, $\emptyset$, "à pieds"),
({"nager"}, {"piscine"}, $\emptyset$),
({"se déplacer"}, $\emptyset$, "vélo"),
({"faire des achats"}, {"boulangerie"}, $\emptyset$),
({"se déplacer"}, $\emptyset$, "vélo"),
({"écouter de la musique", "peindre"}, {"maison"}, $\emptyset$),
({"écouter de la musique"}, {"maison"}, $\emptyset$)$\rangle$ □

Cet exemple montre la richesse sémantique du modèle APM.

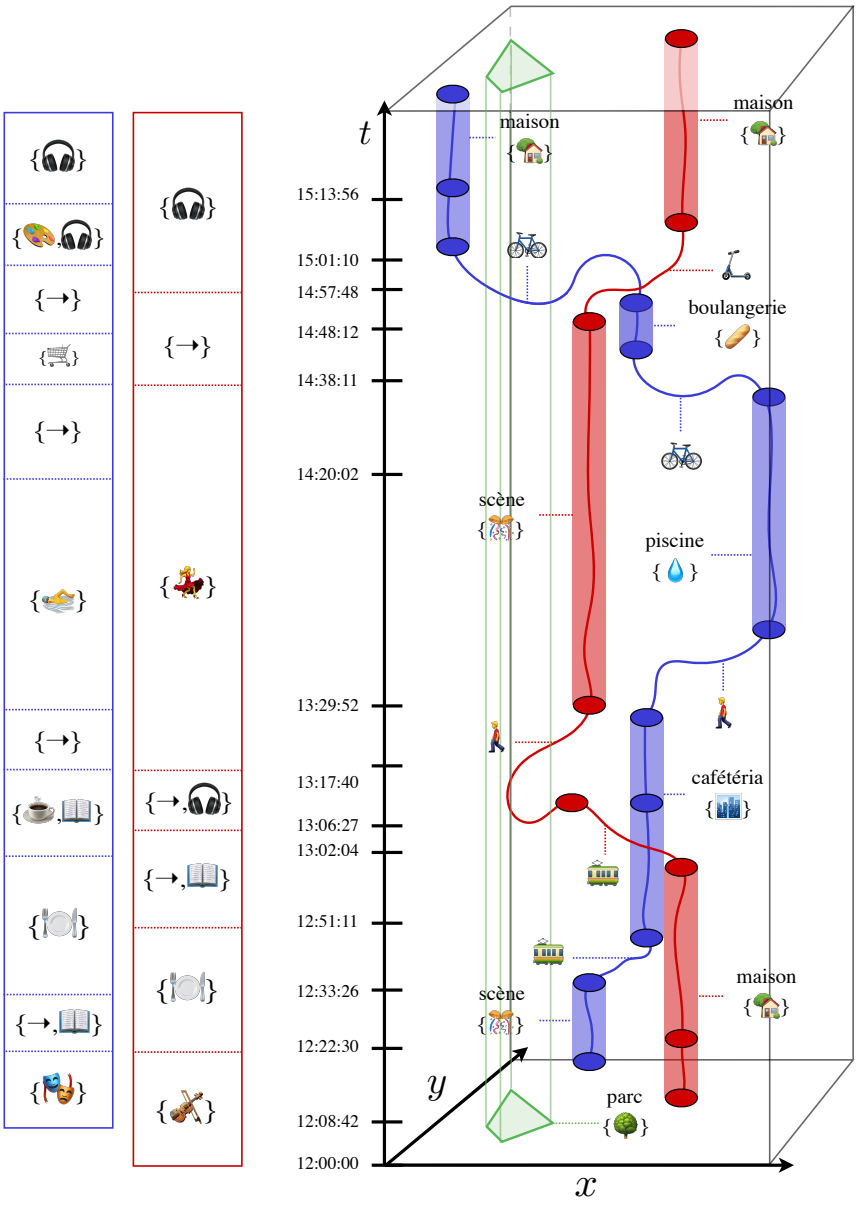


FIGURE 3. *Exemple de deux trajectoires sémantiques selon APM*

3. Distance entre trajectoires sémantiques

3.1. Similarité entre symboles sémantiques

Dans sa dimension sémantique, une trajectoire sémantique TS peut-être résumée par une séquence sémantique $S \in \Sigma^n$ représentant une suite d'annotations temporellement ordonnées comme vu dans l'exemple 7. Par conséquent, une similarité doit être créée entre les différents composants de S, par exemple via une matrice des coûts M_Σ. Cette tâche peut être réalisée à l'aide des ontologies comme le suggère Harispe (2014). Il existe de nombreuses méthodes afin d'évaluer la similarité de concepts au sein d'une ontologie. L'approche structurelle qui est envisagée ci-après considère que plus le chemin entre deux concepts est court dans le graphe conceptuel, et plus ces concepts partagent d'ancêtres en commun, plus ces concepts sont semblables. Il existe de nombreuses métriques afin de calculer la similarité entre concepts d'un point de vue structurel, plusieurs d'entre elles sont répertoriées et commentées par Deza, Deza (2009); Aime (2011); Harispe (2014). On note aussi l'existence d'un framework, Semantch, pour l'élaboration, l'évaluation, l'application et le calcul de scores de similarité sémantique entre concepts présentés dans Zhu, Iglesias (2017). Notons enfin qu'il existe des approches qui mesurent la similarité conceptuelle sans utiliser d'ontologie, c'est le cas par exemple de la Google similarity présentée dans Cilibrasi, Vitanyi (2007). Retenons ici que ce dont nous avons besoin est une fonction de similarité $sim : O^2 \rightarrow [0, 1]$ afin de pouvoir mesurer la ressemblance de deux concepts, non plus de façon binaire par une distance discrète, mais fine, en accord avec un l'intuition humaine.

Soit un couple $(e_1^{sem}, e_2^{sem}) \in \Sigma^2$. La difficulté d'établir la similarité entre ces symboles tient dans le fait que chaque élément de Σ est un triplet [d'ensembles] de concepts, avec possiblement le concept $\emptyset$ et où chaque composant (activité, POI, déplacement) peut avoir un intérêt différent pour l'utilisateur. Par exemple, il est possible qu'un utilisateur place beaucoup d'importance dans l'activité pratiquée alors que pour un autre, c'est le lieu qui doit être pris en compte de façon affirmée. Aussi, il est primordial que la mesure reflète avec le plus de force possible l'idée de similarité que porte l'utilisateur. Ainsi, l'approche privilégiée est d'utiliser des outils issus de la logique floue connue pour refléter l'intuition à travers certains agrégateurs et pouvant facilement prendre en compte les préférences utilisateurs quant à la manière de résumer un ensemble de données. Un panorama des différents opérateur d'agrégation est donné dans [Xu (2007); Xu, Da (2003); Dubois, Prade (1985)].

Soit une première équation générique de forme :

$$M_\Sigma(e_1^{sem}, e_2^{sem}) = \underset{k\in\{a,p,m\}}{Agg} \left(\underset{x\in C_k^{(1)}\,;\, y\in C_k^{(2)}}{Agg} sim(x, y) \right) \tag{1}$$

Où $Agg : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction d'aggrégation et $sim : O^2 \rightarrow [0, 1]$ une similarité sémantique.

Pour l'opérateur d'agrégation interne, il peut-être suggéré le choix d'une T-conorme pour ces bonnes propriétés, notamment le fait que 0 soit l'élément neutre (c'est-à-dire qu'il n'impacte pas la mesure).
Pour l'opérateur d'agrégation externe, des opérateurs de compromis comme une moyenne pondérée, min/max pondéré ou opérateurs OWA sont proposés par Yager, Kacprzyk (2012) afin d'exprimer les préférences sur chaque composant du modèle APM. Une explication détaillée des différents opérateurs d'agrégation suggérés précédemment est donnée dans [Grabicsh, Perny (2003)].
Enfin, concernant la similarité sémantique sim, sa nature dépend de l'ontologie. Comme abordé précédemment, et en l'absence de cardinalité des concepts (ce qui permettrait de pondérer la présence de certains concepts rares), l'approche structurelle est à privilégier. Au sein des graphes conceptuels, la mesure de Leacook, Chodorow (1998) normalisée $sim_{LC} : \mathcal{O}^2 \rightarrow [0, 1]$ se base sur la distance du plus court chemin entre deux concepts :

$$sim_{LC}(c_1, c_2) = -\log_2 \left(\frac{D + \text{len}(c_1, c_2)}{2 \times D} \right)$$

avec len : $O^2 \rightarrow \mathbb{N}$, la distance du plus court chemin entre deux concepts du graphe (par l'algorithme de Dijkstra par exemple) et $D = \max_{(c_1,c_2)\in\mathcal{O}^2} \text{len}(c_1, c_2)$ est la distance la plus grande pouvant être obtenue dans le graphe d'ontologie.
Un inconvénient de cette mesure est qu'elle n'évalue pas l'héritage des concepts, c'est-à-dire les liens de parenté potentiels. Une possibilité peut-être de s'inspirer de la mesure de Jaccard Jaccard (1901) de telle manière que :

$$sim_{Jac}(c_1, c_2) = \begin{cases} 0 & \text{si } |\Gamma(c_1) \cap \Gamma(c_2)| = 1 \\ \frac{|\Gamma(c_1)\cap\Gamma(c_2)|}{|\Gamma(c_1)\cup\Gamma(c_2)|} & \text{sinon} \end{cases}$$

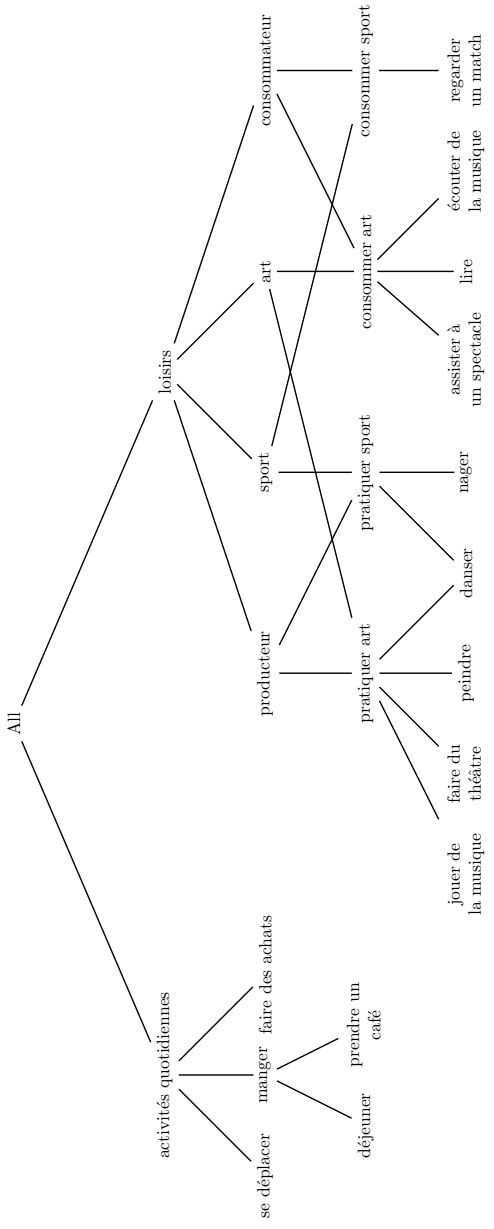
Où $\Gamma : O \rightarrow \mathcal{P}(O)$ est l'ensemble des ancêtres d'un noeud concept du graphe. Ainsi, nous proposons l'équation suivante afin de calculer la similarité de deux symboles sémantiques :

$$M_\Sigma(e_1^{sem}, e_2^{sem}) = \sum_{k\in\{a,p,m\}} \alpha_k^* \left(\max_{x\in C_k^{(1)}\,;\, y\in C_k^{(2)}} sim_\beta^*(x, y) \right) \tag{2}$$

Où $\alpha_a + \alpha_p + \alpha_m = 1$ sont les poids des coefficients respectivement pour les activités, les POI et les déplacements et α_i^* est le poids normalisé de α_i tel que :

$$\alpha_i^* = \frac{\alpha_i \times \left\lfloor sim\left(C_i^{(1)}, C_i^{(2)}\right)\right\rfloor}{\sum_{k\in\{a,p,m\}} \alpha_k \times \left\lfloor sim\left(C_k^{(1)}, C_k^{(2)}\right)\right\rfloor}$$

et $sim_\beta^* : O^2 \rightarrow [0, 1]$ est la mesure de similarité entre deux concepts dans une ontologie Cette mesure combine à la fois une approche par plus court chemin (sim_{LC}) et par héritage (sim_{Jac}). Cette mesure peut-être pondérée par un coefficient $\beta \in [0, 1]$:


All
activités quotidiennes
loisirs
se déplacer
manger
faire des achats
déjeuner
prendre un café
producteur
sport
art
consommateur
pratiquer art
pratiquer sport
consommer art
consommer sport
jouer de la musique
faire du théâtre
peindre
danser
nager
assister à un spectacle
lire
écouter de la musique
regarder un match

(*a*)

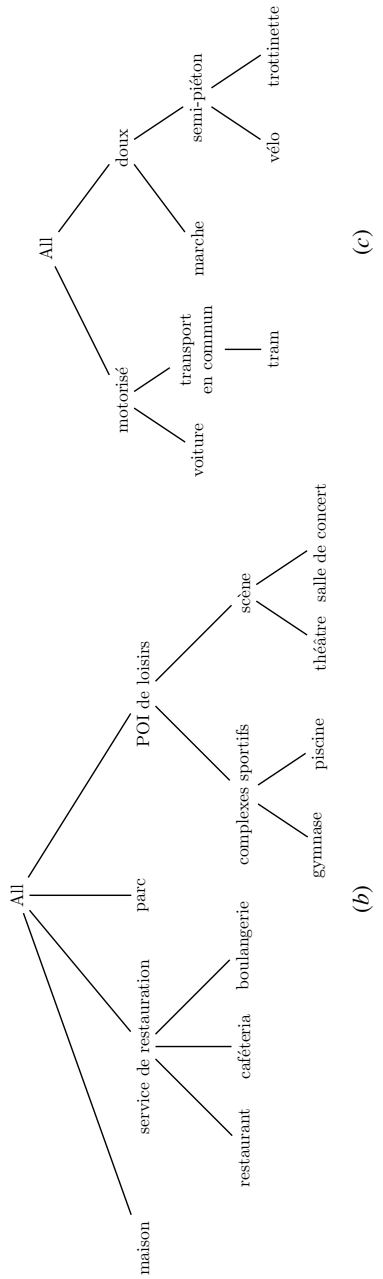


FIGURE *4. Exemple de trois graphes de concepts (a) Ontologie des activités O_a (b) Ontologie des POI O_p (c) Ontologie des déplacements O_d*

$$sim^*_\beta(c_1, c_2) = \beta \times sim_{LC}(c_1, c_2) + (1 - \beta) \times sim_{Jac}(c_1, c_2)$$

Enfin, et comme le concept $\emptyset$ est autorisé, il est nécessaire de déterminer une façon de comparer un concept $c \in O$ avec $\emptyset$. La manière la plus simple de procéder est de considérer le concept $\emptyset$ comme l'élément neutre de l'opérateur d'agrégation. Par exemple, pour l'équation (2), l'élément neutre de $\max$ est 0, dès lors : $\forall x \in C_k^{(i)}, sim(x, \emptyset) = 0$. De même, et comme 0 est également l'élément neutre de l'opérateur '$+$' utilisé pour la moyenne pondérée, il est possible de normaliser le poids des α_k si la similarité des composants est 0, dans ce cas, il est assumé que l'information ne doit pas être prise en compte.

EXEMPLE 8. — On considère les symboles sémantiques suivants :

- $e_1^{sem} = (\{\text{nager}\}, \{\text{cafétéria, piscine}\}, \emptyset)$,
- $e_2^{sem} = (\{\text{nager}\}, \{\text{piscine}\}, \emptyset)$,
- $e_3^{sem} = (\{\text{danser}\}, \{\text{gymnase}\}, \emptyset)$,
- $e_4^{sem} = (\{\text{assister à un spectacle}\}, \{\text{théâtre, restaurant}\}, \emptyset)$,

On considère également les graphes conceptuels suivants représentant respectivement les ontologies O_a, O_p et O_d.

On considère l'attribution de poids suivante :

- α_i : $\alpha_a = 0.5, \alpha_p = 0.25, \alpha_m = 0.25$,
- $\beta = 0.5$ Afin de donner autant d'importance à la structure du graphe qu'aux liens de parenté entre les concepts.

Comme c_m est toujours égal au concept $\emptyset$, les poids sont ré-évalués comme suit : $\alpha_a^* = \frac{0.5}{0.75} = \frac{2}{3}, \alpha_p^* = \frac{0.25}{0.75} = \frac{1}{3}, \alpha_m^* = 0$.
Par exemple, pour a_1^{sem}, on a :

- $M_\Sigma(e_1^{sem}, e_2^{sem}) = \frac{2}{3} \times \max(1) + \frac{1}{3} \times \max(0, 1) = 1$
- $M_\Sigma(e_1^{sem}, e_3^{sem}) = \frac{2}{3} \times \max(0.68) + \frac{1}{3} \times \max(0, 1) = 0.79$
- $M_\Sigma(e_1^{sem}, e_4^{sem}) = \frac{2}{3} \times \max(0.18) + \frac{1}{3} \times \max(0, 0.33, 0.76, 0) = 0.37$

Ainsi, M_Σ est telle que : $M_\Sigma = \begin{array}{c} \\ e_1^{sem} \\ e_2^{sem} \\ e_3^{sem} \\ e_4^{sem} \end{array} \begin{array}{c} \begin{array}{cccc} e_1^{sem} & e_2^{sem} & e_3^{sem} & e_4^{sem} \end{array} \\ \begin{pmatrix} 1 & 1 & 0.79 & 0.37 \\ 1 & 1 & 0.79 & 0.23 \\ 0.79 & 0.79 & 1 & 0.34 \\ 0.37 & 0.23 & 0.34 & 1 \end{pmatrix} \end{array}$

□

LEMME 9. — *Soit $|O_a| = N_a, |O_p| = N_p$ and $|O_d| = N_d$, alors, la matrice de coûts M_Σ entre tous les symboles sémantiques de Σ contient $(2^{N_a} - 1)^2 \times 2^{2N_p} \times (N_d + 1)^2$ éléments.*

PREUVE 10. —

La taille de l'alphabet Σ est telle que :
$|\Sigma| = |\mathcal{P}(O_a)\backslash\{\emptyset\}| \times |\mathcal{P}(O_p)| \times (|O_d| + 1)$
$= (2^{N_a} - 1) \times 2^{N_p} \times (N_d + 1)$
La matrice de coût M_Σ entre tous les symboles sémantiques Σ est de taille $|\Sigma| \times |\Sigma|$, alors elle contient : $((2^{N_a} - 1) \times 2^{N_p} \times (N_d + 1))^2 = (2^{N_a} - 1)^2 \times 2^{2N_p} \times (N_d + 1)^2$ éléments. ■

COROLLAIRE 11. — *La complexité d'un algorithme pour la construction de M_Σ est en $\Omega\left(N_d^2 2^{(N_a+N_p)}\right)$.*

Il convient toutefois de noter que cette complexité demeure théorique et qu'en pratique, il ne sera pas nécessaire de calculer tous les sous-ensembles des ontologies, par exemple, les activités comme le 'sommeil' et la 'natation' sont incompatibles ; ainsi, M_Σ est une matrice largement creuse.
Ainsi, le coût réel de calcul pour construire M_Σ est, considérant un ensemble de séquences sémantiques $\mathcal{S}$, le nombre K de symboles sémantiques différents au sein de $\mathcal{S}$, i.e. $K = \left| \bigcup_{S_i \in \mathcal{S}} \{e^{sem} | e^{sem} \in S_i\} \right| \leq \sum_{s_i \in \mathcal{S}} |S_i|$.

3.2. Une métrique pour comparer les séquences sémantiques

Puisqu'une séquence sémantique est semblable à la structure de chaîne de caractères dans le sens où il s'agit d'une séquence de symboles issus d'un alphabet, il est possible d'appliquer une distance d'édition afin évaluer la proximité de deux séquences sémantiques. Dans sa version classique, la distance d'édition consiste à transformer une séquence S_1 en une séquence S_2 grâce à des opérations d'édition. Cependant, les opérations d'édition utilisées (en général addition, suppression, modification) ne tiennent pas compte du contexte, c'est-à-dire des autres symboles présents dans la séquence ainsi que de leur ordre, ni de la similitude entre les symboles qui est évaluée bien souvent de manière binaire à l'aide de la distance discrète.

Ainsi, nous proposons une distance d'édition modifiée, appelée Distance d'édition sémantique. Certains opérateurs classiques de la distance d'édition sont modifiés pour tenir compte du contexte de la séquence, entre autres, les opérations d'ajout et de suppression se font de manière contextuelle tout au long de la séquence. Par exemple, l'insertion d'un symbole x, très différent de tous les symboles de la séquence sémantique, sera très coûteuse ; réciproquement, le coût d'insertion d'un symbole y similaire aux symboles de la séquence sémantique sera faible. De même, seul la notion de précédence est considérée ici, ainsi un symbole x peut être dupliqué dans la séquence $\langle x, x \rangle$ sans coût. Un nouveau symbole peut ensuite être ajouté entre ces deux symboles x avec le même coût qu'une opération `add`. Nous nommons `split` cette nouvelle opération. Cette dernière est très utile dans un contexte de déplacement Stop-and-Move tel que décrit par Spaccapietra *et al.* (2008) car il permet de ne prendre en compte que l'ajout du concept y. Cette mesure est générique et adaptable à tout type de séquence de symboles pour laquelle une matrice des coûts M_Σ entre les symboles est définie

comme celle définie dans la section 3.1.

DÉFINITION 12. — (OPÉRATEURS D'ÉDITION)
Soient deux séquences sémantiques $S \in \Sigma^n$ et $S' \in \Sigma^p$. On veut transformer S en S'. Soient $x, y \in \Sigma$ tels que $x \neq y$ et $(x, y) \neq (\varepsilon, \varepsilon)$ où ε est le symbole vide.
On considère l'ensemble d'opérateurs d'édition suivant : $E = \{\texttt{add}, \texttt{del}, \texttt{mod}, \texttt{trans}, \texttt{split}, \texttt{unsplit}\}$. Pour tout les opérateurs d'édition $e \in E$, on considère les signatures suivantes :

– $\texttt{add} : \varepsilon_i \rightarrow x$
Ajout de x dans S à la position i.

– $\texttt{del} : x_i \rightarrow \varepsilon$
Suppression de x dans S à la position i.

– $\texttt{modif} : x_i \rightarrow y$
Modification de x_i en y dans S à la position i.

– $\texttt{trans} : (x_i, x_{i+1}) \rightarrow (x_{i+1}, x_i)$
Transposition de deux symboles dans S.

– $\texttt{split} : x_i \rightarrow (x_i, y_{i+1}, x_{i+2})$
Ajout de deux symboles y_{i+1} et x_{i+2} dans S tel que $x_i = x_{i+2}$ et $y_{i+1} \neq x_i$.

– $\texttt{unsplit} : (x_i, y_{i+1}, x_{i+2}) \rightarrow x_i$
Fonction inverse de split.

Ainsi, une séquence d'édition $(e_1, e_2, ..., e_N)$ transformant la séquence S en S' est appelée un *chemin d'édition* $c(S, S')$ et correspond à la composition successive des opérateurs e_1 à e_N. On note $\mathcal{C}(S, S')$ l'ensemble de tous les chemins d'édition pour éditer S en S'. Le coût d'un chemin d'édition est évalué par la somme de chaque coût individuel d'édition.
La distance d'édition sémantique $d_{\mathcal{S}} : \Sigma^n \times \Sigma^p \rightarrow \mathbb{R}^+$ est alors telle que :

$$d_{\mathcal{S}}(S, S') = \min_{(e_1,...,e_N) \in \mathcal{C}(S,S')} \sum_{i=1}^{N} \gamma(e_i) \tag{3}$$

Où $\gamma : E \rightarrow \mathbb{R}^+$ est la fonction de coût d'utilisation de chaque opérateur d'édition.

PROPOSITION 13. — *$d_{\mathcal{S}}$ est symétrique (i.e $d_{\mathcal{S}}(S, S') = d_{\mathcal{S}}(S', S)$) si et seulement si $\forall e \in E, \exists e^{-1} \in E, \gamma(e) = \gamma(e^{-1})$.*

À propos de la complexité de calcul de la distance d'édition, Wagner, Fisher (1994) proposent une approche par programmation dynamique avec une complexité en $O(n \times p)$, où n et p sont les tailles respectives des trajectoires. Cependant, les opérateurs split et unsplit ne sont pas considérés. Cet algorithme offre cependant la possibilité de modifier facilement la fonction de coût γ. Dans le cadre d'une démarche d'édition contextuelle et sémantique des séquences, nous proposons la définition de γ suivante :

DÉFINITION 14. — (FONCTION DE COÛT γ)

En vertu de la Proposition13, on associe le même coût pour l'opérateur e et son inverse e^{-1}.

1. Si $e \in \{\texttt{add}, \texttt{del}\}$
On considère une fenêtre $k \in [\![1, |S| - 1]\!]$ *et un coût constant* $\lambda \in [0, 1]$*, alors :*

$$\gamma(e) = \lambda + (1 - \lambda) \times \left(1 - \max_{\forall j \in [\![i-k, i+k]\!]} M_\Sigma(e_j^{sem}, x)\right) \tag{4}$$

Ainsi, si un symbole sémantique similaire au symbole x à ajouter est observé en S et qu'il s'est produit à moins de k symboles avant ou après la position d'addition ; alors l'ajout de x est considéré comme faible.

2. Si $e = \texttt{modif}$
Comme M_Σ est symétrique, alors:

$$\gamma(e) = 1 - M_\Sigma(x_i, y) \tag{5}$$

3. Si $e = \texttt{trans}$
Soit un coût $\lambda \in [0, 1]$*, alors :*

$$\gamma(e) = 2\lambda \times (1 - M_\Sigma(x_i, x_{i+1})) \tag{6}$$

De ce fait, si $\lambda = 1$, alors, l'opérateur de transposition revient à effectuer deux modifications.

4. Si $e \in \{\texttt{split}, \texttt{unsplit}\}$
Soit une séquence sémantique S telle que $S = \langle x_1, ...x_i, x_{i+1}, ..., x_n\rangle$, alors $S' = \langle x_1..., x_i, x_i, x_{i+1}, ..., x_n\rangle$ est équivalente à S. Dès lors, il est possible d'ajouter y entre les deux symboles x_i. Ainsi pour les opérateurs `split` *et son inverse* `unsplit`*, γ est du même coût que pour les opérateurs* `add` *et* `del` *décrits équation 4.*

EXEMPLE 15. — On considère les deux trajectoires sémantiques décrites figure 3 et dont la séquence sémantique respective à chacune d'elles est présentée à l'exemple 7. Le chemin d'édition $c(S_2, S_1)$ afin de transformer S_2 en S_1 et qui minimise la somme total de fonction de coût γ appliquée aux opérateurs de $c(S_2, S_1)$ est présenté figure 5. Les coût γ qui figurent sont calculés en accord avec les définitions des opérateurs donnés précédemment ; on calcule la similarité entre chaque épisode sémantique e^{sem} qui compose les deux séquences selon l'équation 2 et comme présenté dans le précédent exemple 8. On note que les paramètres pour cette évaluation sont fixés tels que : $\alpha_a = 0.5, \alpha_p = 0.25, \alpha_m = 0.25$. On place ici l'importance sur l'activité effectuée et on met au même niveau lieu et moyen de déplacement.
On pose également $\beta = 0.5$ pour exprimer la même pondération entre les similarités sim_{LC} et sim_{Jac}. Enfin k est fixé à 2 afin de considérer un contexte à deux symboles près lors des opérations d'insertion/suppression ; un coût constant $\lambda = 0.5$ est également pris en compte.
Ainsi, $d_{\mathcal{S}}(S_2, S_1) = 0.4 + 0.2 + 0.58 + 0 + 0.38 + 0.07 + 0.42 + 0.59 + 0.5$
$= 3.14$ □

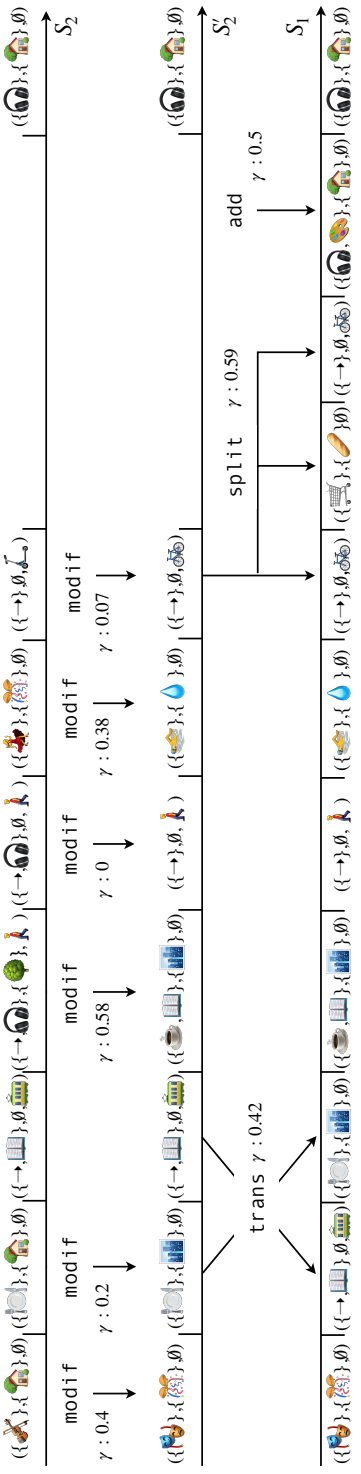


FIGURE 5. *Exemple de chemin d'édition* $c(S_2, S_1)$

3.3. À propos des autres dimensions

La métrique proposée $d_{\mathcal{S}}$ fonctionne sur la dimension sémantique de la trajectoire ; cependant, la trajectoire sémantique est la symbiose des trois dimensions : spatiale, temporelle et sémantique.
Pourtant, ces dimensions sont étroitement liées. L'espace évolue avec le temps et l'aspect sémantique peut être corrélé selon l'axe spatial (par exemple, une zone commerciale, une zone de bureaux, etc. sont des enclaves spatiales pour la sémantique). La sémantique et le temps sont également liés, certaines activités sont typiquement diurnes, par exemple le shopping, et d'autres nocturnes (par exemple le sommeil). Ainsi, et sans une information plus détaillée sur les corrélations entre ces différents axes, il est difficile d'établir une métrique qui puisse relier ces trois dimensions tout en respectant les liens fondamentaux qu'elles peuvent entretenir. Furtado *et al*. (2016) proposent la Mesure de Similarité Multidimensionnelle (MSM) qui pondère la similarité dans toutes les dimensions et propose de nombreuses clarifications sur la méthodologie de prise en compte des différentes dimensions ; en particulier le temps et l'espace. La principale objection pouvant être faite à MSM est que les mesures dimensionnelles proposées sont souvent triviales et que l'évaluation ainsi effectuée manque de finesse. Il est cependant possible de généraliser MSM à d'autres distances et remplacer la pondération des différentes dimensions par un opérateur d'agrégation Agg plus générique.
Enfin, en considérant une distance géométrique $d_{\mathcal{T}}$ (DTW [Berndt, Clifford (1994)], Fréchet [Devogele *et al*. (2017)], EDR [Chen *et al*. (2005)].) pour les trajectoires spatiales, la distance $d_{\mathcal{S}}$ de l'équation 3, et un opérateur d'agrégation Agg, il est possible de définir une métrique $d_{\mathcal{TS}} : \mathcal{TS} \times \mathcal{TS} \rightarrow \mathbb{R}^+$ pour des trajectoires sémantiques telle que :

$$d_{\mathcal{TS}}(TS_1, TS_2) = Agg(d_{\mathcal{T}}(TS_1^{(T)}, TS_2^{(T)}), d_{\mathcal{S}}(TS_1^{(S)}, TS_2^{(S)})) \qquad (7)$$

Toutefois, il convient de noter que la dimension temporelle demeure sans importance dans cette équation ; $d_{\mathcal{S}}$ ne tient compte que de la notion de précédence. Dans la littérature, et en particulier dans la branche de la fouille de trajectoires selon l'approche de pattern-mining, la dimension temporelle et la dimensions spatiale est considérée comme des contraintes [Gianotti *et al*. (2007), Zhang *et al*. (2014)] et gérées à l'aide de seuil; un événement doit se produire dans un intervalle de temps donné et un périmètre spatial défini pour être pris en compte. Ainsi, l'agrégation des métriques issues des différentes dimensions n'est pas l'unique solution. Le gel des différentes dimensions ou l'analyse de chacune des dimensions de manière séparée est également très pertinente. À propos de la dimension temporelle, on notera que Siabato *et al*. (2015) proposent une bibliographie complète sur la notion de temps dans les SIG.

4. Conclusion

Dans cet article, un nouveau modèle générique de trajectoires sémantiques, APM, a été présenté. Ce modèle s'inspire du concept de la Time Geopgraphy et utilise

des ontologies pour construire une mesure de similarité sémantique entre les symboles/épisodes sémantiques, composants de APM.
La mesure de similarité présentée pour la dimension sémantique des instances APM est basée sur des métriques de graphes conceptuels selon une approche structurelle et utilise également des opérateurs d'agrégation issus de la logique floue. Enfin, il est alors possible de pratiquer une distance d'édition entre séquences sémantiques issues de trajectoires sémantiques. Afin de respecter le caractère contextuel de la trajectoire, de nouveaux opérateurs basés sur une matrice des coûts M_Σ entre les symboles sémantiques et impliquant une contextualisation d'ajout/suppression de la séquence ont été proposés.
Même si la complexité théorique de la construction de M_Σ est élevée, l'hypothèse que M_Σ est une matrice creuse est très raisonnable. En effet seuls les sous-ensembles e^{sem} rencontrés dans l'ensemble des séquences sémantiques $\mathcal{S}$ sont à considérés ce qui limite considérablement la taille de M_Σ et donc le coût computationnel. De plus, la taille des séquences sémantiques reste relativement faible, ce qui permet de calculer des trajectoires sémantiques dans un temps CPU adapté au processus de data mining. Cette métrique est générique et peut être réutilisée pour tout type de séquences conceptuelles multidimensionnelles.
Dans les travaux futurs, la dimension temporelle sera explorée plus en détails pour compléter ce travail. Plus généralement, l'exploration de la relation entre les dimensions spatiale, temporelle et sémantique doit être poursuivie. L'agrégation et la prise en compte des différentes dimensions d'un point de vue utilisateur peuvent être envisagées selon la logique floue, proche du raisonnement humain et flexible, ce qui permet de conserver une approche intuitive pour un humain quant à la paramétrisation du modèle. Enfin, ces travaux seront testés sur différents ensembles de données provenant de divers contextes (mobilité touristique et étude sociologique sur la mobilité des enfants) afin de tester leur robustesse.

Bibliographie

Aggarwal J., Ryoo M. (2011). Human activity analysis: A review. *ACM Computing Surveys*, vol. 16, p. 1–43.

Aime X. (2011). *Gradients de prototypicalité, mesures de similarité et de proximité sémantique : une contribution à l'ingénierie des ontologies*. Thèse de doctorat non publiée, Université de Nantes.

Baglioni M. ., Macêdo J., , Renso C., Trasarti R., Wachowicz M. (2009). Towards semantic interpretation of movement behavior. In *Advances in giscience*, p. 271–288. Springer.

Beber M., Ferrero C., Fileto R., Bogorny V. (2017). Individual and group activity recognition in moving object trajectories. *Journal of Information and Data Management*, vol. 8, n° 1, p. 50–66.

Berndt D., Clifford J. (1994). Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd international conference on knowledge discovery and data mining*, p. 359–370.

Bogorny V., Renso C., Aquino A. R. de, Lucca Siqueira F. de, Alvares L. (2014). Constant - a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, vol. 18, p. 66–88.

Chen L., Özsu M. T., Oria V. (2005). Robust and fast similarity search for moving object trajectories. *Proc. of the 2005 ACM SIGMOD*, p. 491–502.

Cilibrasi R., Vitanyi P. (2007). The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, n° 3, p. 370–383.

Devogele T., Etienne L., Esnault M., Lardy F. (2017). Optimized discrete fréchet distance between trajectories. *Proceedings of the 6th ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data*, p. 11-19.

Deza M., Deza E. (2009). Encyclopeadia od distances. In, p. 371–382. Springer.

Dubois D., Prade H. (1985). A review of fuzzy set aggregation connectives. *Information Systems*, vol. 36, p. 85–121.

Eurostat. (2019). *Harmonised european time use surveys (hetus) 2018 guidelines*. Rapport technique.

Ferrero C., Alvares L., Bogorny V. (2016). Multiple aspect trajectory data analysis: Research challenges and opportunities. *GeoInfo*, p. 56–67.

Fileto R., May C., Renso C., Pelekis N., Klein D., Theodoridis Y. (2015). The baquara^2 knowledge-based framework for semantic enrichment and analysis of movement data. *Data & Knowledge Engineering*, vol. 98, p. 104–122.

Furtado A., Kopanaki D., Alvares L., Bogorny V. (2016). Multidimensional similarity measuring for semantic trajectories. *Transactions in GIS*, vol. 20, p. 280–298.

Gianotti F., Nanni M., Pedreschi D., Pinelli F. (2007). Trajectory pattern mining. *ACM SIGKDD*, p. 330–339.

González M., CA.Hidalgo, Barabási A.-L. (2008). Understanding individual human mobility patterns. *Nature*, vol. 453, p. 779–782.

Grabicsh M., Perny P. (2003). Logique floue, principes, aide à la décision,. In, p. 81–120. Hermès-Lavoisier.

Hägerstrand T. (1970). What about people in regional science ? *Papers in Regional science*, vol. 24, p. 6–21.

Harispe S. (2014). *Knowledge-based semantic measures: From theory to applications*. Thèse de doctorat non publiée, Université de Montpellier.

Hu B., Ester M. (2013). Spatial topic modeling in online social media for location recommendation. *Proc. of the 7th ACM conference on Recommender systems*.

Jaccard P. (1901). étude comparative de la distribution florale dans une portion des alpes du jura. *Bulletin de la Société Vaudoise*, vol. 31, p. 547–579.

Leacook C., Chodorow M. (1998). Wordnet: An electronic lexical database. In, p. 265–283. Cambrige MA.

Noël D., Villanova-Oliver M., Gensel J., Quéau P. L. (2017). Design patterns for modelling life trajectories in the semantic web. In *Web and wireless geographical information systems*, p. 51–65. Springer.

Parent C., Spaccapietra S., Renso C., Andrienko G., Bogorny V., Damiani M. *et al*. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys*, vol. 45, p. 1–32.

Renso C., Trasarti R. (2013). Mobility data : Modeling, management and understanding. In, p. 129–151. Cambridge University Press.

Shoaib M., Bosh S., Incel O., Scholten H., Havinga P. (2015). A survey of online activity recognition using mobile phones. *Sensors*, vol. 15, p. 2059–2085.

Siabato W., Claramunt C., Manso-Callejo M., Bernabé-Poveda M. (2015). Timebliography: A dynamic and online bibliography on temporal gis. *Transactions in GIS*, vol. 18, p. 799–816.

Song C., Qu Z., Blumm N., Barabási A.-L. (2010). Limits of predictability in human mobility. *Science*, vol. 327, p. 1018–1021.

Spaccapietra S., Parent C., Damiani M., Macedo J. de, Porto F., Vangenot C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, vol. 65, p. 126–146.

Tu W., Cao J., Yue Y., Shaw S.-L., Zhou M., Wang Z. *et al*. (2017). Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of GIS*.

Valdés F., Güting R. (2018). A framework for efficient multi-attribute movement data analysis. *The VLDB Journal*, p. 1–23.

Wagner R., Fisher M. (1994). The string-to-string correction problem. *Journal of the ACM*, vol. 21, p. 168–173.

Xu Z. (2007). Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems*.

Xu Z., Da Q. (2003). An overview of operators for aggregating information. *International Journal of intelligent systems*, vol. 18, p. 953–969.

Yager R., Kacprzyk J. (2012). *The ordered weighted averaging operators: Theory and applications*. Springer-Science.

Yan Z. (2009). Towards semantic trajectory data analysis: A conceptual and computational approach. *VLDB Endowment*.

Yan Z., Chakraborty D., Parent C., Spaccapietra S., Aberer K. (2011). Semitri: A framework for semantic annotation of heterogeneous trajectories. *Proc. of the 14th International Conference on Extending Database Technology*, p. 259–270.

Yan Z., Parent C., Spaccapietra S., Chakraborty D. (2010). A hybrid model and computing platform for spatio-semantic trajectories. *Proc. of the 7th international conference on The Semantic Web*, p. 60-75.

Zhang C., Han J., Shou L., Lu J., Porta T. L. (2014). Splitter: Mining fine-grained sequential patterns in semantic trajectories. *VLDB Endowment*, vol. 7, p. 769-780.

Zheng K., Shang S., Yuan N., Yang Y. (2013). Towards efficient search for activity trajectories. *International Conference on Data Engineering (ICDE)*, p. 230-241.

Zhu G., Iglesias C. (2017). Sematch: Semantic similarity framework for knowledge graphs. *Knowledge-Based Systems*, vol. 130, p. 30–32.