

Reconnaissance de modes de transport avec capteurs embarqués

Felipe MASCULO^{1,2}, Andréa VASSILEV¹, Michel DESVIGNES², Olivier MICHEL²

¹Univ. Grenoble Alpes, F-38000 Grenoble France
CEA, LETI, MINATEC Campus, F-38054 Grenoble, France

²Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble France
Felipe.Masculo@cea.fr, Andrea.Vassilev@cea.fr
Michel.Desvignes@gipsa-lab.grenoble-inp.fr
Olivier.Michel@gipsa-lab.grenoble-inp.fr

Résumé – Grâce à l’augmentation de la capacité de calcul et du nombre de capteurs des dispositifs portables, ainsi qu’à leur pénétration croissante du marché, les applications de capture de contexte ont gagné un nouvel intérêt. Cette étude s’intéresse à un type particulier de contexte, le mode de transport utilisé par une personne, et synthétise une méthode capable de classifier automatiquement différents modes de transport avec un smartphone. Elle prend en compte les contraintes liées à l’autonomie de la batterie par le choix de capteurs à faible consommation et par la mise en œuvre d’un algorithme à faible complexité. La méthode a été évaluée sur des données réelles et présente des résultats prometteurs : une précision d’environ 95% pour la classification de 8 classes différentes a été obtenue avec une forêt aléatoire.

Abstract – Thanks to the increase in processing power and in the number of sensors present in today’s mobile devices as well as their growing market penetration, context-aware applications have gained a renewed interest. This study focuses on a particular type of context, the transportation mode used by a person, and it summarizes a method for automatically classifying different transportation modes with a smartphone. It takes into account battery life constraints by using low-power sensors and by the implementation of a low-complexity algorithm. The method was evaluated with real data presenting promising results: an accuracy of around 95% was obtained when classifying 8 different classes with a random forest.

1 Introduction

L’intégration permanente de nouveaux capteurs dans les smartphones et récemment dans les objets connectés, tels que les montres, ainsi que l’omniprésence de ces objets les ont transformé en instruments idéaux pour la capture de contexte. Les intérêts de la capture de contexte sont nombreux : offrir de manière transparente le bon service au bon moment à l’utilisateur, réduire sa charge cognitive, permettre la mesure précise et continue de ses données personnelles (*quantified self*), etc. Le contexte d’une personne englobe plusieurs dimensions telles que sa localisation, son niveau d’activité physique, son état de santé et ses émotions. Nous nous concentrons ici sur la reconnaissance automatique des modes de transport. Cette détection a des applications dans différents domaines comme l’automatisation des enquêtes de mobilité, l’estimation de l’impact écologique du transport ou les applications fitness. L’objectif est de proposer une méthode pour la classification fine en ligne des modes de transport basée sur les capteurs embarqués à faible consommation et sous des conditions réalistes d’usage.

2 État de l’art

Les travaux antérieurs se répartissent en 3 catégories [1] : a) les approches basées sur une infrastructure extérieure ; b) les

approches basées sur des capteurs inertiels c) les approches mixtes.

Le premier type (a) utilise des systèmes tels que le GPS, des cartes géographiques, les signaux du réseau cellulaire ou encore les réseaux Wi-Fi. Par exemple, [2] utilise les caractéristiques des fluctuations des signaux GSM pour reconnaître les classes « stationnaire », « marche » et « en voiture » avec 85% de précision. [3] utilise le GPS et des informations du réseau de transport pour reconnaître 6 classes différentes avec une précision de 93%. En dépit d’assez bonnes performances, ces approches se heurtent à l’accès, la qualité et la quantité des informations fournies par l’infrastructure externe : le réseau cellulaire ne fournit pas suffisamment d’information pour une classification fine, le signal GPS n’est pas toujours disponible et son usage intensif entraîne une forte consommation énergétique.

Le deuxième type (b) essaie d’éviter les inconvénients mentionnés ci-dessus par l’utilisation de capteurs inertiels présents dans la plupart des smartphones actuels. Par exemple, [4] et [6] utilisent uniquement l’accéléromètre pour classifier 7 et 8 classes différentes respectivement, mais les précisions atteintes d’environ 80% ne sont pas encore au niveau souhaité. En revanche, [1] utilise l’accéléromètre, le magnétomètre et le gyroscope pour atteindre 92% de précision sur la reconnaissance de 5 classes, mais les différents modes de transport motorisés

ne peuvent pas être distingués dans ce travail, ce qui peut être inacceptable pour certaines applications (e.g., estimation d'impact écologique).

Le troisième type (c) fusionne les stratégies des deux autres. Ainsi, [5] utilise conjointement l'accéléromètre avec le GPS pour la classification. Cette approche, bien que moins affectée par les inconvénients liés à l'utilisation d'une infrastructure extérieure, reste limitée par une consommation élevée.

En résumé, les méthodes proposées ne sont pas encore suffisamment performantes pour répondre aux attentes des utilisateurs. Les principaux problèmes sont la précision de classification insuffisante, une consommation énergétique élevée, une classification pas assez fine (nombre de modes reconnus insuffisant) ou hypothèses d'usage non réalistes du type « GPS toujours disponible » ou « téléphone toujours placé à la même position ».

Indépendamment de l'approche utilisée, la plupart des méthodes proposées procèdent en 3 étapes :

1. acquisition des données et segmentation en fenêtres d'analyse ;
2. extraction de caractéristiques ;
3. classification supervisée, suivie d'une étape de post traitement.

Les travaux antérieurs utilisent tous des classifieurs performants sur des données plus ou moins similaires. Les différences proviennent essentiellement des caractéristiques extraites (coefficients de la FFT, variances, etc.) et de la possibilité de les interpréter physiquement afin d'explicitier les résultats obtenus. Plutôt que de calculer un grand nombre de caractéristiques avant de les sélectionner par des outils de réduction de dimension linéaire telle que l'ACP, ou non linéaire telle que ISOMAP, nous proposons la construction d'un espace de caractéristiques riche en informations interprétables physiquement et de faible dimension (évitant la malédiction de la dimensionnalité) dans un contexte de faible consommation.

3 Méthode proposée

La consommation des différents capteurs présents dans un smartphone a été étudiée dans [1]. L'accéléromètre et le magnétomètre sont parmi les capteurs de plus faible consommation avec 0,1mA et 0,4mA pour une fréquence d'échantillonnage de 30Hz. À titre de comparaison, le GPS consomme 30mA lors du suivi de satellite. Étant donné que la consommation énergétique est une contrainte importante, notre méthode se base sur ces 2 capteurs inertiels. D'autre part, nous proposons un espace de caractéristiques de faible dimension et physiquement interprétable, dont le but est de réduire les risques de surapprentissage.

3.1 Prétraitement des données

Les données sont acquises à la fréquence d'échantillonnage de 100Hz et traitées dans des fenêtres glissantes de 5 secondes avec 80% de chevauchement entre 2 fenêtres consécutives. Une

fenêtre de petite taille permet de détecter les changements de mode plus rapidement, mais certains motifs liés au mode de transport ne sont visibles que dans des fenêtres plus longues. Les valeurs choisies sont un bon compromis entre précision, coût de calcul et latence en milieu urbain.

Comme les smartphones ne sont pas des systèmes temps-réel et qu'une faible priorité est attribuée aux capteurs par le système d'exploitation, les fréquences d'échantillonnage de l'accéléromètre et du magnétomètre ne sont pas constantes. Comme les caractéristiques fréquentielles des accélérations mesurées sont pertinentes, une étape de ré-échantillonnage est réalisée par une interpolation linéaire toutes les 10ms ($f_s = 100\text{Hz}$).

Par ailleurs, les données mesurées par les capteurs sont exprimées dans le repère du téléphone (x, y, z) , dont l'orientation est inconnue et peut changer dans le temps. L'accéléromètre ne mesure pas directement son accélération linéaire (la dérivée de sa vitesse), mais la différence entre sa propre accélération et le champ gravitationnel local. Par exemple, un accéléromètre au repos sur une table mesure une accélération de $1g$ vers le haut. Ce comportement est décrit par l'équation :

$$\mathbf{m}_t = \mathbf{a}_t - \mathbf{g}_t + \mathbf{b}_t \quad (1)$$

où $\mathbf{m}_t = (m_x, m_y, m_z)_t$: mesure de l'accéléromètre à l'instant t dans le repère (x, y, z) ; $\mathbf{a}_t = (a_x, a_y, a_z)_t$: l'accélération linéaire ; $\mathbf{g}_t = (g_x, g_y, g_z)_t$: champ gravitationnel $\mathbf{b}_t = (b_x, b_y, b_z)_t$: bruit de mesure.

Pour extraire des caractéristiques invariantes par rapport à l'orientation du téléphone, la plupart des travaux précédents ([1] [5] [6] [7]) utilisent uniquement la norme des mesures, perdant ainsi des informations précieuses. [8] [4] [9] extraient les accélérations verticales et horizontales de l'accélération linéaire. Nous estimons l'accélération dans un repère (v, h_1, h_2) où v correspond à la direction verticale, h_1 , à la direction horizontale parallèle au déplacement et h_2 correspond à la direction horizontale perpendiculaire à h_1 .

Les hypothèses faites sont les suivantes : a) la gravité reste constante dans une fenêtre, b) les accélérations moyennes sont faibles devant la gravité, c) la moyenne du bruit est nulle et d) l'accélération horizontale principale est dans la direction du mouvement [10].

Sous ces hypothèses, le vecteur gravité est estimé par la moyenne des échantillons dans la fenêtre

$$\hat{\mathbf{g}} = \frac{-1}{N} \sum_{n=1}^N \mathbf{m}_n \quad (2)$$

où N est le nombre d'échantillons dans la fenêtre. L'accélération linéaire estimée vaut donc

$$\hat{\mathbf{a}}_n = \mathbf{m}_n + \hat{\mathbf{g}}, \text{ avec } n = 1, 2, \dots, N. \quad (3)$$

Le calcul des composantes verticale et horizontale de l'accélération linéaire devient (nous supprimons l'indice d'échantillon n pour ne pas alourdir la notation) :

$$a_v = \hat{\mathbf{a}} \cdot \frac{-\hat{\mathbf{g}}}{\|\hat{\mathbf{g}}\|} \quad (4)$$

$$\mathbf{a}_h = \hat{\mathbf{a}} - a_v \frac{-\hat{\mathbf{g}}}{\|\hat{\mathbf{g}}\|}. \quad (5)$$

Pour estimer les composantes a_{h_1} et a_{h_2} de l'accélération horizontale, nous effectuons une analyse en composantes principales sur les vecteurs \mathbf{a}_h . La première composante principale correspond à la composante a_{h_1} qui est la direction du mouvement et la deuxième correspond à a_{h_2} .

3.2 Extraction de caractéristiques

Alors que des études précédentes (e.g., [1] [6] [5]) ont souvent utilisé directement les coefficients de la transformée de Fourier discrète (TFD) comme caractéristiques, nous utilisons 9 caractéristiques facilement interprétable physiquement, 2 provenant du magnétomètre et 7 de l'accéléromètre. Ces caractéristiques, extraites sur chaque fenêtre, sont :

1. L'écart-type de la norme du champ magnétique mesuré ;
2. La valeur moyenne de la norme du champ magnétique mesurée ;
- 3-5. Les écarts-type des trois composantes (a_v, a_{h_1}, a_{h_2}) de l'accélération ;
6. L'énergie des fréquences supérieures à 10Hz pour la composante a_v de l'accélération ;
7. Le rapport entre (6) et l'énergie totale pour la composante a_v de l'accélération ;
8. La périodicité de la composante a_{h_1} de l'accélération ;
9. La fréquence fondamentale du mouvement estimée à partir de a_{h_1} .

Les caractéristiques (1) et (2) fournissent des informations sur l'environnement magnétique de l'utilisateur, ce qui est utile puisque chaque mode de transport produit des perturbations magnétiques différentes. Par exemple, le tramway provoque des accélérations relativement faibles, mais il présente l'environnement magnétique le plus perturbé.

Les caractéristiques (3-5) sont liées à l'hétérogénéité des mouvements dans chaque direction : un piéton change plus facilement de direction qu'un véhicule.

(6) et (7) mesurent des vibrations de haute fréquence qui sont souvent produites par les modes de transport motorisés et qui, selon nos expérimentations, sont plus facilement détectables dans la direction verticale.

(8) et (9) sont utiles pour les classes « marche », « course » et « vélo », lesquelles sont caractérisées par des mouvements quasi-périodiques et par des niveaux importants d'accélération. L'estimation de la périodicité évite, par exemple, que les mouvements brusques faits pendant l'usage normal du téléphone soient classifiés par erreur comme une activité physique.

Nous avons observé que l'allure des TFD (notamment le nombre de composantes harmoniques et l'harmonique dominante) calculées pour une même activité varie de manière significative selon l'utilisateur et selon la position du téléphone. Nous avons, donc, fait le choix d'estimer la fréquence fondamentale du mouvement, laquelle est étroitement liée à la cadence de l'activité réalisée. La périodicité et la fréquence fondamentale du mouvement sont estimées à partir de la fonction

d'autocorrélation $R_{h_1}[\tau]$ de la composante a_{h_1} de l'accélération.

$$R_{h_1}[\tau] = \frac{1}{(N - \tau)\sigma_{h_1}^2} \sum_{n=1}^{N-\tau} a_{h_1}[n]a_{h_1}[n + \tau] \quad (6)$$

La fonction d'autocorrélation estimée à partir de l'équation (6) a un maximum global à $\tau = 0$ avec $R_{h_1}[0] = 1$. Si le signal a_{h_1} est périodique avec une période fondamentale T_0 , la fonction $R_{h_1}[\tau]$ a aussi des maximums globaux aux multiples de la période T_0 , c'est-à-dire, $R_{h_1}[kT_0] \approx 1$. Grâce à cette caractéristique, la fonction d'autocorrélation est très bien adaptée à l'estimation de la fréquence fondamentale : il suffit de chercher sa valeur maximale dans une région d'intérêt, cette valeur étant en plus une bonne mesure de la périodicité du signal.

En pratique, nous cherchons tous les pics de $R_{h_1}[\tau]$ pour $\tau \in [50, 200]$ ce qui correspond aux fréquences entre 0, 5Hz et 2Hz. Ensuite une interpolation quadratique comme proposée en [11] est effectuée autour de chaque pic trouvé pour obtenir une estimation plus précise de sa localisation et valeur. Enfin, notre estimation de la périodicité du mouvement correspond à la valeur du plus grand pic trouvé après interpolation.

3.3 Classification

Pour l'apprentissage et la classification automatique, nous avons considéré trois algorithmes différents : arbre de décision (J48), naïve Bayes et forêt aléatoire. L'arbre de décision et le naïve Bayes sont parmi les algorithmes de classification les plus utilisés dans la littérature. Cependant, nous nous limitons ici aux forêts aléatoires car cette méthode a fourni les meilleurs résultats et présente deux caractéristiques importantes dans notre contexte de faible quantité de données : l'algorithme fournit une estimation interne de l'erreur de généralisation et est robuste vis-à-vis du surapprentissage [13].

L'algorithme proposé par Breiman [13] se compose d'un ensemble de N_a arbres de décision où chaque arbre i est appris à partir d'un sous-ensemble S_i de l'ensemble d'apprentissage A , i.e. $S_i \subset A$ avec $i = 1, 2, \dots, N_a$. Chacun des S_i est construit à partir d'un échantillonnage avec remplacement de l'ensemble A . Chaque arbre est construit jusqu'à sa taille maximale (aucun élagage), en considérant, à chaque nœud, un sous-ensemble aléatoire des caractéristiques. Puis chaque arbre vote et la classe la plus populaire est la sortie de l'algorithme. Nous utilisons WEKA [12] pour construire une forêt composée de 20 arbres, chacun généré en considérant 2 caractéristiques aléatoirement choisies à chaque nœud.

4 Résultats et discussion

4.1 Acquisition de données

Pour développer et évaluer notre méthode, une application Android a été développée permettant l'enregistrement automatique des données capteurs et du mode de transport annoté par l'utilisateur, lequel définit la vérité terrain pour l'apprentissage.

8 heures de données ont été acquises entre novembre 2013 et février 2015 dans la ville de Grenoble par 3 utilisateurs avec 2 modèles de smartphone différents sous Android 4.3 (Samsung Galaxy Nexus GT-I9250 et S3). Les smartphones ont été portés de façon à simuler des différents cas d’usage (téléphone à la main, poche pantalon, poche veste, etc.) et 8 classes différentes sont représentées : « stationnaire (*still*) », « marche (*walk*) », « course à pied (*run*) », « vélo (*bike*) », « voiture (*car*) », « voiture électrique (*ecar*) », « bus électrique (*ebus*) » et « tram ».

4.2 Évaluation

Pour évaluer notre méthode et la comparer aux travaux antérieurs, nous utilisons le taux de réussite obtenu par validation croisée (*10-fold*) et l’erreur *out-of-bag* (*OOB*), qui est une estimation de l’erreur de généralisation interne à la forêt aléatoire. Comme mentionné en section 3.3, chaque arbre i d’une forêt aléatoire est appris à partir d’un sous-ensemble des données (S_i). Ainsi, il est possible de classifier chaque échantillon x de l’ensemble d’apprentissage, en comptant uniquement les votes des arbres i' pour lesquels $x \notin S_{i'}$. L’erreur *OOB* est défini comme le taux d’erreur qu’on obtient en réalisant cette procédure. Selon Breiman [13], cette mesure est une aussi bonne estimation de l’erreur de généralisation que l’utilisation d’un ensemble de test indépendant de même taille que l’ensemble d’apprentissage.

Avec la validation croisée, 95,4% des fenêtres d’analyse ont été correctement classifiées, ce qui est clairement supérieur aux environ 80% obtenus par [4] et [6], seuls travaux basés uniquement sur des capteurs inertiels avec un nombre de classes comparable au nôtre. Notons que cette comparaison est délicate, car les bases de données ne sont pas comparables en nombre d’échantillons.

La matrice de confusion obtenue est présentée dans la table 1. On remarque notamment que la classe « stationnaire » est particulièrement difficile à classifier (et souvent confondue avec « tram ») ce qui s’explique partiellement par le fait que tous les modes de transport présentent des périodes stationnaires (e.g., tram à l’arrêt ou cycliste qui attend devant un feu rouge) ce qui diminue la distance entre la classe « stationnaire » et les autres dans l’espace des caractéristiques.

TABLE 1 – Matrice de confusion de la validation croisée.

a	b	c	d	e	f	g	h	← classif.
2024	17	0	55	41	143	56	0	a = still
7	5443	2	38	4	0	0	1	b = walk
0	19	1787	0	0	0	0	0	c = run
35	47	0	3698	14	25	98	11	d = bike
98	1	0	17	1261	72	84	0	e = ecar
71	13	0	46	31	8748	39	9	f = tram
61	19	0	40	44	72	2324	0	g = ebus
3	2	0	2	0	19	1	2906	h = car

Enfin, nous avons obtenu une erreur *OOB* de 5,65% ce qui est compatible avec les résultats de validation croisée.

5 Conclusion

Nous avons proposé une méthode capable de détecter à l’aide d’un smartphone avec une bonne précision 8 classes différentes liées à la mobilité d’une personne. En se basant sur des capteurs inertiels à faible consommation avec des algorithmes de complexité relativement faible, l’impact prévu sur l’autonomie de la batterie du téléphone reste acceptable. La validation de la méthode a été réalisée avec des données acquises sous conditions réalistes et une précision d’environ 95% a été obtenue.

Ces travaux se poursuivent avec l’acquisition d’une base plus importante et un traitement par chaîne de Markov afin de modéliser les successions de déplacements. .

Références

- [1] M. Yu et al. *Big Data Small Footprint : The Design of A Low-Power Classifier for Detecting Transportation Modes*. VLDB Endowment, 2014.
- [2] T. Sohn et al. *Mobility detection using everyday gsm traces*. UbiComp, 2006.
- [3] L. Stenneth et al. *Transportation mode detection using mobile phones and GIS information*. SIGSPATIAL, 2011.
- [4] S. Hemminki et al. *Accelerometer-based transportation mode detection on smartphones*. 11th ACM Conference on Embedded Networked Sensor Systems, 2013.
- [5] S. Reddy et al. *Using mobile phones to determine transportation modes*. Transactions on Sensor Networks, 2010.
- [6] V. Manzoni et al. *Transportation mode identification and real-time CO2 emission estimation using smartphones*. Rapport technique, SENSEable City Lab - MIT, 2010.
- [7] T. Nick et al. *Classifying means of transportation using mobile sensor data*. Int. Conf. on Neural Networks, 2010.
- [8] S. Wang et al. *Accelerometer based transportation mode recognition on mobile phones*. Asia-Pacific Conference on Wearable Computing Systems, 2010.
- [9] H. Lu et al. *The Jigsaw continuous sensing engine for mobile phone applications*. 8th ACM Conference on Embedded Networked Sensor Systems, 2010.
- [10] K. Kunze et al. *Which way am I facing : Inferring horizontal device orientation from an accelerometer signal*. Int. Symposium on Wearable Computers, 2009.
- [11] P. Boersma. *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*. Institute of phonetic sciences, 1993.
- [12] M. Hall et al. *The WEKA data mining software : an update*. SIGKDD explorations newsletter, 2009.
- [13] L. Breiman. *Random forests*. Machine learning, Springer, 2001.