

# Les transitions de positions du smartphone, quelles conséquences sur la reconnaissance de l'activité humaine ?

The transitions of smartphone positions what consequences on the recognition of the human activities?

Hamdi Amroun<sup>1</sup>, Nizar Ouarti<sup>2</sup>, M'Hamed (Hamy) Temkit<sup>3</sup>, Mehdi Ammi<sup>4</sup>

<sup>1,4</sup> LIMSI-CNRS, Université Paris Sud, France, hamdi.amroun@limsi.fr, ammi@limsi.fr

<sup>2</sup> IPAL (Sorbonne UPMC, CNRS, ASTAR, NUS, IMT, UJF), Singapore, nizar.ouarti@ipal.cnrs.fr

<sup>3</sup> Mayo Clinic, Division of Health Sciences Research, USA, temkit.hamy@mayo.edu

**RÉSUMÉ.** Cet article présente une expérience montrant l'impact des transitions des positions d'un smartphone (d'une position fixe à une autre position fixe sur le corps) sur la reconnaissance de l'activité humaine, dans un environnement non contrôlé. Les activités étudiées sont : marcher, allonger, debout et assis. Dans cet environnement, les utilisateurs sont libres d'utiliser leur smartphone et de le faire passer d'une position à une autre, par exemple de la main à la poche. Notre méthodologie comprend la combinaison de différents capteurs. La classification est faite en utilisant l'algorithme du Deep Neural Network (DNN). Deux ensembles de données ont été considérés, la seule différence entre ces deux ensembles de données est que les transitions des positions du smartphone sont supprimées dans le deuxième ensemble de données. Les résultats montrent une amélioration de la précision de la classification des activités 93,87% pour l'ensemble total comprenant les transitions de positions et 95,07% pour l'ensemble de données avec les transitions supprimées. Nous montrons également qu'avec aucune transition, la convergence du DNN est plus rapide et plus robuste. Par conséquent, notre méthode consistant à supprimer des informations correspondant aux transitions devrait permettre d'économiser de l'espace mémoire et du temps de calcul tout en offrant une haute précision de classification.

**ABSTRACT.** This paper presents an experiment that shows the impact of the transitions of smartphone positions (from one position to another on the human body) on the recognition of the human activity, in an uncontrolled environment. The studied activities are walking, laying, standing and sitting. In this environment, users are free to switch smartphone position, for instance from hand to pocket. Our methodology includes the combination of different sensors and a classification with a Deep Neural Network algorithm. Two datasets were considered, the only difference between the two datasets is that the transition are removed in the second dataset. Results show an improvement of accuracy of the classification of activities 93.87% for the control dataset compare to 95.07% for the dataset with removed transitions. We also show that with no transitions the convergence is quicker and more robust. Consequently, our method including removing information corresponding to transitions should save memory space and computing time while offering a high classification precision

**MOTS-CLÉS.** Reconnaissance de l'activité, CNN, transition de positions.

**KEYWORDS.** Activity recognition, CNN, transients.

## 1. Introduction

Les techniques d'apprentissage en profondeur (DNN) ont été employées avec succès dans le domaine du traitement d'images, plus spécifiquement en reconnaissance faciale [1], la classification de caractères manuscrits [2]. Mais leur exploitation dans le domaine des objets connectés n'est pas encore très répandue. Ceci est probablement dû à la méthodologie qui n'est pas encore suffisamment mature.

La reconnaissance de l'activité humaine, en utilisant un réseau d'objets connectés dans un environnement non contrôlé, est un sujet qui reste encore sans réponse satisfaisante. En effet, de nombreux problèmes peuvent surgir lors du processus de reconnaissance de l'activité: le choix de la plateforme de collecte des données d'activité, la méthode de prétraitement des données, la sélection des descripteurs ; la fusion des données multi capteurs et enfin le modèle d'apprentissage.

Avec l'accroissement de la quantité de données échangée par les objets connectés, il apparaît indispensable aujourd'hui de réfléchir sur des approches exploitant au mieux ces objets connectés avec une maîtrise de la mémoire de stockage des données et un gain en temps de calcul.

Les techniques d'apprentissage en profondeur connaissent aujourd'hui un grand développement dans divers domaines dont la reconnaissance de l'activité humaine. Le problème avec ce genre de modèles est qu'ils sont gourmands en espace mémoire et demandent un temps de calcul assez important.

Nous proposons dans ce papier une approche pouvant utiliser un algorithme d'apprentissage en profondeur tout en permettant de réduire le temps de calcul et l'espace mémoire grâce à la suppression des données de transitions du smartphone d'une position fixe sur le corps humain vers une autre position fixe. Par exemple, pendant le processus de reconnaissance de l'activité humaine, reconnaître l'activité *debout* en ayant le smartphone dans la poche du pantalon et en le passant de la poche du pantalon vers la poche de la veste revient à reconnaître cette même activité (*debout*) mais en supprimant les données relatives au passage du smartphone de la poche du pantalon vers la poche de la veste.

Nous proposons donc dans ce qui suit de reconnaître les activités humaines en utilisant un DNN avec un prétraitement préalable aux données d'apprentissage en utilisant la DCT (Discret Cosinus Transform) comme méthode de calcul des descripteurs dans deux configurations: a) en considérant les données de transition et b) en supprimant les données de transition.

La partie suivante présente un bref état de l'art sur les techniques d'apprentissage utilisées en reconnaissance de l'activité en mettant l'accent sur le calcul des descripteurs.

## 2. Etat de l'art

L'apparition de l'Internet des objets (IdO) dans la vie quotidienne ouvre de nombreuses perspectives dans le domaine de la reconnaissance de l'activité humaine. De nombreux chercheurs ont étudié la reconnaissance de l'activité humaine en utilisant des objets intelligents standards (smartphone, smart watch, ...) [1]. En fait, cette approche fournit des plateformes souples, simples à utiliser, peu onéreuses et légères pour une surveillance transparente de l'activité quotidienne avec différents types de capteurs embarqués (ex. capteur inertielle, capteur de force, de pression, capteur de vitesse ...) [2] [3]. Généralement, les méthodes de classification sont appliquées après un prétraitement des données extraits, extraction et éventuellement fusion des descripteurs pour la reconnaissance de l'activité. Plusieurs méthodes sont couramment utilisées telles que les chaînes de Markov (HMM), SVM, arbres de décision [4, 11, 12].

Cependant, les méthodes de recherche actuelle n'utilisent pas efficacement les plates-formes d'objets connectés. En fait, la plupart des travaux portent sur la reconnaissance de l'activité dans des environnements contrôlés où les participants sont invités à exécuter des tâches spécifiques (par ex. debout, monter l'escalier) [4], ou pour annoter l'emplacement du dispositif au début de l'activité (ex. dans la poche, dans le sac) [5] [6]. En outre, de nombreuses études ont étudié la reconnaissance de l'activité en utilisant des capteurs dans des configurations spécifiques en fixant les dispositifs sur certains endroits sur le corps (ex: poignet, poche) [7] [8]. Les transitions des positions du smartphone d'un endroit sur le corps à l'autre n'ont pas été étudiées. Par exemple, lorsqu'une personne tient le smartphone dans la main en marchant, puis elle le met dans la poche, ou quand elle déplace le smartphone de la main à la poche du pantalon en position debout. Ce changement d'emplacement correspond à l'utilisation réelle des téléphones intelligents et est très récurrente dans la vie quotidienne. Les rares études qui tiennent compte de la transition entre les emplacements de dispositifs ont montré une diminution substantielle de la précision de la reconnaissance d'activité [9] [10]. De plus, la répétitivité du changement d'emplacement nécessite généralement plus de mémoire de stockage des données d'enregistrements et du temps de calcul surtout quand il s'agit d'utiliser des algorithmes comme le DNN pour la reconnaissance de l'activité.

Cet article propose d'aborder la question de la transition des positions du smartphone dans un environnement non contrôlé (figure 02). L'objectif principal est de trouver un compromis entre la précision de la reconnaissance de l'activité humaine et le gain en temps de calcul et en espace mémoire. Basé sur des capteurs intégrés dans un smartphone standard (accéléromètre, gyroscope, magnétomètre et capteurs de proximité (Pr)), nous développons une série de traitements efficaces pour la reconnaissance de quatre états d'activité humaines (marcher, debout, assis et allonger) en utilisant un algorithme de classification (Convolutional neural network).

L'étude a été réalisée dans un environnement non contrôlé. Trois participants équipés d'un smartphone ont été observés, pendant trois semaines, effectuer leurs activités quotidiennes sans instructions. Un algorithme DNN (Deep Neural Network) a été utilisé pour la reconnaissance de l'activité [13].

Nous avons étudié l'effet des transitions de positions lors du changement de la configuration du dispositif, que nous appelons: *les transitions*. Nous étudions l'effet de la configuration du smartphone sur le taux de reconnaissance et nous concluons sur le gain en espace mémoire et en temps de calcul.

### 3. Déroulement de l'expérience

L'expérience de collecte de données a eu lieu dans une pièce d'une maison pendant trois semaines. Trois participants âgés de 30 à 48 ans (deux hommes et une femme) ont participé à l'expérience pendant une semaine chacun. Trois caméras IP ont été fixées à différents endroits de la salle pour enregistrer les activités quotidiennes des participants. La fréquence d'échantillonnage des capteurs a été réglée à 120 Hz. La durée d'enregistrement était jusqu'au déchargement du smartphone soit en moyenne de 3 h et 50 min puis on effectue le rechargement de la batterie et on refait l'enregistrement et ce pendant trois semaines. Les vidéos et l'enregistrement des capteurs ont été synchronisés et programmés pour un démarrage simultané. Les enregistrements vidéo et les signaux des capteurs ont été étiquetés avec le logiciel ELAN. Les participants ont placé le Smartphone soit dans la poche du pantalon ou le tenir dans la main.

### 4. Analyse de l'activité humaine

Dans cette étude, un algorithme d'apprentissage par renforcement (DNN) a été utilisé comme classificateur, qui peut extraire des caractéristiques (features) par lui-même et sans aucune connaissance spécifique des données (accéléromètre, le gyroscope, le magnétomètre et les données du capteur de proximité).

En ignorant la procédure d'extraction des caractéristiques, le modèle pourrait devenir plus réactif. Le processus d'apprentissage du DNN est subdivisé en deux étapes: la pré-formation et le réglage.

La pré-formation est non supervisée et un réseau initial est créé à l'aide d'un algorithme d'apprentissage.

Le réglage fin est supervisé et les paramètres de toutes les couches seront mis à jour en utilisant l'algorithme de rétro-propagation du gradient.

Les notations suivantes sont utilisées pour désigner les paramètres du réseau :

- $I = h_0$  Entrée du réseau
- $h_i$  ( $i=1, 2, \dots, \tau-1$ ),  $i^{ième}$  Couche cachée
- $0 = h_\tau$  Sortie du réseau.

- $w_i$  ( $i=1, \dots, \tau$ ): Matrice de poids de connexion entre  $h_i$  et  $h_{i+1}$ .
- $\rho_i$  ( $i=1, \dots, \tau$ ): Biais pour les neurones de la couche  $h_i$  quand ils sont activés par la couche  $h_{i+1}$ .
- $\zeta_i$  ( $i=1, \dots, \tau$ ): Les biais des neurones de la couche  $h_i$  quand ils sont activés par la couche  $h_{i-1}$ .
- $\Theta$ : Tous les paramètres du réseau.
- $\mathcal{T}$ : l'ensemble de l'apprentissage.
- $[f_{\theta(x)}]_i$ : Le score associé avec le  $i^{ième}$  label par le paramètre du réseau.

D'après [15], deux couches adjacentes:  $h_{i-1}$  et  $h_i$  la fonction d'activation est définie par

$$p(h_{i-1,s} = 1 | h_i) = \rho_{i,s} + \sum_j w_{i,j} h_{i,j} \quad [1]$$

$$S(h_{i,t} = 1 | h_{i-1}) = \zeta_{i,t} + \sum_j w_{i,j} h_{i,j} \quad [2]$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad [3]$$

Tel que  $\Gamma(\cdot)$  est la fonction logistique.

#### 4.1. Pré-formation

L'objectif de la pré-formation est de maximiser la probabilité de générer des données d'apprentissage. La probabilité de chaque donnée d'apprentissage assignée par le réseau a été calculée en utilisant la fonction *énergie* (4):

$$P(I) = \sum_{h \in H} p(v, h) = \frac{\sum_h \exp(-E(v, h))}{\sum_{u, g} \exp(-E(u, g))} \quad [4]$$

Hinton [16] a proposé une méthode basée sur une couche de pré-formation.

Elle est utilisée pour obtenir un réseau de neurones approprié, en plaçant la couche inférieure comme visible  $v$ , et la couche supérieure comme couche cachée  $h$ .

Chaque couple de couches adjacentes peut être considéré comme une machine de Boltzmann restreinte (RBM).

L'ensemble du réseau est construit en formant un RBM qui a la fonction énergétique suivante:

$$E(v, h) = -\sum_{s,t} v_s w_{st} h_t - \sum_s b_s v_s - \sum_t c_t h_t \quad [5]$$

#### 4.2. Fine tuning

Le modèle a été entraîné en utilisant le maximum de vraisemblance par descente de gradient stochastique.

Nous avons maximisé la log-vraisemblance :

$$\Theta \rightarrow \sum_{(x,y) \in \mathcal{T}} \log(p(y|x, \Theta)) \quad [6]$$

Tel que  $x$  est la donnée d'entrée et  $y$  correspond aux labels.

Soit  $x$  un exemple donné, la probabilité  $p$  est calculée depuis les sorties d'un réseau de neurones par le biais d'une fonction softmax :

$$P(i|x, \theta) = \frac{e^{f_{\theta}(x)_i}}{\sum_j e^{f_{\theta}(x)_j}} \quad [7]$$

Cela permet d'exprimer facilement le log vraisemblance :

$$\text{Log } p(y|x, \theta) = f_{\theta}(x)_y - \log \left( \sum_j e^{f_{\theta}(x)_j} \right) \quad [8]$$

La maximisation de la log-vraisemblance à l'aide d'un gradient stochastique est effectuée en sélectionnant au hasard un exemple d'apprentissage  $(x, y)$  et en effectuant une descente du gradient:

$$\theta \rightarrow \theta + \phi \frac{\partial \log p(y|x, \theta)}{\partial \theta} \quad [9]$$

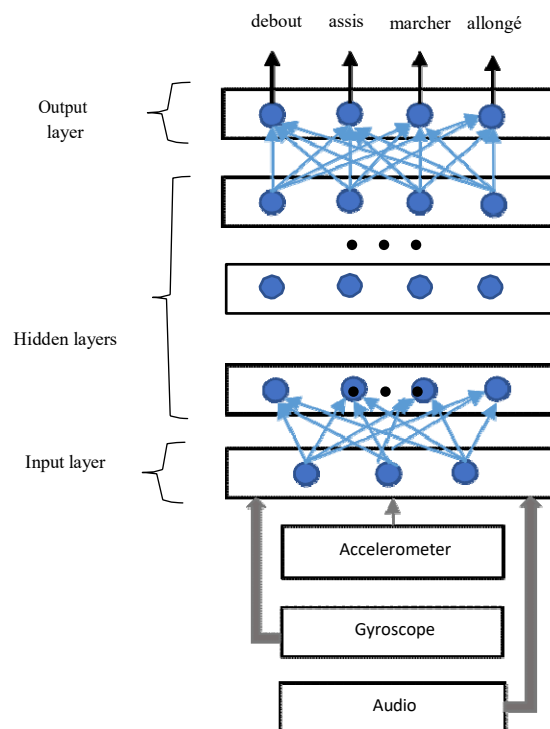
Où  $\phi$  est le taux d'apprentissage.

Toute l'architecture proposée est réalisée à l'aide de Theano Library.

L'efficacité de la méthode proposée a été évaluée sur la base de données des capteurs utilisés et testée par validation croisée (10-fold cross validation).

Le nombre de couches cachées de notre modèle est fixé à cinq et le nombre de neurones dans les couches cachées est fixé à 850-340-430-920-870.

D'autres paramètres de réseaux sont fixés comme paramètre par défaut du DBN de Hinton [16].



**Figure 1.** L'architecture du modèle utilisé

## 5. Etude de la transition de positions :

La figure 2 montre les quatre activités étudiées et les transitions des positions du smartphone correspondant entre la main et la poche du pantalon.

Les flux de données bruts enregistrés, de l'accéléromètre, gyroscope, magnétomètre et les signaux du capteur de proximité sont tous recadrés avec la même taille avec un chevauchement entre les échantillons de 256 points.

La longueur de la fenêtre de découpages des signaux de l'accélération, du gyroscope, du magnétomètre et des données du capteur de proximité était d'environ 4,16 secondes.

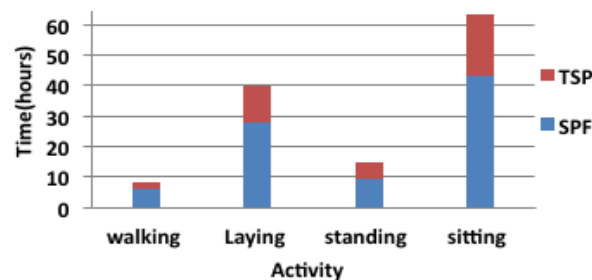
Les signaux de données ont été concaténés puis utilisés comme entrées du classificateur (figure 1), les enregistrements ont été déclenchés automatiquement via une application mobile (IOS) lorsque le participant rentre chez lui.

La figure 3 montre la répartition des données correspondant à 1) l'activité totale pour les quatre états pendant trois semaines (TA), 2) l'activité pendant les transitions des positions du Smartphone (TSP), et

3) l'activité lorsque la position du Smartphone est fixe (SPF). SPF correspond à TA à laquelle on soustrait TSP.



**Figure 2.** Les quatre états d'activité avec les transitions de positions du smartphone



**Figure 3.** La distribution des données correspondants aux : données totaux (TA), de l'activité de transition des positions du smartphone (TSP), et 3) l'activité quand le smartphone est dans une position fixe (SPF).  $TA = SPF + TSP$

Pour prouver l'efficacité de notre approche, nous avons étudié la reconnaissance des quatre activités humaines dans deux cas selon les données considérées: cas A avec TA et cas B avec SPF.

#### 4.1. Cas (A): les données totales de l'activité

Dans ce cas, toutes les données enregistrées (TA) sans prétraitement pour les trois participants ont été considérées. Le modèle DNN a été entraîné pour classer les activités des signaux d'entrée.

Nous avons proposé la procédure suivante: le modèle a d'abord été entraîné avec les signaux de données d'accéléromètre, puis avec les signaux de données d'accéléromètre et de gyroscope, et enfin avec tous les signaux de données du capteur de proximité.

Le tableau 1 montre les résultats de la précision de classification [7] [14] pour tous les participants.

Capteurs	Acc	Acc, gyro	Acc, gyro, mag, Pr
<i>AVG<sub>TA</sub></i>	90.01%	92.72%	93.87%
<i>AVG<sub>SPF</sub></i>	92.88%	94.75%	95.07%

**Tableau 1.** La précision de classification pour tous les participants pour (TA) et (SPF)

De plus, la classification a été appliquée sur les données des capteurs de chaque participant. Les données des activités du premier, deuxième et troisième participant ont été utilisées séparément pour entraîner le modèle DNN avec comme données d'entrée, les signaux des quatre capteurs: accéléromètre, gyroscope et magnétomètre. Le tableau 2 montre les résultats.

	Acc	Acc, gyro	Acc, gyro, mag, pr
<b>Participant N°1</b>	90.06%	91.69%	92.02%
<b>Participant N°2</b>	90.01%	91.73%	93.9%
<b>Participant N°3</b>	89.96%	94.74%	95.69%

**Tableau 2.** Précision de classification pour les trois participants

Pour un examen détaillé, les résultats ont été regroupés sous forme de matrices de confusion pour chaque participant. Les résultats sont présentés dans les tableaux 3, 4 et 5

Activité	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	91.8%		8.2%	
Sitting		93%	2.77%	4.23%
Standing	1.1%	6.8%	92.1%	
Laying		8.82%		91.18%

**Tableau 3.** La matrice de confusion des activités du participant 1

Activité	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	92.8%		7.2%	
Sitting		92.8%	0.1%	7.1%
Standing	0.08%	6.32%	93.6%	
Laying		3.8%		96.2%

**Tableau 4.** La matrice de confusion pour l'activité du participant 2

Activité	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	97.18%		2.82%	
Sitting		96.12%	0.28%	3.60%
Standing	0.72%	3.84%	95.44%	
Laying		5.98%		94.02%

**Tableau 5.** La matrice de confusion des activités du participant 3

#### 4.2. Cas (B) Les données des positions fixes du smartphone

Dans ce cas, seules les données correspondant aux positions fixes du smartphone (SPF) pour les trois participants ont été considérées. Les données TSP ont été supprimées.

Le tableau 1 montre les résultats de précision de la classification pour tous les participants. La classification a également été appliquée aux données des capteurs de chaque participant (voir tableau 6). Les résultats ont été regroupés sous forme matrices de confusion pour chaque participant (voir le tableau 8).

	Acc	Acc, gyro	Acc, gyro, mag, pr
Participant N°1	92.66%	93.92%	93.93%
Participant N°2	92.06%	93.83%	94.56%
Participant N°3	93.92%	96.5%	96.72%

**Tableau 6.** La précision de classification pour les trois participants

Pour un examen détaillé, les résultats ont été regroupés en matrices de confusion pour chaque participant. Les résultats sont présentés dans les tableaux 7, 8 et 9.



Activité	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	95.38%		4.62%	
Sitting		91.66%	2.63%	5.71%
Standing	2.13%	3.85%	94.02%	
Laying		5.34%		94.66%

**Tableau 7.** La Matrice de confusion des activités du participant 1

Activité	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	97.8%		2.2%	
Sitting		93.22%	1.46%	5.32%
Standing	1.91%	7.98%	90.11%	
Laying		4.89%		95.11%

**Tableau 8.** La Matrice de confusion des activités du participant 2

Activity	Classifié comme			
	Walking	Sitting	Standing	Laying
Walking	96.01%		3.99%	
Sitting		95.12%	2.13%	2.75%
Standing	0.22%	3.26%	96.72%	
Laying		3.98%		96.02%

**Tableau 9.** La Matrice de confusion des activités du participant 3

A titre de comparaison, après avoir appliqué un PCA, nous avons choisi de comparer nos résultats avec les résultats d'autres algorithmes: SVM, forêt aléatoire (RF) et arbres de décision (DT).

Les résultats sont regroupés dans le tableau:

	DNN	SVM	DT	RF
CAS A	93.87%	91.01%	90.18%	87.25%
CAS B	95.07%	92.97%	90.9%	88.03%

**Tableau 10.** Précision de classification des deux cas

## 5. Discussion

Le tableau 1 montre que la précision de classification dans le cas B (données de transitions supprimées) est plus élevée que dans le cas A (données totaux (TA)).

Dans le cas A, la précision de la classification, en utilisant uniquement les signaux de données de l'accéléromètre, est de 90,01%, alors que dans le cas B, la précision est de 92,88%. Mais en fusionnant les données des quatre capteurs, nous avons atteint respectivement 93,87% et 95,07% pour les cas A et le cas B. La différence de précision de classification entre le cas A et le cas B est d'environ 1%. De ce résultat, on peut déduire deux points:

- En fusionnant les données des capteurs, la différence de précision entre le cas A et le cas B tend à diminuer. Plus on ajoute des capteurs, plus la différence de précision de classification des activités entre les deux cas diminue (cas A et cas B).
- La suppression des données correspondant à la transition (TSP) n'a pas d'impact significatif sur la perte d'information. Ces données de transition correspondent à l'ensemble de données de transitions du smartphone d'une position fixe sur le corps humain vers une autre position fixe sur le corps.

Par exemple, un participant P, fait passer son téléphone portable de la poche du pantalon vers la poche de veste (ce qui correspond à deux positions fixes), ceci implique la suppression des données relatives à ce passage (de la poche du pantalon vers la poche de la veste). Cela est principalement dû à la fusion des données des quatre capteurs. Nous jugeons que les données correspondantes aux TSP sont négligeables et donc nous pourrions les supprimer.

Si les capteurs sont considérés individuellement, comme pour le participant 3 (voir tableau 2 et 6), la performance de classification est 89,96% et 93,92% dans le cas A et le cas B respectivement, avec une différence de de précision de classification de près de 4%. Dans ce cas, nous ne pouvons pas supprimer les données correspondant à la transition (TSP). Mais en fusionnant les données des capteurs, la précision de reconnaissance est respectivement de 95,69% et 96,72% dans le cas A et le cas B. Ici, la différence entre le cas A et le cas B est de 1,03%. On peut donc négliger les données correspondant à la transition (TSP). Le même raisonnement peut être fait sur les autres participants.

En traçant les matrices de confusion des trois participants (voir tableau 3, 4, 5, 7, 8 et 9), on a constaté que la précision de classification des quatre états est plus élevée dans le cas B que dans le cas A. Les matrices de confusion ont montré que le modèle de fusion proposé distingue clairement les différentes classes d'activité. Par exemple, à la fois dans le cas A et dans le cas B, i) la classe de l'activité « marcher » est confondue avec la classe « debout », ii) la classe « assis » est confondue avec les classes « debout » et « allonger », iii) la classe « debout » est confondue avec des classes « marcher » et « assis » et enfin, iv) la classe « allonger » est confondue avec la classe « assis ».

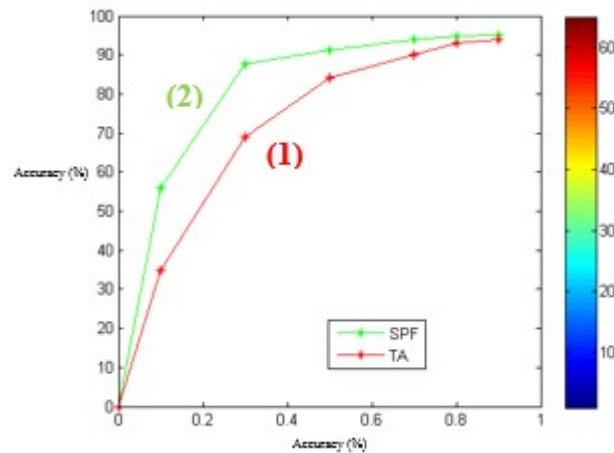
Dans le cas (A) et (B), le modèle conserve les mêmes caractéristiques de classification par rapport à la confusion entre les classes, c'est-à-dire si une classe est confondue avec une autre classe dans le cas A, elle restera aussi confondue avec la même classe dans le cas B. Par exemple, dans le cas A (voir tableau 3): la classe « marcher » présente une précision de 91,8% et n'est confondue qu'avec la classe « debout », avec une précision de 8,2%. Dans le cas B (voir tableau 7), la classe « marcher » présente une précision de 95,38% et est également confondue avec la classe « debout » avec une précision de 4,62%.

Ainsi, l'élimination des données de transition permet de conserver les mêmes caractéristiques de classification entre les classes et de conserver les mêmes structures pour les matrices de confusion. Tant dans le cas A que dans le cas B, l'approche de fusion proposée donne de meilleurs résultats que des travaux similaires pour la classification des activités étudiées. Par exemple, Kazuya Murao et Tsutomu

Terada [14] ont obtenu une précision de 70% pour la marche et de 91,7% pour l'activité « assis » lorsqu'ils tiennent le téléphone dans la main, en environnement contrôlé.

Le tableau 10 montre que le modèle basé sur le DNN fonctionne mieux que d'autres modèles populaires tels que SVM, DT et RF. En outre, les précisions de classification dans le cas A sont plus élevées que dans le cas B.

Pour étudier le gain en espace mémoire, une comparaison entre le jeu de données total (TA) (1) et le jeu de données excepté l'activité pendant la transition (SPF) (2) a été faite dans la figure 4 ci-dessous :



**Figure 4.** La performance de classification sur les données totales (TA) vs. Données des positions fixes du smartphone (SPF)

La précision de la classification avec la méthode DNN augmente rapidement dans le cas 1 (TA). Seulement 20% des données permet d'atteindre près de 70% de précision de classification et 60% des données ont permis d'obtenir une précision de classification de 90%.

Contrairement au cas 2 (SPF), la précision augmente lentement. 20% des données ont permis d'obtenir une précision de 45% à 50% et plus de 80% à 85% de la base de données est nécessaire pour atteindre la précision maximale de 90% à 93%. Cela a permis d'économiser près de 30% à 35% de l'espace mémoire en supprimant les données correspondant à la transition.

Cela signifie également un gain en temps de calcul par conséquent, près de 490.86 (s) et autour de 10.3 Go de mémoire pour chaque participant).

## 6. Conclusion

Dans cet article, nous avons étudié l'effet de transition des positions du Smartphone sur la performance de la classification de l'activité humaine.

Cette étude a montré que, avec notre approche, l'élimination des données de transition de position du smartphone n'entraîne pas de perte significative de l'information. Nous avons également montré quantitativement que la précision de classification augmente plus rapidement et est plus robuste sans transitions. Le DNN peut extraire des caractéristiques par lui-même et sans prétraitement spécifique des signaux de l'accélération, gyroscope, magnétomètre et les données du capteur de proximité. En ignorant la procédure d'extraction des caractéristiques, le modèle pourrait devenir plus réactif. Ainsi, l'utilisation de DNN permet de sauvegarder le calcul des caractéristiques pour chaque donnée de capteur, ce qui est comparable à d'autres méthodes telles que SVM, HMM [18, 19, et 20].

## Bibliographie

- [1] A. Milenković, C. Otto, et E. Jovanov, « Wireless sensor networks for personal health monitoring: Issues and an implementation », *Computer communications*, vol. 29, no 13, p. 2521–2533, 2006.
- [2] J. Lester, T. Choudhury, et G. Borriello, « A practical approach to recognizing physical activities », in *Pervasive Computing*, Springer, 2006, p. 1–16.
- [3] Geroch, M. S. (2004). Motion capture for the rest of us. *J.Comput.Small Coll.*, 19(3), 157-164.
- [4] C. Zhu et W. Sheng, « Multi-sensor fusion for human daily activity recognition in robot-assisted living », in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, p. 303–304.
- [5] L. Bao et S. S. Intille, « Activity recognition from user-annotated acceleration data », in *Pervasive computing*, Springer, 2004, p. 1–17.
- [6] A. J. Brush, A. K. Karlson, J. Scott, R. Sarin, A. Jacobs, B. Bond, O. Murillo, G. Hunt, M. Sinclair, K. Hammil, et others, « User experiences with activity-based navigation on mobile devices », in *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, 2010, p. 73–82.
- [7] L. Gao, A. K. Bourke, et J. Nelson, « A comparison of classifiers for activity recognition using multiple accelerometer-based sensors », in *Cybernetic Intelligent Systems (CIS), 2012 IEEE 11th International Conference on*, 2012, p. 149–153.
- [8] K. Ouchi et M. Doi, « Smartphone-based monitoring system for activities of daily living for elderly people and their relatives etc. », in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, p. 103–106.
- [9] G. M. Weiss et J. W. Lockhart, « The impact of personalization on smartphone-based activity recognition », in *AAAI Workshop on Activity Context Representation: Techniques and Languages*, 2012.
- [10] J. R. Kwapisz, G. M. Weiss, et S. A. Moore, « Activity recognition using cell phone accelerometers », *ACM SigKDD Explorations Newsletter*, vol. 12, no 2, p. 74–82, 2011.
- [11] A. Ghosh et G. Riccardi, « Recognizing human activities from smartphone sensor signals », in *Proceedings of the ACM International Conference on Multimedia*, 2014, p. 865–868.
- [12] S. A. Rahman, C. Merck, Y. Huang, et S. Kleinberg, « Unintrusive eating recognition using Google Glass », in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, 2015, p. 108–111.
- [13] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, et A. Y. Ng, « Multimodal deep learning », in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, p. 689–696.
- [14] K. Murao et T. Terada, « A recognition method for combined activities with accelerometers », in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, p. 787–796.
- [15] T. Liu, M. Li, S. Zhou, et X. Du, « Sentiment classification via l2-norm deep belief network », in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, p. 2489–2492.
- [16] G. E. Hinton et R. R. Salakhutdinov, « Reducing the dimensionality of data with neural networks », *Science*, vol. 313, no 5786, p. 504–507, 2006.
- [17] R. Salakhutdinov et G. E. Hinton, « Deep boltzmann machines », in *International conference on artificial intelligence and statistics*, 2009, p. 448–455.
- [18] N. Ravi, N. Dandekar, P. Mysore, et M. L. Littman, « Activity recognition from accelerometer data », in *AAAI*, 2005, vol. 5, p. 1541–1546.
- [19] T. Van Kasteren, A. Noulas, G. Englebienne, et B. Kröse, « Accurate activity recognition in a home setting », in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, p. 1–9.
- [20] C. W. Han, S. J. Kang, et N. S. Kim, « Implementation of hmm-based human activity recognition using single triaxial accelerometer », *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 93, no 7, p. 1379–1383, 2010.