Apprentissage en profondeur des données brutes de l'activité humaine

In-Depth Learning of Raw Human Activity Data

Hamdi Amroun[1], M'Hamed (Hamy) Temkit[2], Mehdi Ammi[3]

[1,3] LIMSI-CNRS, Université Paris Sud, France, hamdi.amroun@limsi.fr, ammi@limsi.fr
[2] Mayo Clinic, Division of Health Sciences Research, USA, temkit.hamy@mayo.edu

RÉSUMÉ. Dans cet article, nous proposons une approche pour reconnaitre certaines activités physiques en utilisant un réseau d'objets connectés. L'approche consiste à classer certaines activités humaines : marcher, debout, assis et allonger. Cette étude utilise un réseau d'objets connectés usuels: une montre connectée, un smartphone et une télécommande connectée. Ces objets sont portés par les participants lors d'une expérience non contrôlée. Les données des capteurs des trois dispositifs ont été classées par un algorithme du DNN (Deep Neural Networks) sans prétraitement préalable des données d'entrée. Nous montrons que (DNN) fournit de meilleurs résultats par rapport aux autres algorithmes classiques de type arbres de décision (DT) et Support Vector Machine (SVM). Les résultats montrent également que les activités des participants ont été classées avec une précision de plus de 98,53%, en moyenne.

ABSTRACT. This paper proposes to study the recognition of certain daily physical activities by using a network of smart objects. The approach consists in the classification of certain participants' activities, the most common ones and those that are carried out with smart objects:Make a phone call (Call), open the door (Open), close the door (Close) and watch its smartwatch (Watch). The study exploits a network of commonly connected objects: a smart watch and a smartphone, transported by participants during an uncontrolled experiment. The sensors' data of the two devices were classified by a deep neural network (DNN) algorithm without prior data pre-processing. We show that DNN provides better results than Decision Tree (DT) and Support Vector Machine (SVM) algorithms. The results also show that some participants' activities were classified with an accuracy of more than 98%, on average.

MOTS-CLÉS. Reconnaissance de l'activité, DNN, environnement non contrôlé.
KEYWORDS. Activity recognition, DNN, Non controlled environnement, IOT.

1. Introduction

Avec les développements d'Internet des objets (IdO) dans la vie quotidienne, de nouveaux usages ont émergé, ce qui ouvre les portes à de nombreuses perspectives dans le domaine de la reconnaissance de l'activité humaine, en termes d'utilisations et d'applications.

De nombreuses études ont travaillé sur la reconnaissance de l'activité humaine à l'aide d'objets connectés standards et portables [1]. Ces méthodes permettent de fournir des dispositifs et des plates-formes faciles à utiliser, flexibles et surtout légères pour un contrôle efficace de l'activité humaine sur une base quotidienne, et comprennent de nombreux capteurs (centrale inertielle, capteur de pression, mesure d'oxygène, etc.) [2] [3].

Les travaux actuels n'utilisent pas ces technologies et plateformes de manière suffisamment exhaustive et mature. En effet, la majorité des travaux de recherche sur la reconnaissance de l'activité se font dans des environnements contrôlés où les participants exécutent des tâches quotidiennes spécifiques (par exemple, se préparer à manger, se mettre debout, monter l'escalier) [4], ou simplement leur demander d'annoter leurs activités ou de donner l'emplacement d'un dispositif au début de l'expérience. (Par exemple déposer le smartphone dans la poche du pantalon, dans le sac) [5] [6].

En général, l'application d'un algorithme de reconnaissance de l'activité humaine est conditionnée par un prétraitement préalable des données d'apprentissage avant qu'elles soient données en entrées de ces algorithmes (SVM, Random Forest, HMM,) [4, 11, 12]. Cependant, tous ces travaux antérieurs ont été principalement réalisés dans des environnements contrôlés et ne sont pas très robustes, avec des

taux de reconnaissance faibles et parfois très exigeants en termes de temps de calcul et d'espace mémoire.

Deep Neural Networks (DNN) sont une alternative prometteuse. En fait, ils ont été appliqués avec succès pour diverses problématiques comme la classification des caractères manuscrits et la reconnaissance des gestes [8] où les DNN ont été appliqués directement au flux de données sans prétraitement des données ni sélection des caractéristiques (features). Les résultats montrent un très bon taux de reconnaissance.

Dans cet article, nous proposons d'étudier la reconnaissance de l'activité humaine, dans un environnement non contrôlé, en utilisant un iPhone, une Apple Watch et une télécommande Apple TV. L'IPhone et la télécommande Apple TV contiennent un accéléromètre, un gyroscope et un microphone. La télécommande Apple TV contient un accéléromètre et un gyroscope.

Après avoir extrait les données de capteur de ces trois dispositifs, ils sont donnés comme entrée à un DNN sans prétraitement spécifique et préalable de ces données, c'est à dire des données brutes.

L'iPhone est placé dans la poche de pantalon des participants tandis que la télécommande Apple TV est gardée à la main. Le reste de cet article est structuré comme suit :

Une description de l'expérience a été donnée dans la section 2. Dans la section 3, nous présenterons en détail le processus de reconnaissance de l'activité en utilisant le DNN avec données brutes en entrée. La section 4 présente les résultats obtenus. La section 5 discute ces résultats et les compare aux travaux existants. Nous finissons avec une conclusion et quelques références.

2. Configuration expérimentale

L'expérience a eu lieu dans une maison pendant une semaine. Sept participants, âgés entre 25 ans à 48 ans (4 hommes et 3 femmes), ont participé à l'expérience pendant une semaine chacun. Trois caméras IP ont été fixées à différents endroits de la salle pour enregistrer l'activité des participants. Les vidéos ont été enregistrées sur un serveur local. Les participants ont été invités à porter un IPhone, une Apple Watch et une télécommande Apple TV pendant l'expérience. L'IPhone et la Apple Watch contiennent trois capteurs : un accéléromètre, un gyroscope et un microphone, chacun. La télécommande Apple TV embarque un accéléromètre et un gyroscope. La fréquence d'échantillonnage des capteurs de l'IPhone, Apple Watch et Apple TV a été fixée à 120 Hz, 128 Hz et 132 Hz respectivement et à une fréquence d'échantillonnage de 8 KHz pour le microphone. La durée d'enregistrement était de 3 heures et 50 minutes deux fois par jour pendant une semaine. Nous avons développé une application IOS, Apple Watch et Apple TV pour accéder, enregistrer et envoyer les données des capteurs via Wi-Fi à un serveur local, qui était stocké dans une base de données SQL SERVER. Un entrepôt de données a été créé pour intégrer les données automatiquement à partir de la base de données, à la fin de chaque enregistrement. Les enregistrements vidéo et capteurs ont été synchronisés pour démarrer simultanément. Les enregistrements vidéo et les signaux des capteurs ont été étiquetés avec le logiciel ELAN Software. L'IPhone est placé dans la poche des pantalons, la télécommande Apple TV est placée dans la main, l'Apple Watch est placée au niveau bras.

3. Analyse de l'activité

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 microphone). 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 une étape 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, …, τ-1), i^{ieme} Couche cachée
- O = h_τ Sortie du réseau.
- w_i(i=1, …, τ) : Matrice de poids de connexion entre h_i et h_{i+1}.
- ρ_i(i=1, …, τ) : Biais pour les neurones de la couche $h_i$quand ils sont actives par la couche h_{i+1}.
- ς_i (i=1, …, τ) : Les biais des neurones de la couche h_i quand ils sont activés par la couche h_{i-1}.
- Θ : Tous les paramètres du réseau.
- Ϯ : l'ensemble d'apprentissage.
- $[f_{\theta(x)}]_i$: Le score associé avec le i^{eme} 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

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

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

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

Tel que Γ (.) est la fonction logistique.

Pre-training ou 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) :

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

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 :

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

Fine tuning:

Le modèle a été entrainé en utilisant le maximum de vraisemblance de par décente de gradient stochastique. Nous avons maximisé la log-vraisemblance :

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

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 :

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

Cela permet d'exprimer facilement le log vraisemblance :

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

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 :

$$\Theta \rightarrow \Theta + \varphi \frac{\delta \log py|x,\Theta)}{\delta\Theta} \quad [9]$$

Où φ 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].

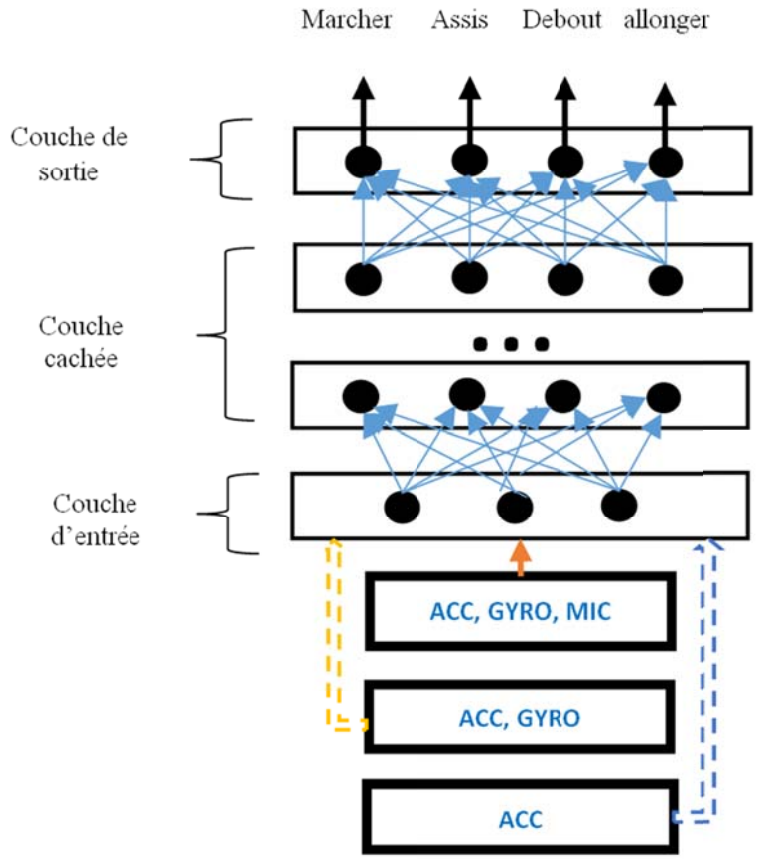

Figure 1. *L'architecture du modèle utilisé*

4. Résultats

La figure 2 montre les quatre activités étudiées :

Figure 2. *Les quatre activités étudiées*

Les flux des données des signaux d'accélération, de gyroscope et des capteurs audio enregistrés ont été tous recadrés de telle sorte qu'elles aient la même taille avec un chevauchement entre les échantillons des signaux de 256 points. La longueur de la fenêtre du signal d'accélération, du gyroscope et des données audio est d'environ 4,16 secondes chaque capteur. Les signaux de données ont été concaténés, puis utilisés comme entrées au classificateur (figure 1).

Les enregistrements ont été déclenchés automatiquement via une application lorsque les participants retournent chez eux. Le tableau 1 présente les résultats de classification des quatre activités par l'algorithme du DNN.

Activité	Iphone	Iphone,Watch	Iphone, watch, TV
debout	95.26%	96.08%	98.44%
assis	94.84%	95.99%	97.89%
allonger	95.88%	96.47%	99.26%
marcher	98.01%	98.66%	98.98%

Tableau 1. *Précision de la classification*
(IPhone dans la poche du pantalon, télécommande Apple TV à la main)

A noter que toutes les activités ont été reconnues avec une bonne précision de classification. La concaténation de ces signaux améliore nettement la précision de reconnaissance.

5. Discussion

La reconnaissance de l'activité physique a été réalisée en utilisant un iPhone, une Apple Watch et une télécommande Apple TV, classique comme défini sur le site officiel d'Apple. Les différents appareils ont été utilisés avec certaines configurations.

L'iPhone est resté dans la poche du pantalon des participants, tandis que la télécommande Apple TV est restée entre les mains des participants.

La précision de la classification à l'aide de l'algorithme DNN donne une précision moyenne de 95,32% pour l'IPhone, puis cette précision s'améliore sensiblement et s'élève à 96,18% en concaténant les données des deux appareils (IPhone et Apple Watch), puis, elle atteint presque 98,53%) en utilisant la télécommande Apple TV. Ceci confirme la contribution de la concaténation des différentes sources de données sur l'amélioration de la précision de reconnaissance.

Pour prouver la pertinence de nos résultats, nous avons comparé nos résultats de reconnaissance en appliquant deux algorithmes différents : SVM et DT. Les résultats sont résumés dans le tableau 2.

Pour appliquer ces deux algorithmes et les comparer avec le DNN, nous avons dû sélectionner des descripteurs comme dans [7] et appliquer une ACP pour réduire la dimensionnalité des données d'apprentissage.

	SVM	DT	DNN
debout	91.66%	92.66%	98.44%
assis	91.84%	93.49%	97.89%
allonger	92.89%	94.05%	99.26%
marcher	91.01%	93.66%	98.98%

Tableau 2. *Comparaison des précisions de classification de DNN, SVM et DT*

Nous remarquons que la précision de classification des trois algorithmes est relativement bonne. En effet, la SVM a une précision de classification moyenne de 92.13% alors que la précision moyenne de 93.4%. L'algorithme DT est meilleur que l'algorithme SVM, nous pouvons l'expliquer, entre autres, par le fait que le DT ne subit pas les problèmes de sur-apprentissage mais aussi, l'application de SVM nous oblige à sélectionner des descripteurs.

La sélection des caractéristiques joue un rôle important dans la réduction des données et donc des informations pertinentes pour la reconnaissance de l'activité humaine, mais aussi le fait de réduire la dimensionnalité pénalisent dans une certaine mesure la précision de la classification. Contrairement au modèle DNN qui extrait ces descripteurs automatiquement et donc le risque de perdre des informations pertinentes à la classification n'est pas prévu.

L'approche 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 “s'asseoir” lorsqu'ils tiennent un téléphone dans un environnement contrôlé.

6. Conclusion

Dans cet article, nous avons montré que nous pouvons reconnaître les activités humaines (debout, assis, couché et marcher) avec une très bonne précision de classification en utilisant un réseau d'objets intelligents et en appliquant un algorithme DNN sans prétraitement préalable des signaux d'entrée. Le DNN peut extraire des fonctions par lui-même et sans aucun prétraitement spécifique pour l'accélération, le gyroscope et les données de capteurs audio. Par conséquent, l'utilisation de DNN a permis d’économiser le calcul des descripteurs et donc la perte d'information en réduisant la taille des données, comme c'est le cas pour l'algorithme SVM [18, 19, et 20]. À l'avenir, nous nous concentrerons sur le calcul des peuilleures dscripteurs pour cette étude.

7. 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.