

Modélisation du processus de génération de données pour la reconnaissance d'activités

Massinissa Hamidi et Aomar Osmani

LIPN-UMR 7030 CNRS, Univ. Sorbonne Paris Nord
{hamidi,ao}@lipn.univ-paris13.fr

Résumé. La dynamique des mouvements liés à une activité (marche, vélo, etc.) est souvent déterminée par des interactions complexes impliquant diverses parties du corps. Ces dynamiques font partie d'un processus sous-jacent de génération de données et l'intégration de celui-ci dans les systèmes de reconnaissance d'activités basés sur des données peut améliorer leur robustesse et leur efficacité. Dans cet article, nous proposons de modéliser le processus de génération de données et de l'utiliser afin de contraindre l'apprentissage de modèles plus simples au travers de la sélection d'exemples. Nous présentons les résultats des expériences menées sur le jeu de données SHL collecté dans des conditions réelles à partir d'un environnement riche en capteurs. Comparativement au cadre de base, notre approche permet une amélioration des performances de reconnaissance tout en réduisant simultanément de moitié le nombre de sources de données nécessaires.

1 Introduction

La prolifération des objets connectés portatifs (*wearables*) permet l'émergence d'environnements riches en capteurs qui génèrent diverses formes d'informations. Ces sources offrent un large éventail de perspectives, permettant une reconnaissance plus robuste des activités humaines (Hamidi et Osmani, 2020; Radu et al., 2018). En effet, positionnées à différents endroits et générant des modalités variées, ces sources d'information si elles sont exploitées proprement, pourraient offrir de nombreux avantages tels qu'un meilleur rapport signal/bruit, une réduction de l'ambiguïté et une fiabilité accrue (Kurle et al., 2019) et surtout l'acquisition de connaissances de meilleure qualité répondant aux besoins exprimés.

Les tâches d'apprentissage qui émergent dans ces environnements riches en capteurs sont profondément structurées. C'est le cas des *wearables* avec le jeu de données SHL (Gjoreski et al., 2018) considéré dans cet article. Notre travail se concentre sur la reconnaissance des activités humaines à partir de sources de données matérialisées par des capteurs placés à différents endroits du corps suivant une topologie prédéfinie et fixe. Il a été observé que pour une activité donnée, des dynamiques impliquant des parties spécifiques du corps émergent. La combinaison de différentes modalités peut fournir des informations complémentaires permettant de capturer ces dynamiques et par conséquent de reconnaître précisément l'activité en question (Mantjarvi et al., 2001; Reddy et al., 2010; Bevilacqua et al., 2018).

La dynamique des mouvements du corps fait partie d'un processus de génération de données (PGD) et une longue lignée de recherche, par exemple (Papadopoulos et al., 2014; Parisi et al., 2017), a proposé d'intégrer ce type de connaissances a priori dans les modèles de reconnaissance d'activité. Plus précisément, Parisi et al. (2017) dérivent des représentations 3D basées sur le squelette du corps tandis que d'autres travaux codent des connaissances d'experts du domaine en utilisant des représentations basées sur l'ontologie (Ousmer et al., 2019; Rodríguez et al., 2013). Ces représentations sont ensuite utilisées pour contraindre l'entraînement des modèles de reconnaissance d'activité. Alors que l'incorporation de connaissances a priori sur la dynamique des mouvements du corps dans les systèmes d'apprentissage améliore les performances et est attrayante en termes d'interprétabilité, il est difficile de se fier uniquement à l'expertise humaine pour dériver des modèles de ces dynamiques (Vatavu et Pentiu, 2012). En effet, celles-ci sont le résultat d'interactions complexes impliquant diverses parties du corps.

Dans cet article, nous proposons une approche novatrice pour dériver et incorporer le PGD dans les modèles de reconnaissance d'activités. Notre approche améliore les performances de ces modèles grâce à deux étapes majeures. On commence par dériver un modèle du PGD par le biais d'une exploration à grande échelle d'un modèle de substitution (espace d'architecture neuronale). Sur la base du modèle ainsi dérivé, on sélectionne des sources de données très fiables pour les inclure dans l'échantillon d'apprentissage en utilisant un algorithme d'estimation de l'importance basé sur la variance. Les expériences que nous avons menées montrent l'efficacité de la combinaison du processus de génération de données par la sélection de sources de données hautement fiables. En particulier, nous obtenons une amélioration des performances de reconnaissance allant jusqu'à 17,84% par rapport à la configuration de référence, ce qui s'accompagne d'une réduction substantielle des données requises.

2 Formulation de la problématique

Nous considérons des configurations où une collection \mathcal{S} de M capteurs (également appelés générateurs de données ou sources de données), notées $\{s_1, \dots, s_M\}$, sont portés par l'utilisateur au cours des activités quotidiennes et capturent les mouvements du corps. Chaque capteur s_i génère un flux $\mathbf{x}^i = (x_1^i, x_2^i, \dots)$ d'observations d'une certaine modalité (ex. accéléromètre). L'objectif est de reconnaître un ensemble \mathcal{Y} d'activités, comme la course à pied, la marche ou le vélo.

Definition 1 (Source de données) *Une source de données (ou capteur) est caractérisée par deux attributs principaux : le premier est la **modalité** produite par le capteur et le second est la **position** où la source de données est située sur le corps. Une source de données est alors définie de manière unique avec ces deux attributs.*

Les activités humaines sont largement déterminées par la dynamique des gestes. En effet, chaque activité est caractérisée par un ensemble de gestes différents qui, à leur tour, impliquent des parties spécifiques du corps. Dans le cas des *wearables*, où ces parties du corps sont équipées de sources de données, souvent, le fait de se concentrer sur ces sources de données spécifiques permet de reconnaître précisément une activité donnée. Par conséquent, notre approche tente de sélectionner des sous-ensembles de sources de données qui sont très informatives par

rapport à ces dynamiques. Dans ce travail, nous nous concentrons sur deux notions qui codent ces dynamiques : l'importance d'une source de données unique et le degré d'interaction entre un ensemble de sources de données.

Definition 2 (Importance) *Étant donné une source de données s_i qui est attachée à une partie du corps et une activité y , l'importance de s_i par rapport à l'activité y , notée $\mu_i^y \in [0, 1)$, est définie comme une quantité qui représente l'implication relative de cette partie du corps dans la dynamique des gestes liés à cette activité.*

Definition 3 (Interaction) *Une interaction implique deux ou plusieurs sources de données et se définit comme leur degré de dépendance par rapport à l'implication relative des parties du corps auxquelles elles sont attachées, dans la dynamique des gestes. Plus le degré est élevé, plus les sources de données interagissent. Étant donné un ensemble de sources de données en interaction, $S \subset \mathcal{S}$, leur degré d'interaction est donné par $\mu_S^y \in [0, 1)$.*

Problem 1 (Sélection de sources de données basée sur le PGD) *Soit $PGD : \mathcal{O}(\mathcal{S}) \times \mathcal{Y} \rightarrow [0, 1)$ le processus de génération de données, qui donne, pour chaque activité $y \in \mathcal{Y}$ l'influence d'un ensemble de sources de données $S \subset \mathcal{S}$. L'objectif est d'utiliser PGD comme fonction indicatrice pour sélectionner des sources de données très fiables et informatives. Soit $\tau_{imp} \in [0, 1)$ et $\tau_{int} \in [0, 1)$ deux paramètres qui déterminent les seuils au-dessus desquels un ensemble donné de sources de données $S \subset \mathcal{S}$ peut être sélectionné. Il s'ensuit que les sous-ensembles de sources de données en interaction se rapportant à l'activité $y \in \mathcal{Y}$, notés \mathcal{S}_y , sont définis comme : $\mathcal{S}_y := \{s_i \in \mathcal{S} | PGD(\{s_i\}, y) = \mu_{s_i}^y \geq \tau_{imp}\} \cup \{S \subseteq \mathcal{S} | PGD(S, y) = \mu_S^y \geq \tau_{int}\}$*

3 Approche proposée

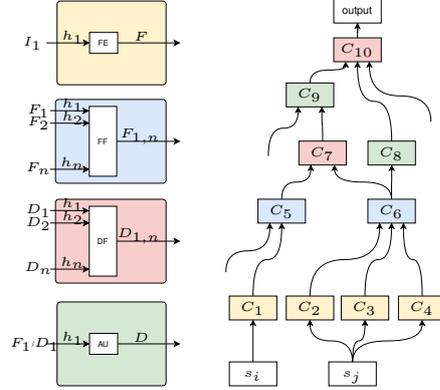
Notre approche améliore les performances des modèles de reconnaissance d'activité grâce à deux étapes majeures : (1) construction d'un modèle du PGD tel que décrit dans le problème 1 en utilisant un espace d'architecture comme modèle de substitution (proxy), et (2) sélection des sources de données très fiables et informatives à inclure dans l'échantillon d'apprentissage en utilisant un algorithme d'estimation de l'importance basé sur la variance. Ces deux étapes sont décrites ci-après.

3.1 Espace des architectures comme modèle de substitution

Nous utilisons l'espace défini par les architectures d'analyse multimodale comme un modèle de substitution pour la dynamique des mouvements du corps. L'exploration de cet espace d'architecture servira à dériver le PGD tel que défini dans le problème 1.

Une architecture est définie comme un ensemble de composants architecturaux chargés d'extraire, à partir des observations, des informations pertinentes, sous forme de caractéristiques, et de fusionner efficacement différentes sources de données (voir Fig 1). Il est pratique de représenter une architecture sous la forme d'un graphe acyclique dirigé où les composants architecturaux sont reliés entre eux par des arêtes valuées. Nous associons une valeur (hyperparamètre) $h_{u,v}^o$ à chaque arête du graphe qui relie deux composants C_u et C_v . Ces valeurs contrôlent la façon dont les composants architecturaux traitent chaque entrée et par la même

FIG. 1: (A gauche) composants d'extraction de caractéristiques et de fusion multimodale définis dans (Atrey et al., 2010). Extraction de caractéristiques (FE), fusion de caractéristiques (FF), fusion de décisions (DF) et unité d'analyse (AU). Ces éléments peuvent être combinés afin de former une analyse multimodale au niveau des caractéristiques, au niveau de la décision mais aussi à un niveau hybride. En outre, les hyperparamètres $h_i \in \mathcal{H}$ contrôlant les effets de chaque entrée sont représentés. (À droite) Illustration d'une architecture où chaque nœud correspond à un composant. Une arête qui lie le composant C_u au composant C_v indique que C_v reçoit la sortie de C_u comme entrée.



occasion leur influence sur la performance globale de l'architecture. L'ensemble des hyperparamètres d'une architecture est noté \mathcal{H} .

Nous nous concentrons sur ce qui découle de l'ajustement et de l'adaptation de ces architectures, à travers leurs hyperparamètres et plus particulièrement ceux qui contrôlent l'influence des sources de données. À chaque couche d'une architecture donnée, il est essentiel de définir la bonne combinaison d'hyperparamètres. En particulier, le choix de la bonne instantiation pour les composants d'extraction des caractéristiques et de fusion des sources peut conduire à une architecture capable de construire, à partir des différentes sources de données, un ensemble original de caractéristiques qui convient à la reconnaissance d'une activité donnée. Nous prenons en compte l'hypothèse suivante : Soit $\mathcal{H}_s \subsetneq \mathcal{H}$ l'ensemble des hyperparamètres contrôlant l'impact d'une source de données s . L'impact global de \mathcal{H}_s sur les performances de reconnaissance représente l'impact de la source de données s .

Le problème de la modélisation du PGD devient alors une exploration de l'espace des architectures. Cette exploration est déterminée par trois aspects : (i) un espace de recherche qui définit les composantes architecturales (ex., les couches convolutives); (ii) une stratégie de recherche qui définit comment l'espace doit être exploré (ex., optimisation Bayésienne des hyperparamètres); et (iii) une stratégie d'estimation des performances (ex. problème de la classification des séquences) (Elsken et al., 2019). Étant donné un budget d'exploration de B , la stratégie d'exploration produit une série de scores de validation ν_1, \dots, ν_B . La tâche de modélisation du PGD se réduit donc à trouver un lien entre ces scores et l'impact de chaque source de données.

3.2 Estimation de l'importance basée sur la Variance

Soit \mathcal{V} un ensemble de scores de validation où chaque score ν_k représente la performance estimée d'une instantiation particulière des hyperparamètres. Afin d'estimer l'importance de chaque source de données individuelle, nous décomposons la relation non linéaire f décrite par \mathcal{V} comme suit $f(\mathcal{S}, y) = \mu_0^y + \sum_{i=1}^M \mu_i^y(s_i) + \sum_{i \neq j} \mu_{ij}^y(s_i, s_j) + \dots + \mu_{1..M}^y(s_1, \dots, s_M)$, une moyenne μ_0^y à laquelle s'ajoutent les effets du premier ordre (μ_i^y), les effets du second

ordre (μ_{ij}^y) et ainsi de suite. Plus la variance induite par une source de données est faible, plus son influence sur la relation non linéaire f est importante. Cette formulation correspond à une expansion additive et la variance de chaque terme peut être estimée en utilisant l’analyse fonctionnelle de la variance (fANOVA) (Hoeffding, 1948). Elle peut être quantifiée en utilisant l’implémentation efficace proposée dans (Hoos et Leyton-Brown, 2014) qui est basée sur un algorithme de temps linéaire pour marginaliser les prédictions d’arbres de décisions. Comme nous avons accès à l’ensemble des scores de validation indexés par les instanciation d’hyperparamètres, afin d’estimer la décomposition, nous devons déterminer la correspondance entre chaque source de données et l’ensemble des hyperparamètres qui contrôlent leur influence.

Correspondance source de données/hyperparamètres. Étant donné une architecture A , nous déterminons une correspondance, $Corr_A : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{H} \times \mathbb{R})$, entre chaque source de données et les hyperparamètres qui contrôlent leurs effets, comme suit : $Corr_A(s) = \bigcup_{(u,v) \in s \rightarrow^* t} \langle h_u^v, w \rangle$, où $s \rightarrow^* t$ désigne tous les chemins dans l’architecture qui ont s comme source et t comme destination, h_u^v l’hyperparamètre associé à l’arête (u, v) , et w correspond à un poids calculé comme $w = \frac{\omega_1 \cdot dist(s,v) + \omega_2 \cdot \delta^-(v)}{\omega_1 + \omega_2}$ qui pondère la correspondance d’un hyperparamètre donné h_u^v en fonction de son éloignement ($dist(s, v)$) par rapport à l’entrée s et du nombre d’arêtes incidentes du noeud v ($\delta^-(v)$). Les paramètres de pondération $\omega_1, \omega_2 \in [0, 1)$ sont tous deux fixés à $\frac{1}{2}$. Dans le cas où une arête est partagé par plusieurs chemins, nous additionnons les poids attribués à l’hyperparamètre correspondant suivant chaque chemin.

4 Expériences et Résultats

Dans cette section, nous présentons les axes de l’évaluation empirique extensive de l’approche proposée. Nous évaluons différentes stratégies d’exploration de l’espace des architectures pour la dérivation du PGD, avec entre autre *Tree Parzen Estimator*, *Hyperband*, *Grid Search*, etc. En effet, différentes stratégies d’exploration tendent à favoriser différentes régions de l’espace des architectures ce qui entraîne la dérivation de PGD variés. À des fins de comparaison, nous dérivons, à l’aide d’experts du domaine, un autre modèle du PGD que l’on désigne dans ce qui suit par *HExp*. Les modèles ainsi obtenus sont comparés entre eux par rapport à la plausibilité des ensembles de sources de données dérivés. Nous évaluons ensuite l’efficacité de l’incorporation du modèle ainsi dérivé sur quatre jeux de données de reconnaissance d’activité, à savoir *USC-HAD*, *HTC-TMD*, *US-TMD* en plus du SHL. Nous détaillons dans ce qui suit les résultats de l’incorporation du PGD dans le jeu de données SHL. De plus amples analyses peuvent être retrouvées dans (Hamidi et Osmani, 2020).

Jeu de données SHL. Le SHL (Gjoreski et al., 2018) est un jeu de données très polyvalent et précisément annoté consacré à la reconnaissance de l’activité humaine. Ce jeu de données a été collecté dans des conditions réelles et fournit, simultanément, des données de locomotion multimodales issues de plusieurs positions du corps : main, torse, hanches et dos. Ces quatre positions définissent la topologie qui nous permet de modéliser et d’exploiter la dynamique des mouvements du corps pour des modèles de reconnaissance d’activité. Parmi les 16 modalités de l’ensemble de données original, nous sélectionnons les modalités de mouvements corporels à

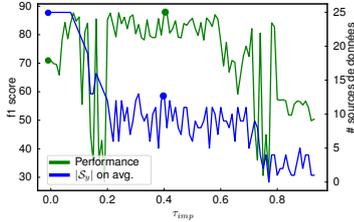


FIG. 2: Performances de reconnaissance en fonction du seuil d'importance de la source de données τ_{imp} . En parallèle, la cardinalité en moyenne des sous-ensembles $|S_y|$ utilisés pour l'entraînement des modèles est indiquée. Les points les plus à gauche correspondent à une configuration où toutes les sources de données sont utilisées (s-PGD).

Dataset	s-PGD	a-HExp	a-PGD
USC-HAD	72.1%	75.38%	89.33%
HTC-TMD	74.4%	77.16%	78.9%
US-TMD	71.32%	80.28%	83.64%
SHL	70.86 %	77.18%	88.7%

TAB. 1: Comparaison des différentes configurations d'incorporation du PGD en termes de performances de reconnaissance. Les scores de la colonne a-PGD correspondent aux modèles les plus performants, sélectionnés tout en faisant varier le seuil d'importance de la source de données τ_{imp} .

inclure dans nos expériences, à savoir : accéléromètre, gyroscope, magnétomètre, accélération linéaire, orientation, gravité, et en plus, pression ambiante.

Évaluation. Nous incorporons le modèle du PGD que nous avons dérivé dans des modèles de reconnaissance d'activité par la sélection d'échantillons. Nous sélectionnons des sources de données hautement informatives pour former des ensembles d'entraînement. Pendant la phase d'apprentissage, les modèles de reconnaissance d'activité sont encouragés à se concentrer sur les sous-ensembles de sources de données dérivées. Nous désignons cette configuration *a-PGD*, ce qui signifie, avec processus de génération de données. À titre de comparaison, nous apprenons aussi des modèles en utilisant les données générées par la totalité des sources de données. Ces modèles constituent notre configuration de référence que nous désignons *s-PGD* (sans processus de génération de données). En outre, nous évaluons également des modèles auxquelles nous incorporons le modèle HExp. Cette configuration est désignée *w-HExp*.

Fig. 2 montre l'évolution des performances de reconnaissance obtenues en fonction des paramètres τ_{int} et τ_{imp} . De plus, cette figure illustre le nombre moyen de sources de données, qui sont incluses dans les sous-ensembles, en fonction de ces deux seuils. En particulier, lorsque, par exemple, les paramètres τ_{imp} et τ_{int} sont fixés à 0, toutes les sources de données sont incluses. Nous constatons que les modèles appris avec des sous-ensembles restreints de sources de données ont de meilleures performances que la configuration de référence. Cette domination apparaît aussi par rapport à la plupart des configurations qui reposent sur un nombre plus élevé de sources de données. Il est à noter que nous obtenons un score de reconnaissance de $88,7\% \pm 0,6$, mesurée par la mesure f1, en utilisant des sous-ensembles contenant en moyenne 12 sources de données. Il s'agit donc d'une amélioration par rapport au niveau de référence de 17,84% en termes de performances de reconnaissance et d'une réduction de moitié concernant les quantités de données requises. Ce que l'on observe par ailleurs est qu'il n'y a pas beaucoup de sous-ensembles de sources de données qui entraînent de mauvaises performances de reconnaissance pour $0,2 \leq \tau_{imp} \leq 0,6$, où le nombre de sources de données par sous-ensemble est limité entre 5 et 12.

5 Discussion

Dans ce papier, nous avons proposé de dériver le PGD et de l'intégrer dans les modèles de reconnaissance d'activités. L'intégration de la connaissance du domaine dans ce type de modèles est particulièrement prometteuse et a attiré de nombreuses recherches (Vatavu et Pentiu, 2012; Kovalenko et al., 2014; Papadopoulos et al., 2014; Parisi et al., 2017; Dhiman et al., 2019). Si l'intégration des connaissances du domaine dans les processus d'apprentissage est bénéfique, la manière de procéder diffère sensiblement d'une approche à l'autre. Dans (Kovalenko et al., 2014) par exemple, l'ontologie dérivée par les auteurs leur sert de base à la construction d'un réseau d'inférence Bayésien tandis que dans (Parisi et al., 2017), les représentations construites aident les réseaux neuronaux à s'auto-organiser.

Au-delà de la reconnaissance des activités, de nombreuses autres applications s'appuient sur des modèles du domaine, telles que des lois de la physique, afin d'imposer certaines conditions ou équations dans les modèles d'apprentissage (Stewart et Ermon, 2017; Nabian et Meidani, 2020). Dans la même lignée, nous avons proposé dans (Osmani et al., 2019) d'exploiter un modèle plus expérimenté, un proxy, qui est chargé de sélectionner des échantillons afin de former de nouvelles générations de modèles dans le cadre de la surveillance industrielle. De nouveaux paradigmes, comme l'apprentissage de Vapnik en utilisant des informations privilégiées (Vapnik et Izmailov, 2015) et les connaissances distillées de Hinton (Hinton et al., 2015), proposent d'incorporer des modèles à haute capacité, similaires aux modèles de substitution que nous proposons ici, appelés *intelligent teachers* dans les modèles d'apprentissage.

Un volume de travaux en constante croissance propose d'exploiter les connaissances du domaine pour améliorer les performances des modèles d'apprentissage. Nos expériences encouragent encore plus l'exploration de cette voie, où des modèles plus complexes et plus expérimentés, comme l'espace d'architecture neurale que nous avons proposé, constituent des substituts aux connaissances du domaine et fournissent des indications pour des modèles plus simples par le biais de la sélection d'échantillons.

Références

- Atrey, P. K. et al. (2010). Multimodal fusion for multimedia analysis : a survey. *Multimedia systems* 16(6), 345–379.
- Bevilacqua, A. et al. (2018). Human activity recognition with convolutional neural networks. In *ECML-PKDD*, pp. 541–552. Springer.
- Dhiman, C. et al. (2019). Skeleton based activity recognition by fusing part-wise spatio-temporal and attention driven residues. *arXiv :1912.00576*.
- Elsken, T. et al. (2019). Neural architecture search : A survey. *JMLR* 20(55), 1–21.
- Gjoreski, H. et al. (2018). The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*.
- Hamidi, M. et A. Osmani (2020). Data generation process modeling for activity recognition. In *ECML-PKDD*. Springer.
- Hinton, G. et al. (2015). Distilling the knowledge in a neural network. *arXiv :1503.02531*.
- Hoeffding, W. (1948). A non-parametric test of independence. *Ann. of math. stat.*, 546–557.

- Hoos, H. et K. Leyton-Brown (2014). An efficient approach for assessing hyperparameter importance. In *International Conference on Machine Learning*, pp. 754–762.
- Kovalenko, M. et al. (2014). Real-time hand tracking and gesture recognition using semantic-probabilistic network. In *Int. Conf. on Comp. Model. and Simul.*, pp. 269–274. IEEE.
- Kurle, R., S. Günnemann, et P. van der Smagt (2019). Multi-source neural variational inference. In *AAAI*, Volume 33, pp. 4114–4121.
- Mantjarvi, J., J. Himberg, et T. Seppanen (2001). Recognizing human motion with multiple acceleration sensors. In *SMC*, Volume 2, pp. 747–752. IEEE.
- Nabian, M. A. et H. Meidani (2020). Physics-driven regularization of deep neural networks for enhanced engineering design and analysis. *J. of Comput. and Inf. Sci. in Eng.* 20(1).
- Osmani, A., M. Hamidi, et S. Bouhouche (2019). Monitoring of a dynamic system based on autoencoders. In *IJCAI*.
- Ousmer, M. et al. (2019). An ontology for reasoning on body-based gestures. In *SIGCHI EICS*.
- Papadopoulos, G. T. et al. (2014). Real-time skeleton-tracking-based human action recognition using kinect data. In *Int. Conf. on Multimedia Modeling*, pp. 473–483. Springer.
- Parisi, G. I., J. Tani, C. Weber, et S. Wermter (2017). Emergence of multimodal action representations from neural network self-organization. *Cogn. Syst. Res.* 43, 208–221.
- Radu, V., C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, et F. Kawsar (2018). Multimodal deep learning for activity and context recognition. *IMWUT* 1(4), 157.
- Reddy, S., M. Mun, J. Burke, D. Estrin, M. Hansen, et M. Srivastava (2010). Using mobile phones to determine transportation modes. *TOSN* 6(2), 13.
- Rodríguez, N. D. et al. (2013). Understanding movement and interaction : an ontology for kinect-based 3d depth sensors. In *Int. Conf. on UCAM*, pp. 254–261. Springer.
- Stewart, R. et S. Ermon (2017). Label-free supervision of neural networks with physics and domain knowledge. In *AAAI*, Volume 1, pp. 1–7.
- Vapnik, V. et R. Izmailov (2015). Learning using privileged information : similarity control and knowledge transfer. *JMLR* 16(2023-2049), 2.
- Vatavu, R. D. et S. G. Pentiu (2012). Multi-level representation of gesture as command for human computer interaction. *Computing and Informatics* 27(6), 837–851.

Summary

The dynamics of body movements are often driven by large and intricate low-level interactions involving various body parts. These dynamics are part of an underlying data generation process and incorporating them into data-driven activity recognition systems has the potential to enhance their robustness and data-efficiency. In this paper, we propose to model the underlying data generation process and use it to constrain training of simpler learning models via sample selection. We report on experiments conducted on the SHL dataset featuring a sensor-rich environment. Compared to the basic setting, our approach achieves substantial improvements, by simultaneously reducing the number of required data sources by one-half.