

Détection des activités quotidiennes à l'aide des Séparateurs à Vaste Marge

Gaëlle Loosli, Stéphane Canu, Alain Rakotomamonjy

Perception Système Informations - FRE CNRS 2645 - INSA de Rouen Avenue de l'Université 76801 Saint-Etienne-du-Rouvray Cedex - France.

{gloosli scanu arakotom}@insa-rouen.fr et

<http://asi.insa-rouen.fr/gloosli>

Résumé : Une problématique actuelle, induite par la complexité croissante des appareils technologiques, est la capacité d'un système à s'adapter à la situation de l'utilisateur, sans que celui-ci n'ait à s'en inquiéter. Dans cette optique, nous nous proposons de déterminer le comportement de déplacement d'une personne à partir de capteurs non intrusifs (accéléromètres). La méthode globale consiste à créer des caractéristiques susceptibles d'apporter de l'information à partir des signaux issus des capteurs, à garder les meilleures et à appliquer une méthode de discrimination. L'étape du processus sur laquelle nous avons insisté est le problème de la sélection des variables pertinentes : comment parvenir à ne garder que les caractéristiques utiles ? Pour répondre à cette question, nous avons mis en concurrence notre approche (sélection de variables par approche globale (Grandvalet & Canu, 2002)) avec celles servant de références dans le domaine. Nos méthodes obtiennent jusqu'à 99% de bonne classification hors ligne. Ces résultats permettent d'envisager une extension en mode en ligne pour une application dans le domaine de la détection de contexte.

Mots clef Apprentissage – Sélection de Variables – Séparateurs à Vaste Marge – Cartes Auto-Organisatrices de Kohonen

1 Introduction

Le sujet de cette étude s'inscrit dans le thème général de la détection du comportement d'un individu à partir de capteurs biométriques ou biomécaniques pouvant à terme être intégrés dans des vêtements de manière transparente pour l'utilisateur. En particulier nous nous intéressons ici à la détection des mouvements ou positions (tels que assis, debout, en train de marcher, de courir, ou de faire du vélo) à partir de deux accéléromètres placés au niveau des genoux. On peut distinguer à ce niveau deux problématiques. La première, que l'on peut appeler la détection hors ligne, consiste à s'intéresser à la détection des classes connues et pré-définies du système. La deuxième, la détection en ligne, s'intéresse davantage à la découverte des classes inconnues et à leur reconnaissance. Il nous semble important d'avoir dans un premier temps une bonne maîtrise de la détection hors ligne avant de s'intéresser aux classes inconnues, aussi cette étude

se restreint-elle à la première problématique. L'objectif de ce travail est de déterminer le cadre général (pré-traitement, représentation et algorithme d'apprentissage) qui nous permettra ensuite de mener à bien l'étude en ligne.

Contexte Il existe beaucoup de travaux sur la détection des activités humaines et du contexte de l'utilisateur. La plupart, tels que les travaux de C. Stauffer (Stauffer & Grimson, 2000) utilisent des caméras et font du traitement d'image. D'autres utilisent plutôt l'analyse de la parole (Zhang *et al.*, 2002) ou encore combinent les deux (Clarkson *et al.*, 2000). Ces travaux, compte tenu des temps de traitements des séquences vidéos et sons, fonctionnent hors-ligne et nécessitent des capteurs difficilement intégrables à des vêtements. Dans l'optique d'intégrer le système de détection aux vêtements de manière invisible et de faire de la détection en ligne, nous avons plutôt regardé les travaux utilisant des capteurs pouvant passer inaperçus tels que des capteurs biologiques (EEG, ECG, EMG...), des accéléromètres ou des gyroscopes. Ces capteurs permettent de déterminer le contexte de l'utilisateur et donc entre autres l'activité et les types de déplacements. On retrouve des dispositifs analogues dans le cadre médical (Najafi *et al.*, 2002). Il s'agit alors avant tout de déterminer si le patient est malade ou non. Ces travaux ne traitent pas l'aspect « découverte de nouveaux comportements ». A notre connaissance seuls les travaux de Kristof Van Laerhoven et son équipe partagent nos objectifs (Laerhoven, 2001). Dans ce travail, le contexte est déterminé à partir de deux accéléromètres en utilisant des caractéristiques issues des signaux, des cartes auto-organisatrices de Kohonen et une discrimination à l'aide de modèles gaussiens pour chaque classe, et ce pour chaque neurone de la carte, ce que les auteurs appellent de manière confuse les *k*-moyennes dans l'article. Notre approche consiste à utiliser les résultats de K. Van Laerhoven comme référence afin de démontrer sur les mêmes données qu'il est possible de faire mieux à partir d'hypothèses de travail différentes.

Hypothèses de travail Les trois points importants dans le processus permettant de passer des données issues des accéléromètres au choix de la classe, sont la définition des caractéristiques potentiellement pertinentes à partir des données, la sélection des variables réellement pertinentes, et la discrimination (voir figure 1). Dans la plupart des études du domaine, et en particulier dans (Laerhoven, 2001), la discrimination est faite sur l'ensemble de données, sans sélection de variables. Nous nous sommes donc proposés de travailler plus particulièrement sur cet aspect. En effet, dans l'optique d'un traitement en ligne, il est important de limiter le temps de traitement. Cela passe entre autres par une bonne représentation des données. Cela signifie que l'on cherche un bon compromis entre l'information utile et la taille du vecteur traité. Notre objectif ici est de vérifier notre hypothèse de départ, à savoir que les SVM (*Séparateurs à Vaste Marge*) associés à une sélection de variables répondent mieux au problème de discrimination que les cartes de Kohonen utilisées par K. Van Laerhoven. Nous avons donc cherché à reproduire les résultats qui nous servent de référence et ensuite nous avons appliqué nos méthodes afin de démontrer qu'elles sont plus performantes.

L'article est organisé de la manière suivante : dans un premier temps nous allons exposer brièvement les méthodes mises en œuvre dans l'ensemble du projet, en portant une attention plus particulière sur les SVM et la sélection de variables par approche

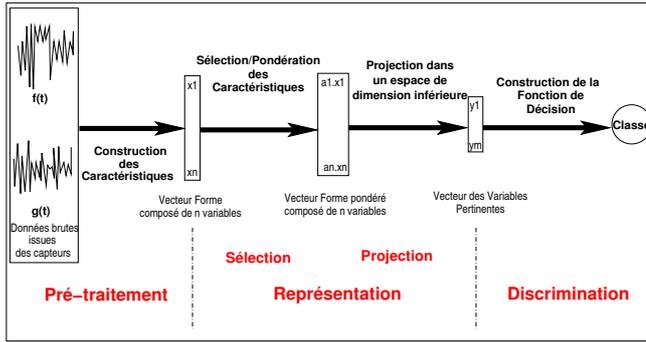


FIG. 1 – Chaîne des traitements, des signaux bruts à la décision. On voit ici les trois macro-tâches utiles à la prise de décision, à savoir le pré-traitement, la représentation des données, et enfin la discrimination.

globale, qui s’appuie sur les SVM. Dans un second temps, nous expliquerons l’étude expérimentale menée et les résultats obtenus et enfin nous discuterons et commenterons ces résultats.

2 Méthodes

Dans cette partie nous allons passer en revue les méthodes utilisées dans les quatre phases du processus. La première partie concernera la définition des caractéristiques issues des signaux. Les deux suivantes, plus détaillées puisqu’elles concernent le point clef de notre étude, seront à propos de la représentation de l’information, c’est-à-dire à propos de la sélection de variables et de la projection. Enfin nous aborderons les méthodes de discrimination que nous avons utilisées.

2.1 Pré-traitement - Définition des caractéristiques

Les données brutes issues des capteurs ne suffisent généralement pas pour faire un apprentissage des classes. L’idée est d’utiliser des caractéristiques que l’on peut construire à partir des signaux de base. Notre point de départ là encore est (Laerhoven, 2001) dans lequel K. Van Laerhoven utilise pour chaque signal la moyenne mobile, la variance mobile, le minimum et le maximum sur une fenêtre mobile. Mais on peut également penser aux transformées de Fourier, aux transformées en ondelettes, ou encore à la corrélation entre les signaux. On s’aperçoit vite que le problème n’est pas tellement de trouver des caractéristiques issues des signaux, mais plutôt de parvenir à déterminer lesquelles sont utiles à la détection des classes. Notre approche consiste à créer en premier lieu un grand nombre de caractéristiques puis d’éliminer celles qui sont inutiles. Dans les deux parties sur le traitement de la dimension, nous nous poserons la question de la sélection des variables pertinentes et de la projection dans un espace de dimension restreinte. Notons ici également que la taille des fenêtres mobiles est un hyper-paramètre duquel dépend la qualité des caractéristiques et par conséquent la qualité de la reconnaissance.

2.2 Sélection de Variables

Le problème majeur qui se pose à nous dans le traitement des caractéristiques que nous avons créées est connu sous le nom de malédiction de la dimensionnalité (Bellman, 1961). En effet, plus on a de caractéristiques et plus il est probable d'y trouver l'information dont on a besoin, mais également d'y trouver du bruit, voire des caractéristiques dites nuisibles. Par ailleurs, plus le nombre de caractéristiques est grand, plus il nous faudra d'entrées dans la base d'exemple pour des raisons de précision. Notre but est par conséquent d'éliminer les variables nuisibles ou inutiles et de mettre en valeur les autres.

2.2.1 Sans traitement

Le plus simple des traitements à appliquer concernant la dimension des caractéristiques est de tout garder ou encore de faire appel à un expert. De cette façon, il est sûr de ne pas éliminer de variables pertinentes. En revanche, dans le cas où l'on garde toutes les caractéristiques, le problème peut être de grande dimension et les traitements devenir très longs. Par ailleurs, comme nous venons de le voir, il est possible que certaines des variables créées non seulement n'apportent aucune information utile mais en plus soient nuisibles.

2.2.2 Sélection pas à pas

La sélection pas à pas (voir *stepwise method* dans (Leray & Gallinari, 1999)) est une procédure qui combine la sélection ascendante de variables (ajouter une à une les variables tant que l'ajout apporte significativement de l'information) et la sélection descendante (supprimer une à une les variables tant que la suppression n'engendre pas de perte significative d'information). La sélection pas à pas alterne ces deux phases de façon à pouvoir revenir sur une décision (ajout ou suppression) précédente.

2.2.3 Approche globale

L'approche globale présentée dans (Grandvalet & Canu, 2002) permet d'effectuer la sélection de variables en même temps que l'apprentissage par SVM. Toutefois cette méthode peut également être mise en œuvre uniquement pour la sélection de variables (comme ce sera le cas ici) et non pour l'apprentissage. En d'autres termes, une fois les variables pondérées par cet algorithme, nous appliquerons ou non une méthode de projection et ensuite une méthode de discrimination. L'approche globale, déclinaison des SVM, sera présentée plus en détail dans la partie 2.4.2, quand la méthode des SVM en elle-même sera expliquée.

2.3 Projection

Une autre manière de réduire la dimensionnalité du problème consiste à projeter les données dans un espace de plus petite dimension. Il existe deux types de méthodes, les méthodes linéaires (analyse de données) et les méthodes non-linéaires. Certaines méthodes classiques de projection telles que l'ACP (*Analyse en Composantes Principales*)

ou l'AFD (*Analyse Factorielle Discriminante*) ne seront pas présentées ici. En effet, lors d'une étude préliminaire, il s'est avéré qu'elles ne donnaient pas de résultats satisfaisants. Cela se comprend aisément du fait que se sont des méthodes linéaires et que nos données ne sont pas linéairement séparables.

2.3.1 Cartes Auto-Organisatrices de Kohonen

L'algorithme des cartes de Kohonen permet de réduire la dimension d'un problème par projection (Kohonen, 1989) (voir (Cottrell *et al.*, 2003) pour des applications avancées). La couche de sortie (c'est-à-dire la carte) est composée d'un nombre important de neurones, disposés dans un espace de \mathbb{R} , \mathbb{R}^2 ou parfois \mathbb{R}^3 , selon une topologie fixée. Notons les poids d'activation d'un neurone de la couche de sortie ω_k et Φ le vecteur en entrée du réseau (avec $\phi_1, \phi_2, \dots, \phi_n$ les caractéristiques d'un point). A chaque étape de l'apprentissage, c'est-à-dire à chaque nouvelle arrivée d'un vecteur Φ , chaque neurone de la couche de sortie calcule son activation : $A_k = \Phi^t \omega_k$. A l'issue de ce calcul, le neurone dont l'activation est la plus grande est déclaré vainqueur (on le notera y^*) et ainsi ses poids sont mis à jour selon la formule suivante :

$$\omega_{ki}(t+1) = \omega_{ki}(t) + \eta(t)\Lambda(|y - y^*|)\Phi_i,$$

avec $\eta(t)$ le taux d'apprentissage qui décroît avec le temps. La fonction $\Lambda(|y - y^*|)$ correspond à l'influence du neurone gagnant sur son voisinage. Ainsi, non seulement le neurone gagnant se spécialise dans la reconnaissance d'un type d'entrée, mais son voisinage évolue aussi, dans une moindre mesure. Cela permet de spécialiser des zones de la carte. Le principal avantage d'un tel algorithme est d'être non linéaire. Toutefois il ne garantit pas une solution unique au problème d'optimisation et l'apprentissage est long et instable.

2.3.2 Espace issu de la carte

A partir de la carte obtenue et de la matrice des poids des neurones de sortie, on peut engendrer un espace de dimension 2 où les points seront projetés. Lorsque la carte de Kohonen est de dimension 2, elle peut être représentée selon une topologie fixée (par exemple un voisinage carré). L'idée de la projection consiste à donner des coordonnées à chaque neurone dans l'espace de représentation de la carte (notées $Coor_k$). Ensuite, plutôt que d'attribuer à l'entrée Φ les coordonnées du neurone gagnant, on calcule ses coordonnées de la façon suivante :

$$x_{proj} = \frac{\sum_{k=1}^p A_k \cdot Coor_k}{\sum_{k=1}^p A_k}.$$

2.4 Discrimination

Les méthodes de discrimination s'appliquent sur des données i.i.d. tirées aléatoirement qui suivent une loi de probabilité. Connaissant cette loi de probabilité, nous pouvons appliquer la discrimination de Bayes. En l'absence d'hypothèses sur la nature de la loi sous-jacente, il nous faut avoir recours à des techniques d'apprentissage de type

universel ou non paramétrique. La méthode d'apprentissage principale de cette étude est celle des Séparateurs à Vaste Marge (SVM-*Support Vector Machines*). Pour comparaison avec une méthode plus classique, nous avons également mis en œuvre l'algorithme des k -Plus Proches Voisins (k PPV).

2.4.1 k PPV - Méthode des k -Plus Proches Voisins

La méthode k PPV est un algorithme supervisé classique (Duda *et al.*, 2001), dont le principe est de classer tout nouveau point dans la classe majoritairement présente dans son entourage. L'entourage est constitué des k points les plus proches au sens d'une distance pré définie en fonction de la nature des points à classer. La distance la plus couramment utilisée est la distance euclidienne. Cette méthode est d'autant plus performante que l'espace de représentation choisi sépare au mieux les différentes classes.

2.4.2 SVM - Séparateurs à Vaste Marge

Principe général Les séparateurs à vaste marge sont basés sur un algorithme d'apprentissage ayant comme fonction de décision un hyperplan optimal (2) dans un espace de grande dimension que nous allons appeler \mathcal{F} . A partir des données d'apprentissage $\{x_i, y_i\}_{i=1 \dots \ell}$, où $x_i \in \mathbb{R}^n$ et $y_i \in \{-1, 1\}$, l'algorithme consiste à représenter les points x_i dans cet espace \mathcal{F} en utilisant une transformation non-linéaire, construite à partir d'une famille de fonctions $\{\Phi_k\}_{k \in \mathbb{N}}$.

\mathcal{F} peut être vu comme l'espace vectoriel généré par cette famille $\{\Phi_k\}$. Sous certaines conditions très générales, il se trouve que la série $\Phi_k(x)\Phi_k(y)$ converge vers une fonction de \mathcal{F} appelée le noyau, que l'on note :

$$K(x, y) = \sum_{k \in \mathbb{N}} \Phi_k(x)\Phi_k(y) \tag{1}$$

Dans ce cas, la fonction de décision est donnée par le signe de la fonction de discrimination suivante qui ne dépend plus que du noyau K :

$$f(x) = \sum_{i=1}^{\ell} y_i \alpha_i K(x, x_i) + b \tag{2}$$

où les α_i et b sont des coefficients à déterminer, en maximisant la distance, appelée marge, entre la frontière de décision $f(x) = 0$ et le nuage de point dans \mathcal{F} (voir figure 2 pour une illustrations de ces concepts). Le problème à résoudre s'écrit alors :

$$\left\{ \begin{array}{l} \min_{f \in \mathcal{F}} \|f\|_{\mathcal{F}}^2 + C \sum_{i=1}^{\ell} |\xi_i|, \\ \text{avec } y_i f(x_i) \geq 1 - \xi_i, \quad i = 1, \ell \\ \text{et } \xi_i \geq 0, \quad i = 1, \ell \end{array} \right.$$

avec C un paramètre qui permet de régler le taux d'erreur admissible dans la solution (si $C = 0$, on peut faire autant d'erreurs que possible, si $C = \infty$, on n'admet aucune

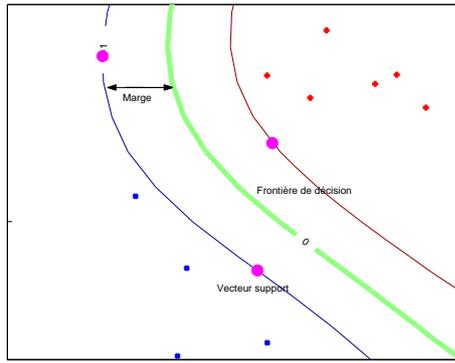


FIG. 2 – Illustration des notions de marge et de vecteurs supports sur un problème de discrimination à deux classes. La courbe la plus épaisse représente la frontière de décision du SVM, les courbes les plus fines montrent l'ensemble des points à une distance de 1 de la frontière de décision. Les points se trouvant sur ces deux lignes sont les vecteurs supports (seuls points utiles à la discrimination). La marge est la distance entre la frontière de décision et un vecteur support, et elle vaut $\frac{1}{\|f\|_{\mathcal{F}}}$

erreur). En pratique, ce problème d'optimisation se résout de la manière suivante :

$$\begin{cases} \max_{\alpha_i} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{avec } 0 \leq \alpha_i \leq C, i = 1, \ell \\ \text{et } \sum_{i=1}^{\ell} \alpha_i y_i = 0. \end{cases}$$

On peut se demander comment déterminer $K(x, y)$ à partir des fonctions Φ_k , mais on pose en fait le problème dans l'autre sens : pour un noyau donné $K(x, y)$, existe-t-il une famille des Φ_k vérifiant (1) ? Pour vérifier cela, il est nécessaire et suffisant que le noyau $K(x, y)$ soit symétrique et semi défini positif, c'est-à-dire que :

$$\left\{ \forall (\alpha_i)_{i \in \mathbb{N}} \in \mathbb{R}, \forall (z_i)_{i \in \mathbb{N}} \in \mathbb{R}^n, \sum_i \sum_j \alpha_i \alpha_j K(z_i, z_j) \geq 0 \right\},$$

ce qui est la cas des noyaux gaussiens que nous utilisons (Canu *et al.*, 2003; Vapnik, 1998).

Cas multi classes Les SVM sont des machines de discrimination binaire seulement. La généralisation de leur utilisation dans le cas multi classe peut se faire de trois façons différentes : par stratégie de type *un contre les autres* ou *un contre un*, et il existe également une manière de formaliser le problème en prenant en compte cet aspect multi-classes (Weston & Watkins, 1999).

Pour la stratégie *un contre les autres*, k frontières de décision $f_k(x)$ séparant une classe des autres sont générées. x est affecté à la classe v si :

$$v = \arg \max_k f_k(x)$$

Pour la stratégie *un contre un*, on apprend $k(k - 1)/2$ classifieurs où chaque classifieur apprend en fait à discriminer les points de deux classes différentes. Dans ce cas, l'affectation d'une classe à un nouveau point x_t consiste à évaluer la classe associée à ce point par les différentes fonctions de décision et de décider ensuite par une règle de vote majoritaire.

sélection de variables issue des SVM L'approche globale (dont on se sert pour la sélection de variables (Grandvalet & Canu, 2002)) ajoute une contrainte sur l'influence d'une variable dans la solution. On notera σ_j l'influence de la $j^{ième}$ variable. Le système à résoudre devient alors :

$$\left\{ \begin{array}{l} \min_{f \in H, \sigma \in \mathbb{R}^d} \|f(x_i)\|_H^2 + C \sum_{i=1}^{\ell} |\xi_i| + \mu \sum_{j=1}^n |\sigma_j|, \\ \text{avec } y_i f(x_i) \geq 1 - \xi_i, \quad i = 1, \ell \\ \text{et } \xi_i \geq 0, \quad i = 1, \ell \\ \text{et } \sigma_j \geq 0, \quad j = 1, n \end{array} \right.$$

avec μ un paramètre qui permet de régler le niveau de sélection de variables (si $\mu = 0$, on ne fait pas de sélection, si μ grand, on élimine beaucoup de variables).

La résolution de ce système se fait par une méthode itérative en deux étapes. La première consiste à calculer les SVM à σ fixé et la deuxième à optimiser σ . L'optimisation de σ entraîne la modification des vecteurs supports et des coefficients α et b . Les σ peuvent alors venir pondérer l'importance des variables avant d'appliquer d'autres méthodes de discrimination.

3 Etude expérimentale

3.1 But de l'étude

Nous cherchons ici à comparer différentes méthodes visant à réduire la dimensionnalité du problème, que ce soit par le biais de la sélection de variables ou par celui de la projection dans un espace de dimension restreinte. Les critères que nous nous sommes donnés pour cette comparaison sont la qualité de la classification (taux de bonne classification et stabilité), la vitesse d'exécution ainsi que la faisabilité en ligne.

3.2 Démarche expérimentale

3.2.1 Données

Les données utilisées sont celles de K. Van Laerhoven dans l'article (Laerhoven, 2001). Nous disposons des signaux de deux accéléromètres placés au niveau des genoux. Les cinq classes présentes sont *assis, debout, marche, cours, à vélo*. Les données

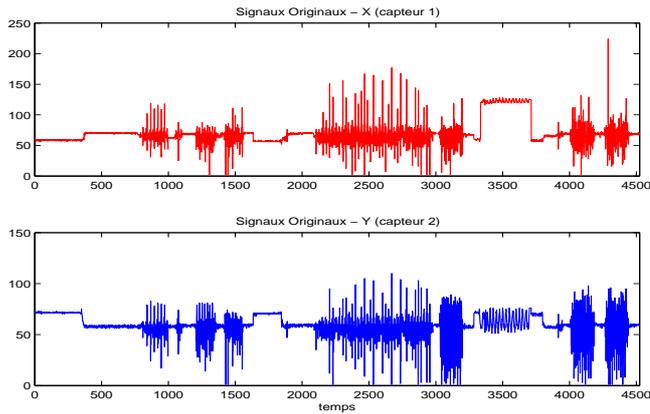


FIG. 3 – *Signaux de départ. Les différents déplacements apparaissent ici : assis, debout, marche, cours et fait du vélo*

sont partiellement étiquetées. Les signaux, composé chacun de 4524 points, sont représentés sur la figure 3.

3.2.2 Pré-traitement

Les caractéristiques utilisées sont les deux signaux de départ et pour chaque signal, la moyenne, la variance, le minimum et le maximum sur une fenêtre mobile, ainsi que les transformées de Fourier ; une autre caractéristique est la différence entre les deux signaux. Par ailleurs nous avons ajouté une variable de type bruit blanc pour vérifier la validité des méthodes de sélection de variables. Nous partons donc sur une base de 14 variables. Ce choix de caractéristiques est arbitraire et le nombre de variables aurait pu être augmenté et diversifié, mais ce n'était pas là le point clef de cette étude. Par conséquent nous nous sommes limités aux caractéristiques évoquées dans l'article (Laerhoven, 2001) et quelques autres.

3.2.3 Réglage des hyper-paramètres

Le nombre de paramètres à régler est assez conséquent. En effet, nous avons la taille de la fenêtre de lissage (ici choisie identique pour toutes les caractéristiques) pour la construction des caractéristiques. Pour la dimensionalité, on trouve selon les méthodes, la taille de la carte de Kohonen, le type de voisinage, les pas d'apprentissage, les paramètres du noyaux pour l'approche globale. Quant aux méthodes de discrimination, on retrouve le nombre de voisins pour la méthode k PPV et encore les paramètres des SVM, soit les paramètres d'ajustement C et μ . Les réglages de tous ces hyper-paramètres ont été fait lors d'une étude préalable -par validation croisée- sur des ensembles de 1250 points. Ces points sont tirés aléatoirement dans chaque classe (250 points par classe).

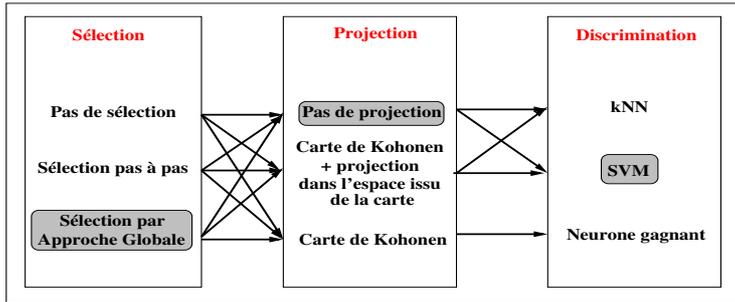


FIG. 4 – Ordonnement des méthodes. Les cases grisées montrent l'enchaînement donnant les meilleurs résultats. Nous retrouvons ici les deux phases de représentation (divisée en sélection et projection) et de discrimination.

3.2.4 Validation

Les méthodes de discrimination (*k*PPV et SVM multi classes *un contre un*) sont utilisées avec un ensemble d'apprentissage et un ensemble de test. L'ensemble d'apprentissage est constitué d'un nombre pré-défini *n* d'éléments de chaque classe (ici, 250 points par classe, soit 1250 points au total). Les tableaux 2 et 3 rapportent les résultats obtenus sur l'ensemble des points étiquetés restants, soit 2500 points. Dans un premier temps, ces éléments sont tirés au hasard dans les classes. Cela induit une certaine corrélation entre les données d'apprentissage et les données de test et donc des performances biaisées à la hausse, mais cela permet un bon aperçu global de la performance des méthodes. Dans un second temps, la base d'apprentissage est constituée des *n* premiers points de chaque classe, de manière à simuler les données que l'on pourrait avoir dans un contexte réel (par exemple on n'aurait pas dès le départ tous les styles de marche que l'on serait amené à rencontrer lors du fonctionnement).

3.2.5 Ordonnement des méthodes

Les combinaisons de méthodes comparées ici sont au nombre de 15. Ces 15 combinaisons sont réparties en trois groupes de sélection de variables (sans sélection, pas à pas et approche globale). Dans chacun de ces groupes, on applique une des trois projections (sans projection, carte de kohonen, espace issu de la carte). Le taux de bonne reconnaissance est calculé directement sur la carte de Kohonen (affectation du point à la classe du neurone gagnant). Dans le cas sans projection ou dans l'espace issu de la carte, on applique les *k*PPV et les SVM multi classes (voir figure 4).

3.3 Résultats

Les variables sélectionnées par chacune des deux méthodes sont reportées dans le tableau 1. Lors de la sélection pas à pas, les variables qui sont gardées sont le maximum mobile du premier capteur, les minimum, maximum, moyenne et variance mobiles du deuxième capteur. On remarque que les deux signaux de départ sont éliminés, ainsi

Variabes	Pas à Pas	Approche Globale
X	0	0.728
Y	0	0.183
bruit blanc	0	0
min X	0	0.0541
max X	1	2.807
moy X	0	0.556
var X	0	2.562
fft X	0	0
min Y	1	2.168
max Y	1	2.702
moy Y	1	0.496
var Y	1	3.496
fft Y	0	0
X-Y	0	0.620

TAB. 1 – Illustration de la sélection des variables. Les résultats sont les coefficients appliqués à chaque variable, suivant la méthode de sélection utilisée. Notons ici que ces valeurs sont un exemple de résultats et dépendent de l'échantillon d'apprentissage.

que le bruit blanc que l'on avait ajouté. De plus, on note que les données issues du premier capteur n'apportent pas une grande quantité d'information. Lors de la sélection par approche globale, le bruit blanc et les transformées de Fourier sont éliminés de la même façon. Bien que ces résultats soient une illustration de ce que peuvent donner les méthodes de sélection, on note la cohérence des résultats pour les variables $max(X)$, $min(Y)$, $max(Y)$ et $var(Y)$, qui apparaissent comme les variables les plus utiles. Si l'on souhaitait avoir une caractéristique supplémentaire à ces quatre prépondérantes, la méthode de sélection pas à pas ajouterait plutôt $moy(Y)$ et l'approche globale $var(X)$.

Le tableau 2 présente les résultats moyens obtenus lors des différentes combinaisons de méthodes. Le pourcentage entre parenthèses représente l'écart type et les résultats sont obtenus sur 50 essais. Les paramètres utilisés pour obtenir ces résultats sont :

- Construction des caractéristiques : taille de la fenêtre = 45
- Cartes de Kohonen : taille 12x12, pas d'apprentissage 0.03, voisinage gaussien
- SVM approche globale : noyau gaussien, $C = 5$, $\sigma = 0.17$
- SVM discrimination : noyau gaussien, $C = 500$, $\sigma = 0.17$
- k PPV : $k = 1$

Le tableau 3 présente les résultats obtenus à partir de la base d'apprentissage composée des 250 premiers points de chaque classe (l'ordre est défini de manière temporelle). Pour les résultats marqués d'une étoile (*) le nombre de points de chaque classe pour la base d'apprentissage est 100 et pour ceux marqués de deux étoiles (**), nous avons pris 50 points par classe). Le choix de la taille de la base d'apprentissage a été fait en fonction de l'étude préalable.

La figure 5 représente la matrice de confusion pour l'application des SVM après sélection de variables par approche globale, sans aucune projection et avec des données d'apprentissage tirées aléatoirement. La matrice de confusion montre ici que les principales erreurs se font entre les classes *marcher* et *courir* et que par ailleurs la classe *fait*

	Sans Projection		SOM	Espace SOM	
	kPPV	SVM	N. Gagnant	kPPV	SVM
SS	97.86% (0.28%)	98.99% (0.27%)	70.80% (1.58%)	87.27% (0.58%)	81.31% (0.58%)
PP	98.84% (0.25%)	96.62% (0.35%)	87.97% (1.43%)	94.45% (1.01%)	85.14% (0.59%)
AG	98.58% (0.6%)	99.34% (0.5%)	76.91% (5.2%)	89.15% (2.7%)	84.36% (2.6%)

TAB. 2 – Résultats avec les données de la base d'apprentissage tirées aléatoirement. SS correspond à « aucune sélection de variables », PP à la sélection pas à pas et AG à la sélection par approche globale.

	Sans Projection	
	kPPV	SVM
Sans sélection	87.07% (*) -	88.89% (*) -
Pas à Pas	83.20% (*) -	84.21% (*) -
Approche Globale	86.78% (**) (0.24%)	88.91% (**) (1.36%)

TAB. 3 – Résultats avec données de la base d'apprentissage ordonnées.

du vélo est systématiquement bien reconnue. On remarque également ici que les classes n'ont pas toutes des tailles équivalentes.

4 Discussion

Notre méthode référence (à savoir le neurone gagnant sur carte de Kohonen, sans sélection de variables - (Laerhoven, 2001)) donne environ 70% de bonne classification, ce qui est comparable aux résultats obtenus par K. Van Laerhoven sans optimisation. Sur les mêmes données dans son article sus-cité, il obtient environ 80% de bonne classification en ajoutant un algorithme des *k*-moyennes - c'est-à-dire en faisant de la discrimination dans l'espace de départ. Ces résultats sont dépassés (87%) en appliquant les *k*PPV dans l'espace issu de la carte de Kohonen. Nos meilleures performances dépassent les 99% avec sélection de variables par approche globale et SVM multi classes. Cette même combinaison de méthodes obtient les meilleurs résultats (de peu) à la fois sur les données corrélées (tirées au hasard) et sur les données ordonnées.

Les résultats que nous obtenons montrent que les performances entre *k*PPV et SVM sont statistiquement équivalentes dans notre cas et qu'il en est de même pour la discrimination sans sélection de variables et après sélection par approche globale. Toutefois, dans l'optique d'utiliser ce système en ligne, l'une des contraintes est la vitesse de calcul, une autre étant la place mémoire. Or le fait de sélectionner les variables réduit la dimension du problème et donc diminue le temps de traitement. De la même façon, les SVM permettent de classer les nouveaux points en ne tenant compte que des vecteurs

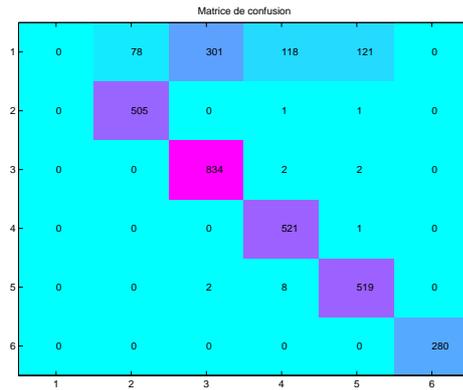


FIG. 5 – Matrice de confusion pour SVM après une sélection de variables par approche globale (données d'apprentissage tirées au hasard). La première ligne correspond aux données non étiquetées et la première colonne représente les données non classées. La case 2*2 indique donc le nombre de points de la classe 1 bien classés et la case 5*4 montre que 8 points de la classe 4 ont été attribués à la classe 3. La classe 1 est la classe assis, la classe 2 est debout, 3 est marche, 4 est cours et 5 est fait du vélo.

supports et par conséquent on gagne ici encore en vitesse d'exécution.

Concernant la sélection de variables, on remarque sur l'ensemble de résultats que l'approche globale augmente la variance des résultats, mais que la sélection pas à pas peut engendrer un phénomène de sur-apprentissage. Il apparaît que les cartes de Kohonen ne sont pas les plus adaptées à ce type de problème, mais que si l'on tient à les utiliser, il vaut mieux utiliser la carte obtenue pour engendrer un espace de projection dans lequel on fera la discrimination. Ici on peut noter le résultat de 94.45% obtenu par la méthode des k PPV après sélection de variables pas à pas et projection dans l'espace issu de la carte de Kohonen, pour lequel nous n'avons pas d'explications a priori. Pour ce qui est des méthodes de discrimination, bien que statistiquement équivalentes, il semble que les SVM multi classes soient légèrement plus performantes que les k PPV. Notons également que les résultats du tableau 3 sont obtenus dans des conditions défavorables (seulement 250 points d'apprentissage pour l'approche globale au lieu de 1250 dans le tableau 2) et qu'ils sont tout de même qualitativement comparables à ceux de ce tableau 2.

5 Conclusion et perspectives

A l'issue de cette étude hors ligne nous avons acquis la conviction que les SVM plus la sélection de variables par approche globale permet d'obtenir des résultats significativement meilleurs (99%) que ceux de la littérature (80%). Les résultats obtenus montrent que notre approche résout le problème de manière très satisfaisante. La qualité de la méthode proposée nous permet d'envisager son adaptation à l'approche en ligne avec confiance. Avant d'en arriver là, des études restent à faire sur la composition de la base

d'apprentissage (on pourra par exemple travailler sur les différents segments de classes dans le signal et non sur les n premières valeurs de chacune). En outre, la mise en œuvre de ces méthodes sur des signaux enregistrés sur différents sujets permettrait d'avoir une vision beaucoup plus proche de la réalité et de cerner des problèmes que nous ne pouvons pas traiter ici. Enfin, pour la poursuite de cette étude, il nous semble crucial de prendre en compte l'aspect temporel des données et de penser à un système global de supervision de l'apprentissage qui prendrait en charge le réglage des hyper-paramètres et assurerait la stabilité de l'ensemble des processus.

6 Remerciements

Nous remercions Kristof Van Laerhoven pour nous avoir permis d'utiliser ses données et Yves Grandvalet pour le code Matlab dont nous nous servons.

Références

- BELLMAN R. E. (1961). Adaptive control processes. *Princeton Univ. Press*.
- CANU S., MARY X. & RAKOTOMAMONJY A. (2003). *Advances in Learning Theory : Methods, Models and Applications NATO Science Series III : Computer and Systems Sciences*, chapter Functional learning through kernel. IOS Press, Amsterdam.
- CLARKSON B., MASE K. & PENTLAND A. (2000). Recognizing user context via wearable sensors. In *ISWC*, p. 69–76.
- COTTRELL M., IBBOU S., LETRÉMY P. & ROUSSET P. (2003). Cartes auto-organisées de kohonen pour l'analyse exploratoire de données et la visualisation. *Journal de la Société Française de Statistiques - A paraître*.
- DUDA R., HART P. & STORK D. (2001). *Pattern Classification*. Wiley Interscience - 2e édition.
- GRANDVALET Y. & CANU S. (2002). Adaptive scaling for feature selection in SVMs. *NIPS*.
- KOHONEN T. (1989). *Self-Organization and Associative Memory*. Springer Verlag, berlin, third edition.
- LAERHOVEN K. V. (2001). Combining the kohonen self-organizing map and k-means for on-line classification of sensordata. *Artificial Neural Networks*, vol 2130, pp. 464–470.
- LERAY P. & GALLINARI P. (1999). Feature selection with neural networks. *Behaviormetrika*, 26(1), 145–166.
- NAJAFI B., AMINIAN K., LOEW F., BLANC Y. & ROBERT P. (2002). Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly. *IEEE Transactions on Biomedical Engineering*, 49(8), 843–851.
- STAUFFER C. & GRIMSON W. (2000). Learning patterns of activity using real-time tracking. *PAMI*, 22(8), 747–757.
- VAPNIK V. (1998). *Statistical Learning Theory*. Wiley.
- WESTON J. & WATKINS C. (1999). Support vector machines for multiclass pattern recognition. In *Proceedings of the Seventh European Symposium On Artificial Neural Networks*.
- ZHANG J., WARD W. & PELLOM B. (2002). Phone based voice activity detection using online bayesian adaptation with conjugate normal distributions. *International Conference on Acoustics, Speech and Signal Processing (ICASSP), Orlando, Florida*.