
Perception d'états affectifs et apprentissage

Gaëlle Loosli, Sang-Goog Lee^(*), Vincent Guigue, Stéphane Canu, Alain Rakotomamonjy

*PSI, CNRS FRE2645, INSA de Rouen, FRANCE
gaelle.loosli@insa-rouen.fr*

() Interaction Lab./Context Awareness TG,
Samsung Advanced Institute of Technology, Korea*

RÉSUMÉ. Le problème abordé dans ce travail est celui de la reconnaissance des états affectifs d'un utilisateur à partir de mesures physiques (accéléromètres) et physiologiques (ECG, EMG...) issues de capteurs portés. Etant donné la nature complexe de la relation entre les signaux dont nous disposons et les états affectifs à reconnaître, nous proposons d'utiliser une méthode d'apprentissage statistique. Nous commençons par discuter des états de l'art dans les domaines de l'apprentissage statistique et de la reconnaissance d'émotions. Nous présentons ensuite un cadre permettant de comparer les différents algorithmes d'apprentissages et leurs conditions d'utilisation. A l'issue de cette pré-étude, nous proposons une architecture globale d'un système embarqué de reconnaissance en temps réel fondé sur la notion de détection de changement d'état. Nous démontrons enfin l'intérêt de notre approche sur deux exemples réels.

ABSTRACT. This article deals with the problem of affective states recognition from physical and physiological wearable sensors. Given the complex nature of the relationship between available signals and affective states to be detected we propose to use a statistical learning method. We begin with a discussion about the state of the art in the field of statistical learning algorithms and their application to affective states recognition. Then a framework is presented to compare different learning algorithms and methodologies. Using the results of this pre-study, a global architecture is proposed for a real time embedded recognition system. Instead of directly recognizing the affective states we propose to begin with detecting abrupt changes in the incoming signal to segment it first and label each segment afterwards. The interest of the proposed method is demonstrated on two real affective state recognition tasks.

MOTS-CLÉS : Reconnaissance d'émotions, états affectifs, apprentissage, machines à vecteurs supports, SVM, contexte, SVM à une classe, détection de ruptures

KEYWORDS: Emotion recognition, affective computing, affective states, learning systems, learning algorithms, support vector machines, SVM, 1-class SVM, context retrieval, sequential detection, novelty detection

1. Introduction

Le détecteur de mensonges pourrait être considéré comme un bon exemple, sinon le plus ancien, de machine censée percevoir les émotions humaines. Preuve scientifique pour certains, gadget non fiable pour d'autres, la controverse qui l'entoure illustre une partie des enjeux et des problèmes liés à la conception d'un système capable de reconnaître une émotion. Car pour qu'un agent puisse interagir émotionnellement avec son environnement, il doit être capable de percevoir l'état affectif de son interlocuteur. Le but de nos recherches est de concevoir un système qui permette cette reconnaissance ; il pose trois questions préalables :

- les entrées - de quel type de capteurs et de mesures va-t'on disposer ?
- les sorties - comment définir l'état affectif d'une personne ?
- comment établir la relation entre les mesures et les émotions ?

Établir cette relation est difficile : il n'existe pas de modèle explicite permettant d'expliquer les émotions à partir de mesures. Cette relation est non linéaire, fortement bruitée et sujette à une grande variabilité selon les situations et les individus concernés. Enfin il est difficile de définir objectivement la notion d'émotion ; d'où notre point de départ selon lequel la perception d'émotion requiert l'utilisation de données propres à la situation et d'algorithmes d'apprentissage de type « boîte noire ». En effet, les progrès récents dans le domaine de la théorie de l'apprentissage statistique permettent d'envisager la conception de méthodes adaptatives, génériques, robustes et efficaces pour reconnaître automatiquement l'état affectif d'une personne à partir d'un ensemble de mesures.

Le but du travail présenté ici est de montrer ce que les techniques d'apprentissage statistique peuvent apporter au problème de la perception d'émotions.

Dans le domaine de l'apprentissage statistique, il est commun de distinguer le problème d'apprentissage dit « supervisé » de l'apprentissage « non supervisé ». Dans le premier cas les exemples sont étiquetés (ici des couples mesures-émotions), alors que dans le second, ils ne le sont pas (mesures seules). Il faut alors d'abord rassembler les observations en groupes homogènes (ou classes) pour pouvoir ensuite les étiqueter. Selon que l'on considère la relation entre mesures et émotions comme stable et fixée à l'avance ou sujette à d'imprévisibles dérives, l'une ou l'autre de ces approches sera utilisée. Tout dépend des « sorties » de l'algorithme d'apprentissage.

La nature de ces « sorties » fait l'objet d'une controverse entre ceux qui pensent que les émotions peuvent être détectées automatiquement [PIC 99, HUD 03] et ceux qui ne le pensent pas, considérant comme H. Bergson [BER 88] (p 20) qu'il faut voir « dans l'état affectif autre chose que l'expression consciente d'un ébranlement organique ». Nous allons, dans ce travail, nous en tenir à ces « ébranlements organiques » car ils sont objectivement observables en les appelant, peut être par abus de langage, « états affectifs ». Si le langage courant utilise indifféremment les termes « état émotionnel » et « état affectif », nous utiliserons plutôt le second en lui donnant un sens plus large et plus proche de ce que nous cherchons : affectif s'utilise *en parlant des réactions qui affectent l'être humain*. Ainsi l'état affectif peut être d'ordre émotionnel (stressé, calme...) ou autre (marche, est assis...). De notre point de vue, la perception de ces deux types d'états est conceptuellement de même nature. De plus, puisque notre objectif est d'intégrer notre système de perception d'états affectifs dans une ap-

plication pour l'améliorer, nous proposons de définir ces états affectifs en relation avec l'application considérée et plus précisément, à terme, de construire cette relation par apprentissage. Ainsi, les états affectifs que nous considérons ne sont pas subjectifs comme des sentiments personnels ou tout autre intimité réveillée. En revanche ils doivent être observables et « utiles » à l'application considérée. Il s'agit dans ce travail d'une suite d'états d'une personne pouvant être détectés (observabilité) et étiquetés dans un cadre précis.

Notre travail vise à trouver comment apprendre le lien entre ce qui affecte un utilisateur (tout en étant observable) et ce qui est utile pour une application (un programme ou une machine).

Cette définition de la nature des « états affectifs » à détecter conditionne, dans une certaine mesure, le choix des « entrées » considérées dans notre étude. En effet, les capteurs utilisés se répartissent en deux catégories selon qu'ils sont extérieurs à l'utilisateur comme une caméra ou un microphone, ou portés par lui comme une centrale d'acquisition physiologique (ECG, EMG...). Dans le premier cas, le dispositif d'acquisition est indépendant de l'utilisateur et peut être utilisé en toute occasion alors que l'utilisation de capteurs du deuxième type exige l'instrumentation de l'utilisateur. En revanche, les capteurs extérieurs nous permettent d'accéder de manière indirecte et plus subjective aux émotions, alors que les mesures physiologiques contiennent une part d'information objective. On sait par exemple qu'une sensation de bien-être est liée à un ralentissement du rythme cardiaque et que la colère augmente la température corporelle. C'est la raison principale qui nous a fait choisir, dans un premier temps, de ne travailler qu'avec des capteurs portés par l'utilisateur. Ce n'est toutefois pas la seule. Dans certains cas, l'utilisateur est déjà équipé (ou pourrait l'être dans un futur proche) d'un dispositif d'acquisition intégré dans ses vêtements. C'est le cas par exemple des pilotes, des secouristes, dans le domaine de la surveillance médicale ou des « cobayes » qui acceptent volontairement l'instrumentation (comme un chef d'orchestre ou un sportif souhaitant mieux se connaître). Enfin l'étude de la vision et de l'audio relève de problématiques spécifiques qui ne sont pas les nôtres ici.

Les objectifs de notre étude peuvent maintenant être précisés. Il s'agit de concevoir un système permettant d'apprendre la relation entre des signaux mesurés sur un utilisateur (physiques et physiologiques) et la suite de ses états affectifs. Cela nous amène à considérer les points suivants :

- démontrer l'utilité des techniques d'apprentissage pour ce problème et préciser le cadre de leur utilisation,
- définir quels sont les états d'une personne accessibles (observables à partir de signaux physiques et physiologiques),
- concevoir un système robuste (indépendant de l'utilisation et de l'utilisateur) permettant l'apprentissage de ces états,
- préciser quelle est l'utilité des états ainsi détectés, c'est-à-dire leur interaction avec l'application considérée.

Pour atteindre ces objectifs nous avons suivi deux stratégies différentes. D'une part nous avons recherché dans la littérature des études analogues, récupéré des données déjà traitées pour comparer différents algorithmes d'apprentissage. Ces cas ont pu être considérés comme des problèmes d'apprentissage supervisé. D'autre part, nous nous sommes dotés de notre propre système d'acquisition avec lequel nous avons pu déve-

opper et tester nos idées sur la question concernant notamment l'utilisation de l'apprentissage non supervisé. Notre approche du problème est la suivante : d'abord se donner un cadre et définir ce que nous entendons par reconnaissance d'un état affectif. Ensuite, suivant l'analyse proposée dans la huitième section de [CRO 05] pour aller du signal jusqu'à l'émotion, utiliser des capteurs pertinents, segmenter le signal, puis dans chaque séquence homogène, extraire des caractéristiques discriminantes permettant de l'étiqueter.

L'article est organisé de la manière suivante : après une revue de l'existant en matière d'apprentissage et de détection d'états affectifs, nous montrons comment traiter la détection d'états affectifs comme un problème d'apprentissage supervisé. Nous nous intéressons ensuite au cas de l'apprentissage non-supervisé (et nous le verrons au cas semi-supervisé) en décrivant une architecture originale de perception des états affectifs et un dispositif expérimental permettant d'illustrer l'intérêt de la méthode proposée. Nous concluons par une discussion des résultats et leur mise en perspective.

2. État de l'art et travaux connexes

Afin de positionner notre approche de l'apprentissage des états affectifs à partir de mesures physiques et physiologiques, nous allons rappeler l'état de l'art dans le domaine de l'apprentissage puis dans celui de la perception d'états affectifs. Nous allons voir que les enjeux croisés de ces deux domaines révèlent des difficultés fondamentales qui sont autant de blocages de la théorie de l'apprentissage et donc de ses applications [PIC 04].

2.1. Mise en perspective succincte de la théorie statistique de l'apprentissage

Le cadre de l'apprentissage statistique qui nous intéresse ici est celui de la reconnaissance des formes dans lequel on cherche à estimer, à partir d'un échantillon, une dépendance fonctionnelle entre des variables explicatives (observées) $\mathbf{x} \in \mathbb{R}^d$ et une variable qualitative à expliquer $y \in \{1, \dots, C\}$. Il est classique de distinguer les méthodes de reconnaissance des formes dites « paramétriques » des méthodes « non paramétriques » [HAS 01, DUD 01]. Dans le cadre paramétrique, la problématique consiste à faire des hypothèses sur la nature du modèle, à en estimer les paramètres puis à vérifier ces hypothèses. Dans le cas non paramétrique, qui est aussi celui de l'apprentissage statistique, on ne fait pas (du moins explicitement) référence à un modèle. On recherche des méthodes à caractère « universel » qui fonctionnent, sinon pour tout, du moins sur une large classe de problèmes. On parle aussi de modèle de type « boîte noire ». La principale qualité d'une méthode d'apprentissage est sa faculté de généralisation : son aptitude à associer en moyenne à une nouvelle observation, sinon la bonne étiquette du moins une décision raisonnable et ce, quel que soit le problème initial à résoudre. Les méthodes d'apprentissage qui nous intéressent possèdent la capacité de pouvoir approcher correctement (au sens d'un coût) n'importe quel type de fonction de décision (c'est leur caractère d'universalité). Si leur ensemble d'hypothèses doit être assez vaste pour leur permettre de tout apprendre (avec suffisamment d'exemples) c'est qu'elles peuvent tout aussi bien apprendre n'importe quoi (surapprentissage). Il est donc indispensable de disposer d'un mécanisme permettant d'ajuster la capacité

du modèle à la complexité du problème. Les méthodes d'apprentissage statistiques se distinguent par :

- la nature de la dépendance à estimer,
- la nature des données disponibles,
- l'ensemble des hypothèses considérées,
- le critère qualifiant la qualité de l'apprentissage (un coût),
- le mécanisme de contrôle de la complexité définit comme un *a priori* sur la nature de la solution.

Afin d'illustrer certains débats actuels relatif à l'utilisation des techniques d'apprentissage, nous allons présenter une taxinomie de quelques méthodes récentes ayant démontré leur intérêt dans des applications (notamment la reconnaissance d'états affectifs).

Nous avons déjà souligné en introduction la distinction fondamentale entre apprentissage supervisé et apprentissage non supervisé. Cette distinction traite des deux premiers points de la liste. Pour schématiser, on peut dire que dans le premier cas on dispose d'un échantillon $(\mathbf{x}_i, y_i), i = 1, n$, indépendant et identiquement distribué (i.i.d.) selon une loi inconnue $\mathbb{P}(\mathbf{x}, y)$, à partir duquel on cherche à inférer une règle. Celle-ci vise à décider à partir d'une observation \mathbf{x} quelle classe lui associer. Dans le second cas, on ne dispose que d'un échantillon d'observations $\mathbf{x}_i, i = 1, n$ (sans étiquettes y_i), toujours i.i.d. selon une loi inconnue $\mathbb{P}(\mathbf{x})$, que l'on cherche à répartir en un certain nombre de groupes à étiqueter ensuite. Lorsque l'on cherche à effectuer cet étiquetage des groupes automatiquement à partir d'autres données, on peut parler d'une certaine manière d'apprentissage « semi supervisé » bien que *stricto sensu*, l'apprentissage semi supervisé traite du cas où l'on dispose de peu de données étiquetées et d'un grand nombre d'observations non étiquetées. Il existe entre apprentissage supervisé et non supervisé un autre mode : l'apprentissage par renforcement. Il traite du cas où l'on cherche à contrôler un système (par exemple apprendre à jouer au *backgammon*). Les observations sont des suites de données temporellement liées $\mathbf{x}_i = \{\mathbf{x}_1, \dots, \mathbf{x}_{t_i}\}$ (par exemple une partie de *backgammon*) et les étiquettes une information indiquant la qualité globale de la séquence (par exemple si la partie a été gagnée ou perdue).

Théorie statistique de l'apprentissage non supervisé

Pour l'apprentissage non supervisé, l'objectif est de résumer l'information disponible. Deux approches sont possibles : soit regrouper les individus analogues (c'est une forme d'estimation de la densité de probabilité sous-jacente) ou soit représenter les variables par un espace de plus petite dimension. Parmi les méthodes permettant de regrouper les individus (on parle alors de création de classes, de coalescence ou de *clustering*) on trouve l'algorithme des k moyennes, les nuées dynamiques, les méthodes associées aux modèles de mélanges (et à l'algorithme EM ou ses variantes) et les méthodes à « noyaux » (dont l'algorithme des *machines à vecteur support* ou SVM pour une seule classe [SCH 00]). Dans ce cadre, un noyau est une fonction de deux variables mesurant d'une certaine manière une forme de leur affinité, leur corrélation ou de leur proximité. Un exemple typique est le noyau gaussien $k_b(\mathbf{u}, \mathbf{v}) = \exp(-\|\mathbf{u} - \mathbf{v}\|^2/b)$ où b est un (hyper)paramètre à déterminer. En utilisant les noyaux, l'apprentissage peut être vu comme un problème de sélection des

individus pertinents pour le problème considéré. Cette approche a donné lieu au développement de méthodes théoriquement justifiées, pratiques et efficaces [SCH 01]. Les méthodes de représentation des variables se subdivisent elles-mêmes en deux. On peut décrire un espace par un ensemble de points topologiquement liés (sur une grille par exemple) : c'est le cas des cartes auto organisatrices ou cartes de Kohonen. Une autre manière de faire consiste à créer de nouvelles variables sur lesquelles on projette les données disponibles : il s'agit des méthodes factorielles comme l'analyse en composantes principales ou indépendantes et leur généralisation au cas non linéaire notamment par l'utilisation de noyaux (*kernel PCA – ICA*)[SCH 01].

Théorie statistique de l'apprentissage supervisé

En ce qui concerne l'apprentissage supervisé, une première distinction est à faire entre les méthodes estimant des probabilités pour en déduire la fonction de décision et les méthodes estimant directement la fonction de décision sans passer, du moins explicitement, par les probabilités.

Parmi les approches probabilistes, les plus utilisées se distinguent suivant la nature du modèle sous-jacent. La méthode de Bayes naïve permet d'estimer des probabilités discrètes alors que les modèles de mélange associés à l'algorithme EM sont utilisés pour l'estimation de densités et de lois mixtes. Les méthodes à noyaux ont aussi été utilisées dans le cadre bayésien sous différentes formes : processus gaussiens, *relevance vector machines* ou *Bayes point machines* [HER 02]. Lorsque le modèle est plus complexe, la question centrale est celle de l'indépendance conditionnelle des variables. Le formalisme des modèles graphiques a été développé pour représenter ce type de relation. Voir par exemple [JOR 04] pour une présentation détaillée de ces modèles et des algorithmes associés pour l'estimation des probabilités. En particulier si l'on s'intéresse à la prise en compte des contraintes temporelles, les différents modèles graphiques employés sont les réseaux bayésiens dynamiques (RBD) qui sont des graphes orientés, les modèles de Markov caché (MMC) qui en plus vérifient l'hypothèse de Markov, les modèles de Markov d'entropie maximum, et les champs aléatoires conditionnels (CAC ou CRF en anglais). Les approches du type CRF ont été introduites récemment pour résoudre le problème « étiquette-biais » lié à l'utilisation de modèles markoviens [WAL 02]. Le problème typique partiellement résolu grâce à l'utilisation des modèles de Markov est celui du traitement de la parole. Soulignons aussi, pour prendre en compte le temps, l'utilisation du filtre de Kalman là encore généralisé avec des noyaux [RAL 05]. Dans ce type de modèle, l'apprentissage est réalisé par l'estimation de probabilités et aussi parfois par l'évolution du modèle (ajout ou suppression d'un état). Il peut être argumenté qu'il ne s'agit pas là de méthodes d'apprentissage à proprement parler mais plutôt de méthodes permettant la modélisation d'un phénomène et donc l'introduction de connaissance *a priori* dans la solution.

Une manière de présenter les méthodes d'apprentissage supervisé dites « directes » consiste à les classer suivant la nature de leur ensemble d'hypothèses. Conceptuellement, cet ensemble d'hypothèses peut être vu comme l'ensemble de combinaisons linéaires de fonctions de référence [HAS 01]. On distingue trois catégories selon la nature des fonctions de références. Soit elles sont fixées indépendamment de l'échantillon en prenant une base (d'ondelettes par exemple). Soit elles ne dépendent que des observations \mathbf{x}_i et pas des étiquettes. C'est le cas des méthodes à base de noyaux et notamment des séparateurs à vaste marge (SVM), des k plus proches voisins, des arbres de décisions, des méthodes additives et celles d'agrégation comme le *boosting*. Enfin

les fonctions de références peuvent dépendre à la fois les observations et des étiquettes comme pour les réseaux de neurones de type « perceptron multicouche » ou « fonctions de base radiales ». Les méthodes du deuxième type conduisent à des algorithmes d'apprentissage rapides et souvent bien posés alors que les réseaux de neurones sont connus pour être associés à des algorithmes d'apprentissage au comportement chaotique. En revanche ils donnent souvent des modèles plus concis que les algorithmes de la deuxième catégorie. Une des problématiques de recherche dans le domaine de l'apprentissage est donc de trouver des modèles parcimonieux facilement identifiables. C'est le cas des SVM, ce qui explique en partie leur succès actuel. Une autre explication est leur efficacité dans le traitement de problèmes concrets comme celui de la reconnaissance de caractères manuscrits [LEC 98]. Sur ce problème, les SVM donnent de très bons résultats, et surtout offrent un compromis – performance/facilité de mise en œuvre – très intéressant.

Le débat reste d'actualité entre les tenants des approches probabilistes et ceux qui prônent les méthodes directes. Si les applications ont dicté leur choix (probabiliste pour la reconnaissance de la parole et directe pour la reconnaissance de l'écriture) au niveau théorique deux conceptions s'opposent. D'un côté on défend les modèles probabilistes comme la manière convenable d'introduire les a priori sur la nature de la solution pour obtenir des modèles concis et de l'autre on met en avant le fait que l'estimation de probabilités est un problème plus difficile que celui qui nous intéresse *in fine* : l'estimation d'une fonction de décision [VAP 98].

Au-delà de cette question, il existe bien d'autres manières de distinguer les algorithmes d'apprentissage. A chaque ensemble d'hypothèses peuvent être associés différents critères qualifiant la qualité de l'apprentissage et différentes manières de minimiser ces critères. Parmi elles, une autre distinction peut être faite suivant que l'apprentissage est effectué en ligne d'une manière incrémentale ou hors ligne, une fois pour toute avant sa mise en œuvre (c'est le mode *batch*). L'aspect dynamique lié à l'apprentissage incrémental pose de nombreux problèmes difficiles loin d'être résolus aujourd'hui. En effet, à partir du moment où un modèle appris évolue dans le temps se pose le dilemme « stabilité plasticité » : comment garantir la stabilité du modèle tout en l'autorisant à se modifier pour suivre les évolutions imprévisibles d'un système. Un algorithme évoluant dans le temps doit être autonome et disposer d'une sorte d'auto calibration, un mécanisme de réglage de sa « capacité ». Par ailleurs un algorithme en ligne se doit d'être capable d'« oublier » des exemples pour rester réalisable en taille et en temps dans une situation réelle sans pour autant tout oublier. Il se pose alors le problème de la sélection des exemples pertinents ou qui peuvent le devenir. Enfin, pour des raisons d'efficacité, un algorithme en ligne est beaucoup plus pénalisé par le bruit et les exemples mal étiquetés que s'il est entraîné hors ligne. Cela incite à réfléchir différemment sur la façon de sélectionner les exemples pertinents en s'inspirant des méthodes d'*active learning* [BOR 05].

2.2. Apprentissage pour la perception d'états affectifs

Si l'idée d'utiliser des techniques de reconnaissance des formes pour identifier les états affectifs d'une personne à partir de signaux physiques ou physiologiques est ancienne [SCH 87, CAC 90] (voir [LIS 04] pour une revue détaillée), il faut attendre les travaux relativement récents de R. Picard et son équipe [PIC 01] pour voir les premiers

résultats significatifs dans le domaine. Cette attente est symptomatique de réelles difficultés (« *the manifold of related theoretical and practical open problems* » [PAN 03]) à commencer par celle de la définition de la variable à reconnaître : qu'entend-on par « état affectif » ?

L'utilisation d'une technique de reconnaissance des formes nécessite la connaissance a priori d'un certain nombre d'exemples étiquetés. Il faut se donner la liste des émotions à reconnaître et pour chaque émotion un certain nombre d'exemples et donc définir les états affectifs auxquels on s'adresse. On constate trois manières différentes de construire ces exemples étiquetés. Soit ils sont définis avant l'expérience, soit ils sont provoqués pendant l'expérience ou bien ils sont retrouvés après l'expérience. Pour définir les émotions avant l'expérience, certains ont recours à des acteurs professionnels qui jouent les émotions [BAN 96, PIC 01]. Si cette technique se justifie lorsqu'il s'agit de reconnaître des émotions dans la voix ou sur un visage, elle est, pour le moins, sujette à caution en ce qui concerne la physiologie.

D'autres ont préféré contrôler les expériences en provoquant les états affectifs sur le sujet. De nombreuses techniques ont été utilisées. Des modifications du comportement de l'ordinateur avec lequel un sujet interagit ont provoquées la frustration par des blocages intempestifs [KLE 02] ou par la perte d'une saisie fastidieuse [QI 02]. Dans [NAS 03] ce sont des films et des questions de mathématiques difficiles qui sont utilisés pour provoquer sept différentes émotions alors que dans [RAN 05] c'est l'anxiété des sujets qui est provoquée par des tâches de difficulté croissante. Pour s'assurer de l'effet escompté [KIM 04b] ont eu recours à des événements multiples combinés (son, lumière, cognitif). Dans [BID 03] le comportement de haut niveau à reconnaître est induit par la question à laquelle cet utilisateur doit répondre. D'autres ont utilisées des photos dont la vue est supposé déclencher certaines émotions [HAA 04]. Dans l'étude menée [LI 05] c'est la fatigue qui est l'état affectif provoqué par privation de sommeil (la même expérience étant reproduite au réveil et après 25 heures d'éveil). D'une manière générale, cette façon de procéder conduit à des données et des états dépendants de l'application ciblée. Ce qui est mesuré c'est la réaction à un stimulus et non pas véritablement un état affectif.

Une autre manière de construire des exemples étiquetés est de le demander au sujet lui même avant, pendant ou après l'expérience. C'est par exemple le sujet qui choisit des chansons correspondant à quatre différents types d'émotions avant l'expérience [KIM 04c]. Certains [HEA 05] ont utilisé un questionnaire associé à un score établi a posteriori après visionnage de la vidéo pour mesurer le stress d'un conducteur alors que d'autres proposent des interfaces spécialisées pour saisir son humeur [ZIM 03]. Mais cette manière de faire n'est pas non plus une panacée car l'intervention du sujet contient en soit des risques de subjectivité. Si l'on veut absolument revenir à des étiquettes objectives, on peut s'en tenir aux comportements objectifs d'une personne [LAE 01b]. Mais on dispose là que d'une information non « émotive ».

Une démarche radicalement différente consiste à non plus s'intéresser aux états affectifs eux mêmes, mais à ce concentrer sur les effets de ce type d'outils, par exemple par une mesure objective de l'amélioration de système prenant en compte les états affectifs [KLE 02].

Nous allons retrouver dans notre étude ce triple étiquetage : les états affectifs provoqués par le contexte (les différents événements d'un jeu), les états affectifs identifiés *a posteriori* par le sujet (après visionnage de l'expérience) et les états détectés par

notre système. Reste à savoir quel est le bon étiquetage ? Nous pensons qu'à terme, ce sont les effets sur la qualité globale de l'application qui donneront la mesure de la pertinence des choix effectués.

Une fois le cadre expérimental précisé avec la définition des états affectifs à reconnaître, tout est loin d'être résolu car le choix d'une méthode d'apprentissage exige la réponse à d'autres questions, analogues à celles qui se posent dans les domaines connexes de la modélisation de l'utilisateur [WEB 01] ou de la reconnaissance d'émotions dans la parole et sur les visages [PAN 03]. Une de ces questions est liée à la prise en compte du caractère dynamique du phénomène. En effet la modélisation des suites d'états affectifs est déjà complexe en soit. Une **première génération** de systèmes fait l'impasse sur cet aspect et considère la reconnaissance d'états affectifs sans prise en compte du temps. Dans ce cas, la reconnaissance d'un état affectif est vu comme un problème de classification multi-classes avec un nombre fixe et connu de classes. Pratiquement toutes les méthodes d'apprentissage supervisé « directes » que nous avons mentionnées ont été utilisées dans différentes études. Il s'agit de l'analyse discriminante [ARK 99, PIC 01], l'utilisation de modèles de mélanges [QI 02] avec l'algorithme EP (*expectation propagation*) et leur comparaison avec les séparateurs à vaste marge (SVM) qui sont aussi utilisés dans [KIM 04b]. On retrouve aussi les réseaux de neurones type perceptron multicouche [NAS 03, HAA 04] avec des comparaisons avec les k plus proches voisins [NAS 03]. Généralement les auteurs ne concluent pas sur la supériorité d'une méthode en terme de « taux de reconnaissance ». En revanche, d'autres arguments sont mis en avant pour préconiser l'une ou l'autre de ces méthodes et notamment la simplicité en terme de mise en œuvre. [WEB 01] ont déjà souligné qu'il s'agit là d'une difficulté fondamentale liée à l'utilisation des méthodes d'apprentissage. Elles doivent être rapides, auto configurables pour pouvoir d'adapter et peu gourmandes en ressources. C'est l'argument mis en avant par [QI 02] pour rejeter les SVM qui sont jugées lentes, nécessitant des ressources importantes et ne permettant pas l'apprentissage en ligne (elle est qualifiée de méthode globale). Ce qui pouvait être vrai à l'époque de l'écriture de cet article ne l'est plus aujourd'hui car nous disposons maintenant d'algorithmes SVM rapides, adaptatifs, relativement peu gourmand en terme d'espace mémoire [LOO 04] et utilisés avec succès par exemple pour la reconnaissance des expressions du visage en temps réel [LIT 04].

De part sa nature contrôlée, ce genre de dispositif doit permettre de répondre à certaines questions préalables à la reconnaissance d'états affectifs :

- y a t'il des formes à apprendre ?
- quelle méthode employer ?
- comment choisir et de la sélectionner des caractéristiques pertinentes ?

Si l'on peut répondre par l'affirmative à la première question [CHR 04], les deux autres restent ouvertes.

Sous l'influence du domaine de la modélisation de l'utilisateur où l'on doit considérer des suites d'actions, la **deuxième génération** d'architecture d'apprentissage regroupe les méthodes prenant en compte le temps pour effectuer l'apprentissage dynamique des états affectifs à partir de signaux biologiques. [SCH 02] ont été les premiers à proposer l'utilisation d'un modèle de Markov caché puis à suggérer celle de CRF. Par analogie avec le traitement de la parole en général et en particulier la reconnaissance d'émotion à partir de la parole ou de séquences vidéo, le problème est vu comme une

tâche d'étiquetage de séquences. Les réseaux bayésiens dynamiques ont été préférés par [LI 05] dans le cadre d'une modélisation globale des émotions, de l'utilisateur et de ses actions (en relation avec l'application). Ce type d'approche, permettant la prise en compte du temps, nous semble a priori mieux adapté à la modélisation de la succession des états affectifs. Mais une première critique peut être adressée aux modèles probabilistes. S'il est vrai qu'ils permettent de prendre en compte les connaissances a priori dont on dispose sur la nature des dépendances entre les variables considérées, ce que l'on cherche à apprendre c'est une fonction de décision et non pas une loi de probabilité (cf la discussion du paragraphe précédent sur les différentes méthodes d'apprentissage). La seconde critique est liée au problème de variabilité intra et inter utilisateur. Si pour un utilisateur donné, dans une condition d'utilisation donnée ce type d'approche est très efficace, les performances se dégradent notablement, comme dans le cas de la reconnaissance de la parole, lorsque l'on modifie les conditions d'utilisation et que l'on change de sujet. En effet, ils font l'hypothèse que la distribution de probabilité du signal sous-jacent est stationnaire, ce qui n'est pas le cas lorsque l'on considère différents utilisateurs. Il y a là une autre question difficile posée par notre application au domaine de l'apprentissage. Comment apprendre sur certains individus (avec une loi de probabilité donnée) et généraliser sur d'autres (la loi de probabilité ayant changé).

Une réponse à cette question a été proposée par l'utilisation de cartes de Kohonen pour suivre l'évolution du comportement d'une personne [LAE 01b]. L'originalité de cette approche c'est l'utilisation d'une technique d'apprentissage non supervisée pour représenter l'information utile. Mais dans cette application, le caractère dynamique du problème n'est pas vraiment pris en compte. L'un des intérêts de cette méthode est de s'adapter, pour un utilisateur donné, aux changements de contextes (dérive de concept) et de permettre le passage d'un utilisateur à un autre. Un autre avantage lié à l'utilisation d'un modèle non supervisé, est que l'on n'est plus obligé de spécifier a priori tous les états que l'on souhaite détecter. Nous pouvons définir ceux que nous connaissons et laisser au système la possibilité d'en créer de nouveaux. Se pose alors le problème de l'étiquetage de ces nouveaux états. En terme d'apprentissage, on parle de système incrémental. L'étiquetage des états est un problème lui-même complexe pour lequel nous proposerons une solution semi-supervisée dans la présentation de notre architecture. Toutefois cette proposition reste à l'heure actuelle une question à approfondir.

Pour faire face à ces problèmes, nous proposons une **troisième génération** qui n'existe pas encore qui devrait être capable de reconnaître en temps réel une suite états affectifs indépendamment du sujet mais en adéquation et interaction avec les besoins de l'application considérée. Ses spécifications seraient les suivantes :

- prise en compte du temps, traitement des non stationnarités du signal, de son caractère fortement bruité et du caractère multimodal des données (variabilité intra utilisateur),
- insensibilité face à la dérive de concept et adaptation au sujet (variabilité inter utilisateur),
- efficacité, robustesse et stabilité de la mise en œuvre,
- définition évolutive des concepts à apprendre (apprentissage incrémental partiellement étiqueté),

– intégration dans l'application (association entre les états reconnus et les actions possibles et prise en compte d'un signal de qualité),

On retrouve là les questions posées par l'introduction de l'apprentissage pour la modélisation de l'utilisateur [WEB 01] en partie aussi celles évoquées pour l'apprentissage des états affectifs [PIC 04]. Plus généralement, il s'agit de questions difficiles, fondamentales et non résolues posées par l'apprentissage des facultés humaines.

Enfin, dans un tout autre registre, il nous semble important d'évoquer les problèmes éthiques soulevés par la recherche de systèmes de reconnaissance d'état affectifs. Notre position est double. D'abord souligner l'importance de ces questions (il s'agit d'une des six principales critiques adressées aux recherches dans ce domaine [PIC 03]) et ensuite essayer, autant que faire se peut, d'avoir recours le plus souvent possible à des éléments objectifs. C'est aussi pour cette raison que nous ne prétendons pas reconnaître des émotions, mais une suite d'états, dont l'étiquetage nous importe moins que l'utilisation pratique que l'on peut en faire.

3. Pré-étude sur le cas supervisée

Si aborder le problème de reconnaissance d'états affectifs comme un problème d'apprentissage supervisé (de première génération suivant notre terminologie) n'est pas réaliste pour concevoir un système tel que nous le souhaitons, les simplifications introduites par ce formalisme autorisent des études préalables, difficiles à mener sur un système complet. Nous avons utilisé ce cadre simplifié pour répondre à certaines questions :

- les méthodes d'apprentissages sont elles utiles ici ?
- quelles sont les méthodes les plus adaptées au problème ?
- quelles variables et quelles caractéristiques utiliser ?

Pour répondre à ces questions, nous avons choisi de reprendre deux études significatives dans le domaine et de comparer, sur ces données déjà connues, différentes méthodes d'apprentissage. Le premier jeu de données est constitué de signaux physiologiques correspondant à des états émotionnels simulés par un acteur et utilisés par J. Healey dans sa thèse [PIC 02]. Le second utilise les signaux physiques (issus d'accéléromètres) associés à différentes activités (marche, cours...) mis à disposition par K. Van Laeroven [LAE 01a]. L'analyse conjointe de signaux physiques et physiologiques permet d'élargir le spectre des données à traiter, ce qui étend la portée de nos conclusions.

3.1. Données

3.1.1. Reconnaissance d'émotions

Les signaux physiologiques correspondant à des états émotionnels sont ceux utilisés par J. Healey dans sa thèse [PIC 02]. Ils sont issus de 4 capteurs (résistance de la peau, respiration, pression sanguine et électromyogramme) auxquels s'ajoute le rythme cardiaque (dédit des variations de pression sanguine). Ces données représentent 25 minutes d'enregistrement par jour sur une période de 20 jours. Chaque jour

contient 5 signaux qui représentent les 8 émotions basiques à identifier (pas d'émotion, énervement, haine, peine, amour, amitié, joie, révérence) qu'une actrice c'est évertuée à simuler.

3.1.2. Détection des déplacements quotidiens

Les données utilisées dans cette partie de l'étude sont mise à disposition par K. Van Laerhoven [LAE 01a]. Elles sont issues de deux accéléromètres disposés sur le genoux et les classes étudiées sont des déplacements, à savoir *être debout*, *marcher*, *courir*, *faire du vélo* et *être assis*. Nous disposons des données brutes issues des capteurs et de la classe courante à chaque instant.

3.2. Méthodes et chaîne des traitements

Pour être efficace, l'utilisation d'un algorithme d'apprentissage doit se faire au sein d'une chaîne de traitements dont chacune des composantes est critique pour le bon fonctionnement de l'application visée. La figure 1 rappelle les étapes principales à suivre pour aller des données à la décision : définition de caractéristiques a priori sans apprentissage, sélection des caractéristiques pertinentes, représentation et discrimination avec apprentissage. Nos travaux visent à déterminer pour chaque étape si elle est utile ou non et quand elle l'est, quelle est l'approche la mieux adaptée à notre problème.

Nous allons commencer par détailler les principales méthodes utilisées pour chacune des étapes pour les deux jeux de données.

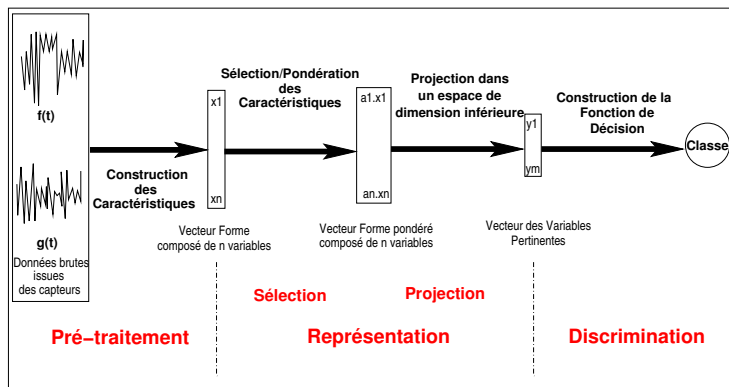


Figure 1. Chaîne des traitements à mettre en œuvre pour faire de l'apprentissage supervisé.

3.2.1. Pré-traitements

Dans ces deux applications, les signaux sont représentés par un ensemble de caractéristiques allant des plus simples (min, max, moyenne...) aux plus compliquées

(coefficients de la FFT). Ces caractéristiques doivent être universelles pour pouvoir s'adapter à tous les cas.

3.2.2. Réduction de la dimensionalité

Analyse discriminante linéaire (ADL) L'ADL vise à réduire la dimension du problème tout en séparant au mieux les différentes classes [DUD 01]. De manière formelle, le problème est de trouver la matrice de projection qui maximise le rapport de la variance inter-classes et la variance globale. Le but est de favoriser le regroupement des points d'une même classe et de maximiser la distance entre les groupes obtenus.

Cartes Auto-Organisatrices de Kohonen L'algorithme des cartes de Kohonen permet de réduire la dimension d'un problème par projection. La couche de sortie (c'est-à-dire la carte) est composée d'un nombre important de « neurones formels », disposés selon une grille. A chaque étape de l'apprentissage, c'est-à-dire à chaque nouvelle arrivée d'une observation, chaque neurone de la couche de sortie calcule son activation. A l'issue de ce calcul, le neurone dont l'activation est la plus grande est déclaré vainqueur. Ses poids ainsi que ceux de ses voisins sont mis à jour en fonction de l'influence du neurone gagnant sur son voisinage.

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. L'idée de la projection consiste à donner des coordonnées à chaque neurone dans l'espace de représentation de la carte. Ensuite, plutôt que d'attribuer à l'entrée les coordonnées du neurone gagnant, on calcule ses coordonnées en faisant une moyenne pondérée des neurones activées.

3.2.3. Sélection de variables

Critère de Lambda Wilks La sélection de variables consiste en une suppression séquentielle de variables (*backward selection*) visant à minimiser le critère de Lambda Wilks [SAP 90]. Ce critère permet de mesurer la séparabilité des données : $CritLW = \frac{\det(W)}{\det(V)}$ où W est la matrice de covariance intraclasse et V la matrice de covariance globale. L'objectif est de choisir un sous-ensemble de variables qui minimise le volume moyen de chaque classe et maximise le volume total.

Sélection pas à pas La sélection pas à pas (voir *stepwise method* dans [DUD 01]) 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.

Critère de maximisation de la marge SVM La sélection de variables dans les méthodes à noyaux peut aussi s'effectuer de manière globale par l'introduction des variables unitaires ν_i au niveau de l'expression du noyau, ce qui donne dans le cas gaussien $\forall x, y \in \mathbb{R}^d$:

$$k_b(\nu \cdot x, \nu \cdot y) = \exp\left(-\frac{\sum_{i=1}^d (\nu_i(x_i - y_i))^2}{2b}\right) \quad (1)$$

où \cdot est le produit terme à terme de 2 vecteurs. Différents critères, comme par exemple la sensibilité de chaque variable sur la marge du classifieur [RAK 03], ont ensuite été

proposés pour régler automatiquement les coefficients ν_i qui représentent l'influence de la variable i pour la discrimination. Choisir $\nu_i = 0$ revient à éliminer la variable i .

3.2.4. Algorithmes de discrimination

Une fois les caractéristiques déterminées, il reste à trouver une règle de décision, telle que $f(x)$ soit la meilleure prédiction possible de l'étiquette de x au sens d'un critère donné.

Maximum a posteriori sur une modélisation gaussienne (MAP gaussien) Dans cette méthode (retenue par J. Healey [PIC 02]), chaque classe c est assimilée à une gaussienne dont la moyenne μ_c et la variance Σ_c sont déterminées sur les points d'apprentissage. La classe des points de test est celle de probabilité maximale, en admettant que les classes sont équiprobables [DUD 01].

K plus proches voisins Le classifieur des k plus proches voisins (k ppv) est un classifieur universel de référence défini à partir d'une règle. Cette règle stipule que chaque point de test prend l'étiquette de la classe dominante parmi ses k ppv.

SVM multi-classes L'algorithme des *Support Vector Machines* (SVM) décrit par V. Vapnik [VAP 98] propose de projeter les points de l'ensemble d'apprentissage dans un espace de Hilbert à noyau reproduisant \mathcal{H} muni d'un produit scalaire $\langle \cdot, \cdot \rangle$ grâce à un noyau $k_b(x, y)$. On peut ensuite montrer que pour un ensemble d'apprentissage $(\mathbf{x}_i, y_i), i = 1, n$, la frontière de décision est de la forme :

$$f(x) = \langle g(\cdot), k_b(x, \cdot) \rangle + b = \sum_{i=1}^n \alpha_i k(x_i, x) + b \quad (2)$$

Dans le cas où les données sont séparables, la frontière $f(x)$ optimale (qui maximise la marge entre les classes) est obtenue en résolvant le problème quadratique :

$$\begin{cases} \min_{g,b} \|g\|_{\mathcal{H}}^2 = \langle g(\cdot), g(\cdot) \rangle \\ \text{avec } y_i (\langle g(\cdot), k_b(x_i, \cdot) \rangle + b) \geq 1 \quad i \in \{1, \dots, n\} \end{cases} \quad (3)$$

Outre ses qualités statistiques, la méthode est intéressante du fait de sa parcimonie. En effet, contrairement aux k -ppv qui exigent le stockage de tous les exemples, avec les SVM seuls les exemples pertinents pour la décision sont conservés (ceux pour lesquels $\alpha_i \neq 0$). Dans la pratique ce nombre peut être très petit.

Pour pouvoir traiter plus de deux classes, il convient d'apporter les modifications nécessaires. C. W. Hsu [HSU 02] compare trois approches du problème SVM multi-classes. Les deux premières méthodes sont basées sur une multiplication des classifieurs bi-classes tandis que la dernière propose une résolution globale.

– **Un-contre-tous.** Le un-contre-tous a été la première réponse proposée pour faire face aux problèmes multi-classes. Chaque classe est opposée à toutes les autres. Il faut donc poser autant de problèmes binaires que de classe. Le résultat final est celui du classifieur de sortie maximale.

– **Un-contre-un.** Cette solution consiste à créer tous les classifieurs bi-classes envisageables du problème, c'est à dire pour C classes $C(C-1)/2$ classifieurs binaires. Chaque classifieur vote pour tous les points et chaque point se voit attribuer la classe qui a reçu le plus de suffrages.

– **Méthode globale.** Plusieurs solutions ont été proposées pour résoudre formellement le problème multiclasse dans le cadre des SVM. Nous avons mis en œuvre la solution proposée par J. Weston [SCH 01] réputée efficace.

Dans cette étude nous avons utilisé notre boîte à outils *Matlab*¹.

Bayes Point Machines (BPM) Du point de vue Bayésien, si les SVM choisissent un hyperplan maximisant une sphère (définie par $\|g\|_{\gamma}^2$), il est préférable de calculer un hyperplan moyen cohérent avec les données. Ce calcul étant assez lourd, la méthode des *Bayes Point Machines* se propose d'approcher la solution bayésienne en calculant une moyenne dans l'espace des caractéristiques [HER 02, QI 02]. Ce faisant on obtient une méthode statistiquement robuste mais toujours gourmande en ressource car elle exige le stockage de tous les exemples de l'ensemble d'apprentissage. Dans cette étude nous avons utilisé la boîte à outils *Matlab* réalisée par T. Minka².

3.3. Résultats expérimentaux

3.3.1. Apprentissage des données émotionnelles

L'évaluation des performances a été effectuée par estimation non biaisée de l'erreur en généralisation (méthode du *leave-one-out* [HAS 01]). Les résultats présentés dans le tableau 1 sont des résultats optimisés. Les paramètres ont été échantillonnés puis toutes les combinaisons ont été évaluées (toujours en *leave-one-out*) afin de trouver les paramètres optimaux.

Prétraitements Nous avons construit 56 variables explicatives pour chaque portion de signal représentant une émotion. Ces 56 variables contiennent les moyennes, écarts-types et densités spectrales (sur une bande de 0 à 0,6 Hz) des cinq signaux issus des capteurs. Finalement, la base d'exemples comporte 160 points repartis dans 8 classes.

Paramètres utilisés Les algorithmes de réduction de la dimensionnalité nécessitent différents réglages : il faut leur donner le nombre de variables à éliminer (pour la sélection de variables) et le nombre d'axes de projection pour l'analyse discriminante linéaire. Les résultats optimaux ont été obtenus en sélectionnant entre 20 et 30 variables puis en projetant les points sur 3 à 5 axes discriminants. Sur ces 2 plages de valeurs, les performances sont constantes. Le réglage optimal est le même lorsque plusieurs algorithmes sont conjugués. Les SVM utilisés sont basés sur des noyaux gaussiens. Le couple de paramètres ($\sigma = 0.08$, $C = 1000$) est optimal lorsque les données sont traitées par ADL. Dans le cas contraire, c'est le couple ($\sigma = 2$, $C = 200$) qui a donné les meilleurs résultats. Le paramètre σ du noyau est beaucoup plus sensible que C . Les k ppv se sont montrés particulièrement robustes. Les performances sont optimales pour k appartenant à $\{5, (\dots), 15\}$, elles faiblissent lentement en s'éloignant de cet intervalle. L'algorithme du MAP gaussien est adaptatif, il ne nécessite aucun réglage.

Dans ce cas, sans doute à cause des huit classes à reconnaître, nous avons été surpris par la lenteur de la méthode BPM contrairement à ce qui avait été rapporté dans la littérature [QI 02]. Nous avons décidé de pousser en avant nos expérimentations et de

1. asi.insa-rouen.fr/~gloosli/simpleSVM.html

2. www.stat.cmu.edu/~minka/papers/ep/bpm/

	sans trait.	SV LW	SV SVM1/1	ADL seule	ADL+ SV LW	ADL+SV (SVM 1/1)	ADL+SC +SV LW
SVM 1/	53,12%	66,87%	68,75%	80,63%	87,50%	88,13%	90,62%
SVM 1/t	52,50%	66,25%	-	79,37%	85,00%	-	88,12%
SVM k-c	55,00%	67,50%	-	80,63%	86,68%	-	88,75%
kppv	37,50%	39,37%	-	81,25%	88,75%	-	90,62%
MAP	41,88%	41,25%	-	77,50%	83,75%	-	85,00%
BPM			-	75,00%	85,62%	-	

Tableau 1. Meilleurs résultats obtenus (en pourcentage de reconnaissance). SV = sélection de variables, LW = Lambda Wilks, ADL = Analyse discriminante linéaire.

comparer les deux méthodes sur des exemples classiques (à partir des boîtes à outils *Matlab*). Pour les performances en temps de calcul, nous avons travaillé sur un problème jouet à 2 classes gaussiennes à 2 dimensions. Les paramètres des noyaux sont identiques pour les SVM et les BPM tandis que le paramètre C a été arbitrairement fixé à 100. Nous avons mesuré le temps d'exécution de chaque algorithme d'apprentissage en fonction de la taille de la base d'apprentissage. Les résultats reportés à la figure 2 sont des résultats moyennés sur 50 tirages aléatoires des données. Les courbes de performances montrent clairement que les SVM sont nettement moins gourmands en temps de calcul et en ressources. Par ailleurs, nous avons comparé les performances de ces 2 algorithmes sur 3 données classiques disponibles à l'UCI Repository. Nous avons comparé les taux d'erreurs des algorithmes sur 50 tirages aléatoires des données d'apprentissage (75% des données) et de tests. Les performances présentées dans la Figure 2 sont obtenues pour le meilleur paramètre du noyau gaussien. Un test statistique sur l'égalité des médians des erreurs (pour les 50 tirages) a également été réalisé. Nos résultats confirment le *no free lunch theorem* [DUD 01] à savoir qu'en terme de généralisation aucun algorithme n'est meilleur que tous les autres.

Résultats Le pourcentage de reconnaissance global masque des disparités dans les résultats. Toutes les émotions ne sont pas aussi bien reconnues, certaines prêtent particulièrement à confusion (*pas d'émotion*). De plus, les résultats ne sont pas identiques d'un classifieur à l'autre cf Tab. 2.

données	BPM	SVM	p-val
spectf	0.18 ± 0.03	0.20 ± 0.03	0.000
ionosphere	0.11 ± 0.03	0.05 ± 0.02	0.000
credit	0.13 ± 0.02	0.13 ± 0.02	0.667

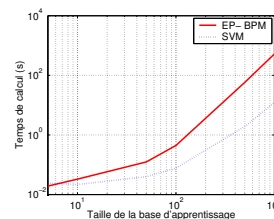


Figure 2. Comparaison des BPM et des SVM en terme d'erreur de classification (tableau de gauche) et de temps de calcul (figure de droite).

3.3.2. Apprentissage des données physiques

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

obtenu réel	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	16	0	3	0	0	0	0	1
(2)	0	17	0	0	0	3	0	0
(3)	0	0	19	0	0	0	0	1
(4)	0	0	0	18	2	0	0	0
(5)	0	0	0	0	20	0	0	0
(6)	0	2	0	0	0	17	1	0
(7)	0	0	0	1	0	0	19	0
(8)	0	0	1	0	0	0	0	19

Matrice de confusion pour les SVM un-contre-un. Résultat global : 90.62% de bonne classification.

obtenu réel	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	16	0	3	0	0	0	0	1
(2)	0	18	0	0	0	2	0	0
(3)	0	0	19	0	0	0	0	1
(4)	0	0	0	18	2	0	0	0
(5)	0	0	0	1	19	0	0	0
(6)	0	2	0	0	0	17	1	0
(7)	0	0	0	1	0	0	19	0
(8)	0	0	1	0	0	0	0	19

Matrice de confusion des k ppv. Résultat global : 90.62% de bonne classification.

Tableau 2. Comparaison de 2 matrices de confusion. Émotions :(1) pas d'émotion, (2) énervement, (3) haine, (4) peine, (5) amour, (6) amitié, (7) joie, (8) révérence. Les paramètres utilisés sont les suivants : $\sigma = 0.08$, $C = 1000$, $k = 7$, $nb_{var} = 20$, $nb_{axes} = 5$.

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.

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 la largeur de bande et le paramètre d'ajustement C . Les réglages de tous ces hyper-paramètres ont été fait lors d'une étude préalable sur des ensembles de 1250 points. Ces points sont tirés aléatoirement dans chaque classe (250 points par classe).

Résultats 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 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 3 présente les résultats obtenus avec les paramètres suivants :

- Construction des caractéristiques : taille de la fenêtre = 45

	Sans Projection		SOM	Espace SOM	
	<i>k</i> PPV	SVM	N. Gagnant	<i>k</i> PPV	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%)

Tableau 3. Résultats. SS correspond à « aucune sélection de variables », PP à la sélection pas à pas et AG à la sélection par approche globale.

- Cartes de Kohonen : taille 12×12, 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 : $k = 1$

obtenu réel	(1)	(2)	(3)	(4)	(5)
(1)	505	0	1	1	0
(2)	0	834	2	2	0
(3)	0	0	521	1	0
(4)	0	2	8	519	0
(5)	0	0	0	0	280

Tableau 4. Matrice de confusion obtenue par la meilleure méthode - SVM après une sélection de variables par approche globale. La classe 1 est la classe assis, la classe 2 est debout, 3 est marche, 4 est cours et 5 est fait du vélo.

3.4. Bilan et discussion

La principale conclusion de cette pré-étude concerne l'utilisation des méthodes d'apprentissage. Grâce à elles, sur les deux exemples traités, les performances ont été significativement améliorées, atteignant plus de 90% contre 81% de bonne classification dans les travaux originaux de J. Healey [PIC 01] et passant de 80 % dans les travaux initiaux [LAE 01a] à plus de 99 % dans notre étude. Plus qu'à l'emploi des SVM, cela est dû à la mise en œuvre de toute la méthodologie liée à l'apprentissage et notamment à la sélection des variables pertinentes, qui dans les deux cas permet d'obtenir les meilleurs résultats.

Il reste néanmoins des différences entre les deux cas et notamment concernant l'utilisation ou non d'une technique de réduction de la dimensionalité. Sur les données émotionnelles, il faut utiliser une projection pour améliorer la qualité des résultats alors qu'elle n'est pas utile sur les données de K. Van Laerhoven. Cela s'explique à notre sens plus par la taille des ensembles d'apprentissages que par la nature des problèmes. En effet, nous ne disposons que de 160 exemples d'émotions contre plus

de 2500 pour les cinq classes de K. Van Laerhoven. Lorsque l'on ne dispose que de peu de données, l'effet des prétraitements est important alors qu'il tend à disparaître lorsque la taille de l'échantillon augmente.

4. Proposition d'une architecture globale

Nous définissons maintenant l'architecture du système de détection d'états affectifs. Les entrées de ce système sont d'une part les signaux issus des capteurs et d'autre part les étiquettes (ou informations de pertinence des réponses) données par l'utilisateur (par l'intermédiaire de l'application qui utilise les états affectifs). Le flux d'information doit donc être à double sens. Le premier sens (ascendant sur la figure 3) doit être capable de transformer des signaux bruts en état affectif. Le deuxième sens quant à lui (descendant) doit permettre l'ajustement du flux ascendant.

Notre hypothèse de travail est qu'il est excessivement difficile de caractériser une émotion à partir de signaux, de manière générale et applicable à tout utilisateur. *A fortiori* il est encore plus difficile de le faire pour l'ensemble des états affectifs que nous pouvons rencontrer. Par ailleurs, chaque application utilisant une connaissance de l'état affectif n'a besoin en fait que d'un nombre limité d'états pertinents. Par conséquent il n'est pas utile de chercher à étiqueter chaque état : l'apprentissage supervisé n'est pas la solution.

Les flux ascendant et descendant passent par quatre modules. Chaque module est double : d'une part un traitement de l'information et d'autre part un contrôle de la méthode de traitement.

– Module 1 :

- Ascendant : Des signaux bruts aux caractéristiques. *La construction des caractéristiques est idéalement faite pour tout type de capteurs. Cette phase est référée comme « prétraitements ».*

- Descendant : Sélection de capteurs. *L'élimination des caractéristiques issues d'un capteur par le module 2 peut conduire à l'élimination pure et simple du capteur - défectueux par exemple.*

– Module 2 :

- Ascendant : Des caractéristiques à la séquence d'états non étiquetés. *La détection de changement permet de résumer l'information des capteurs de manière non supervisée - détails dans la section suivante*

- Descendant : Sélection de caractéristiques et adaptation des paramètres de la détection de rupture. *L'information par le module 3 d'une mauvaise rupture permet d'ajuster la sensibilité de la détection ou d'identifier quelle caractéristique induit en erreur - ce qui conduit à son élimination.*

– Module 3 :

- Ascendant : De la séquence d'états au réseau d'états regroupés partiellement étiquetés. *Le regroupement des segments similaires (clustering) permet d'organiser l'information et d'étiqueter les états similaires à des états précédemment étiquetés*

- Descendant : Détection de fausse alarme et de non détection. *Les fausses alarmes sont caractérisées par l'attribution des segments précédents et suivants au*

même regroupement. Les non-détections peuvent être remarquées par un retour de l'utilisateur qui donne une étiquette non concordante. De la même manière, le critère de regroupement peut être ajusté.

– Module 4 :

- Ascendant : Des états regroupés à l'application. Dépendant de l'application, informer en permanence de l'état courant ou bien prévenir d'un état particulier lorsqu'il est détecté.

- Descendant : Etiquetage et indication de pertinence. Retour de l'utilisateur, dépendant de l'application.

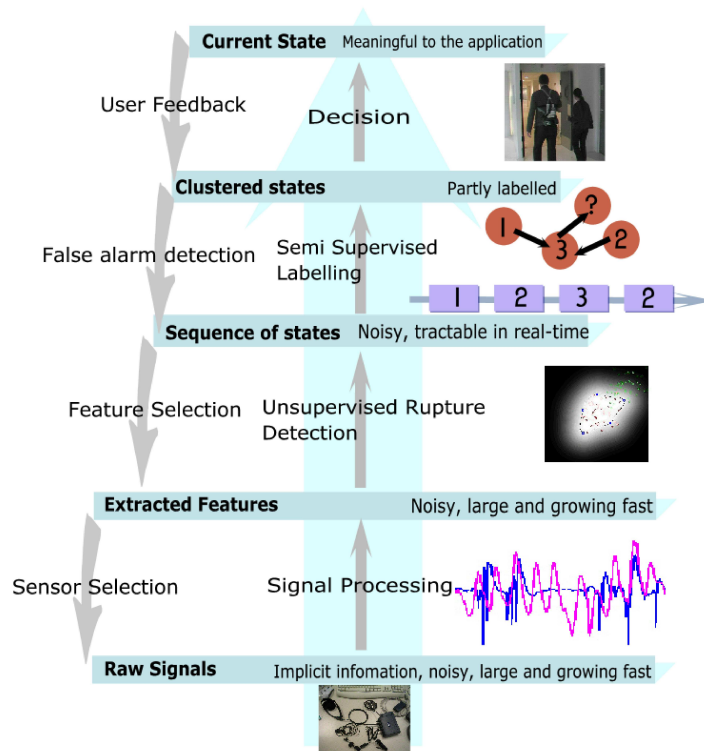


Figure 3. Architecture globale. L'état de l'utilisateur est obtenu à partir des signaux issus des capteurs qu'il porte. Les étapes intermédiaires visent à simplifier et extraire l'information reçue.

Nous étudions dans la suite des travaux présentés ici la phase ascendante du deuxième module - la détection de changement dans les signaux. Pour cela nous nous intéressons aux tests statistiques et aux méthodes d'apprentissage à noyau.

5. Approche non supervisée

Il existe beaucoup de travaux sur la segmentation de signaux du fait du grand nombre d'applications possibles, en reconnaissance vocale, indexation musicale ou détection de défauts par exemple. Ces travaux ont conduit à de nombreux algorithmes pour la détection de changement (voir [BAS 93] pour une revue complète). Jusqu'à maintenant les principales méthodes utilisent des modèles paramétriques de la distribution. Du côté des méthodes non paramétriques, plusieurs modèles ont été proposés (voir [MAR 03] pour une revue détaillée dans le contexte de la détection de nouveauté) tels que les réseaux de neurones [FAN 00], les modèles de Markov cachés, les modèles bayésiens [ROB 04], les SVM [DAV 02] (et autres méthodes à noyaux [NGU 05]) et les mesures de croyance [LEN 05].

5.1. Tests Statistiques et détection de rupture

Soit $X_i, i = 1, 2t$ une suite de variables aléatoires distribuées selon une distribution \mathbb{P}_i . Nous cherchons à savoir si un changement est arrivé au temps t . Pour commencer dans un cadre simple, nous supposons que la séquence est stationnaire de 1 à t et de $t + 1$ à $2t$, c'est-à-dire qu'il existe des distributions \mathbb{P}_0 et \mathbb{P}_1 telles que $P_i = P_0, i \in [1, t]$ et $P_i = P_1, i \in [t + 1, 2t]$. Notre problème est de savoir si $\mathbb{P}_0 = \mathbb{P}_1$ (aucun changement) ou si au contraire $\mathbb{P}_0 \neq \mathbb{P}_1$ (un changement est arrivé). La reformulation statistique peut se faire comme suit :

$$\begin{cases} \mathcal{H}_0 & : \mathbb{P}_0 = \mathbb{P}_1 \\ \mathcal{H}_1 & : \mathbb{P}_0 \neq \mathbb{P}_1 \end{cases}$$

Notre objectif est d'avoir une méthode universelle qui fonctionne avec n'importe quelle densité. Nous faisons quelques approximations pour clarifier le cadre de travail. Tout d'abord nous considérons que les densités \mathbb{P}_0 et \mathbb{P}_1 appartiennent à la famille exponentielle généralisée et donc qu'il existe un espace de Hilbert à noyaux reproduisant \mathcal{H} avec le produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ et un noyau reproduisant k tel que [SMO 04] :

$$\mathbb{P}_0(x) = \mu(x) \exp\langle \theta_0(\cdot), k(x, \cdot) \rangle_{\mathcal{H}} - g(\theta_0), \quad \mathbb{P}_1(x) = \mu(x) \exp\langle \theta_1(\cdot), k(x, \cdot) \rangle_{\mathcal{H}} - g(\theta_1)$$

avec μ une mesure de probabilité appelée le support et $g(\theta)$ la fonction de log-partition.

La seconde hypothèse est que les paramètres fonctionnels θ_0 et θ_1 de ces densités seront estimés sur les données, respectivement sur la première et la seconde moitié de l'échantillon considéré en utilisant un SVM à une classe. Ce faisant nous suivons notre hypothèse initiale, à savoir qu'avant le temps t nous savons que la distribution est constante et égale à \mathbb{P}_0 . L'algorithme du SVM à une classe (OC-SVM - [SCH 00]) nous donne une bonne estimation de cette densité. $\hat{\mathbb{P}}_1(x)$ peut être vu comme une approximation robuste de l'estimateur du maximum de vraisemblance. Utiliser le SVM à une classe et le modèle des familles exponentielles s'écrit de la façon suivante :

$$\hat{\mathbb{P}}_j(x) = \mu(x) \exp\left(\sum_{i=1}^t \alpha_i^{(j)} k(x, x_i) - g(\theta_0)\right) \quad j = 1, 2$$

où $\alpha_i^{(0)}$ est obtenu en résolvant le SVM à une classe sur la première moitié des données (x_1 to x_t) tandis que $\alpha_i^{(1)}$ est obtenu de la même façon sur la seconde moitié des données (x_{t+1} to x_{2t}). Forts de ces trois hypothèses, après simplification des calculs, la région d'acceptation d'un test de type « vraisemblance généralisé » (VG) est donnée par :

$$\sum_{j=t+1}^{2t} \left(\sum_{i=1}^t \alpha_i^{(0)} k(x_j, x_i) \right) < s''$$

Cela aboutit à l'algorithme de détection de nouveauté proposé dans [SCH 00]. La mise en œuvre de ce test peut se faire de plusieurs façons, selon l'interprétation que l'on en fait.

– Tests sur la variance des sorties de l'OC-SVM

$$\text{Var}_j \left(\sum_{i=t}^t \alpha_i^{(0)} k(x_j, x_i) \right) < s, \quad j \in [t+1, 2t]$$

Le seuil reste à déterminer. Nous utilisons une heuristique pour cela. Nous laissons à l'algorithme une période d'initialisation correspondant à quelques secondes de signaux de manière à observer l'ordre de grandeur moyenne et en déduire un seuil. Une autre heuristique utilisée consiste à travailler avec un seuil mobile (s) et un seuil minimal (s_m) en dessous duquel on considère que toute variation est du bruit). On utilise aussi un délai pendant lequel on stabilise s . Les règles utilisées sont les suivantes :

- s et s_m ont pour valeur initiale 0
- pendant la phase d'initialisation, s_m est égal à la moyenne des variances obtenues de puis le début,
- à la fin de la phase d'initialisation, s_m est fixé,
- si la valeur courante de s est inférieure à s_0 : alors $s = s_m$,
- si la variance à dépassé le seuil s : s prend la valeur maximale des dernières valeurs de la variance,
- si la valeur de s est stable depuis un temps supérieur au délai défini : alors s est égal à la moyenne entre les dernières valeurs des variances obtenues et lui-même, sans jamais être inférieur à s_m .

Cette heuristique fonctionne bien sous réserve de disposer d'une phase d'initialisation représentative des signaux à venir (par exemple un état stable de « non activité » est un bon critère pour déterminer ce qui peut être considéré comme du bruit).

– Tests sur le taux de mauvaise classification

Nous venons de donner une heuristique pour utiliser une méthode dont le seuil de détection dépend des données. Afin d'éviter ces étapes nous proposons d'approximer le test par un critère qui permet d'utiliser un seuil facile à régler. Nous remplaçons la variance par la proportion de points non attribués à la classe courante. Ainsi nous pouvons fixer le seuil comme étant un pourcentage fixe.

– Tests sur la somme cumulée des sorties de l'OC-SVM

Inspirée de la méthode CUSUM, nous avons aussi utilisé la sortie du SVM comme un score qui se cumule négativement. Quand l'accumulation atteint un seuil on prend une décision. On pose donc deux seuils, un négatif et un positif. Si le seuil négatif est

atteint on décide qu'il n'y a pas eu de rupture, si le seuil positif est atteint on décide qu'il y a eu rupture. Si aucun des deux seuils n'est atteint on regarde le point suivant

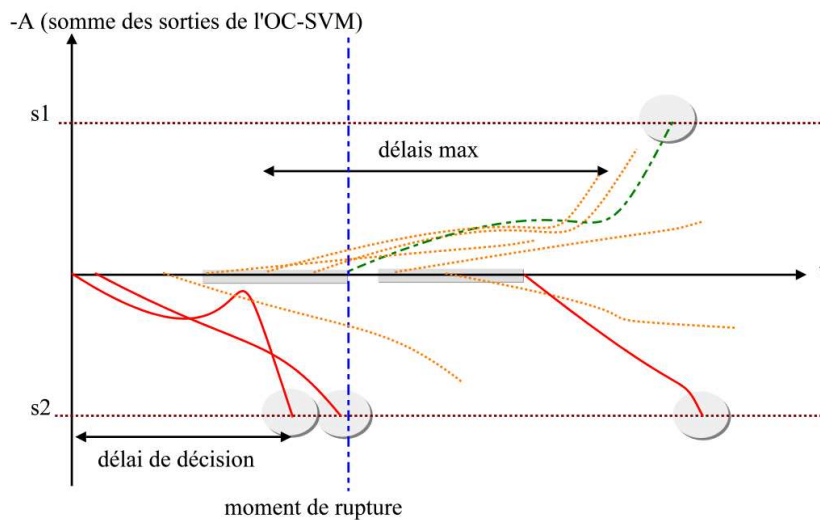


Figure 4. Illustration du fonctionnement de CUSUM. A chaque point, on regarde l'évolution de la somme de la sortie du SVM (concrètement, on classe le prochain point non vu et on ajoute sa sortie à la somme des points précédents). Trois événements arrêtent ce processus. 1/ on atteint le seuil de décision de stabilité de la classe (courbes pleines rouges). 2/ on atteint le seuil de décision de rupture (courbe verte en pointillés irréguliers). 3/ on atteint le délai maximal sans pouvoir prendre de décision (courbes oranges en pointillés).

5.2. Expérimentations

Dans cette partie nous allons d'abord donner quelques exemples sur des données synthétiques et sur des données utilisées dans notre pré-étude. Ensuite nous expliquerons notre démarche expérimentale et enfin nous donnerons les résultats obtenus pour deux jeux de données réelles³.

5.2.1. Illustration de la méthode sur des données synthétiques

Nous générons des données aléatoires distribuées selon des lois gaussiennes dont on fait varier les paramètres. Les données sont représentées par la moyenne et la variance sur une fenêtre mobile (figure 5.2.1).

5.2.2. Illustration sur signaux réels

Nous avons testé notre méthode sur les signaux utilisés pour la première étude, à savoir les données issues d'accéléromètres pour la classification de mouvements.

Cette expérience (figure 6) montre qu'il est possible de segmenter utilement les données. Il n'y a pas de non détection sur des changements « à long terme » et les

3. disponibles sur <http://asi.insa-rouen.fr/~gloosli/contextaware.html>

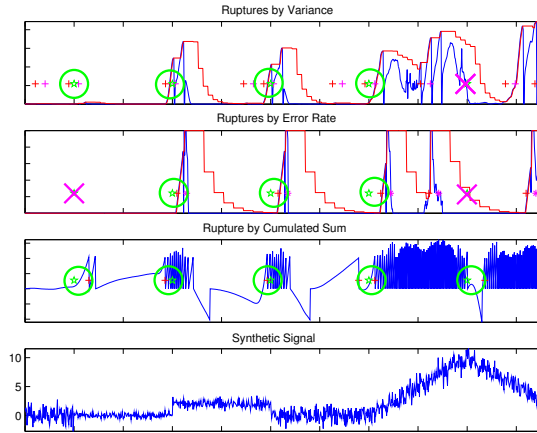


Figure 5. Exemple d'utilisation de l'algorithme sur des données issues de lois gaussiennes dont on fait varier la moyenne et la variance. La courbe du bas montre le signal de départ. Les trois premières montrent les résultats obtenus par les trois méthodes de détection de rupture.

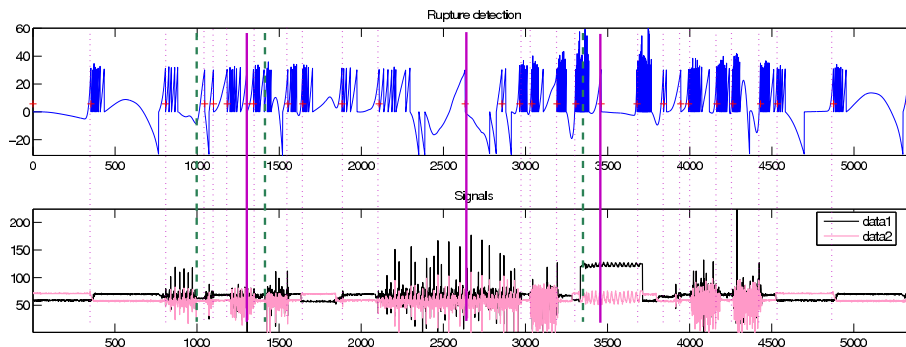


Figure 6. Résultats sur les données de mouvements avec la méthode de somme cumulée. On remarque qu'il y a peu d'erreurs (bonnes détections en pointillés fins roses). Les non détections (traits épais pointillés verts) apparaissent à des moments où les changements s'enchaînent rapidement. Les fausses alarmes sont notées en traits épais pleins roses.

fausses alarmes seraient facilement décelées lors de la classification des tronçons de signaux. Pour deux des fausses alarmes (les deux dernières) on voit aussi qu'elles ont une caractéristique qui pourrait permettre de les éliminer d'office : elles apparaissent entre deux moments où la somme cumulée est négative, ce qui n'est jamais le cas pour les bonnes détections. Toutefois cette heuristique n'a pas été validée sur d'autres données jusqu'à présent.

5.2.3. Acquisition de signaux

Le ProComp est un système d'acquisition sur lequel nous disposons de cinq capteurs biologiques, à savoir un capteur de conductance de la peau, un capteur de respiration, de pression sanguine, un électromyogramme et un capteur de température périphérique.

Dans un cadre médical ou pour rechercher une information spécifique, les capteurs doivent être placés à des endroits soigneusement choisis et la personne équipée doit veiller à ne pas faire de mouvements qui perturbent les signaux. Dans notre approche nous cherchons à reproduire des données réalistes pour une utilisation dans la vie quotidienne. Par conséquent l'utilisateur ne doit pas « prendre garde » aux capteurs lors des prises de données. Dans le même ordre d'idée, les caractéristiques issues de signaux brutes sont généralistes et identiques pour tous les capteurs.

5.2.4. Utilisateur monitoré dans un contexte anodin

L'utilisateur est équipé des cinq capteurs biologiques ainsi que de trois accéléromètres tri-axiaux. Il peut ensuite aller et venir à sa convenance. Les données sont recueillies sur quelques minutes pendant lesquelles l'utilisateur est filmé. A partir de la vidéo, nous définissons à priori les moments où nous nous attendons à détecter un changement dans les signaux. Notre objectif est de déterminer à la fois si la détection automatique donne des résultats sensés mais aussi de voir jusqu'à quel point une observation extérieure est fiable.

5.2.4.1. Obtention de résultats

Les signaux obtenus sont traités « en ligne ». Il n'y a donc pas de normalisation globale possible. Par ailleurs les caractéristiques extraites sont la variance, la moyenne, le minimum, le maximum, et la transformée de Fourier sur une fenêtre glissante. La taille de la fenêtre est fixée arbitrairement de manière à la faire correspondre à une demi-seconde. La largeur de bande du noyau (gaussien) est obtenue par validation croisée.

5.2.4.2. Résultats et discussion

A posteriori	Manuel (à priori)	Auto (Variance)	Auto (Taux d'erreur)	Auto (Somme cumulée)
Nb de ruptures	16	30	23	17
Détection	16	15 (94%)	15(94%)	15(94%)
Fausse alarme	0	15 (50%)	8 (35%)	2 (12%)
Non Détection	1	2 (13%)	2 (13%)	2 (13%)

Tableau 5. Résultats obtenus pour les différentes méthodes. Les ruptures détectées sont les mêmes mais les taux de fausse alarme diffèrent.

Les résultats présentés dans le tableau 5 valident notre approche. Entre les trois variantes la somme cumulée donne les résultats les moins bruités. Nous utiliserons donc préférentiellement cette méthode sur les données plus complexes (à savoir obtenues uniquement à partir de capteurs physiologiques).

5.2.5. Utilisateur monitoré dans le cadre d'un jeu vidéo

Pour cette expérience nous n'utilisons que les capteurs biologiques, car l'utilisateur reste assis. On enregistre les signaux pendant qu'il joue à un jeu vidéo impliquant des réflexes et créant joie ou frustration. Le jeu (Xblast) se joue en réseau et consiste à poser des bombes pour tuer tous ses adversaires avant la fin du temps imparti. Une caractéristique intéressante du jeu est son organisation par niveau. Au court de la partie, le joueur ne joue pas tout le temps s'il ne gagne pas tous les niveaux. Une des choses que l'on va donc chercher à retrouver dans les signaux est l'état « en jeu/hors jeu », ce qui se traduit dans les ruptures par « commence à jouer/arrête de jouer ». Les autres ruptures qui sont relevées et que l'on essaiera de retrouver sont les événements tels que « tuer/blesser un adversaire » ou « être tué/blessé ».

5.2.5.1. Obtention de résultats

Les signaux, moins nombreux que précédemment, sont traités de façon similaire. La vidéo et les enregistrement des parties permettent au joueur de donner une liste d'événements significatifs pour lui, tandis qu'un observateur dresse lui aussi une liste d'événements. Ce double étiquetage a pour objectif de vérifier une de nos hypothèses de travail, à savoir qu'il est impossible, dans une situation réelle, de connaître avec certitude l'état que le système doit découvrir. On constate sur ce double étiquetage que le joueur a noté environ 20% d'événements significatifs en plus par rapport à l'observateur extérieur.

5.2.5.2. Résultats et discussion

A posteriori	Manuel Joueur (à priori)	Manuel Observateur (à priori)	Auto (Somme cumulée)
Nb de ruptures	50	39	55
Détection	50	39	36 - 32 (72% - 82%)
Fausse alarme	0	0	19 - 23 (35% - 42%)
Non Détection	0	11	14 - 7 (22% - 18%)

Tableau 6. Résultats obtenus pour le joueur de jeux vidéos.

Les résultats (tableau 6) beaucoup plus mitigés sur cette expérience confirment le fait qu'avoir uniquement des données physiologiques est un problème plus complexe. Les accéléromètres fournissent en effet des informations « faciles » à segmenter, ce qui n'est pas le cas pour la conductivité de la peau (qui est lente) par exemple. Par ailleurs nous constatons que ce qui apparaît notable par l'utilisateur ou par l'observateur ne correspond pas nécessairement à une variation physiologique. En revanche nous détectons tout de même des changements qui correspondent au débuts de fins de parties ainsi qu'aux victoires ou défaites. Les événements plus intermédiaires comme blesser un adversaire ne semblent pas être détectables (et le plus souvent n'ont pas été notés par l'observateur).

6. Conclusion et Perspectives

Les résultats présentés dans cet article démontrent la faisabilité de l'utilisation de l'apprentissage pour la reconnaissance d'états affectifs à partir de signaux physiques et physiologiques. L'étude de deux jeux de données de la littérature a révélé que l'emploi de caractéristiques adaptées associées à une technique de sélection des variables permet d'obtenir de meilleurs résultats que l'état de l'art. Nous avons ensuite montré qu'avec ces prétraitements pertinents, plusieurs algorithmes d'apprentissage donnent des résultats statistiquement comparables (k -ppv, approche bayésienne et SVM). Le choix d'une méthode doit alors être guidé par d'autres critères : la robustesse, l'adaptabilité et la complexité en terme de temps de calcul et d'espace mémoire. Selon ces critères, ce sont les SVM qui sont les mieux adaptées au problème. L'importance des prétraitements s'explique en partie par le relativement faible nombre d'exemples dont on dispose (160 cas pour huit classes contre par exemple 60.000 exemples sur la base NIST de reconnaissance de caractères manuscrits). Il s'agit là d'une des difficultés fondamentales du problème qui se retrouve dans le domaine de l'apprentissage statistique : comment apprendre sur certains sujets dans des conditions données puis généraliser sur d'autres sujets alors que les conditions ont changé ?

Pour répondre à cette question nous avons proposé une architecture de reconnaissance basée sur la notion de détection de changement, utilisant une approche de type « non supervisée ». Au lieu de reconnaître à chaque instant l'état affectif du sujet, nous proposons de commencer par segmenter les signaux disponibles en tronçons homogènes qui seront étiquetés dans un second temps. Les algorithmes de segmentation disponibles n'étant pas adaptés aux contraintes de notre application (cadre non paramétrique, temps réel, signaux hétérogènes), nous avons proposé en la justifiant une nouvelle approche basée sur l'utilisation de SVM adaptées au problème de détection de rupture. La faisabilité et l'intérêt de l'approche proposée a été démontrée sur deux jeux de données réels que nous avons générés nous même.

A ce stade, il nous reste à déterminer comment étiqueter les états ainsi détectés par notre approche. Plutôt que d'essayer une approche « générique » à ce problème nous pensons à une approche spécialisée : une méthode permettant de faire le lien entre les états détectés et ce qui est utile à une application donnée. Nous pouvons suggérer un scénario possible permettant la mise en œuvre d'une application prenant en compte l'état affectif d'un utilisateur. D'abord il faut pour cette application définir les actions possibles ainsi qu'une fonction d'utilité. Puis l'algorithme de segmentation proposé dans cet article pourra être utilisé. Les segments ainsi obtenus devront ensuite être réunis au sein de groupes homogènes (*clustering*). Enfin, la relation entre chacun des groupes et les actions sera apprise de sorte à maximiser la fonction d'utilité. Il faudrait pré-régler le système en laboratoire à partir d'un dispositif expérimental complet, et lui permettre ensuite d'évoluer pour s'adapter en ligne au destin propre de chaque utilisation. Il s'agit là du futur que nous entendons donner à nos recherches.

Remerciements : Ce travail est financé en partie par le Programme IST de la Communauté Européenne, avec le réseau d'excellence PASCAL, IST-2002-506778. Cette publication reflète uniquement le point de vue des auteurs.

7. Bibliographie

[ARK 99] ARK W., DRYER D., LU D., « The emotion mouse », *Proceedings of HCI International. Munich, Germany, 1999.*

- [BAN 96] BANSE R., SCHERER K., « Acoustic profiles in vocal emotion expression », *Journal of Personality and Social Psychology*, vol. 70, n° 3, 1996, p. 614–636.
- [BAS 93] BASSEVILLE M., NIKIFOROV I. V., *Detection of Abrupt Changes - Theory and Application*, Prentice-Hall, 1993.
- [BER 88] BERGSON H., « Essai sur les données immédiates de la conscience », http://www.uqac.quebec.ca/zone30/Classiques_des_sciences_sociales/classiques/bergson_henri/essai_conscience_immediate/essai_conscience.pdf, 1888.
- [BID 03] BIDEL S., LEMOINE L., PIAT F., ARTIÈRES T., GALLINARI P., « Apprentissage de comportements utilisateurs de produits Hypermédias », *RIA*, vol. 17, 2003, p. 423–436.
- [BOR 05] BORDES A., ERTEKIN S., WESTON J., BOTTOU L., « Working Document : Fast Kernel Classifiers with Online and Active Learning », May 2005, <http://leon.bottou.com/publications/pdf/huller3.pdf>.
- [CAC 90] CACIOPPO J., TASSINARY L., « Inferring psychological significance from physiological signals », *Am Psychol*, vol. 45, 1990, p. 16–28.
- [CHR 04] CHRISTIE I., FRIEDMAN B., « Autonomic specificity of discrete emotion and dimensions of affective space : A multivariate approach », *International Journal of Psychophysiology*, vol. 51, 2004, p. 143–153.
- [CRO 05] CROWIE R., SCHRODER M., « Piecing together the emotion jigsaw », *Lecture Notes in Computer Science*, vol. 3361, Springer-Verlag Heidelberg, 2005.
- [DAV 02] DAVY M., GODSILL S., « Detection of abrupt spectral changes using support vector machines », *Proc. IEEE ICASSP-02*, 2002.
- [DUD 01] DUDA R., HART P., STORK D., *Pattern Classification*, Wiley Interscience - 2e édition, 2001.
- [FAN 00] FANOURT C., PRINCIPE J. C., « On the use of neural networks in the generalized likelihood ratio test for detecting abrupt changes in signals », *Intl. Joint Conf. on Neural Networks*, pp. 243-248, at Como, Italy, 2000.
- [HAA 04] HAAG A., GORONZY S., SCHAICH P., WILLIAMS J., « Emotion Recognition Using Bio-sensors : First Steps towards an Automatic System », *Lecture Notes in Computer Science*, vol. 3068, p. 36–48, Springer-Verlag, 2004.
- [HAS 01] HASTIE T., TIBSHIRANI R., FRIEDMAN J., *The Elements of Statistical Learning*, Springer, 2001.
- [HEA 05] HEALEY J., PICARD R., « Detecting Stress During Real-World Driving Tasks », *IEEE Trans. on Intelligent Transportation Systems*, , 2005, To appear.
- [HER 02] HERBRICH R., *Learning Kernel Classifiers*, MIT press, Cambridge, Massachusetts, 2002.
- [HSU 02] HSU C.-W., LIN C.-J., « A comparison of methods for multi-class Support Vector Machines », *IEEE Transactions on Neural Networks*, vol. 13, 2002, p. 415-425.
- [HUD 03] HUDLICKA E., MCNEESE M., « Special issue on the Applications of Affective Computing in Human-Computer Interaction », *International Journal of Human-Computer Studies*, vol. 59, n° 1-2, 2003, p. 1-255, Elsevier Ltd.
- [JOR 04] JORDAN M. I., « Graphical models », *Statistical Science (Special Issue on Bayesian Statistics)*, vol. 19, 2004, p. 140–155.
- [KER 03] KERN N., SCHIELE B., SCHMIDT A., « Multi-sensor Activity Context Detection for Wearable Computing », *Lecture Notes in Computer Science : Ambient Intelligence*, vol. 2875, Springer-Verlag Heidelberg, 2003.
- [KIM 04a] KIM J., BEE N., WAGNER J., ANDRÉ E., « Emote toWin : Affective Interactions with a Computer Game Agent », *GI Jahrestagung*, vol. 1, 2004, p. 159–164.

- [KIM 04b] KIM K. H., BANG S. W., KIM S. R., « Emotion recognition system using short-term monitoring of physiological signals », *Medical and biological engineering and computing*, vol. 42, 2004, IFMBE.
- [KIM 04c] KIM S., ANDRÉ E., « Composing Affective Music with a Generate and Sense Approach », *Proceedings of Flairs 2004 - Special Track on AI and Music*, AAAI Press, 2004.
- [KLE 02] KLEIN J., MOON Y., PICARD R. W., « This Computer Responds to User Frustration : Theory, Design, Results, and Implications », *Interacting with Computers*, vol. 14, 2002, p. 119–140.
- [KOR 03] KORPIPÄÄ P., KOSKINEN M., PELTOLA J., MÄKELÄ S.-M., SEPPÄNEN T., « Bayesian approach to sensor-based context awareness », *Personal Ubiquitous Comput.*, vol. 7, n° 2, 2003, p. 113–124, Springer-Verlag.
- [LAE 01a] LAERHOVEN K. V., « Combining the Kohonen Self-Organizing Map and K-Means for On-line Classification of Sensordata », *Artificial Neural Networks*, vol. vol 2130, 2001, p. pp. 464-470.
- [LAE 01b] LAERHOVEN K. V., AIDOO K., « Teaching Context to Applications », *Personal Ubiquitous Comput.*, vol. 5, n° 1, 2001, p. 46–49, Springer-Verlag.
- [LEC 98] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P., « Gradient-Based Learning Applied to Document Recognition », *Proceedings of the IEEE*, vol. 86, n° 11, 1998, p. 2278-2324.
- [LEN 05] LENSER S., VELOSO M., « Non-Parametric Time Series Classification », *Under review for ICRA'05*, 2005.
- [LI 05] LI X., JI Q., « Active affective State detection and user assistance with dynamic bayesian networks », *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 35, 2005, p. 93– 105.
- [LIS 04] LISETTI C. L., NASOZ F., « Using Non-invasive Wearable Computers to Recognize Human Emotions from Physiological Signals », *EURASIP Journal on Applied Signal Processing - Special Issue on Multimedia Human-Computer Interface*, vol. 2004, n° 11, 2004, p. 1672–1687.
- [LIT 04] LITTLEWORT G., BARTLETT M. S., FASEL I., CHENU J., KANDA T., ISHIGURO H., MOVELLAN J. R., « Towards social robots : Automatic evaluation of human-robot interaction by face detection and expression classification », *Advances in Neural Information Processing Systems 16*, p. 97-104, MIT Press, Cambridge, MA, 2004.
- [LOO 04] LOOSLI G., CANU S., VISHWANATHAN S., SMOLA A. J., CHATTOPADHYAY M., « Une boîte à outils rapide et simple pour les SVM », , 2004, p. 113-128, Presses Universitaires de Grenoble.
- [MAR 03] MARKOU M., SINGH S., « Novelty detection : a review, part 2 : neural network based approaches », *Signal Process.*, vol. 83, n° 12, 2003, p. 2499–2521, Elsevier North-Holland, Inc.
- [NAS 03] NASOZ F., ALVAREZ K., LISETTI C. L., FINKELSTEIN N., « Emotion Recognition from Physiological Signals for Presence Technologies », *International Journal of Cognition, Technology, and Work – Special Issue on Presence*, vol. 6, n° 1, 2003.
- [NGU 05] NGUYEN X., WAINWRIGHT M. J., JORDAN M. I., « Nonparametric Decentralized Detection using Kernel Methods », *IEEE Transactions on Signal Processing (accepted for publication)*, , 2005.
- [PAN 03] PANTIC M., ROTHKRANTZ L., « Toward an affect-sensitive multimodal human-computer interaction », *Proceedings of the IEEE*, vol. 91, n° 9, 2003.
- [PIC 99] PICARD R., « Response to Sloman's Review of Affective Computing », *AI Magazine*, vol. 20, n° 1, 1999, p. 134–137.

- [PIC 01] PICARD R. W., VYZAS E., HEALEY J., « Toward Machine Emotional Intelligence : Analysis of Affective Physiological State », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, n° 10, 2001, p. 1175–1191, IEEE Computer Society.
- [PIC 02] PICARD R., HEALEY J., « Eight-emotion Sentics Data », MIT Affective Computing Group <http://affect.media.mit.edu>, 2002.
- [PIC 03] PICARD R. W., « Affective computing : challenges. », *Int. J. Hum.-Comput. Stud.*, vol. 59, n° 1-2, 2003, p. 55-64.
- [PIC 04] PICARD R. W., PAPERT S., BENDER W., BLUMBERG B., BREAZEAL C., CAVALLO D., MACHOVER T., RESNICK M., ROY D., STROHECKER C., « Affective Learning : A Manifesto », *BT Technology Journal*, vol. 22, n° 4, 2004, p. 253–269, Kluwer Academic Publishers.
- [QI 02] QI Y., PICARD R. W., « Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification », *Proceedings of International Conference on Pattern Recognition*, 2002.
- [RAK 03] RAKOTOMAMONJY A., « Variable Selection Using SVM-based Criteria », *Journal of Machine Learning Research*, vol. 3, 2003, p. 1357-1370.
- [RAL 05] RALAIVOLA L., D'ALCHÉ BUC F., « Time Series Filtering, Smoothing and Learning using the Kernel Kalman Filter », *Proc. of IEEE Int. Joint Conference on Neural Networks, Montreal, Canada*, 2005.
- [RAN 05] RANI P., SARKAR N., SMITH C. A., ADAMS J. A., « Affective Communication for Implicit Human-Machine Interaction », *IEEE Transactions on Systems, Man, and Cybernetics*, , 2005, Under Review.
- [ROB 04] ROBERTS S., ROUSSOS E., CHOUDREY R., « Hierarchy, priors and wavelets : structure and signal modelling using ICA », *Signal Process.*, vol. 84, n° 2, 2004, p. 283–297, Elsevier North-Holland, Inc.
- [SAP 90] SAPORTA G., *Probabilités, Analyse de données et Statistique*, Editions Technip, 1990.
- [SCH 87] SCHERER K. R., « Toward a dynamic theory of emotion : The component process model of affective states », http://www.unige.ch/fapse/emotion/publications/pdf/tcte_1987.pdf, 1987, Unpublished manuscript.
- [SCH 00] SCHOLKOPF B., WILLIAMSON R., SMOLA A., SHAW-TAYLOR J., « Support Vector Method for Novelty Detection », SOLLA S., LEEN T., MULLER K., Eds., *NIPS*, MIT Press, 2000, p. 582–588.
- [SCH 01] SCHÖLKOPF B., SMOLA A., *Learning with Kernels*, MIT Press, 2001.
- [SCH 02] SCHEIRER J., FERNANDEZ R., KLEIN J., PICARD R. W., « Frustrating the User on Purpose : A Step Toward Building an Affective Computer », *Interacting with Computers*, vol. 14, n° 2, 2002, p. 93–118.
- [SMO 04] SMOLA A., « Exponential Families and Kernels », Berder summer school, 2004, <http://users.rsise.anu.edu.au/smola/teaching/summer2004/>.
- [VAP 98] VAPNIK V., *Statistical Learning Theory*, Wiley, 1998.
- [WAL 02] WALLACH H. M., « Efficient Training of Conditional Random Fields », Master's thesis, Division of Informatics, University of Edinburgh, 2002.
- [WEB 01] WEBB G. I., PAZZANI M. J., BILLSUS D., « Machine Learning for User Modeling », *User Modeling and User-Adapted Interaction*, vol. 11, n° 1-2, 2001, p. 19–29, Kluwer Academic Publishers.
- [ZIM 03] ZIMMERMANN P., GUTTORMSEN S., DANUSER B., GOMEZ P., « Affective Computing A Rationale for Measuring Mood with Mouse and Keyboard », *Journal of Occupational Safety and Ergonomics (JOSE)*, vol. 9, n° 4, 2003, p. 539–551, <http://www.ciop.pl/7901.html>.